Citation for published version (APA):
Cannabis is the most widely produced and consumed illicit psychoactive substance worldwide. Occasional cannabis use can progress to frequent use, abuse and dependence with all known adverse physical, psychological and social consequences. Individual differences in cannabis initiation are heritable (40–48%). The International Cannabis Consortium was established with the aim to identify genetic risk variants of cannabis use. We conducted a meta-analysis of genome-wide association data of 13 cohorts \((N = 32330)\) and four replication samples \((N = 5627)\). In addition, we performed a gene-based test of association, estimated single-nucleotide polymorphism (SNP)-based heritability and explored the genetic correlation between lifetime cannabis use and cigarette use using LD score regression. No individual SNPs reached genome-wide significance. Nonetheless, gene-based tests identified four genes significantly associated with lifetime cannabis use: NCAM1, CADM2, SOCC, and KNT2. Previous studies reported associations of NCAM1 with cigarette smoking and other substance use, and those of CADM2 with body mass index, processing speed and autism disorders, which are phenotypes previously reported to be associated with cannabis use.
INTRODUCTION

Cannabis is the most widely produced and consumed illicit psychoactive substance worldwide.1 Following initiation, occasional cannabis use can progress to frequent use, abuse and dependence. About 1 in 10 occasional users becomes dependent, which is associated with physical, psychological, social and occupational consequences.2–3 Despite the increasing use of cannabis for medicinal purposes,4 associations with adverse health effects have been reported.5,6 These include increased risk for psychiatric outcomes, including psychosis, schizophrenia, schizotypal personality disorder and mania.7,8 Early cannabis use appears to moderate relationship between polygenic risk scores for schizophrenia and brain maturation.9 In view of expanding medicalization and decriminalization, the potential consequences, and the debate surrounding the benefits versus adverse consequences associated with cannabis use,10 understanding the genetics of cannabis use should be a public health priority.11

The risk of lifetime cannabis use, defined as any use of cannabis during the lifetime, varies between individuals. Previous studies have shown that individual differences in lifetime cannabis use can be partly explained by genetic differences between individuals; a meta-analysis of twin studies reported significant heritability estimates of lifetime cannabis use of 48% for males and 40% for females.12 Shared environmental factors, such as cannabis availability and parental monitoring,13,14 also have a role accounting for 25 and 39% of the risk for males and females, respectively.15 Also, there is substantial overlap in the genetic risks underlying lifetime cannabis use and cannabis use disorder.15

Several studies have sought to identify specific genetic risk factors associated with cannabis use phenotypes. Genome-wide linkage studies have revealed suggestive evidence for linkage across many chromosomes.16–20 With very little consistency across studies, nearly all findings failed to meet genome-wide significance. The one study examining lifetime cannabis use16 reported a nonsignificant linkage locus on chromosome 18 (LOD score = 1.97).

Candidate gene studies, including reports examining the CNR1, GABRA2, FAAH and ABCB1 genes have detected some significant associations with cannabis use but again, replication has been inconsistent.21–25 On the basis of a sample of 7452 Caucasian individuals, Verweij et al.21 found no gene-based associations between the frequency of cannabis use and 10 candidate genes identified by Agrawal and Lynskey.24 Overall, the results of candidate-gene studies are inconclusive; some associations have been replicated a few times, but failed to replicate in other studies. Moreover, findings may be further distorted due to publication bias favouring significant results.

As an alternative to the candidate-gene approach, the genome-wide association study (GWAS) is a hypothesis-free method that aims to detect novel genetic variants involved in complex traits. To date, three GWASs of cannabis use phenotypes have been published: one GWAS of cannabis dependence in 708 cannabis-dependent individuals and 2346 controls;25 a GWAS meta-analysis of lifetime cannabis use based on two studies with a combined sample size of 10 091 individuals (40.7% users);26 and a recent GWAS of lifetime cannabis use and age of cannabis use onset based on a sample of 6744 individuals (of whom 20% were users).27 None of the studies identified any genome-wide significant associations. This was likely due to the small effect sizes typical of common variants underpinning highly polygenic traits,28 thereby indicating a need for larger sample sizes. In this context, the success of larger GWASs and international consortia examining a variety of complex traits is encouraging.29 For example, multiple large meta-analyses of GWAS results for number of cigarettes smoked per day have independently identified associations on chromosome 15q25 spanning the α5, α3 and β4 nicotinic receptor subunit gene clusters (CHRNA5, CHRNA3, CHRNB4).30–32 These and other recent GWAs findings32 clearly illustrate the need for larger sample sizes. In response to this need, the International Cannabis Consortium was established to combine the results of multiple GWASs to identify the genetic variants underlying individual differences in cannabis use phenotypes. Our rationale for focusing on lifetime cannabis use (yes/no) is because this phenotype is heritable and shares significant genetic risks with that risk for cannabis abuse or dependence.14,15,33 In contrast to frequency of use or abuse and dependence, which are not commonly assessed in large-scale genetic studies, most general population studies have assessed lifetime cannabis use, thereby increasing our sample size and power to detect associations. Currently, the combined International Cannabis Consortium sample size for lifetime cannabis is 32 330 individuals from 13 cohorts from Europe, the United States and Australia, along with four independent replication samples comprising 5627 individuals. This sample size is considerably larger than the sample size of the previous GWAS investigating lifetime cannabis use in two samples from Australia and the UK, thereby providing substantially greater power to detect genetic variants of small effect size. The aims of the present study is to identify genetic variants associated with lifetime cannabis use by meta-analysis of the GWAS results from all contributing International Cannabis Consortium samples. The tests of association for individual genetic variants will be complemented with gene-based tests of association. In addition, we will investigate which proportion of the heritability inferred by twin studies is explained by common SNPs captured on GWAS arrays. Finally, we will estimate the genetic correlation between lifetime cannabis and smoking initiation based on the analysis of our summary statistics and those from the publicly available Tobacco Alcohol and Genetics consortium.

MATERIALS AND METHODS

Cohorts

We performed a meta-analysis of GWAS results from 13 discovery samples from Europe, USA and Australia including a total of 32 330 individuals of European ancestry. The size of the samples ranged from 721 to 6778 individuals. The age of the participants ranged from 16 to 87 years with an average of 34 years. The percentage of females ranged from 30 to 66% with an average of 53%. Owing to the differences in recruitment strategies, cultural and temporal difference, combined with likely variation in the drug availability between countries, there was a wide range in the prevalence of lifetime use (that is, never/ever used cannabis), which varied from 1 to 92% with an average of 44.5%.

Four additional independent samples with a total of 5627 subjects were used for replication. One sample (n = 2660) consisted of African American
subjects. The other three included subjects of European ancestry. See Table 1 for individual sample characteristics. The procedures for data collection per sample are described in the Supplementary Information 1.

Phenotype and covariates
For all individuals, the data were available on whether or not the subject reported having ever used cannabis during their lifetime: yes (1) versus no (0). Although phrasing of the question slightly differed between samples (see Supplementary Information 3), our unit of analysis reflected lifetime cannabis use in all the samples. Covariates included age at the time of phenotypic assessment, sex, birth cohort and principal components (obtained from the genome-wide genotype data). Spanning 20-year intervals, birth cohort was dummy coded, with the lowest birth cohort (that is, oldest age group) used as the reference group. The details about phenotypic assessment and individual sample characteristics are located in Supplementary Information 1 and Supplementary Table 1.

Genotyping and imputation
Genotype imputation was based on the 1000 Genomes phase 1 reference panel.34 Allelic dosage data were used to account for genotype uncertainties. See Supplementary Table 2 for the genotyping platform, imputation program and quality control thresholds used.

Statistical analyses
**GWA analysis in each discovery cohort.** The GWA analyses were performed by each group separately. Associations between the binary phenotype and the genotypes were tested genome-wide using a logistic regression model including covariates (see above). For family-based samples, familial relatedness was taken into account by using a sandwich correction as implemented in PLINK.16 The analyses plan can be found in Supplementary Information 3. It should be noted that some groups did do the analyses in a slightly different manner based on the characteristics of their sample. The analyses plan that was sent to the participating groups is included in Supplementary Information 3. It should be noted that some groups did do the analyses in a slightly different manner based on the characteristics of their sample. Supplementary Table 2 lists the program used by each group.

**Meta-analysis of GWA results.** Before performing the meta-analysis, we applied a set of filters to each GWA results set independently. First, we removed insertions and deletions, ensuring that all base pair positions were unique and referred to the same genetic variant (that is, SNP). Second, we removed genotyped SNPs that were not in Hardy–Weinberg equilibrium ($P \leq 10^{-5}$). Third, we removed SNPs with minor allele frequency (MAF) < $\sqrt{(5/N)}$, which under the assumption of Hardy–Weinberg equilibrium corresponded to less than five estimated individuals in the least frequent genotype group. In the EGCUT1 sample, due to very low prevalence of lifetime cannabis use (1.3%), we excluded SNPs with MAF < 0.2. Fourth, regardless of the quality score type used, we excluded SNPs with imputation quality scores below 0.6. Finally, SNPs present in only one sample and SNPs with alleles or allele frequencies inconsistent with the 1000 Genomes phase 1 European reference panel (absolute MAF difference $>0.15$) were removed.

We performed a fixed-effects meta-analysis based on the cohort’s effect sizes and standard errors using METAL.36 Our meta-analysis combined association summary statistics for 6 444 471 SNPs that passed all the filters. We applied the conventional threshold of $5 \times 10^{-8}$ as an indication of genome-wide significance (see ref. 37). Although the combined sample size of the meta-analysis based on the discovery samples is 32 330, the sample size per SNP varies due to missingness across subsamples.

**Gene-based test.** Results of the GWAS were then used as part of gene-based tests of association in the Knowledge-based mining system for Genome-wide Genetic studies (KGG) software package Version 3.2.38,39 This approach uses an extended Simes test that integrates prior functional information and the meta-analysis association results when combining the SNP $P$-values within a gene to obtain an overall association $P$-value for each entire gene. We conducted 24 576 gene-based tests of association. The genome-wide significance level according to the Knowledge-based mining system for Genome-wide Genetic studies default setting of Benjamini and Hochberg false discovery rate threshold of 0.05 (ref. 40) was $9.38 \times 10^{-6}$.

**Estimation of SNP-based heritability and genetic overlap with lifetime cigarette smoking.** The proportion of phenotypic variance that could be explained by the SNPs was estimated using the density estimation method developed by So et al.31 The density estimation method estimates the genome-wide distribution of effect sizes based on the difference between the observed distribution of test statistics in the meta-analysis and the corresponding null distribution. Before estimation, the SNPs present in at least 25% of the meta-analysis samples were pruned for LD. We used the $r^2 = 0.15$ pruning level as the primary result for consistency with other applications of this method. Additional details are located in the Supplementary Information 2. LD Score regression32,43 was used as an alternative method to estimate the SNP-based heritability, as well as to estimate the degree of genetic covariance between lifetime cannabis use (present study) and lifetime cigarette smoking31 (see Supplementary Information 2).

<table>
<thead>
<tr>
<th>Table 1. Discovery and replication sample characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample</strong></td>
</tr>
<tr>
<td>-------------</td>
</tr>
<tr>
<td>Discovery</td>
</tr>
<tr>
<td>ALS PAC</td>
</tr>
<tr>
<td>BLTS</td>
</tr>
<tr>
<td>CADD</td>
</tr>
<tr>
<td>EGCUT1</td>
</tr>
<tr>
<td>EGCUT2</td>
</tr>
<tr>
<td>FinnTwin</td>
</tr>
<tr>
<td>HUVH</td>
</tr>
<tr>
<td>MCTFR</td>
</tr>
<tr>
<td>NTR</td>
</tr>
<tr>
<td>QIMR</td>
</tr>
<tr>
<td>TRAILS</td>
</tr>
<tr>
<td>Utrecht</td>
</tr>
<tr>
<td>Yale Penn EA</td>
</tr>
<tr>
<td>Replication</td>
</tr>
<tr>
<td>Radar</td>
</tr>
<tr>
<td>SYS</td>
</tr>
<tr>
<td>Twins UK</td>
</tr>
<tr>
<td>Yale Penn EA</td>
</tr>
</tbody>
</table>

Abbreviations: N, sample size; N SNPs, number of SNPs used for the meta-analysis; SNP, single-nucleotide polymorphism; % female, percentage of females; % users, percentage of users that ever used cannabis.
RESULTS

Meta-analysis

No genome-wide significant associations between individual SNPs and lifetime cannabis use were observed (see Manhattan plot, Supplementary Figure 1a). However, the QQ plot (Supplementary Figure 1b) reveals strong enrichment of SNPs with $P < 10^{-4}$. Supplementary Figures 2a–m and 3a–m illustrate the Manhattan and QQ plots for each sample. Table 2 illustrates the top 10 independent ($R^2 < 0.1$) SNPs associated with lifetime cannabis use. None of these 10 SNPs were replicated in the four independent replication samples (Supplementary Table 3). In a combined meta-analysis of the 10 top SNPs (that is, discovery plus replication samples), none of the SNPs reached genome-wide significance. Local plots of the most strongly associated regions, including neighboring genes, are provided in Supplementary Figures 4a–j. The most statistically significant marker ($P = 4.6 \times 10^{-7}$) was rs4984460 located on chromosome 15 (see Supplementary Figure 5 for the forest plot). The SNP is located in an intergenic region between LOC400456/LOC145820 and NR2F2 and MIR1469 genes. Supplementary Table 4 includes the 153 SNPs identified with $P$-values $< 10^{-5}$. Because not all SNPs passed the post-imputation quality control steps in all the samples, this table includes the effective sample size per SNP.

Gene-based tests

The gene-based tests of associations were run on 24,576 genes/genetic regions (see Supplementary Table 6). Stricter LD pruning (that is, $r^2 < 0.05$), or restricting analyses to SNPs present in all studies substantially decreased the estimate of variance explained. Both SNP heritability estimates confirmed that lifetime cannabis use has a significant heritable component (13–20%), indicating that GWAS should be able to identify these common SNPs (but effect sizes are small and large sample sizes are thus required). However, because these estimates are only based on common SNPs, the total heritability of lifetime cannabis use is likely to be higher.

The LD score regression analyses revealed a strong and highly significant genetic correlation ($r_g = 0.83$, s.e. = 0.15, $P = 1.85 \times 10^{-40}$) between lifetime cannabis use and lifetime cigarette smoking (based on the Tobacco Alcohol and Genetics consortium31 summary results), implying that SNPs for lifetime cannabis use and lifetime cigarette smoking are highly correlated.

**DISCUSSION**

To date, this is the largest GWA study of lifetime cannabis use. We performed meta-analysis of the GWA results based on a discovery sample comprising 32,330 individuals from 13 cohorts, and a replication sample comprising 5627 subjects from four cohorts (including one African American cohort). There were no genomewide significant SNP associations. However, heritability analyses revealed that between 13 and 20% of the variation in lifetime cannabis use could be explained by common SNPs. Moreover, allele is associated with a decreased risk of lifetime cannabis use. The forest plot for two SNPs with lowest $P$-values in the other significant gene regions can be found in Supplementary Figure 5.

Of the five genes included in our replication analyses, none were replicated in two of the independent replication samples (see Table 3). In the African American replication sample, suggestive associations with SCOC-AS1 ($P = 0.044$) and SCOC ($P = 0.027$) were found.

SNP-based heritability and genetic overlap with lifetime cigarette smoking

Using the density estimation method (see ‘Materials and Methods’ section for a description), all the SNPs available in at least 25% of the samples when combined explained 20% of the total variance in lifetime cannabis use ($P < 0.001$). Alternative estimation with LD score regression also yielded a significant heritable component of $13\% (h^2_{LD} = 0.13$, s.e. = 0.02, $P = 1.4 \times 10^{-7}$). These variance estimates were robust across pruned sets with similar $r^2$ thresholds (see Supplementary Table 6). Stricter LD pruning (that is, $r^2 = 0.05$), or restricting analyses to SNPs present in all studies substantially decreased the estimate of variance explained. Both SNP heritability estimates confirmed that lifetime cannabis use has a significant heritable component (13–20%), indicating that GWAS should be able to identify these common SNPs (but effect sizes are small and large sample sizes are thus required). However, because these estimates are only based on common SNPs, the total heritability of lifetime cannabis use is likely to be higher.

The LD score regression analyses revealed a strong and highly significant genetic correlation ($r_g = 0.83$, s.e. = 0.15, $P = 1.85 \times 10^{-40}$) between lifetime cannabis use and lifetime cigarette smoking (based on the Tobacco Alcohol and Genetics consortium31 summary results), implying that SNPs for lifetime cannabis use and lifetime cigarette smoking are highly correlated.

**Table 2.** Top 10 SNPs with meta-analysis results of discovery samples, and results of combined discovery and replication samples

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>BP (hg19)</th>
<th>A1</th>
<th>A2</th>
<th>Freq</th>
<th>Beta (s.e.)</th>
<th>Discovery</th>
<th>Direction</th>
<th>Combined&lt;sup&gt;a&lt;/sup&gt;</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4984460</td>
<td>15</td>
<td>96424399</td>
<td>T</td>
<td>G</td>
<td>0.75</td>
<td>−0.11 (0.23)</td>
<td>$4.6 \times 10^{-7}$</td>
<td>++ + + + + + + + +</td>
<td>−0.11 (0.023)</td>
<td>$2.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs2099149</td>
<td>12</td>
<td>30479358</td>
<td>T</td>
<td>G</td>
<td>0.81</td>
<td>−0.16 (0.032)</td>
<td>$9.8 \times 10^{-7}$</td>
<td>−− −− −− −− −− −− −−</td>
<td>−0.17 (0.034)</td>
<td>$5.1 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs7675351</td>
<td>4</td>
<td>141218757</td>
<td>A</td>
<td>C</td>
<td>0.86</td>
<td>−0.15 (0.031)</td>
<td>$1.4 \times 10^{-6}$</td>
<td>−− −− −− −− −− −− −−</td>
<td>−0.13 (0.033)</td>
<td>$1.1 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs4471463</td>
<td>11</td>
<td>112983595</td>
<td>T</td>
<td>C</td>
<td>0.55</td>
<td>−0.09 (0.020)</td>
<td>$1.5 \times 10^{-6}$</td>
<td>−− −− −− −− −− −− −−</td>
<td>−0.1 (0.021)</td>
<td>$9.0 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs7107977</td>
<td>11</td>
<td>915764</td>
<td>A</td>
<td>G</td>
<td>0.60</td>
<td>0.27 (0.058)</td>
<td>$1.9 \times 10^{-6}$</td>
<td>???? + + + + + + + +</td>
<td>0.29 (0.064)</td>
<td>$6.4 \times 10^{-7}$</td>
</tr>
<tr>
<td>rs86961539</td>
<td>2</td>
<td>52753909</td>
<td>G</td>
<td>T</td>
<td>0.91</td>
<td>−0.09 (0.062)</td>
<td>$2.1 \times 10^{-6}$</td>
<td>−−−− −−−− −−−− −−−−</td>
<td>−0.09 (0.062)</td>
<td>$2.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs2033867</td>
<td>2</td>
<td>175188281</td>
<td>A</td>
<td>C</td>
<td>0.06</td>
<td>0.24 (0.051)</td>
<td>$2.6 \times 10^{-6}$</td>
<td>???? + + + + + + + +</td>
<td>0.23 (0.050)</td>
<td>$4.2 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs35053471</td>
<td>3</td>
<td>47124761</td>
<td>A</td>
<td>T</td>
<td>0.38</td>
<td>−0.10 (0.022)</td>
<td>$2.7 \times 10^{-6}$</td>
<td>−−−− −−−− −−−− −−−−</td>
<td>−0.09 (0.022)</td>
<td>$9.2 \times 10^{-5}$</td>
</tr>
<tr>
<td>rs12518098</td>
<td>5</td>
<td>60864467</td>
<td>C</td>
<td>G</td>
<td>0.68</td>
<td>0.10 (0.022)</td>
<td>$2.6 \times 10^{-6}$</td>
<td>???? + + + + + + + +</td>
<td>0.09 (0.023)</td>
<td>$4.7 \times 10^{-6}$</td>
</tr>
<tr>
<td>rs73067624</td>
<td>1</td>
<td>196333461</td>
<td>T</td>
<td>C</td>
<td>0.90</td>
<td>−0.10 (0.039)</td>
<td>$3.1 \times 10^{-5}$</td>
<td>−−−− −−−− −−−− −−−−</td>
<td>−0.16 (0.041)</td>
<td>$6.3 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

Abbreviations: A1, allele 1; A2, allele 2; BP (hg19), location in base pairs in human genome version 19; Chr, chromosome; Freq A1, frequency of allele 1; SNP, single-nucleotide polymorphism. The combined sample contains the discovery samples and the Radar, SYS and TwinsUK replication samples. Direction per sample: allele A1 increases (+) or decreases (−) liability for cannabis use, or sample did not contribute to this SNP because it did not pass the post-imputation quality control (?). Order of samples: ALSPAC, BLTS, CADD, EGCG1, EGCG2, FinnTwin, HUVH, MCTFR, NTR, QIMR, TRAILS, Utrecht, Yale Penn EA. Sample information can be found in Table 1. SNPs are displayed when not in linkage disequilibrium ($R^2 < 0.1$). For SNPs with $R^2 \geq 0.1$, only the most significant SNP is shown in the top 10.)
gene-based tests of association identified four protein-coding genes and one intergenic region significantly associated with lifetime cannabis use including NCAM1, which has previously been linked to substance use.45–48 Finally, we revealed that the genetic liability to lifetime cannabis use correlated to a large extent (r = 0.83) with the genetic liability to lifetime cigarette smoking. Our results are consistent with the hypothesis that lifetime cannabis use is a highly polygenic trait, comprising many SNPs each with small effects contributing to lifetime risk. Moreover, portions of the genetic risk in lifetime cannabis use likely correlates with other substances including cigarette smoking.

Our top gene associated with lifetime cannabis use was NCAM1, a known candidate for nicotine dependence.45 The role of NCAM1 is to regulate pituitary growth hormone secretion as a membrane-bound glycoprotein that mediates cell–cell contact by hemophilic interactions.46 NCAM1 is part of the NCAM1–TTC12–ANKK1–DRD2 (NTAD) gene cluster, which is related to neurogenesis and dopaminergic neurotransmission. Importantly, the NTAD cluster

Figure 1. The Manhattan (a) and the QQ plot (b) based on results of the gene-based analysis performed in the discovery sample using HYST (hybrid set-based test).

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr</th>
<th>Start position (hg19)</th>
<th>BP length</th>
<th>N SNPs</th>
<th>Nominal P-values discovery</th>
<th>Corrected P-values discovery</th>
<th>Nominal P-values discovery</th>
<th>Nominal P-values replication African Americans</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCAM1</td>
<td>11</td>
<td>112831968</td>
<td>303 952</td>
<td>400</td>
<td>6.26 × 10^{-7}</td>
<td>0.015</td>
<td>0.381</td>
<td>0.302</td>
</tr>
<tr>
<td>CADM2</td>
<td>3</td>
<td>85008132</td>
<td>1115 448</td>
<td>978</td>
<td>2.13 × 10^{-6}</td>
<td>0.026</td>
<td>0.744</td>
<td>0.112</td>
</tr>
<tr>
<td>SCOC-AS1</td>
<td>4</td>
<td>141204879</td>
<td>89 668</td>
<td>81</td>
<td>5.76 × 10^{-6}</td>
<td>0.046</td>
<td>0.681</td>
<td>0.044</td>
</tr>
<tr>
<td>SCOC</td>
<td>4</td>
<td>141264614</td>
<td>39 097</td>
<td>111</td>
<td>7.85 × 10^{-6}</td>
<td>0.046</td>
<td>0.636</td>
<td>0.027</td>
</tr>
<tr>
<td>KCNT2</td>
<td>1</td>
<td>196194909</td>
<td>382 653</td>
<td>237</td>
<td>9.38 × 10^{-6}</td>
<td>0.046</td>
<td>0.269</td>
<td>0.201</td>
</tr>
</tbody>
</table>

Abbreviations: BP length, base pair length; chr, chromosome; hg19, human genome version 19; N SNPs, number of SNPs used for the meta-analysis; SNP, single-nucleotide polymorphism.
has been associated with smoking behavior and nicotine dependence,\textsuperscript{35,47–52} alcohol dependence,\textsuperscript{53,54} heroin dependence,\textsuperscript{55} as well as other substance use disorders.\textsuperscript{56} Although it is plausible that NCAM1 is capturing pleiotropic risks underpinning the liability to licit and illicit substance use in general, we note that NCAM1 was not identified either by the Tobacco Alcohol and Genetics consortium or other consortia for cigarette smoking.\textsuperscript{30–32} The functions of the putative variants responsible for the associations in the candidate-gene studies remain to be determined.

The second gene, CADM2, is a synaptic cell adhesion molecule (SynCAM family) belonging to the immunoglobulin (Ig) superfamily. Variants in the CADM2 gene have been previously associated with body mass index,\textsuperscript{56} processing speed\textsuperscript{57} and autism disorders.\textsuperscript{58} Interestingly, these phenotypes were associated with cannabis use in previous studies.\textsuperscript{59–61} which together suggest that CADM2 can be considered an important gene related to a variety of complex traits. It is possible that the association with lifetime cannabis use may be driven, for example, by differences in personality rather than as a direct relationship with lifetime use.

The third gene, SCOC, encodes a short coiled-coil domain-containing protein that localizes to the Golgi apparatus. Many coiled-coil-type proteins are involved in important biological functions such as the regulation of gene expression through the regulation of transcription factor binding.\textsuperscript{62} The function of SCOC is largely unknown and no previous association studies have linked SCOC to cannabis or other substance use phenotypes. The SCOC antisense RNA1 gene is located in the same chromosomal region.

Finally, KCNT2 encodes a potassium voltage-gated channel (subfamily S, member 2). The sodium-activated potassium channels Slack and Slick are encoded by KCNT1 (potassium channel, subfamily T, member 1) and KCNT2, respectively, which are found in neurons throughout the brain. Suggestive association for SNPs near KCNT2 have previously been found for cocaine dependence and for early-onset, highly comorbid, heavy opioid use.\textsuperscript{63,64} This suggests that potassium signaling may have a role in addiction.

The lack of genome-wide significant associations for individual SNPs is consistent with previous GWA studies of lifetime cannabis use\textsuperscript{26,27} and cannabis dependence.\textsuperscript{25} The difficulty of identifying specific SNPs for lifetime cannabis use may be attributable to several reasons. First, complex traits are known to be influenced by many variants, each with very small effect sizes. Although power calculations reveals suitable power (96%) to detect odds ratios of 1.15 based on common SNPs (MAF = 0.2), the power to detect smaller effect sizes remains lower. For example, there is only 28% power to detect effect sizes with odds ratio of 1.1 and MAF = 0.2. Therefore, our data suggest that the effect sizes of single variants contributing to lifetime cannabis use are likely to be smaller than 1.15. Combining variants within larger units (that is, genes) did however reveal four significant genes associated with lifetime cannabis use implying that these genes are appropriate targets for future functional studies of cannabis use. Unfortunately, our gene-based results were not replicated in the replication samples, probably due to low sample sizes and therefore low power. In the African American replication sample, we did find suggestive association with SCOC-AS1 and SCOC.

On the basis of twin studies, the heritability of lifetime cannabis use is estimated at 40–50\%.\textsuperscript{12} In our study, all common SNPs combined explained 13–20% of the variance in the liability to use cannabis depending on the method used. Stricter LD pruning (that is, \( r^2 = 0.05 \)) or restricting to SNPs observed (genotyped or imputed) in all the analyses, substantially reduces the estimate of variance explained. Speculatively, this may indicate that much of the variance explained comes from SNPs located in the regions of weak LD. Such effects are likely to be poorly tagged for the estimation of variance explained after strict LD pruning, and are likely to be more difficult to impute owing to a lack of strongly correlated genotyped SNPs (and thus missing from some studies).

Our SNP-based heritability estimates lie in between two previous heritability estimates for lifetime cannabis use based on the Genome-wide Complex Trait Analysis\textsuperscript{65} software package. Verweij et al.\textsuperscript{26} estimated that 6% of the variance in lifetime cannabis use is explained by aggregated common SNPs (MAF > 0.05). Minică et al.\textsuperscript{27} found an estimate of 25%. Provided that the current sample is much larger than the samples used in the previous studies, we conclude that approximately one-third to half of the heritability is explained by common SNPs captured on a GWAS array. Other sources of variation may explain the discrepancy between SNP- and twin-based heritability estimates. For example, age-related genetic differences, non-additive genetic variance, interactions between genetic variants and environmental risk factors, epistasis and/or rare mutations may also have a role.

Our results indicate a very high genetic overlap \((r = 0.83)\) between our measure of lifetime cannabis use and lifetime cigarette use when based on the SNP panel. Twin studies have shown moderate to high genetic correlations of 0.59–0.74 between lifetime cannabis and nicotine use.\textsuperscript{66} Kendler et al.\textsuperscript{67} also reported significant biometrical genetic correlations between the levels of cannabis, nicotine and alcohol use, which were increasingly influenced by common genetic risks detectable in early adulthood.

Our findings should be interpreted in the context of at least four potential limitations. First, our study was underpowered to detect very small effects of individual variants. Power analyses revealed that a twofold increase in sample size is required to detect SNP effect sizes with odds ratios of 1.1. Second, lifetime cannabis use is a dichotomous measure combining single lifetime, regular and chronic users. Consequently, our sample may compromise heterogeneous patterns of use, which has the potential to reduce the power to detect genetic association.\textsuperscript{68} Third, prevalences of lifetime cannabis use varied between 1\% (EGCUT1) and 92\% (Yale Penn EA). This was likely due to differences in the sample characteristics, recruitment strategies and the political differences between countries. Despite these differences, the forest plots of the key SNPs (see Figure 2; see also Supplementary Figure 5) revealed that the 95\% confidence intervals surrounding the effect estimates typically included the estimated meta-analytic effect, which tends to overlap across studies. This indicates that the input
samples were representative of the same population of users. Finally, the average age of participants varied between 18 (ALS PAC) and 45 (QIMR) years. Consequently, some younger participants might have initiated cannabis use at a later age, but have been classified as 'never users' in the current study. This can decrease power, but does not invalidate our results. In addition, we note that the average age of each sample did not correlate with sample prevalences ($r = -0.04, P = 0.91$).

On the basis of our observations, the following recommendations for future studies can be made. We have identified four genes significantly associated with cannabis use, which are candidates for follow-up functional studies. In particular, the role of $NCAM1$ can be examined to determine the functional role of this gene, possibly in combination with other genes in the same gene cluster ($NCAM1$–$TTC12$–$ANKK1$–$DRD2$).

The next goal of the International Cannabis Consortium is to perform a meta-analysis on GWAS studies investigating the age at first cannabis use. Our rationale is based on the observation that early initiation of cannabis use is associated with rapid progression towards cannabis abuse and dependence, polysubstance use and other substance use disorders.69–71 Methods other than GWASs may also be used to reveal the biological pathways of cannabis use, such as rare variant association analyses. The environmental risk factors may be incorporated to investigate gene × environment interactions. Hopefully, the combination of advanced genetic studies and novel statistical approaches with larger samples will further contribute to our understanding of the genetic architecture of cannabis use.

CONCLUSION

We have performed the largest meta-analysis to date of GWASs investigating cannabis use phenotypes. With a sample of over 32 000 individuals, our results implicate four genes as involved in lifetime cannabis use: $NCAM1$, $CADM2$, $SCOC$ and $KCNT2$. Our results illustrated that lifetime cannabis use is under the influence of many common genetic variants. The combined SNPs explained 13–20% of the phenotypic variation, and revealed a high degree of genetic sharing ($r = 0.83$) with lifetime cigarette smoking. Future studies should investigate the impact of these genes on the biological mechanisms leading to lifetime cannabis use.

CONFLICT OF INTEREST

HRK is a consultant or Advisory Board Member for Alkermes, Lilly, Lundbeck, Otsuka, Pfizer, Roche; member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, supported by AbbVie, Ethypharm, Lilly, Lundbeck and Pfizer. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

JNW, CCM and HM are supported by the European Research Council (Beyond the Genetics of Addiction ERC-284167, PI JMW). SS and EMD are supported by the Foundation Völkbond Rotterdam. KJHV is supported in part by the Netherlands Organization for Health Research and Development (ZonMW 31160212) and in part by a 2014 NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation. NAG is supported by US National Institutes of Health, National Institute on Drug Abuse R00DA023549. GHL is supported by NIDA R37 DA-018673 and NSF BCS-1229450. RW is supported by NIH U01 MH094432 and NSF BCS-1229450. Statistical analyses were 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. The study site acknowledgments are as follows: ALS PAC—We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALS PAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, researchers, and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/12/3) and the University of Bristol provide core support for ALS PAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corportation of America) using support from 23andMe. AIS is supported by a Wellcome Trust 4-year PhD studentship in molecular, genetic and lifestyle epidemiology (WT083431MA). JNW is supported by a Postdoctoral Research Fellowship from the Oak Foundation. JNW and MRM are members of the MRC Integrative Epidemiology Unit at the University of Bristol, funded by the UK Medical Research Council (MC_UU_120136/3) and the University of Bristol. JNW and MRM are also members of the UK Centre for Tobacco & Alcohol Studies, a Research Councils UK Centre of Excellence. Funding from British Heart Foundation, Cancer Research UK, Economic and Social Research Council, Medical Research Council and the National Institute for Health Research, under the auspices of the UK Clinical Research Collaboration is gratefully acknowledged. BLTS—The BLTS was supported by grants from the United States National Institute on Drug Abuse (R00DA23549) awarded to NAG, by the Australian Research Council to MJW (Nos. DP0343921, DP0664638 and DP1093900) and by Australian National Health and Medical Research Council. The Australian Government through the National Health and Medical Research Council awarded to Ian Hickie (No. 464914) and GWM (No. 619667). We thank and acknowledge the following project staff: AKH, Leanne Wallace and Lisa Bowdler for the laboratory processing, genotyping and QC; Saod Hancock as Project Coordinator; Lenore Sullivan as Research Editor; our research interviewers Pieta-Marie Shertock and Jill Wood; and David Smyth for IT. We also thank the twins and their siblings for their willing cooperation. SEM was supported by an ARC future fellowship FT110100548. CADD—The Center on Antisocial Drug Dependence (CADD) data reported here were funded by grants from the National Institute on Drug Abuse (P60 DA010115, R01 DA012845, R01 DA021913, R01 DA035804). EGCUT—received funding by FP7 grants (278913, 306031, 313010), Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GVarendra). We acknowledge EGCUT technical personnel, especially Mr V Soo and S Smit. Data analyses were carried out at the High Performance Computing Center of Tartu. FinnTwin—We warmly thank the participating twin pairs and their families for their contribution. We express our appreciation to the skilled study interviewers A-M Livonen, K Karhu, H-M Kuha, U Kulmala-Grähn, M Mantere, K Saanakorpi, M Saarinen, R Ripila, L Viljanen and E Voipio. Anja Häppölä and Kauko Heikkilä are acknowledged for their valuable contribution in recruitment, data collection and data management. Phenotyping and genotyping of the Finnish twin cohorts has been supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 213506, 129680), the Academy of Finland (grants 100499, 205585, 118555, 141054, 265240, 263278 and 264146 to JK), National Institute of Alcohol Abuse and Alcoholism (grants AA-12352, AA-00145, and AA-09203 to RJR and AA-15416 and K02AA018755 to DM Dick), Sigrid Juselius Foundation (to JK), and the Welcome Trust Sanger Institute, UK. Anti-Pelka Sarin and Samuli Ripatti are acknowledged for genotype data quality controls and imputation. GWAS analyses were run at the ELIXIR Finland node hosted at CSC—IT Center for Science for ICT resources. HUHV—Financial support was received from ‘Instituto de Salud Carlos III’ (F11/00571, P11/01629, P11/011339, ‘Plan Nacional Sobre Drogas’ (PNSD2011-0080), ‘Agencia de Gestión de Ayudas Universitarias i de Recerca-AGAUR, Generalitat de Catalunya’ (2014SGR1537) and ‘Departament de Salut’, Government of Catalonia, Spain. MR is a recipient of a Miguel de Servert contract from the ‘Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación’, Spain. MCTFR—This research was supported in part by USPHS Grants from the National Institute on Alcohol Abuse and Alcoholism (AA09367 and AA111886), the National Institute on Drug Abuse (DA05147, DA13240 and DA024417) and the National Institute on Mental Health (MH066140).

NTR—We thank the Netherlands Twin Register participants whose data we analyzed in this study. This work was supported by grants from the Netherlands Organization for Scientific Research (ZonMW Addiction 31160008; ZonMW 940-37-024; NWO/SPI 56-464-14192; NWO-400-05-717; NWO-MW 904-61-19; NWO-MagW 480-40-004; NWO-Veni 016-115-035), the European Research Council (Beyond the Genetics of Addiction ERC-284167; Genetics of Mental Illness: ERC-230374), the Centre for Medical Systems Biology (NWO Genomics), Netherlands Bioinformatics Center/ BioAssist/RK/2008.002. We acknowledge the EMGO+ Institute for Health and Care Research, the Neuroscience Campus Amsterdam, BBMRI-NL (184.021.007: Biobanking and Biomolecular Resources Research Infrastructure), the Avera Institute, Sioux Falls, South Dakota (USA) for support. Genotyping was funded in part by grants from the National Institutes of Health (4R37DA018673-06, RC2 MH089951), Rutgers University Cell and DNA Repository cooperative agreement (National Institute of Mental Health U24 MH068457-06), and the National Institutes of Health (NH R01 HD02157-01A, MCHD000138 and OPP0188951) and the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health. The statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is supported by the Netherlands Scientific Organization (NWO 480-05-003), the Dutch Brain Foundation and the Department of Psychology and Education of the VU University Amsterdam. QIMR—This is supported by National Institutes of Health AA078775, AA14999, AA13320, AA14961, AA19927-01, AA19985, DA012854, DA018267, DA018660, DA23668 and DA019951; by grants from the Australian National Health and Medical Research Council (241994, 339462, 389927, 389875, 389891, 389938, 442915, 442981, 496739, 552485, 552498 and 628911); by...
Grants from the Australian Research Council (A7960034, A79906588, A79801419, DP0770096, DP0212016 and DP0343921); and by the 5th Framework Programme (FP-5) GenomEUtwin Project (QLG2-CT-2002-01254). This research was further supported by the Centre for Research Excellence on Suicide Prevention (CRESPP—Australia). We thank AKH, Richard Parker, Soad Hancock, Judith Mr, Sally Rodda, Pietar-Maree Shorck, Heather Park, Jill Wood, Pam Barton, Fran Husband, Adele Somervielle, Ann Eldridge, Marina Greene, Sophia McPhedran, Lisa Songe, Alexey Todorov, Steven Cooks, David Smyth, Harry Bæby and Daniel Park. Last, we thank the twins and their families for their participation. RADAR—We thank all the adolescents and their families and friends for their participation. Moreover, we thank the various assistants that helped in recruiting participants as well as collecting and cleaning the data. The research was funded partly by the Netherlands Organisation for Scientific Research (Brain and Cognition, 056-21-010). Data of the RADAR study were used. RADAR has been financially supported by main grants from the Netherlands Organisation for Scientific Research (GB-MAGW 480-03-005), and Stichting Achmea Slachtoffener en Sammelvinding (SASS) and various other grants from the Netherlands Organisation for Scientific Research, the VU Amsterdam and Utrecht University. ACH is supported by the Netherlands Organisation for Health Research and Development, ZonMW 31160212. Saguennay Youth Study—The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the SYS (TP, ZP). TP is the Tanenbaum Chair in Population Neuroscience (University of Toronto) and the Dr John and Consuela Phelan Scholar (Child Mind Institute). TRAILS—TRAILS (Tracking Adolescents’ Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen and the Pamiass Bavo, all in the Netherlands. TRAILS has been financially supported by grants from the Netherlands Organisation for Scientific Research NWO (Medical Research Council program grant GB-MW 940-38-011; ZonMW Brainpower grant 100-001-004; ZonMW Risk Behavior and Dependence grant 60-06060-97-118; ZonMW Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MAGW 480-01-006 and GB-MAGW 480-07-001; Social Sciences Council project grants GB-Magw 452-04-314 and GB-Magw 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013; the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-IL (CP 32), the participating universities and Accare Center for Child and Adolescent Psychiatry. We are grateful to all the adolescents, their parents and teachers who participated in this research and to everyone who worked on this project and made it possible. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. TwinsUK—The study was funded by the Wellcome Trust; European Community’s Seventh Framework Programme (FP7-2007-2013). The study also received support from the National Institute on Drug Abuse (NIDA, U01 DA00647), the National Institute on Alcohol Abuse and Alcoholism (NIAAA, U01 AA01330), and the National Institute on Aging (NIA, P01 AG 12046, R01 AA10677, R01 AA12609, R01 DA18289, R01 AA18322, R01 AA11330, R01 AA017535, and the VA Connecticut and Philadelphia VA MIRECC.

REFERENCES

33 Kendler KS, Karkowski LM, Corey LA, Prescott CA, Neale MC. Genetic and environ-


35 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.


37 Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet 2014; 15: 335–346.


41 So HC, Li M, Sham PC. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genet Epidemiol 2011; 35: 447–456.


69 Grant JD, Lynskey MT, Scherrer JF, Agrawal A, Heath AC, Bucholz KK. A cotwin-control analysis of drug use and abuse/dependence risk associated with early-onset cannabis use. Addict Behav 2010; 35: 35–41.


Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp)