nature neuroscience

ARTICLES

https://doi.org/10.1038/s41593-018-0206-1

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

Joëlle A. Pasman[1,37], Karin J. H. Verweij[1,2,37], Zachary Gerring[3], Sven Stringer[4], Sandra Sanchez-Roige[5], Jorien L. Treur[6], Abdel Abdellaoui[2], Michel G. Nivard[7], Bart M. L. Baselmans[7], Jue-Sheng Ong[3], Hill F. Ip[7], Matthijs D. van der Zee[7], Meike Bartels[7], Felix R. Day[8], Pierre Fontanillas[9], Sarah L. Elson[9], the 23andMe Research Team[10], Harriet de Wit[11], Lea K. Davis[12], James MacKillop[13], The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium[14], International Cannabis Consortium[15], Jaime L. Derringer[16], Susan J. T. Branje[17], Catharina A. Hartman[18], Andrew C. Heath[19], Pol A. C. van Lier[20], Pamela A. F. Madden[19], Reedik Mägi[21], Wim Meeus[17], Grant W. Montgomery[22], A. J. Oldehinkel[18], Zdenka Pausova[23], Josep A. Ramos-Quiroga[24,25,26,27], Tomas Paus[28,29], Marta Ribases[24,25,26], Jaakko Kaprio[30], Marco P. M. Boks[31], Jordana T. Bell[32], Tim D. Spector[32], Joel Gelernter[33], Dorret I. Boomsma[7], Nicholas G. Martin[3], Stuart MacGregor[3], John R. B. Perry[8], Abraham A. Palmer[5,34], Danielle Posthuma[4], Marcus R. Munafò[6,35], Nathan A. Gillespie[3,36,38], Eske M. Derks[3,38] and Jacqueline M. Vink[1,38]*

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.

Cannabis is a widely used psychoactive substance, and its use is associated with various adverse mental health outcomes, including psychosis and schizophrenia[1–3]. Successful prevention and intervention efforts aimed at reducing cannabis use, misuse and related outcomes require a better understanding of why some people use cannabis whereas others do not. Lifetime cannabis use, defined as any use of cannabis during lifetime, is a heritable trait: a meta-analysis of twin studies[4] estimated the heritability to be approximately 45%. Twin studies have shown there is substantial overlap in the genetic factors influencing cannabis use and those underlying problematic cannabis use (abuse or dependence)[5,6].

Several GWASs have tried to identify genetic variants underlying cannabis use phenotypes[7–11]. Recently, Demontis et al.[11] performed the largest GWAS for cannabis use disorder to date, with a discovery sample of 2,387 cases and almost 50,000 controls, plus a replication sample of 5,501 cases and ~300,000 controls. They found one genome-wide significant risk locus for cannabis use disorder, a single nucleotide polymorphism (SNP) that is a strong marker for *CHRNA2* expression. Their follow-up analyses showed that cannabis-dependent individuals had a decreased expression of this gene in the cerebellum, as well as in other brain regions.

The largest GWAS of lifetime cannabis use to date is from the International Cannabis Consortium (ICC) and is based on a sample size of 32,330 individuals in the discovery sample along with 5,627 individuals in the replication sample[10]. Although no individual SNPs reached genome-wide significance, gene-based tests identified four genes significantly associated with lifetime cannabis use: *NCAM1*, *CADM2*, *SCOC* and *KCNT2*. Notably, *NCAM1* has previously been linked to other substance use phenotypes (for example, refs [12,13]), and following publication of the study, *CADM2* was found to be associated with alcohol consumption[14], personality[15], reproductive success and risk-taking behavior[16] in other GWASs. These results indicate that *CADM2* may play a role in a broader personality profile of sensation-seeking and risk-taking behavior in general.

A full list of affiliations appears at the end of the paper.

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[3]. 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[17,18]. A substantial body of evidence supports the hypothesis that cannabis use increases the risk for developing psychoses and schizophrenia[19], but other hypotheses (namely, schizophrenia increases the use of cannabis, or the association is due to (genetic) pleiotropy) have also been posed. Previous studies have shown that genetic risk factors for cannabis use and schizophrenia are positively correlated[20,21]. 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[22]. Previous Mendelian randomization studies have investigated the link between lifetime cannabis use and schizophrenia, but findings were inconsistent. Vaucher et al.[23] 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.[24] 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, $P=9.38\times10^{-17}$; rs1448602, $P=6.55\times10^{-11}$; rs7651996, $P=2.37\times10^{-9}$). Other hits were located in *ZNF704*, *SDK1*, *NCAM1*, *RABEP2* or *ATP2A1* and *SMG6* (Fig. 2). All SNPs combined explained 11% ($h^2_{SNP}=0.11$, 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[25] 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[26] 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 ($z=5.88$, $P=4.17\times10^{-9}$). 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 ($r_g=0.24$, s.e.$=0.03$, $P<0.01$). To examine whether there was evidence for a causal effect of cannabis use on schizophrenia risk and vice versa, we performed bidirectional two-sample Mendelian randomization analysis[22]. 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

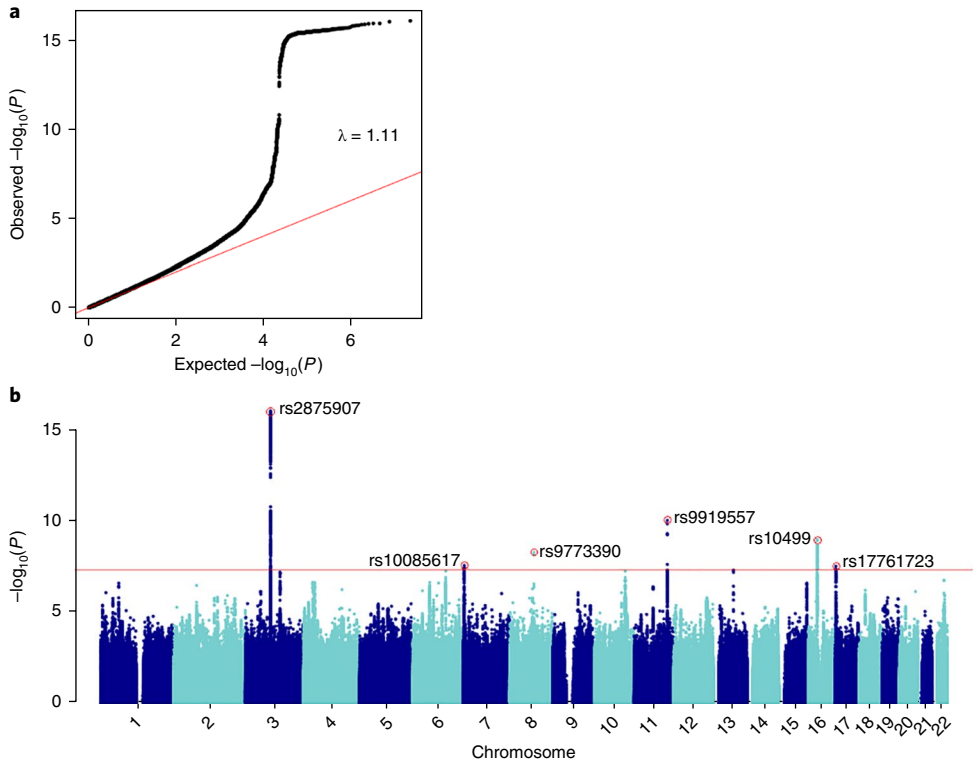

Fig. 1 | Q–Q and Manhattan plot of the GWAS meta-analysis. a, Q–Q plot of the distribution of the $-\log_{10}(P)$ observed for the SNP associations with lifetime cannabis use against those expected under the null hypothesis. Expected $-\log_{10}(P)$ values under the null hypothesis are indicated by the red line. Genomic inflation is indicated by λ in the plot. There was no evidence for population stratification (LD score regression $b_0 = 1.00$, s.e. = 0.007). **b**, Manhattan plot for the SNP-based GWAS meta-analysis. Results are based on $N = 184{,}765$ individuals and $N_{SNPs} = 11{,}733{,}371$ SNPs. The SNP with the lowest *P*-value for each independent ($R^2 < 0.1$, window size 250 kb) genome-wide significant locus is annotated by a red circle with rs-number. The red line represents the conventional genome-wide significance threshold of $P < 5 \times 10^{-8}$. The statistical test comprised linear regression; significance was tested two-sided.

was not significantly different from 0 (Supplementary Table 10), indicating no pleiotropic effects for the SNPs included in the genetic instruments[27], it is likely that this method simply lacked power to be able to reject the null hypothesis of no causal effect[28].

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 estimates[4]. *CADM2* and *NCAM1*, both identified in the original ICC meta-analysis[10], were among the strongest findings in the SNP-based and gene-based tests. The *CADM2* gene (cell adhesion molecule 2) is a synaptic cell adhesion molecule and is part of the immunoglobulin superfamily. Notably, *CADM2* has previously been identified in GWASs of other behavioral phenotypes, including alcohol consumption[14], processing speed[29], and number of offspring and risk-taking behavior[16]. A large-scale phenome-wide scan

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

SNP	Chr	Gene	BP	A1	A2	Freq A1	*N*	β	SE	*P*-value	Direction
rs2875907	3p12.1	*CADM2*	85,518,580	A	G	0.352	181,675	0.070	0.009	9.38×10^{-17}	+++
rs1448602	3p12.1	*CADM2*	85,780,454	A	G	0.756	184,765	-0.062	0.010	6.55×10^{-11}	- - -
rs7651996	3p12.1	*CADM2*	85,057,349	T	G	0.477	184,765	0.049	0.008	2.37×10^{-9}	+++
rs10085617	7p22.2	*SDK1*	3,634,711	A	T	0.416	184,765	0.046	0.008	2.93×10^{-8}	+++
rs9773390	8q21.13	*ZNF704*	81,565,692	T	C	0.933	44,595	-0.171	0.029	5.66×10^{-9}	- -?
rs9919557	11q23.2	*NCAM1*	112,877,408	T	C	0.614	180,428	-0.055	0.009	9.94×10^{-11}	- - -
rs10499	16p11.2	*RABEP2, ATP2A1*	28,915,527	A	G	0.651	179,767	0.053	0.009	1.13×10^{-9}	+++
rs17761723	17p13.3	*SMG6*	2,107,090	T	C	0.346	184,765	0.047	0.009	3.24×10^{-8}	+++

Independent hits were defined as $R^2 < 0.01$, window size 250 kb. The threshold was set at $P < 5 \times 10^{-8}$ (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.

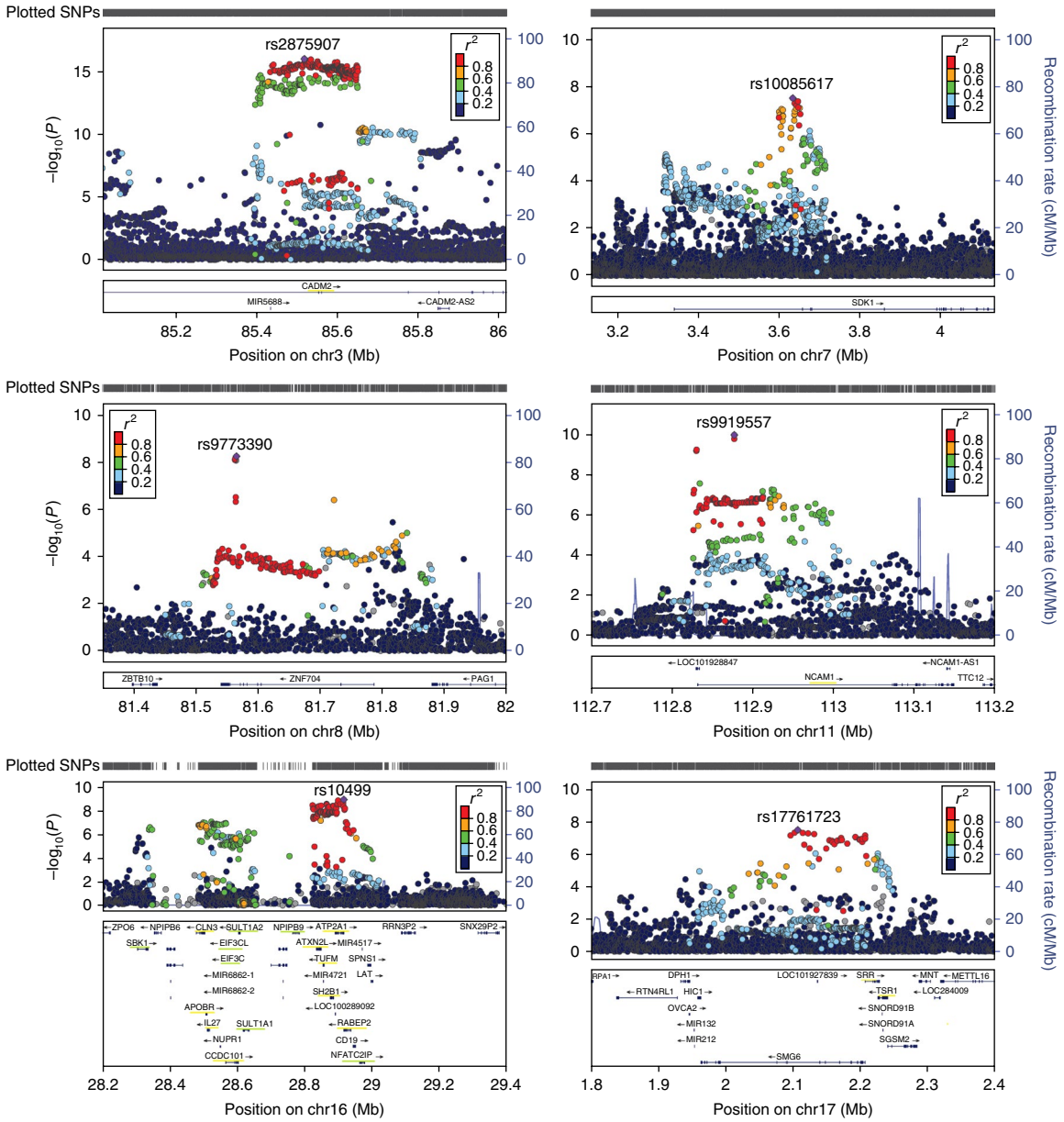

Fig. 2 | Regional plots of the genome-wide significant SNPs. Underlined in yellow are the genes that were significant in the gene-based test (tested two-sided; $P < 2.74 \times 10^{-6}$, Bonferroni corrected threshold of $P < 0.05$ adjusted for 18,293 tests); underlined in green are the genes that were identified in the S-PrediXcan analysis only ($P < 1.92 \times 10^{-7}$, Bonferroni corrected threshold of $P < 0.05$ adjusted for 259,825 tests). Colors of the dots indicate the level of LD (blue for low and red for high LD) with the lead SNP (purple; independent defined as $R^2 < 0.1$, window size 250 kb).

showed that *CADM2* was associated with various personality traits, with the risk variant being associated with reduced anxiety, neuroticism and conscientiousness and with increased risk-taking[15]. Taken together, these findings suggest that risk variants in *CADM2* are associated with a broad profile of a risk-taking, optimistic and care-free personality[15]. Cannabis use has previously been associated with related personality traits, including high levels of impulsivity and novelty seeking[30,31].

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 development[32]. *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 use[12,33–35] and has been implicated in psychiatric disorders, such as schizophrenia and mood disorders[36,37].

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 schizophrenia[38,39], while a common 16p11.2 inversion underlies susceptibility to asthma and obesity[40]. The inversion explains a substantial proportion of variability in expression of multiple genes in this region, including *TUFM* and *SH2B1*[40]. 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*),

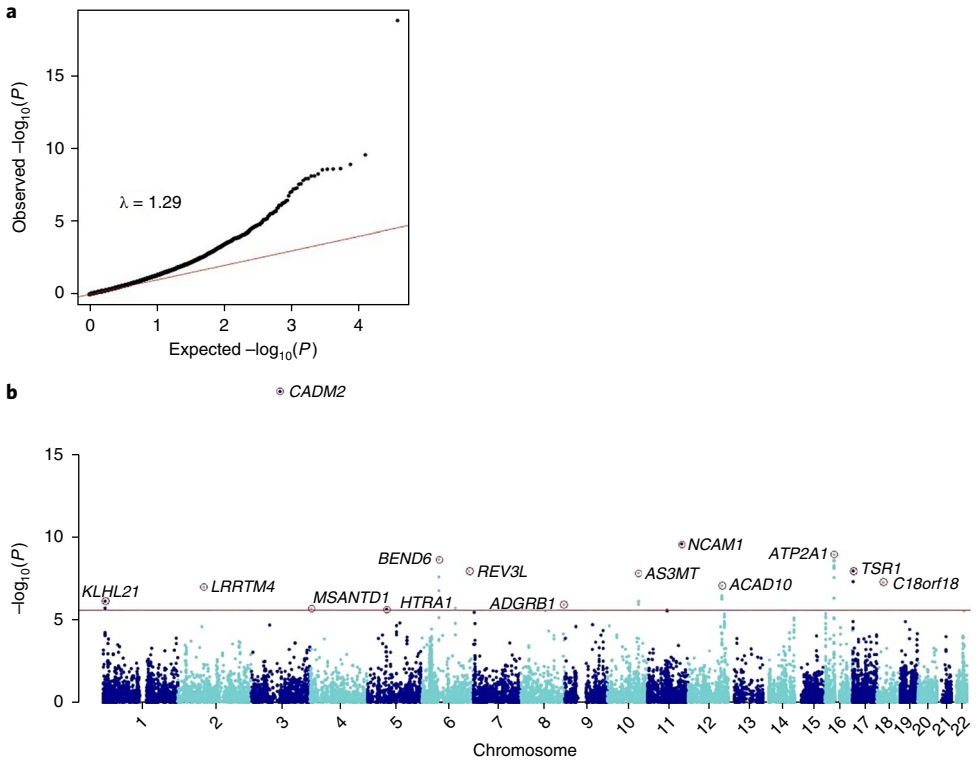

Fig. 3 | Q–Q and Manhattan plot of the gene-based test of association. a, Q–Q plot of the distribution of the $-\log_{10}(P)$ for the gene-based association with lifetime cannabis use against those expected under the null hypothesis. Expected $-\log_{10}(P)$ under the null hypothesis are indicated by the red line. Genomic inflation is indicated by λ. The gene-based test was performed in MAGMA, which uses multiple regression (tested two-sided). **b**, Manhattan plot for the gene-based test of association. The red line represents the genome-wide significance threshold of $P < 2.74 \times 10^{-6}$ (Bonferroni corrected threshold of $P < 0.05$ adjusted for 18,293 tests; $N_{SNPs} = 5{,}710{,}956$ were mapped to at least one gene). The top gene (that with the lowest *P*-value) for each locus is annotated by a red circle and gene symbol.

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[30] are well established.

There are two previous studies that found significant SNP associations for a cannabis use phenotype. Sherva et al.[46] 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 ($t(50{,}856) = 25.13$, $P < 2 \times 10^{-16}$). These 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

Locus	Top genes	BP start	BP stop	SNPs	*z*	*P*-value
1p36.31	***KLHL21***	6,640,784	6,672,958	96	4.81	7.65×10^{-7}
	PHF13	6,663,756	6,694,093	84	4.61	1.99×10^{-6}
2p12	***LRRTM4***	76,969,849	77,754,502	3,621	5.19	1.03×10^{-7}
3p12.1	***CADM2***	85,003,133	86,128,579	4,287	8.96	1.59×10^{-19}
4p16.3	***MSANTD1***	3,240,766	3,283,465	231	4.59	2.22×10^{-6}
5q12.3	***HTR1A***	63,245,875	63,268,119	64	4.57	2.41×10^{-6}
6p12.1	***BEND6***	56,814,773	56,897,450	252	5.22	2.60×10^{-8}
	KIAA1586	56,906,343	56,925,023	58	5.09	1.75×10^{-7}
	RAB23	57,046,790	57,092,112	86	5.86	2.32×10^{-9}
6q21	***REV3L***	111,610,234	111,814,421	539	4.61	1.99×10^{-6}
6q25.3	***ARID1B***	157,093,980	157,536,913	1,344	5.59	1.15×10^{-8}
8q24.3	***ADGRB1***	143,535,377	143,636,369	275	4.71	1.23×10^{-6}
10q24.32–33	*NEURL*	103,493,890	103,592,552	17	5.22	1.83×10^{-7}
	BORCS7	104,603,967	104,634,718	87	4.72	1.19×10^{-6}
	AS3MT	104,624,183	104,666,656	177	5.54	1.53×10^{-8}
	CNNM2	104,673,075	104,843,344	549	4.80	8.02×10^{-7}
	NT5C2	104,842,774	104,958,063	389	4.81	7.64×10^{-7}
11q23.2	***NCAM1***	112,826,969	113,154,158	1,263	6.21	2.63×10^{-10}
12q24.12	***BRAP***	112,069,950	112,133,790	97	4.87	5.48×10^{-7}
	ACAD10	112,118,857	112,199,911	141	5.22	8.96×10^{-8}
	ALDH2	112,199,691	112,252,789	112	4.96	3.61×10^{-7}
	MAPKAPK5	112,275,032	112,336,228	195	4.87	5.58×10^{-7}
	TMEM116	112,364,086	112,456,023	222	4.94	3.96×10^{-7}
16p11.2–16q12.1	*SBK1*	28,303,840	28,335,170	23	5.47	4.52×10^{-8}
	NPIPB7	28,467,693	28,481,868	10	5.44	5.46×10^{-8}
	CLN3	28,483,600	28,510,897	62	5.84	2.56×10^{-9}
	APOBR	28,500,970	28,515,291	49	5.66	7.56×10^{-9}
	IL27	28,505,683	28,523,155	57	5.66	7.48×10^{-9}
	CCDC101	28,560,249	28,608,111	181	4.90	4.87×10^{-7}
	SULT1A2	28,603,264	28,608,391	25	5.40	6.66×10^{-8}
	SULT1A1	28,605,196	28,623,625	51	5.30	1.14×10^{-7}
	CDC37P1	28,700,176	28,701,611	31	5.26	1.42×10^{-7}
	EIF3C	28,722,782	28,747,053	14	5.37	8.08×10^{-8}
	EIF3CL	28,722,785	28,747,053	23	5.47	4.55×10^{-8}
	NPIPB9	28,742,728	28,772,850	8	5.41	6.29×10^{-8}
	ATXN2L	28,829,369	28,853,558	89	5.85	2.50×10^{-9}
	TUFM	28,848,732	28,862,729	55	5.83	2.83×10^{-9}
	SH2B1	28,867,939	28,890,534	71	5.72	5.46×10^{-9}
	ATP2A1	28,884,192	28,920,830	89	5.97	1.20×10^{-9}
	NFATC2IP	28,962,318	28,977,767	8	5.35	8.82×10^{-8}
	RABEP2	28,910,742	28,942,339	71	5.43	2.84×10^{-8}
17p13.3	***SRR***	2,202,244	2,233,553	121	5.33	5.03×10^{-8}
	TSR1	2,220,972	2,245,678	90	5.59	1.12×10^{-8}
18q11.2	***C18orf8***	21,078,434	21,118,311	132	5.30	5.65×10^{-8}
	NPC1	21,081,148	21,171,581	257	5.30	5.87×10^{-8}

For the gene-based test, the *P*-value was set at $P < 2.74 \times 10^{-6}$, Bonferroni corrected threshold of $P < 0.05$ adjusted for 18,293 tests. For the S-PrediXcan analysis, $P < 1.92 \times 10^{-7}$, 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.

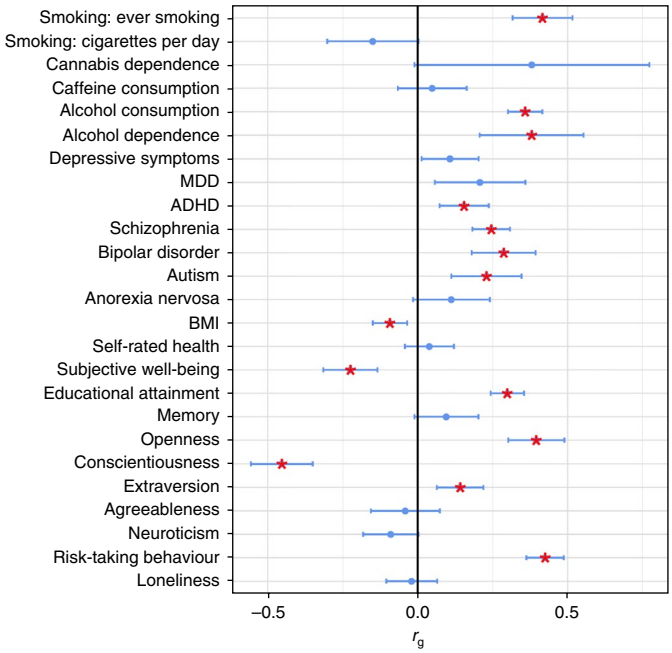

Fig. 4 | Genetic overlap between lifetime cannabis use and other phenotypes. Blue dots represent point estimates of the genetic correlation, blue error bars represent 95% confidence intervals and red asterisks indicate significant associations after correction for multiple testing (two-sided $P < 0.002$,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

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 ($r_g = 0.24$), which is in line with previous findings[20,21], 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.[23], 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.[24], 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[48]. 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 $P < 1 \times 10^{-5}$ and $P < 5 \times 10^{-8}$, 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 rs9773390 (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

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

	Cannabis–schizophrenia ($P < 5 \times 10^{-8}$, 5 SNPs)				Cannabis–schizophrenia ($P < 1 \times 10^{-5}$, 69 SNPs[a])				Schizophrenia–cannabis ($P < 5 \times 10^{-8}$, 109 SNPs[b])			
	B	**SE(*B*)**	**OR**	***P*-value**	***B***	**SE(*B*)**	**OR**	***P*-value**	***B***	**SE(*B*)**	**OR**	***P*-value**
IVW	0.039	0.158	1.04	0.806	0.091	0.051	1.10	0.074	0.151	0.046	1.16	**0.001**
Weighted median	−0.048	0.105	0.95	0.649	0.069	0.049	1.07	0.156	0.163	0.049	1.17	**0.001**
MR-Egger SIMEX	−0.044	0.190	0.96	0.827	0.106	0.110	1.11	0.340	0.071	0.293	1.07	0.810
Weighted mode	−0.084	0.125	0.92	0.536	0.016	0.071	1.02	0.823	0.315	0.178	1.37	0.080
GSMR after HEIDI filtering	–	–	–	–	0.192	0.080	1.21	**0.017**	0.237	0.038	1.27	**5.36×10^{-10}**

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). [a]Number of SNPs in instrument was 74 for the GSMR analysis. [b]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

1. Volkow, N. D., Compton, W. M. & Weiss, S. R. Adverse health effects of marijuana use. *N. Engl. J. Med.* **371**, 879 (2014).
2. Moore, T. H. et al. Cannabis use and risk of psychotic or affective mental health outcomes: a systematic review. *Lancet* **370**, 319–328 (2007).
3. Hall, W. & Degenhardt, L. Adverse health effects of non-medical cannabis use. *Lancet* **374**, 1383–1391 (2009).
4. Verweij, K. J. H. et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction* **105**, 417–430 (2010).
5. Agrawal, A., Neale, M. C., Jacobson, K. C., Prescott, C. A. & Kendler, K. S. Illicit drug use and abuse/dependence: modeling of two-stage variables using the CCC approach. *Addict. Behav.* **30**, 1043–1048 (2005).
6. Agrawal, A. & Lynskey, M. T. The genetic epidemiology of cannabis use, abuse and dependence. *Addiction* **101**, 801–812 (2006).
7. Verweij, K. J. H. et al. The genetic aetiology of cannabis use initiation: a meta-analysis of genome-wide association studies and a SNP-based heritability estimation. *Addict. Biol.* **18**, 846–850 (2013).
8. Agrawal, A. et al. A genome-wide association study of DSM-IV cannabis dependence. *Addict. Biol.* **16**, 514–518 (2011).
9. Minica, C. C. et al. Heritability, SNP- and gene-based analyses of cannabis use initiation and age at onset. *Behav. Genet.* **45**, 503–513 (2015).
10. Stringer, S. et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. *Transl. Psychiatry* **6**, e769 (2016).
11. Demontis, D. et al. Genome-wide association study implicates *CHRNA2* in cannabis use disorder. Preprint at *bioRxiv* https://doi.org/10.1101/237321 (2018).
12. Gelernter, J. et al. Haplotype spanning *TTC12* and *ANKK1*, flanked by the *DRD2* and *NCAM1* loci, is strongly associated to nicotine dependence in two distinct American populations. *Hum. Mol. Genet.* **15**, 3498–3507 (2006).
13. Yang, B. Z. et al. Association of haplotypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in independent case control and family samples. *Hum. Mol. Genet.* **16**, 2844–2853 (2007).
14. Clarke, T.-K. et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank ($N=112\,117$). *Mol. Psychiatry* **22**, 1376–1384 (2017).
15. Boutwell, B. et al. Replication and characterization of *CADM2* and *MSRA* genes on human behavior. *Heliyon* **3**, e00349 (2017).
16. Day, F. R. et al. Physical and neurobehavioral determinants of reproductive onset and success. *Nat. Genet.* **48**, 617–623 (2016).
17. Andréasson, S., Allebeck, P., Engström, A. & Rydberg, U. Cannabis and schizophrenia. A longitudinal study of Swedish conscripts. *Lancet* **2**, 1483–1486 (1987).
18. Smit, F., Bolier, L. & Cuijpers, P. Cannabis use and the risk of later schizophrenia: a review. *Addiction* **99**, 425–430 (2004).
19. Volkow, N. D. et al. Effects of cannabis use on human behavior, including cognition, motivation, and psychosis: a review. *JAMA Psychiatry* **73**, 292–297 (2016).
20. Verweij, K. J. et al. Short communication: Genetic association between schizophrenia and cannabis use. *Drug Alcohol Depend.* **171**, 117–121 (2017).
21. Power, R. A. et al. Genetic predisposition to schizophrenia associated with increased use of cannabis. *Mol. Psychiatry* **19**, 1201–1204 (2014).
22. Burgess, S., Scott, R. A., Timpson, N. J., Davey Smith, G. & Thompson, S. G. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur. J. Epidemiol.* **30**, 543–552 (2015).
23. Vaucher, J. et al. Cannabis use and risk of schizophrenia: a Mendelian randomization study. *Mol. Psychiatry* **23**, 1287–1292 (2018).
24. Gage, S. H. et al. Assessing causality in associations between cannabis use and schizophrenia risk: a two-sample Mendelian randomization study. *Psychol. Med.* **47**, 971–980 (2017).
25. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLOS Comput. Biol.* **11**, e1004219 (2015).
26. Barbeira, A.N. et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Preprint at *bioRxiv* https://doi.org/10.1101/045260 (2017).
27. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).
28. Burgess, S. & Thompson, S. G. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur. J. Epidemiol.* **32**, 377–389 (2017).
29. Ibrahim-Verbaas, C. A. et al. GWAS for executive function and processing speed suggests involvement of the *CADM2* gene. *Mol. Psychiatry* **21**, 189–197 (2016).
30. Walther, B., Morgenstern, M. & Hanewinkel, R. Co-occurrence of addictive behaviours: personality factors related to substance use, gambling and computer gaming. *Eur. Addict. Res.* **18**, 167–174 (2012).
31. Martin, C. A. et al. Sensation seeking, puberty, and nicotine, alcohol, and marijuana use in adolescence. *J. Am. Acad. Child Adolesc. Psychiatry* **41**, 1495–1502 (2002).
32. Nielsen, J., Kulahin, N. & Walmod, P. S. Extracellular protein interactions mediated by the neural cell adhesion molecule, NCAM: heterophilic interactions between NCAM and cell adhesion molecules, extracellular matrix proteins, and viruses. *Adv. Exp. Med. Biol.* **663**, 23–53 (2010).
33. Rubinek, T. et al. The cell adhesion molecules N-cadherin and neural cell adhesion molecule regulate human growth hormone: a novel mechanism for regulating pituitary hormone secretion. *J. Clin. Endocrinol. Metab.* **88**, 3724–3730 (2003).
34. Ducci, F. et al. *TTC12-ANKK1-DRD2* and *CHRNA5-CHRNA3-CHRNB4* influence different pathways leading to smoking behavior from adolescence to mid-adulthood. *Biol. Psychiatry* **69**, 650–660 (2011).
35. Bidwell, L. C. et al. *NCAM1-TTC12-ANKK1-DRD2* variants and smoking motives as intermediate phenotypes for nicotine dependence. *Psychopharmacology (Berl.)* **232**, 1177–1186 (2015).
36. Atz, M. E., Rollins, B. & Vawter, M. P. *NCAM1* association study of bipolar disorder and schizophrenia: polymorphisms and alternatively spliced isoforms lead to similarities and differences. *Psychiatr. Genet.* **17**, 55–67 (2007).
37. Petrovska, J. et al. The *NCAM1* gene set is linked to depressive symptoms and their brain structural correlates in healthy individuals. *J. Psychiatr. Res.* **91**, 116–123 (2017).
38. Weiss, L. A. et al. Association between microdeletion and microduplication at 16p11.2 and autism. *N. Engl. J. Med.* **358**, 667–675 (2008).
39. McCarthy, S. E. et al. Microduplications of 16p11.2 are associated with schizophrenia. *Nat. Genet.* **41**, 1223–1227 (2009).
40. González, J. R. et al. A common 16p11.2 inversion underlies the joint susceptibility to asthma and obesity. *Am. J. Hum. Genet.* **94**, 361–372 (2014).
41. Zuo, L. et al. Genome-wide significant association signals in *IPO11-HTR1A* region specific for alcohol and nicotine codependence. *Alcohol. Clin. Exp. Res.* **37**, 730–739 (2013).
42. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
43. Donaldson, Z. R. et al. The functional serotonin 1a receptor promoter polymorphism, rs6295, is associated with psychiatric illness and differences in transcription. *Transl. Psychiatry* **6**, e746 (2016).
44. Gatt, J. M., Burton, K. L., Williams, L. M. & Schofield, P. R. Specific and common genes implicated across major mental disorders: a review of meta-analysis studies. *J. Psychiatr. Res.* **60**, 1–13 (2015).
45. Takekita, Y. et al. *HTR1A* polymorphisms and clinical efficacy of antipsychotic drug treatment in schizophrenia: a meta-analysis. *Int. J. Neuropsychopharmacol.* **19**, pyv125 (2016).
46. Sherva, R. et al. Genome-wide association study of cannabis dependence severity, novel risk variants, and shared genetic risks. *JAMA Psychiatry* **73**, 472–480 (2016).
47. Patrick, M. E., Wightman, P., Schoeni, R. F. & Schulenberg, J. E. Socioeconomic status and substance use among young adults: a comparison across constructs and drugs. *J. Stud. Alcohol Drugs* **73**, 772–782 (2012).
48. Müller-Vahl, K. R. & Emrich, H. M. Cannabis and schizophrenia: towards a cannabinoid hypothesis of schizophrenia. *Expert Rev. Neurother.* **8**, 1037–1048 (2008).

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

would not have been possible. J.A.P. and J.M.V. are supported by the European Research Council (Beyond the Genetics of Addiction ERC-284167, PI J.M.V.). K.J.H.V. is supported by the Foundation Volksbond Rotterdam. N.A.G. is supported by US National Institutes of Health, National Institute on Drug Abuse R00DA023549. J.L.T. is supported by the Netherlands Organization for Scientific Research (NWO; Rubicon grant 446-16-009). S.M. is supported by an Australian Research Council Fellowship. Statistical analyses 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 PI: 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 U01AA008401 (COGA). J.M.'s contributions were partially supported by the Peter Boris Chair in Addictions Research. S.S.-R. was supported by the Frontiers of Innovation Scholars Program (FISP; #3-P3029), the Interdisciplinary Research Fellowship in NeuroAIDS (IRFN; MH081482) and a pilot award from DA037844. R.M. was supported by the European Union through the European Regional Development Fund (Project No. 2014-2020.4.01.15-0012) and the European Union's Horizon 2020 Research and Innovation Programme under grant agreements No 692065 and 692145. J.K. was supported by Academy Professorship grants by the Academy of Finland (263278, 292782). M.R. is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Spain (CP09/00119 and CPII15/00023).

Author contributions

K.J.H.V., E.M.D. and J.M.V. were responsible for the study concept and the design of the study. I.C.C. contributed existing genome-wide summary data from the International Cannabis Consortium. The PGC-SUD group provided summary statistics of the cannabis dependence GWAS. S.S., S.S.-R., M.G.N., B.M.L.B., J.-S.O., H.F.I., M.D.v.d.Z., M.B., F.R.D., P.F., S.M., J.R.B.P., A.A.P. and D.P. performed or 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., B.M.L.B. and J.M.V. J.A.P., K.J.H.V., Z.G., J.L.T., A.A., M.R.M. and E.M.D. contributed to secondary analyses of the data. J.A.P., K.J.H.V., Z.G., N.A.G., E.M.D. and J.M.V. wrote the manuscript. J.L.D., S.J.T.B., C.A.H., A.C.H., P.A.C.v.L., P.A.F.M., R.M., W.M., G.W.M., A.J.O., Z.P., J.A.R.-Q., T.P., M.R., J.K., M.P.M.B., J.T.B., T.D.S., J.G., D.I.B. and 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.

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.

Correspondence and requests for materials should be addressed to J.M.V.

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, V. Vacic and C.H. Wilson.

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, I. Garcia-Martinez, S. D. Gordon, J. M. Harris, C. A. Hartman, A. K. Henders, A. C. Heath, I. B. Hickie, 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, D. I. Boomsma, W. G. Iacono, N. G. Martin, N. A. Gillespie, E. M. Derks and J. M. Vink.

[1]Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands. [2]Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam, The Netherlands. [3]Genetic Epidemiology, Statistical Genetics, and Translational Neurogenomics Laboratories, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [4]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [5]Department of Psychiatry, University of California San Diego, La Jolla, CA, USA. [6]MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK. [7]Department of Biological Psychology/Netherlands Twin Register, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [8]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. [9]23andMe, Inc., Mountain View, CA, USA. [10]A list of 23andMe Research Team members appears at the end of the paper. [11]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. [12]Vanderbilt Genetics Institute; Division of Genetic Medicine, Department of Medicine, Vanderbilt University, Nashville, TN, USA. [13]Peter Boris Centre for Addictions Research and Michael G. DeGroote Centre for Medicinal Cannabis Research, McMaster University/St. Joseph's Healthcare Hamilton, Hamilton, Ontario, Canada. [14]Psychiatric Genomics Consortium. [15]A list of International Cannabis Consortium members appears at the end of the paper. [16]Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA. [17]Department of Youth and Family, Utrecht University, Utrecht, the Netherlands. [18]Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [19]Department of Psychiatry, Washington University School of Medicine, St. Louis, MO, USA. [20]Department of Developmental Psychology and EMGO Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [21]Estonian Genome Center, University of Tartu, Tartu, Estonia. [22]Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia. [23]Hospital for Sick Children, Toronto, Ontario, Canada. [24]Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Spain. [25]Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain. [26]Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Barcelona, Spain. [27]Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain. [28]Rotman Research Institute, Baycrest, Toronto, Ontario, Canada. [29]Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada. [30]Institute for Molecular Medicine Finland FIMM, HiLIFE Unit, University of Helsinki, Helsinki, Finland. [31]Brain Center Rudolf Magnus, Department of Psychiatry, University Medical Center Utrecht, Utrecht, The Netherlands. [32]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [33]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. [34]Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA. [35]UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, Bristol, UK. [36]Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, VA, USA. [37]These authors contributed equally: Joëlle A. Pasman, Karin J. H. Verweij. [38]These authors jointly supervised: Nathan A. Gillespie, Eske M. Derks, Jacqueline M. Vink. *e-mail: j.vink@bsi.ru.nl

Methods

Samples. Data from three sources were obtained: ICC, 23andMe and UK Biobank (total $N = 184{,}765$). We used existing GWAS summary statistics from the ICC, based on data from 35,297 individuals of European ancestry from 16 cohorts from North America, Europe and Australia[10]. Details regarding ethical approval and informed consents of the ICC cohorts can be found in the original ICC paper[10]. The overall sample included 55.5% females and the age ranged between 16 and 87 years with a mean of 35.7 years. An average of 42.8% of the individuals had used cannabis during their lifetime. The second set of results was derived from the personal genetics company 23andMe Inc. Data were available for 22,683 individuals of European Ancestry who provided informed consent and answered surveys online according to a human subjects protocol approved by Ethical & Independent Review Services, a private institutional review board. The sample included 55.3% females and the age ranged between 18 and 94 years with a mean of 54.0 years. Within the sample, 43.2% had used cannabis during their lifetime. The third sample was obtained from UK Biobank. Data were available for 126,785 individuals of European ancestry. The sample included 56.3% females and the age ranged between 39 and 72 years with a mean of 55.0 years. Within the sample, 22.3% had used cannabis during their lifetime. Ethical approval for UK Biobank data collection procedures has been provided by the North West Multi-center Research Ethics Committee (MREC), the National Information Governance Board for Health & Social Care (NIGB) and the Community Health Index Advisory Group (CHIAG).

Phenotype and covariates. For all participants, self-report data were available on whether the participant had ever used cannabis during their lifetime: yes (1) versus no (0). Measurement instruments and phrasing of the questions about lifetime cannabis use differed across the samples. For the ICC study this has been described for each cohort in the original paper[10]. As part of their online questionnaire, 23andMe used the following phrase to examine lifetime cannabis use: "Have you ever in your life used the following: Marijuana?" The UK Biobank, as part of an online follow-up questionnaire, asked: "Have you taken CANNABIS (marijuana, grass, hash, ganja, blow, draw, skunk, weed, spliff, dope), even if it was a long time ago?"

Genotyping and imputation. Genotyping was performed on various genotyping platforms and standard quality control checks were performed before imputation. Genotype data were imputed using the 1000 Genomes phase 1 release reference set[49] for ICC and 23andMe, and the Haplotype Reference Consortium reference set[50] for the UK Biobank sample. Information about samples, genotyping, imputation and quality control is summarized in Supplementary Tables 11 and 12. After quality control, the ICC sample comprised 35,297 individuals and 6,643,927 SNPs, the 23andMe sample 22,683 individuals and 7,837,888 SNPs, and the UK Biobank sample 126,785 individuals and 10,827,718 SNPs.

Statistics. All statistical tests were two-sided and, unless stated otherwise, we used the conventional P-value of 0.05 for significance testing. When necessary, Bonferroni correction for multiple testing was applied. Randomization and blinding procedures do not apply to our study design.

Genome-wide association analyses and meta-analysis. We conducted the GWASs in 23andMe and UK Biobank samples separately. Associations between the binary phenotype and SNPs were tested using a logistic regression model accounting for the effects of sex, age, ancestry and genotype batch (and age^2 in the UK Biobank sample). The GWAS for UK Biobank was performed in PLINK 1.9[51] and the GWAS for 23andMe using an internally developed pipeline. We then meta-analyzed the GWAS results from ICC, 23andMe and UK Biobank. Prior to conducting the meta-analysis, additional quality control of the summary statistics of each study was conducted in EasyQC[52]. Because of varying GWAS methods and sample characteristics (Supplementary Table 11), slightly different quality control criteria were used for the three samples (Supplementary Table 12). All three samples were aligned with the Haplotype Reference Consortium panel using the EasyQC R-package[52], to ensure that rs-numbers and chromosome-basepair positions referred to the same variants and to correct for strand effects. Variants were deleted if they had a minor allele frequency (MAF) diverging more than 0.15 from that in the reference panel.

We applied genomic control to the three GWAS files before meta-analysis. Inflation due to stratification was estimated using LD-score regression, which can differentiate inflation due to population stratification from that due to real signal. The intercept was used to correct the standard errors (SEs) of the estimated effect sizes as follows: $\mathrm{SE}_{GC} = \sqrt{\text{LDSC intercept} \times \mathrm{SE}^2}$. The intercepts were $b_0 = 1.005$ (s.e. = 0.007) for ICC, $b_0 = 1.004$ (s.e. = 0.007) for 23andMe and $b_0 = 1.022$ (s.e. = 0.008) for UK Biobank. We then performed a fixed effects meta-analysis based on effect sizes (log odds ratios) and standard errors in METAL[53]. We applied the conventional P-value threshold of 5×10^{-8} as an indication of genome-wide significance. The meta-analysis was performed on 11,733,371 SNPs that passed quality control. The combined sample size of the meta-analysis was 184,765 individuals, although the sample size varied per SNP due to differential missingness across samples.

Manhattan and Q–Q plots for the GWASs, meta-analysis and gene-based test results were created using the R-package qqman[54]. Regional plots were created using LocusZoom[55], with varying window size for optimal visualization.

Gene-based test of association. Testing associations on the level of protein-coding genes can be more biologically meaningful and is more powerful (lower multiple testing burden) than testing solely on the level of SNPs. Gene-based analysis was used to test associations for the combined effect of SNPs in protein-coding genes taking into account LD between the SNPs and the size of the gene. The analysis was conducted in MAGMA (v 1.6)[25], which uses the 1000 Genomes reference panel (Phase 3, 2012) to control for LD. SNPs were mapped to genes if they were located in or within 10 kb of the gene; 5,710,956 SNPs (49%) could be mapped to at least one of 18,293 protein-coding genes in the reference panel. The significance threshold was set at $P < 2.74 \times 10^{-6}$ (Bonferroni corrected P-value for 18,293 tests).

Identification of genes with differential expression levels between cannabis users and nonusers. We used S-PrediXcan to integrate eQTL (expression quantitative trait loci) information with our GWAS summary statistics to identify genes whose genetically predicted expression levels are associated with cannabis use[26]. Briefly, S-PrediXcan estimates gene expression weights by training a linear prediction model in samples with both gene expression and SNP genotype data. The weights are then used to predict gene expression from GWAS summary statistics, while incorporating the variance and covariance of SNPs from an LD reference panel. We used expression weights for 48 tissues from the GTEx Project (V7) and the DGN whole blood cohort generated by Gamazon et al.[56], and LD information from the 1000 Genomes Project Phase 3[57]. These data were processed with β values and standard errors from the lifetime cannabis use GWAS meta-analysis to estimate the expression–GWAS association statistic. We used a transcriptome-wide significance threshold of $P < 1.92 \times 10^{-7}$, which is the Bonferroni corrected threshold when adjusting for all tissues and genes (i.e., $N = 259{,}825$ gene-based tests in the GTEx and DGN reference sets).

We used the GTEx Portal (https://www.gtexportal.org/home/; GTEx Analysis Release V7)[58] to obtain gene expression levels of *CADM2* across tissues. We used the same portal to plot a multi-tissue eQTL comparison of the top SNP, rs2875907. The multi-tissue eQTL plot shows both the single-tissue eQTL P-value and the multi-tissue posterior probability from METASOFT[59].

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_{SNPs}) was estimated using LD-score regression analysis[60]. 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, SNPs 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[60]. 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.[60] based on the Hapmap 3 reference panel, restricted to European populations[61].

Genetic correlations with other substances and mental health phenotypes. We used cross-trait LD-score regression[62] 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 GWASs on the LD score. The estimate represents the genetic covariation between the two traits based on all polygenic effects captured by SNPs. Summary statistics from well-powered GWASs 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)[22] to examine whether there was evidence for a causal relationship from cannabis use to schizophrenia and vice versa. Analyses were performed with the R package of database and analytical platform MR-Base[63] and with the gsmr R package, which implements the GSMR (generalized summary-data-based Mendelian randomization) method[64].

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)[65]. Two-sample MR refers to the application of MR methods to well-powered summary association results estimated in non-overlapping sets of individuals[22] 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[66], 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 ($P < 5 \times 10^{-8}$; 5 SNPs) and (ii) an instrument including independent variants with a more lenient significance threshold ($P < 1 \times 10^{-5}$; 69 SNPs). For schizophrenia we used one genetic instrument, including independent genetic variants that were genome-wide significantly associated with schizophrenia (instrument $P < 5 \times 10^{-8}$; 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 ($R^2 \geq 0.8$), 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[67]. In addition, four sensitivity analyses more robust to horizontal pleiotropy were applied, each relying on distinct assumptions regarding instrument validity: weighted median[68], MR-Egger SIMEX[27], weighted mode[69] and generalized summary-data-based Mendelian randomization (GSMR)[64]. These sensitivity analyses rely on orthogonal assumptions, making their inclusion important for triangulation. The weighted median approach provides a consistent estimate of the causal effect even when up to 50% of the weight comes from invalid instruments[68]. MR-Egger regression applies Egger's test to MR instruments that consist of multiple genetic variants[27,28]. MR-Egger provides a consistent estimate of the causal effect, provided that the strength of the genetic instrument (the association between SNPs and exposure) does not correlate with the effect the instrument has on the outcome (i.e., the InSIDE assumption: instrument strength independent of direct effect). This is a weaker assumption than the assumption of no pleiotropy. MR-Egger may, however, be biased when the NOME (no measurement error) assumption is violated—i.e., the assumption that the SNP–exposure associations are known rather than estimated. Violation of NOME can be quantified with the I^2 statistic, which ranges between 0 and 1. A value below 0.9 indicates a considerable risk of bias. This bias can be corrected for with MR-Egger simulation extrapolation, SIMEX[70]. Since I^2 ranged between 0.7 and 0.9 for our analyses, we report results from MR-Egger SIMEX in Table 3. The weighted mode methods can produce an unbiased result, as long as the most common causal effect estimate is a consistent estimate of the true causal effect: the zero modal pleiotropy assumption (ZEMPA)[69]. Finally, we performed GSMR, a method that leverages power from multiple genetic variants while accounting for LD between these variants[64]. Because GSMR accounts for LD, genetic variants that were included in GSMR instruments were pruned at a higher threshold of $R^2 < 0.05$ (instead of $R^2 < 0.001$ for the other MR analyses). Zhu et al.[64] showed that the gain of power from including SNPs in higher LD than 0.05 is limited. GSMR also allows extra filtering for SNPs that are suspected to have pleiotropic effects on both the exposure and the outcome (HEIDI filtering).

To calculate variance explained (R^2) by the instrument, first we selected a single SNP to obtain an estimate of the phenotypic variance, var(y). Assuming effect sizes are normally distributed, we used the quantile function of the Student t-distribution to transform the P-value of the SNP association into an estimate of t, $\hat{t}$. The number of degrees of freedom and N were based on the effective sample size, $4 / (1 / \text{cases} + 1 / \text{controls})$. The effective sample sizes were estimated at $N = 130{,}072$ for schizophrenia and $N = 180{,}934$ for cannabis use. The corresponding value of r was calculated using the formula $t = r / \sqrt{(1-R^2) / (N-2)}$, and we obtained the R^2 that corresponds to t with the online tool http://vassarstats.net/rsig.html. Subsequently, we approximated the variance of the phenotype y using $\text{var}(y) = [2 \times \text{MAF} \times (1 - \text{MAF}) \times \beta^2]/R^2$, in which MAF denotes the minor allele frequency and β the effect size of the specific SNP. Finally, we used the estimated value of var(y) to calculate the R^2 for the remaining SNPs of interest using $R^2 = (2 \times \text{MAF} \times (1 - \text{MAF}) \times \beta^2)/\text{var}(y)$ and summed the R^2 of all SNPs of interest included in the instrumental variable to obtain an estimate of the total R^2 explained by the instrument.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

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) are available via LDhub (http://ldsc.broadinstitute.org/gwashare/) and https://www.ru.nl/bsi/research/group-pages-0/substance-use-addiction-food-saf/vm-saf/genetics/international-cannabis-consortium-icc/. Code and scripts are available upon reasonable request. Full summary statistics can only be provided after permission by 23andMe.

References

49. Abecasis, G. R. et al. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
50. McCarthy, S. & Das, S. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
51. Chang, C. C. et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
52. Winkler, T. W. & Day, F. R. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
53. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
54. Turner, S.D. qqman: an R package for visualizing GWAS results using Q-Q and Manhattan plots. Preprint at *bioRxiv* https://doi.org/10.1101/005165 (2014).
55. Pruim, R. J. et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
56. Gamazon, E. R. & Wheeler, H. E. A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet.* **47**, 1091–1098 (2015).
57. Delaneau, O. & Marchini, J. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat. Commun.* **5**, 3934 (2014).
58. Carithers, L. J. et al. A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv. Biobank.* **13**, 311–319 (2015).
59. Sul, J. H., Han, B., Ye, C., Choi, T. & Eskin, E. Effectively identifying eQTLs from multiple tissues by combining mixed model and meta-analytic approaches. *PLoS Genet.* **9**, e1003491 (2013).
60. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
61. Altshuler, D. M. et al. Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52–58 (2010).
62. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
63. Hemani, G. et al. MR-Base: a platform for systematic causal inference across the phenome using billions of genetic associations. Preprint at *bioRxiv* https://doi.org/10.1101/078972 (2017).
64. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).
65. Davey Smith, G. & Hemani, G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum. Mol. Genet.* **23**(R1), R89–R98 (2014).
66. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
67. Ehret, G. B. et al. Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk. *Nature* **478**, 103–109 (2011).
68. Bowden, J., Davey Smith, G., Haycock, P. C. & Burgess, S. Consistent estimation in Mendelian randomization with some invalid instruments using a weighted median estimator. *Genet. Epidemiol.* **40**, 304–314 (2016).
69. Hartwig, F.P., Davey Smith, G. & Bowden, J. Robust inference in two-sample Mendelian randomisation via the zero modal pleiotropy assumption. Preprint at *bioRxiv* https://doi.org/10.1101/126102 (2017).
70. Bowden, J. et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *Int. J. Epidemiol.* **45**, 1961–1974 (2016).

nature research

Corresponding author(s): Jacqueline M. Vink

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 qqman - 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 transparant 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. Analists were not blind to case-control status.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

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. Stringer, S. et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Translational psychiatry 6, e769, doi:10.1038/tp.2016.36 (2016).