Genome-wide association meta-analysis of age at first cannabis use


ABSTRACT

Background and aims Cannabis is one of the most commonly used substances among adolescents and young adults. Earlier age at cannabis initiation is linked to adverse life outcomes, including multi-substance use and dependence. This study estimated the heritability of age at first cannabis use and identified associations with genetic variants. Methods A twin-based heritability analysis using 8055 twins from three cohorts was performed. We then carried out a genome-wide association meta-analysis of age at first cannabis use in a discovery sample of 24 953 individuals from nine European, North American and Australian cohorts, and a replication sample of 3735 individuals. Results The twin-based heritability for age at first cannabis use was 38% [95% confidence interval (CI) = 19–60%]. Shared and unique environmental factors explained 39% (95% CI = 20–56%) and 22% (95% CI = 16–29%). The genome-wide association meta-analysis identified five single nucleotide polymorphisms (SNPs) on chromosome 16 within the calcium-transporting ATPase gene (ATP2C2) at \( P < 5 \times 10^{-8} \). All five SNPs are in high linkage disequilibrium (LD) \( (r^2 > 0.8) \), with the strongest association at the intronic variant rs1574587 \( (P = 4.09E-09) \). Gene-based tests of association identified the ATP2C2 gene on 16q24.1 \( (P = 1.33e-06) \). Although the five SNPs and ATP2C2 did not replicate, ATP2C2 has been associated with cocaine dependence in a previous study. ATP2B2, which is a member of the same calcium signalling pathway, has been associated previously with opioid dependence. SNP-based heritability for age at first cannabis use was non-significant. Conclusion Age at cannabis initiation appears to be moderately heritable in western countries, and individual differences in onset can be explained by separate but correlated genetic liabilities. The significant association between age of initiation and ATP2C2 is consistent with the role of calcium signalling mechanisms in substance use disorders.

Keywords Age at first use, ATP2C2, cannabis initiation, genome-wide association, heritability, substance use.
INTRODUCTION

Cannabis is one of the most commonly used substances among adolescents and young adults [1]. Annually, approximately 147 million people, or 2.5% of the world’s population, consume cannabis. In the last decade, cannabis use disorders have grown more rapidly than either cocaine or opiate use disorders, with the most rapid growth seen in developed countries in North America, western Europe and Australia [2]. Accompanying these changes, there has also been a global trend towards decreasing age at first cannabis use [3,4].

Globally, younger cohorts are more likely to engage in substance use, including cannabis. In the United States, the mean age at first cannabis use is 18 years, whereas the mean age at first cannabis use among individuals who initiate prior to age 21 is 16 years [1]. European data suggest that age at first cannabis use is lower in countries where prevalence of cannabis use is higher [5]. In addition, the male–female gap observed commonly in older cohorts is closing in more recent cohorts [6,7]. Overall, these trends are due probably to lower risk perception [8] and increased availability due to medicalization and decriminalization.

Early cannabis initiation is linked to a number of maladaptive behaviours. These include educational under-achievement [9,10], possible cognitive decline [11,12], negative life events [13], differences in brain maturation in at-risk adolescents [14], conduct disorder [15], risk-taking behaviours [16], psychosis and other psychopathology [17–20]. Early age at onset of use is also linked to more frequent progression to cannabis misuse and increased likelihood of substance use disorders [21–24].

Despite its widespread use, emerging trends in use and associations with adverse outcomes, very little is known about the genetic aetiology of age at first cannabis use. A meta-analysis of twin studies [25] reported a heritability ($h^2$) of ~45% for life-time cannabis use (ever versus never). In contrast, only a limited number of biometric genetic studies have explored the heritability of age at first cannabis use. In a population-based sample of life-time users, Richmond-Rakerd et al. [26] estimated a non-significant heritability of 19% for age at first cannabis use. Lynskey et al. [27] reported a much larger heritability ($h^2 \approx 80\%$) for early-onset use ($\leq 16$ years), whereas Sartor et al. [28] reported a heritability of 52% when age at first cannabis use was categorized as ‘never’, ‘late’ ($\geq 17$ years) or ‘early’ ($\leq 16$ years). These discrepancies might be due to differences in the biometrical genetic methods employed and the inclusion versus exclusion of never users. To address these limitations, we estimated heritability of age at first cannabis use using three different models to determine if cannabis initiation and age at initiation fall along the same continuum, represent two independent liabilities or two distinct but related liabilities [29].

We are aware of only one genome-wide association study (GWAS) for age at first cannabis use. Minică et al. [30] performed a genome-wide survival analysis in a sample comprising 5148 participants. This study found no single nucleotide polymorphisms (SNPs) or genes associated significantly with age at first cannabis use, due possibly to a lack of statistical power [30]. Because age at first use is likely to be highly polygenic (subjected to the influence of many genetic variants with small effects), identifying genetic variants will require much larger samples than employed previously. The application of survival-based methods [30] is expected to improve statistical power over GWASs limited to cannabis users, or logistic regressions based on samples of users and non-users [31–33]. Therefore, we applied a survival-based approach to nine cohorts from the International Cannabis Consortium (ICC [34]) to detect genetic variants associated with age at first cannabis use.

The ICC was established to identify genetic variants underlying individual differences in cannabis use phenotypes by combining data from numerous cohorts and studies. The ICC has previously identified four genes associated significantly with life-time cannabis use: NCAM1; CADM2; SCOC; and KCNT2 [34]. Interestingly, both NCAM1 and KCNT2 have been linked previously to other substance use phenotypes [34]. Also of note is our novel finding at CADM2, which was associated recently with alcohol consumption [35], personality [36], behavioural reproductive outcomes and risk-taking behaviour [37].

Our aim was to explore the genetic aetiology of age at first cannabis use. First, we performed a biometrical heritability analysis in 8055 twins from three cohorts. Secondly, we performed a GWAS meta-analysis of age at first cannabis use in a discovery sample of 24 953 individuals from nine cohorts from Europe, Australia and the United States. The top findings were tested for replication in a sample of 3735 individuals from three cohorts. The outline of the analyses steps is illustrated in Fig. 1.

MATERIALS AND METHODS

Biometrical heritability

The heritability of age at first cannabis use was estimated based on data from three cohorts: Netherlands Twin Register (NTR), comprising 2027 monozygotic (MZ) and 1771 dizygotic (DZ) twin pairs; QIMR Berghofer Medical Research Institute (QIMR), comprising 1282 MZ and 1969 DZ twin pairs; and Brisbane Longitudinal Twin Study (BLTS), comprising 429 MZ and 577 DZ twin pairs [38]. We applied three models to determine if cannabis initiation and age at initiation fall along the same continuum (single liability), represent two independent liabilities.
(independent model) or two distinct but related liabilities (combined model) [29].

For the best-fitting model, individual differences in liability to early age at initiation of cannabis use were disentangled in additive genetic (A), shared environmental (C) and unshared environmental variation (E) [39] (see Supporting information, Files S2 and S4 for details).

### Study samples

The current discovery meta-analysis was based on genome-wide summary statistics from nine European, North American and Australian cohorts comprising $n = 24,953$ individuals. The mean age ranged from 17.3 to 46.9 years (Table 1). Females represented 53.3% of the sample and 44.4% of the observations were uncensored, i.e. individuals who acknowledged having initiated cannabis use (see Supporting information, Table S1 for more details).

#### Phenotyping

Age at first cannabis use was assessed from questionnaires or clinical interviews (see Supporting information, File S1 for information on the exact phrasing of the question). For individuals who had not initiated cannabis use at the time of the assessment, age at last survey or interview was used. Depending on initiation status, individuals were coded as uncensored (initiated) or censored (did not initiate at the time of the last measurement). Given the young average age of the participating cohorts, we included all available data to maximize sample size, i.e. censored and uncensored observations without imposing age restriction.

### Table 1 Descriptive information on the participating discovery cohorts.

<table>
<thead>
<tr>
<th>Cohort</th>
<th>$n$</th>
<th>% Females</th>
<th>%Uncensored observations</th>
<th>Mean age (SD)</th>
<th>Mean age at first use (SD) (in users)</th>
<th>Number of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALSPAC</td>
<td>6147</td>
<td>51.9</td>
<td>38.4</td>
<td>17.3 (1.7)</td>
<td>14.8 (1.6)</td>
<td>6284749</td>
</tr>
<tr>
<td>BLTS</td>
<td>721</td>
<td>57.1</td>
<td>59.5</td>
<td>26.2 (3.3)</td>
<td>18.8 (2.8)</td>
<td>4093835</td>
</tr>
<tr>
<td>FinnTwin</td>
<td>1029</td>
<td>51.7</td>
<td>27.5</td>
<td>22.8 (1.3)</td>
<td>18.0 (2.5)</td>
<td>4362100</td>
</tr>
<tr>
<td>HUVH</td>
<td>581</td>
<td>31.3</td>
<td>30.3</td>
<td>28.7 (12.5)</td>
<td>16.0 (3.0)</td>
<td>4319651</td>
</tr>
<tr>
<td>NTR</td>
<td>5148</td>
<td>62.3</td>
<td>16.6</td>
<td>46.9 (17.5)</td>
<td>18.9 (5.1)</td>
<td>4773834</td>
</tr>
<tr>
<td>QIMR</td>
<td>6758</td>
<td>53.8</td>
<td>51.3</td>
<td>45.2 (10.9)</td>
<td>19.9 (5.8)</td>
<td>5953917</td>
</tr>
<tr>
<td>TRAILS</td>
<td>1249</td>
<td>53.8</td>
<td>61.7</td>
<td>20.0 (1.6)</td>
<td>16.3 (2.0)</td>
<td>4819504</td>
</tr>
<tr>
<td>Utrecht</td>
<td>958</td>
<td>51.3</td>
<td>59</td>
<td>17.4 (3.2)</td>
<td>15.5 (2.1)</td>
<td>4139839</td>
</tr>
<tr>
<td>Yale-Penn</td>
<td>2362</td>
<td>41.2</td>
<td>92.6</td>
<td>38.2 (10.6)</td>
<td>17.0 (9.4)</td>
<td>5732659</td>
</tr>
</tbody>
</table>

$n =$ Sample size; % uncensored observations (i.e. individuals who have initiated cannabis use). Mean age at first use: mean age at first cannabis use. SD = standard deviation; SNP = single nucleotide polymorphism; ALSPAC = Avon Longitudinal Study of Parents and Children; BLTS = Brisbane Longitudinal Twin Study; FinnTwin = Finnish Twin Cohort Study; HUVH = Hospital Universitari Vall d’Hebron; NTR = Netherlands Twin Register; QIMR = QIMR Berghofer Medical Research Institute; TRAILS = TRacking Adolescents’ Individual Lives Survey.
Genotyping
Genotyping followed by extensive quality control (QC) was performed by each participating cohort (see Supporting information, Table S2 for details). Generally, QC criteria involved removal of SNPs with minor allele frequency (MAF) below 1%, call rates < 90% and Hardy–Weinberg equilibrium (HWE) P-values below 1E-04. SNPs with evidence of poor clustering on visual inspection of intensity plots were also discarded. At the subject level, additional QC criteria involved removal of individuals with low overall call rates, conflicting sex designation or excess autosomal heterozygosity (indicative of genotyping errors). Duplicate samples and unintended first- or second-degree relatives (in samples of unrelated individuals) were removed. In Supporting information, Table S2 the exact QC thresholds used by each cohort can be found.

Imputation
All cohorts performed genotype imputation using the 1000 genomes Phase 1, March 2012 release as reference [40] (see Supporting information, Table S2 for further imputation details). We used best-guess genotypes and restricted analyses to autosomal SNPs.

Quality checks prior to meta-analysis
Prior to the meta-analysis, results for each cohort underwent additional QC pertaining to imputation quality, minor allele frequency and HWE, and only SNPs with high imputation quality (> 0.8) were selected. The average imputation quality for the included SNPs ranged from 0.95 to 0.99 throughout all nine discovery cohorts. Secondly, we retained SNPs with MAF greater than \( \sqrt{5/N} \), where N is the sample size. This ensured that there were at least five individuals in the least frequent genotype group. Thirdly, genotyped SNPs were retained if HWE was not violated (P-value > 1E-04). We also removed SNPs with invalid alleles, or allele frequencies mismatched with the 1000 genomes phase 1 European reference panel (i.e. if the allele frequency difference exceeded \[0.2\]). The discovery meta-analysis included 6 163 759 unique bi-allelic SNPs that passed our QC criteria in at least two cohorts (see Table 1 for the number of SNPs in each input file meeting quality control criteria).

Statistical analysis of individual samples
Cohort-specific analyses were performed using a standardized analysis protocol. Each site performed a Cox proportional hazards regression analysis where age at first cannabis use (or age at the last survey for censored observations) was regressed on the SNP (coded additively codominant as 0, 1, 2) and the following covariates: sex, birth-cohort (to correct for generation effects), the first four principal components (to correct for possible population stratification) and study-specific covariates (to correct for chip and/or batch effects; see Supporting information, Table S2 for details). To account for relatedness in family-based cohorts we used the ‘cluster’ option in the R survival package [41]. This ensured that standard errors were robust to possible misspecification of the familial covariance matrix [42]. The survival package was accessed either directly in R, or called from Plink [43] via the Rserve package [44].

Meta-analysis
The discovery meta-analysis was performed in Metal [45], using a fixed-effects model and the ‘SCHEME STDERR’ option, which weights the beta coefficients by the inverse of their associated standard errors. To ensure that the bulk of the test statistic distribution follows the expectation under a theoretical null model, we applied genomic control to each cohort’s input file prior to meta-analysis. This ensured that none of the input cohorts contributed disproportionately to the meta-analysis results [46]. Similar to the method applied by Furberg et al. [47] and Allen et al. [48], we computed the standard error (and the corresponding P-value) by multiplying the variance of the beta by the lambdaGC (Genomic Control) estimate for each sample (see Supporting information, Table S2). An alpha of 5E-08 was used as the genome-wide significance threshold. Statistical analyses were performed on the Lisa Genetic Cluster computer (http://www.geneticcluster.org).

Gene-based tests of association
Results from the genome-wide meta-analysis were then used to test for gene-based association. We employed the Gene-based Association Test using the Extended Simes procedure (GATES) in the Knowledge-based mining system for Genome-wide Genetic studies (KGG) (version 3.5) [49,50]. GATES combines the P-values of the SNPs within a gene by taking into account the linkage disequilibrium (LD). The SNPs were mapped onto (or within 5 kb) 25 655 genes based on NCBI gene coordinates. LD structure was inferred based on the 1000 genomes haplotypes (version March, 2012). For this analysis, a false discovery rate (FDR) of 0.05 [51] was used as the genome-wide significance threshold.

SNP-based heritability analysis
The proportion of phenotypic variance explained by the retained SNPs was estimated using two different methods. The density estimation (DE) method developed by So et al. [52] 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 (for a detailed overview of the DE method, see [53]). SNPs present in 25% or more of the meta-analysis samples were selected and pruned for LD. We used the $r^2 = 0.15$ pruning level as the primary result for consistency with other applications of this method. The second method used LD score regression analysis [54]. Here, the SNP-based heritability estimate was based only on SNPs present in all cohorts to avoid artefacts resulting from differing $n$s per SNP. In both methods, SNP-based heritability depends upon the relationship between sample size, effect size and the corresponding test statistic. Using a Cox proportional hazards model and applying genomic control affects that relationship. Therefore, we approximated the effective sample size (i.e. the sample size with the intended statistical behaviour for heritability analysis) of the current GWAS (for details see Supporting information, File S3).

Replication analyses

Genes reaching significance and the top eight independent signals in the discovery meta-analysis (present in at least one of the replication samples) were taken forward for replication in a sample of 3735 individuals from three cohorts. In addition, the top SNPs were analysed in the combined discovery and replication samples. Furthermore, we tested whether a polygenic risk score [55] based on the meta-analysis results predicts age at first cannabis use in one of the replication samples (see Supporting information, File S5 for details on the replication analyses). We also evaluated the power to detect a significant association in the replication sample.

RESULTS

Biometrical heritability

The combined model with separate but correlated liabilities provided the best fit to the data (see Supporting information, File S4 for model-fitting details and twin correlations). In this model, the heritability ($A$) of age at first cannabis use was 38% [95% confidence interval (CI) = 19–60%]. Shared ($C$) and unique ($E$) environmental factors explained 39% (95% CI = 20–56%) and 22% (95% CI = 16–29%) of the variance, respectively. $A$, $C$ and $E$ explained 48% (95% CI = 30–65%), 37% (95% CI = 21–52%) and 15% (95% CI = 11–20%), respectively, of the variance in risk of cannabis initiation. We found no evidence for qualitative or quantitative sex differences.

GWAS meta-analysis

The quantile-quantile plot for the fixed effects genome-wide discovery meta-analysis is shown in Supporting information, Fig. S1a. Note that the bulk of the test statistic distribution follows the expectation under a null hypothesis of no association ($\lambda_{GC} = 1$). The test statistic behaved similarly when no genomic control was applied (see Supporting information, Fig. S1b). These results indicate that the meta-analysis is robust to slight deviations of the test statistic distribution from the theoretical null model observed in some of the cohorts. Supporting information, Figs S2a–i and S3a–i show cohort-specific lambda-corrected Manhattan and quantile-quantile plots.

The Manhattan plot in Fig. 2a displays the genome-wide association results. One region on chromosome 16 passed the significance threshold of $P < 5E-08$, with other suggestive signals on chromosomes 6, 10 and 14. Table 2 includes association results and details on the top eight independent SNPs. The top 100 SNPs in the discovery sample are shown in Supporting information, Table S3. Regional association plots and forest plots for the top SNPs are shown in Supporting information, Figs S4a–I, and S5a–k.

The genome-wide significant signals come from a set of six highly correlated SNPs on chromosome 16 ($r^2 > 0.8$) located within the calcium-transporting ATPase (ATP2C2) gene. The strongest predictor of age at onset of cannabis use was rs1574587 (yielding the lowest $P$-value, $P = 4.09E-09$). rs1574587 reached statistical significance regardless of whether or not GC was applied ($P = 1.08e-08$). This SNP has a MAF ranging from 0.105 to 0.185 throughout the discovery samples (commensurate with MAFs reported for European ancestry populations by Ensemble) and an imputation quality $\geq 0.89$ (see Supporting information, Table S4a for more details on this SNP).

The $I^2$ statistic for the top SNP was 32.6% ($I^2 = 10.38$, $P = 0.16$), indicating no evidence of between-cohort heterogeneity in the observed effect. Indeed, the top SNP showed the same direction of the effect in all but one of the discovery cohorts (Fig. 2b).

Gene-based tests of association

Figure 3 provides an overview of the gene-based results. The quantile-quantile plot (Supporting information, Fig. S6) shows that the bulk of the test statistic distribution follows the expectation under the null hypothesis and that several genomic regions are enriched for small $P$-values. Coding genic regions, and not non-coding regions, were enriched for SNPs that yielded strong association signals in the single variant analysis (Supporting information, Fig. S6).

As shown in the Manhattan plot in Fig. 3a, the calcium-transporting ATPase (ATP2C2) gene on chromosome 16 reached the FDR threshold of 0.05 in the gene-based tests of association (nominal $P = 1.33E-06$, $I^2 = 10.38$, $P = 0.16$), indicating no evidence of between-cohort heterogeneity in the observed effect. Indeed, the top SNP showed the same direction of the effect in all but one of the discovery cohorts (Fig. 2b).
corrected $P = 0.034$). See Supporting information, Table S5 for the top 100 genes identified in the discovery meta-analysis and Fig. 3b for the zoom plot of the significant gene.

ATP2C2 is located at 16q24.1 (Fig. 3b) in the vicinity of KCNG4 and COTL1. This gene was also identified in the SNP-based analysis and the top SNP rs1574587 is located in this gene. According to the Gene Ontology annotations [56,57] the ATP2C2 gene is involved in calcium-transporting ATPase activity, calcium ion transmembrane transport, ATP binding and metal ion binding.

SNP-based heritability analyses

The selected SNPs did not contribute significantly to the variance in age at first cannabis use according to either the density estimation method ($h^2 = 0.056; P = 0.29$) or the LD score regression analysis ($h^2 = 0.036; P = 0.22$).

Figure 2 The Manhattan plot of the meta-analysis results for the discovery sample (a) In the Manhattan plot, the y-axis shows the strength of association [−log$_{10}(P)$] and the x-axis indicates the chromosomal position. The blue line indicates suggestive significance level ($P < 1E-05$), while the red line indicates genome-wide significance level ($P < 5E-08$); (b) forest plot of the top SNP (rs1574587) on chromosome 16 in eight discovery cohorts. [Colour figure can be viewed at wileyonlinelibrary.com]
The ATP2C2 gene is associated with language impairment learning [60]. Several studies showed that variation in which, in turn, regulates synaptic plasticity, memory and brain [58] and is involved in calcium homeostasis [59] gene networks associated significance. Gelernter findings was caused probably by insufficiency include ATP2C2-dependent replication sample to replicate the discovery was observed for the intronic variant rs1574587. The replication analyses

To our knowledge, this is the largest biometrical and molecular genetic study investigating the genetic etiology of age at first cannabis use. The biometrical twin analysis of 8055 twin pairs showed that genetic factors explain 38% of the variance in age at first cannabis use (95% CI = 19–60). The discovery genome-wide meta-analysis identified significant associations with five highly correlated SNPs within the calcium-transporting ATPase gene (ATP2C2) on chromosome 16. The strongest association was observed for the intrinsic variant rs1574587. The gene-based tests provided further evidence linking ATP2C2 to age at first cannabis use. The failure of the smaller independent replication sample to replicate the discovery findings was caused probably by insufficient statistical power.

The top associated ATP2C2 gene is expressed in the brain [58] and is involved in calcium homeostasis [59] which, in turn, regulates synaptic plasticity, memory and learning [60]. Several studies showed that variation in the ATP2C2 gene is associated with language impairment (e.g. [61]). ATP2C2 has also been linked to cocaine dependence. Gelernter et al. [62] found that the highest ranked gene networks associated significantly with cocaine dependence include ATP2C2 along with ATPase, Ca2+-transporting and the plasma membrane gene (ATP2B2). Noteworthy is that calcium signalling pathways have also been implicated in opioid dependence [63]. These findings are consistent with observed associations between early-onset of cannabis use and experimentation with other drugs [64] and progression to escalated use/dependence [27]. It is therefore highly plausible that some of the same genetic factors increase the probability of early initiation of substance use and progression to substance use disorders (see, e.g. [65,66]). Taken together, the effects of ATP2C2 are likely to be general rather than substance-specific.

Early age at first cannabis use may be a predictor for more severe phenotypes, such as substance use disorder, and externalizing behaviours, such as conduct disorder. Indeed, we know from previous work that there is high comorbidity between conduct disorder and use of cannabis and other substances (e.g. [67]) and twin studies have shown that part of the covariation is due to overlapping genetic influences [68–70]. It is therefore plausible that genes for age at first cannabis use also play a role in the broader spectrum of externalizing behaviour. Unfortunately, existing GWASs of conduct and antisocial behaviour have not been powered sufficiently to identify genes associated robustly with these behaviours [71,72]. However, using the combined effect of all SNPs, Tielbeek et al. [72] showed a significant genetic correlation between antisocial behaviour and lifetime cannabis use ($r_g = 0.69, P = 0.016$).

The SNP-based heritability for age at first cannabis use was non-significant. Moreover, the polygenic risk score based on a small selection of genotyped SNPs present in at least seven cohorts provided no evidence of association with age at first use of cannabis in the replication sample ($n = 2082, P > 0.10$). These null findings suggest that common SNPs explain a relatively small proportion of total heritability in age at first cannabis use. The difference between the biometric ‘family-based’ and the ‘SNP-based’ heritability estimates suggests that a large proportion of

---

**Table 2** Top eight independent single nucleotide polymorphisms (SNPs) in the meta-analysis of the discovery samples (present in at least one replication sample). SNPs are displayed when not in linkage disequilibrium ($r^2 < 0.1$. For SNPs with $r^2 > 0.1$, only the most significant SNP is shown in the top eight).

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr</th>
<th>BP (hg19)</th>
<th>A1</th>
<th>A2</th>
<th>Freq A1</th>
<th>beta (SE)</th>
<th>P</th>
<th>Direction$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1574587</td>
<td>16</td>
<td>84 453 056</td>
<td>T</td>
<td>C</td>
<td>0.1415</td>
<td>0.09 (0.016)</td>
<td>4.0 × 10⁻⁹</td>
<td>+?+++++ -- +</td>
</tr>
<tr>
<td>rs4935127</td>
<td>10</td>
<td>56 654 986</td>
<td>C</td>
<td>G</td>
<td>0.7741</td>
<td>--0.06 (0.013)</td>
<td>4.6 × 10⁻⁷</td>
<td>-- -- - -- -- +--</td>
</tr>
<tr>
<td>rs2249437</td>
<td>6</td>
<td>1 595 216</td>
<td>T</td>
<td>C</td>
<td>0.4595</td>
<td>0.07 (0.014)</td>
<td>5.1 × 10⁻⁷</td>
<td>++++? ++++</td>
</tr>
<tr>
<td>rs9266245</td>
<td>6</td>
<td>31 325 702</td>
<td>A</td>
<td>G</td>
<td>0.2655</td>
<td>--0.07 (0.015)</td>
<td>1.6 × 10⁻⁶</td>
<td>----+---</td>
</tr>
<tr>
<td>rs28622199</td>
<td>8</td>
<td>5 392 103</td>
<td>T</td>
<td>C</td>
<td>0.8012</td>
<td>0.07 (0.015)</td>
<td>2.7 × 10⁻⁶</td>
<td>+++ - ++++++</td>
</tr>
<tr>
<td>rs215069</td>
<td>16</td>
<td>16 091 237</td>
<td>T</td>
<td>C</td>
<td>0.0685</td>
<td>--0.11 (0.025)</td>
<td>3.8 × 10⁻⁶</td>
<td>--?--?--</td>
</tr>
<tr>
<td>rs4924506</td>
<td>15</td>
<td>41 129 467</td>
<td>A</td>
<td>C</td>
<td>0.7318</td>
<td>0.06 (0.013)</td>
<td>5.5 × 10⁻⁶</td>
<td>++++++-- -- ++</td>
</tr>
<tr>
<td>rs7773177</td>
<td>6</td>
<td>139 143 088</td>
<td>A</td>
<td>G</td>
<td>0.7383</td>
<td>--0.06 (0.013)</td>
<td>8.5 × 10⁻⁶</td>
<td>----- -- --</td>
</tr>
</tbody>
</table>

Chr = chromosome; BP (hg19) = location in base pairs in human genome version 19, A1 = allele 1, A2 = allele 2, Freq A1 = frequency of allele 1, SE = standard error, $P$ = $P$-value. $^a$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 (?). Only SNPs present in at least two samples were included in the meta-analysis. Order of samples in the discovery: Avon Longitudinal Study of Parents and Children (ALSPAC), Brisbane Longitudinal Twin Study (BLTS), Finnish Twin Cohort Study (FinnTwin), Hospital Universitari Vall d’Hebron (HVH), Netherlands Twin Register (NTR), QIMR Berghofer Medical Research Institute (QIMR), TRacking Adolescents’ Individual Lives Survey (TRAILS), Utrecht, Yale Penn EA. Sample information can be found in Table 1.
genetic variation in age at first use of cannabis cannot be captured by current GWAS arrays (e.g. rare genetic variants having a MAF < 0.05) at current sample sizes. Additional sources of discrepancy may be attributable to interactions between genetic loci and environmental factors [73]. Detecting interaction effects also requires larger sample sizes and measures of environmental exposures harmonized across cohorts.

**Strengths and limitations**

**Strengths**

To our knowledge, this is the largest genome-wide study of age at first cannabis. This meta-analytical sample identified \( ATP2C2 \) as a risk gene, which is commensurate with the hypothetical role of calcium signalling mechanisms in substance use. We are unaware of any similarly sized meta-analysis that has fitted a survival-based method to identify genetic loci associated with addiction phenotypes. This approach allowed us to exploit all available information in the participating cohorts, while accounting for the censored nature of observations. Using information from both censored (i.e. individuals who reported not to have initiated cannabis use at the last interview) and uncensored observations for parameter estimation reduces the likelihood of misclassification (i.e. misclassification due to young participants becoming users at later ages), thereby increasing statistical power.
Limitations

Our results should be interpreted in the context of five potential limitations. First, the replication sample was much smaller than the discovery sample. The size of the replication sample was somewhat modest in the context of standard GWAS of highly polygenic traits [74], making it difficult to distinguish false negatives from null effects. Replication sample sizes varied across the loci. The top genome-wide significant SNP rs1574587 met our quality control criteria in only one of the replication samples comprising 593 individuals. We conjecture that the lack of replication was due most probably to lack of statistical power. Secondly, we imposed stringent selection criteria on the SNPs comprising the polygenic scores by selecting only variants present in at least seven discovery samples and genotyped in the NTR2/Research into Antipsychotic Discontinuation and Reduction (RADAR) replication sample (i.e. we removed imputed SNPs). Although this was performed to maximize the prediction accuracy of the polygenic scores, it is possible that SNPs in imperfect linkage disequilibrium with the causal variants were retained, as SNPs GWASs do not tag all causal variants perfectly, in particular those with low frequency and rare variants (see [75]). Rare genetic variants have been shown to explain part of the variation in addiction phenotypes [76]. However, sequencing of much larger samples is required to locate rare variants reliably. For example, we would need to include 80,000 individuals in the discovery sample to detect rare SNPs (MAF = 0.001) with a hazard ratio of 2 and an alpha threshold of 5E-08. Thirdly, because our sample comprised retrospective and longitudinal cohorts, longer intervals between initiation and assessment may result in recall bias. However, when stratified by design, differences in mean age at initiation between retrospective (16.9 years) and longitudinal (17.1 years) studies were minor. Also, the mean age at initiation and the degree of censoring varied between cohorts, due probably to differences in sampling, assessment, drug policy, legality and availability. To the extent to which these discrepancies were driven by age-related differences, the survival analyses were adjusted for the effects of birth cohort if variation in date of assessment spanned 20 or more years. Moreover, despite these differences, the top SNPs generally had an effect in the same direction throughout the samples and there was no evidence of significant between-cohort heterogeneity in the estimated effects (Fig. 2b. Supporting information, Fig. S5 and Table S3 for I² heterogeneity statistic). Furthermore, the forest plots indicate that the 95% confidence intervals surrounding the effect for each cohort mostly overlap and contain the meta-analytic effect. Fourthly, the sample was limited to individuals of European ancestry. Whether or not our conclusions generalize to populations of other ethnicities remains subject to further investigation. Fifthly, we did not collect information on cannabis use opportunities. Recent findings suggest that drug use opportunity should be taken into account when investigating genetic influences on drug use as high genetic risk for drug use may not lead to initiation of use when there is a lack of opportunity to do so.

Conclusion

To date, this study is the largest GWAS meta-analysis of age at first cannabis use. Our SNP-based findings support the involvement of the ATP2C2 gene. The gene-based tests also identified the ATP2C2 gene as a significant predictor of age at onset. Our findings are commensurate with the role of calcium signalling mechanisms in substance use disorders. The failure to replicate is probably attributable to lack of statistical power. Further investigation of these signals in larger samples is warranted, and may yield valuable insights into the genetic aetiology of substance use initiation.

Declaration of interests

H.R.K. has been a consultant, CME speaker or Advisory Board Member for Lundbeck and Indivior and is a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last 3 years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer and XenoPort. Drs. Gelernter and Kranzler are named as inventors on PCT patent application #15/878,640 entitled: “Genotype-guided dosing of opioid agonists.” filed January 24, 2018. The other co-authors have no conflicts of interest.

Acknowledgements

J.M.V., C.C.M. and H.M. are supported by the European Research Council (Beyond the Genetics of Addiction ERC-284167, Principal Investigator J.M.V.). E.M.D. is supported by the Foundation Volksbond Rotterdam. K.J.H.V. is supported by a 2014 NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation. N.A.G. is supported by US National Institutes of Health, National Institute on Drug Abuse R00DA023549. C.C.M. and M.C.N. are supported by NIDA grant DA-018673. R.W. 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 supported financially by the Netherlands Organization for Scientific Research (NWO 480–05-003 Principal Investigator D. Posthumus), along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam.
Study site acknowledgements

ALSIPAC: 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 ALSIPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The National Medical Research Council and the Wellcome Trust (grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSIPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. J.J.W. is supported by a Postdoctoral Research Fellowship from the Oak Foundation. J.J.W. and M.R.M. are members of the MRC Integrative Epidemiology Unit at the University of Bristol, funded by the UK Medical Research Council (MC_UU_12013/6) and the University of Bristol. M.H. is a member of NIHR School of Public Health Research and NIHR Health Protection Research Unit in Evaluation, J.J.W. and M.R.M. are members of UK Centre for Tobacco and Alcohol Studies, and M.H. is a member of DeGEPH (Development and Evaluation of Complex Interventions for Public Health Improvement)—which are both UKCRC Public Health Research: Centres 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 (RO1DA023549) awarded to N.G., by the Australian Research Council to M.W. (grants DP1349212, DP1664641 and DP1093900) and by Australian National Health and Medical Research Council Australia Fellowships awarded to L.A.B. (no. 464914) and G.W.M. (no. 619667). We acknowledge and thank the following project staff: Anjali Henders, Leanne Wallace and Lisa Bowler for the laboratory processing, genotyping and QC; Soad 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.

CADE: The Center on Antisocial Drug Dependence (CADE) data reported here were funded by grants from the National Institute on Drug Abuse (P60 DA011015, R01 DA02845, R01 DA021913, R01 DA021905 and R01 DA035804).

FinnTwin: We warmly thank the participating twin pairs and their family members for their contribution. We would like to express our appreciation to the skilled study interviewers A.-M. Iivonen, K. Karhu, H.-M. Kuha, E. 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 was supported by the Academy of Finland Center 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.

CADD: The Center on Antisocial Drug Dependence (CADE) data reported here were funded by grants from the National Institute on Drug Abuse (RO1DA023549) awarded to N.G., by the Australian Research Council to M.W. (grants DP1349212, DP1664641 and DP1093900) and by Australian National Health and Medical Research Council Australia Fellowships awarded to L.A.B. (no. 464914) and G.W.M. (no. 619667). We acknowledge and thank the following project staff: Anjali Henders, Leanne Wallace and Lisa Bowler for the laboratory processing, genotyping and QC; Soad 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.

RADAR: We thank all adolescents and their families and friends for their participation. Moreover, we want to thank the various assistants who 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, 056-113-2073), and the Dutch Brain Foundation and the Department of Psychology and Education of the VU University Amsterdam. QIMR: Supported by National Institutes of Health Grants AA07535, AA07580, AA07728, AA10249, AA13320, AA14041, AA11998, AA17688, DA012854, DA018267, DA018660, DA23668, DA1019951; by Grants from the Australian National Health and Medical Research Council (241 944, 113 396 2, 139 462, 239 875, 239 891, 389 918, 442 915, 442 981, 496 739, 552 485, 552 499 and 628 911); by Grants from the Australian Research Council (A7960034, A79960588, A79980149, A79070096, D5012106 and D5034921); and by the 5th Framework Programme (FP-5) GenomElwTwin Project (QLG2-CT-2002-01254). This research was further supported by the Centre for Research Excellence on Suicide Prevention (CRESPP–Australia). We thank Anjali Henders, Richard Parker, Sour Hakcock, Judith Moir, Sally Rodda, Pieta-Marie Shertock, Heather Park, Jill Wood, Pam Barton, Fran Husband, Adele Somerville, Ann Eldridge, Markene Grace, Kerrie McAloney, Lisa Bowler, Alexandre Todorow, Steven Crooks, David Smyth, Harry Beeby and Daniel Park. Lastl, we thank the twins and their families for their participation.

RADAR: We thank all adolescents and their families and friends for their participation. Moreover, we want to thank the various assistants who 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); RADAR has been supported financially by main grants from the Netherlands Organisation for Scientific Research (GB-MAGW 480-03-005) and Stichting Achmea Slachtoffer en Samenleving (SASS), a grant from the Netherlands Organisation for Scientific Research to the Consortium Individual Development (CD: 024/011/003) and various other grants from the Netherlands Organisation for Scientific Research, the VU University Amsterdam and Utrecht University. A.J.H. is supported by the Netherlands Organisation for Health Research and Development, ZonMW 31 160 212.

Sauganay Youth Study: The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the SYS (TR.ZP). TP is the Tanenbaum Chair in Population Neuroscience (University of Toronto) and the Dr John and Consuelva Phean 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 University of Utrecht, the Radboud Medical Center Nijmegen and the Piemassla Baro group, all in

© 2018 Society for the Study of Addiction
the Netherlands. Traill2s has been supported financially by grants from the Netherlands Organisation for Scientific Research NWO (Medical Research Council program grant GB-MW 940–38-011; ZenMW Brainpower grant 100-001-004; ZenMW Risk Behavior and Dependence grant 60-60-608–97-118; ZenMw 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 and 481–11–001); the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities and Accare Center for Child and Adolescent Psychiatry. We are grateful to all 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 supported financially by the Netherlands Scientific Organization (NWO 480-05-003), along with a supplement from the Dutch Brain Foundation.

Utrecht: We are grateful to Chris Schubart and Willeminjn van Gastel and numerous students for their work in the study. Foremost, we like to thank our study participants. This study was supported financially by a grant of the NWO (Netherlands Organization for Scientific Research), grant no. 91207039. The study was performed at the University Medical Center Utrecht, the Netherlands.

Yale Penn: Genotyping services for a part of our GWAS study were provided by the Center for Inherited Disease Research (CIDR) and Yale University (Center for Genome Analysis). CIDR is funded fully through a federal contract from the National Institutes of Health to The Johns Hopkins University (contract number N01-HG-65403). This study was supported by National Institutes of Health grants RC2 DA028909, R01 DA12690, and R01 DA028909, R01 DA12690, R01 DA18432, R01 AA11310, R01 AA017535 and the VA Connecticut and Philadelphia VA MIRECCs.

Authors’ Affiliation

Department of Biological Psychology/Netherlands Twin Register; VU University, Amsterdam, the Netherlands;2 Behavioral Science Institute, Radboud University, Nijmegen, the Netherlands;3 Department of Epidemiology, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands;4 Hospital for Sick Children Research Institute, Toronto, Canada;4 Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, the Netherlands;5 Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia;6 Institute for Behavioral Genetics, Department of Psychology and Neuroscience, University of Colorado Boulder; Boulder, CO, USA;7 Department of Psychiatry, University of California San Diego, La Jolla, CA, USA;8 Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia;9 Department of Biostatistics, Yale School of Public Health and VA CT, New Haven, CT, USA;10 Ratman Research Institute, Baycrest, Toronto, Canada;11 Psychology and Psychiatry, University of Toronto, Toronto, Canada;12 Center for the Developing Brain, Child Mind Institute, New York, NY, USA;13 UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK;14 Psychiatry, Genetics, and Neuroscience, Yale University School of Medicine and VA CT, West Haven, CT, USA;15 Department of Psychiatry, Academic Medical Centre, Amsterdam, the Netherlands;16 Translational Neurogenomics group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia;17 and Amsterdam UMC, University of Amsterdam, Department of Psychiatry, Amsterdam, the Netherlands.18

References


© 2018 Society for the Study of Addiction

Addiction. 113. 2073–2086
2084
Camelia C. Minică et al.


52. So H. C., Li M. H., Sham P. C. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. *Genet Epidemiol* 2011; 35: 447–56.


Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1a, b Quantile–quantile (QQ) plot based on (a) lambdaGC corrected input files and (b) uncorrected input files.

Figure S2a–i Manhattan plots of the individual samples in the discovery meta-analysis (lambda-corrected).

Figure S3a–i Quantile–quantile plots of the individual samples in the discovery meta-analysis (lambda-corrected).

Figure S4a–l Regional association plots showing signal around top single nucleotide polymorphisms (SNPs).

Figure S5a–k Forest plots top single nucleotide polymorphisms (SNPs).

Figure S6 The quantile–quantile (QQ) plot of the gene-based test.

Table S1 Descriptives of individual samples.

Table S2 Genotyping and imputation information per sample.

Table S3 Top 100 single nucleotide polymorphisms (SNPs) in the discovery meta-analysis.

Table S4a Association results and descriptive information for the top single nucleotide polymorphism (SNP) rs1574587 based on the discovery samples and the replication samples.

Table S5 Top 100 genes (discovery sample).

File S1 Information about sample collection.

File S2 Heritability study: Methods.

File S3 Effective sample size.

File S4 Heritability in twins: sample description and model-fitting details.

File S5 Genome-wide association meta-analysis of age at first cannabis use – replication analyses.

Table 1-S5 Descriptive information on the participating replication cohorts.

Table 2-S5 Power analysis results.

Table 3-S5 Results for the top eight independent single nucleotide polymorphisms in the meta-analysis of the discovery sample, and results of the meta-analysis of combined discovery and replication samples.