GENOME-WIDE ASSOCIATION META-ANALYSIS OF AGE AT FIRST CANNABIS USE

Camelia C. Minică[1*@], Karin J.H. Verweij[1,2*], Peter J. van der Most[3*], Hamdi Mbarek[1], Manon Bernard[4], Kristel R. van Eijk[5], Penelope A. Lind[6], MengZhen Liu[7], Dominique F. Maciejewski[8,9], Teemu Palviainen[10], Cristina Sánchez-Mora[11,12,13], Richard Sherva[14], Michelle Taylor[15,16], Raymond K. Walters[17,18,19], Abdel Abdellaoui[1], Timothy B. Bigdeli[20], Susan J.T. Branje[21], Sandra A. Brown[22], Miguel Casas[11,12,13,23], Robin P. Corley[7], George Davey Smith[15,16], Gareth E. Davies[24], Erik A. Ehli[24], Lindsay Farrer[25], Iryna O. Fedko[1], Iris Garcia-Martínez[11,12], Scott D. Gordon[26], Catharina A. Hartman[27], Andrew C. Heath[28], Ian B. Hickie[29], Matthew Hickman[16], Christian J. Hopfer[30], Jouke Jan Hottenga[1], René S. Kahn[31], Jaakko Kaprio [10,32], Tellervo Korhonen[10,33], Henry R. Kranzler[34], Ken Krauter[35], Pol A.C. van Lier[8,36], Pamela A.F. Madden[28], Sarah E. Medland[6], Michael C. Neale[37], Wim H.J. Meeus[21,38], Grant W. Montgomery[39], Ilja M. Nolte[3], Albertine J. Oldehinkel[27], Zdenka Pausova[4,40], Josep A. Ramos-Quiroga[11,12,13,23], Vanesa Richarte[11,12,13], Richard J. Rose[41], Jean Shin[4], Michael C. Stallings[7], Tamara L. Wall[42], Jennifer J. Ware[15,16], Margaret J. Wright[43], Hongyu Zhao[44], Hans M. Koot[8], Tomas Paus[45,46,47], John K. Hewitt[7], Marta Ribasés[11,12,13], Anu Loukola[10], Marco P. Boks[31], Harold Snieder[3], Marcus R. Munafò[15,48], Joel Gelernter[49], Dorret I. Boomsma[1], Nicholas G. Martin[26], Nathan A. Gillespie[26,50†], Jacqueline M. Vink[2†@], & Eske M. Derks[51,52†@]

*Shared first author

†Shared last author

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/add.14368

[@] Corresponding authors:

Prof. dr. Eske Derks, Translational Neurogenomics group, QIMR Berghofer, 300 Herston road, Herston QLD, 4006, Australia. Email: eske.derks@qimrberghofer.edu.au

Dr. Camelia Minică, Department of Biological Psychology, Vrije Universiteit Amsterdam, Van der Boechorststraat 1, 1081BT, Amsterdam, the Netherlands. Email: camelia.minica@vu.nl

Prof. dr. Jacqueline Vink, Behavioral Science Institute, Radboud University, Montessorilaan3, 6525 HR, Nijmegen, the Netherlands. Email: j.vink@bsi.ru.nl

Affiliations

1. 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

5. Department of Neurology, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

6. Psychiatric Genetics, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

7. Institute for Behavioral Genetics, Department of Psychology and Neuroscience, University of Colorado Boulder, Boulder, Colorado, USA

8. Vrije Universiteit Amsterdam, Department of Clinical Developmental Psychology, Amsterdam, The Netherlands

9. GGZ inGeest and Department of Psychiatry, Amsterdam Public Health research institute, VU University Medical Center, Amsterdam, The Netherlands

10. Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland

11. Psychiatric Genetics Unit, Group of Psychiatry, Mental Health and Addiction, Vall d'Hebron Research Institute (VHIR), Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain

12. Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain

13. Biomedical Network Research Centre on Mental Health (CIBERSAM), Instituto de Salud Carlos III, Madrid, Spain

14. Biomedical Genetics Department, Boston University School of Medicine, Boston, Massachusetts, USA

15. MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK

16. School of Social and Community Medicine, University of Bristol, Bristol, UK

17. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts, USA

18. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

19. Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA

20. Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia, USA

21. Research Centre Adolescent Development, Utrecht University, Utrecht, the Netherlands

22. Department of Psychology and Psychiatry, University of California San Diego, La Jolla, California, USA

23. Department of Psychiatry and Legal Medicine, Universitat Autònoma de Barcelona, Barcelona, Spain

24. Avera Institute for Human Genetics, Sioux Falls, South Dakota, USA

25. Department of Medicine (Biomedical Genetics), Boston University School of Medicine, Boston, Massachusetts, USA

26. Genetic Epidemiology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

27. Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

28. Department of Psychiatry, Washington University School of Medicine, St Louis, Missouri, USA

29. Brain & Mind Research Institute, University of Sydney, Sydney, NSW, Australia

30. Department of Psychiatry, University of Colorado Denver, Aurora, Colorado, USA

31. Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands

32. Department of Public Health, University of Helsinki, Helsinki, Finland

33. University of Eastern Finland, Institute of Public Health & Clinical Nutrition, Kuopio, Finland

34. Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, USA

35. Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, Colorado, USA

36. Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Rotterdam, the Netherlands

37. Department of Psychiatry and School of Medicine, Virginia Commonwealth University, Richmond, Virginia, USA

38. Developmental Psychology, Tilburg University, Tilburg, The Netherlands

39. Institute for Molecular Bioscience, The University of Queensland, Brisbane, Queensland, Australia

40. Physiology and Nutritional Sciences, University of Toronto, Toronto, Canada

41. Department of Psychological & Brain Sciences, Indiana University, Bloomington, Indiana, USA

42. Department of Psychiatry, University of California San Diego, La Jolla, California, USA

43. Queensland Brain Institute, The University of Queensland, Brisbane, Queensland, Australia

44. Department of Biostatistics, Yale School of Public Health & VA CT, New Haven, Connecticut, USA

45. Rotman Research Institute, Baycrest, Toronto, Canada

46. Psychology and Psychiatry, University of Toronto, Toronto, Canada

47. Center for the Developing Brain, Child Mind Institute, New York, New York, USA

48. UK Centre for Tobacco and Alcohol Studies, School of Experimental Psychology, University of Bristol, Bristol, UK

49. Psychiatry, Genetics, & Neuroscience, Yale University School of Medicine & VA CT, West Haven, Connecticut, USA

50. Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, Virginia, USA

51. Department of Psychiatry, Academic Medical Centre, Amsterdam, The Netherlands

52. Translational Neurogenomics group, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia

Running head: The genetics 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 identify associations with genetic variants.

Methods: A twin-based heritability analysis using 8,055 twins from three cohorts was performed. We then carried-out a genome wide survival 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 3,735 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 survival meta-analysis identified five SNPs on chromosome 16 within the Calcium-transporting ATPase gene (*ATP2C2)* at P < 5E-08. All five SNPs are in high 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 previously associated 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: cannabis initiation, *ATP2C2*, substance use, genome-wide association.

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 commonly observed in older cohorts, is closing in more recent cohorts (6, 7). Overall, these trends are likely due to lower risk perception (8), and increased availability due to medicalisation and decriminalisation.

Early cannabis initiation is linked to a number of maladaptive behaviors. 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 behaviors (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 lifetime 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 lifetime

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=80%) for early-onset use (≤16 years), whereas Sartor et al. (28) reported a heritability of 52% when age at first cannabis use was categorized as 'never', 'late' (≥17 years), or 'early' (≤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 5,148 participants. This study found no single nucleotide polymorphisms (SNPs) or genes significantly associated with age at first cannabis use, possibly due 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 previously employed. 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 significantly associated with lifetime cannabis use: *NCAM1*; *CADM2*; *SCOC*; and *KCNT2* (34). Interestingly, both *NCAM1* and *KCNT2* have been previously linked to other substance use phenotypes (34). Of note is also our novel

finding at *CADM2*, which was recently associated with alcohol consumption (35), personality (36), behavioral reproductive outcomes and risk-taking behavior (37).

Our aim was to explore the genetic etiology of age at first cannabis use. First, we performed a biometrical heritability analysis in 8,055 twins from three cohorts. Second, 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 3,735 individuals from three cohorts. The outline of the analyses steps is illustrated in Figure 1.

-- Figure 1 about here --

MATERIALS AND METHODS

Biometrical heritability

The heritability of age at first cannabis use was estimated based on data from three cohorts: NTR comprising 2027 monozygotic (MZ) and 1771 dizygotic (DZ) twin pairs ; QIMR comprising 1282 MZ and 1969 DZ twin pairs ; and 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 Supplementary File S2 and Supplementary File S4 for details).

Study samples

The current discovery meta-analysis was based on genome-wide summary statistics from 9 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 Supplementary Table S1 for more details).

-- Table 1 about here –

Phenotyping

Age at first cannabis use was assessed from questionnaires or clinical interviews (see Supplementary 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.

Genotyping

Genotyping followed by extensive quality control (QC) was performed by each participating cohort (see Supplementary 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 1st or 2nd degree relatives (in samples of unrelated individuals) were removed. In Supplementary 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 Supplementary 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 across all 9 discovery cohorts. Second, we retained SNPs with MAF greater than $\sqrt{(5/N)}$, where N is the sample size. This ensured that there were at least 5 individuals in the least frequent genotype group. Third, 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 co-dominant 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 Supplementary Table 2 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 weighs 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 Supplementary 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=.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 Ns per SNP. In both methods, SNP-based heritability depends on 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 behavior for heritability analysis) of the current GWAS (for details see Supplemental File S3).

Replication analyses

Genes reaching significance and the top 8 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 3,735 individuals from three cohorts. In addition, the top SNPs were analyzed in the combined discovery and replication samples. Furthermore, we tested whether a polygenic risk score based on the meta-analysis results predicts age at first cannabis use in one of the replication samples (See supplementary 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 Supplementary file S4 for model fitting details and twin correlations). In this model, the heritability (A) of age at first cannabis use was 38% (95% 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 Supplementary Figure 1a. 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 Supplementary Figure 1b). 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. The Supplementary Figures S2a-i and S3a-i show cohort-specific lambda-corrected Manhattan and quantile-quantile plots.

The Manhattan plot in Figure 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 8 independent SNPs. The top 100 SNPs in the discovery sample are shown in Supplementary Table S3. Regional association plots and forest plots for the top SNPs are shown in Supplementary Figures S4a-l, Figure 1b, and Supplementary Figures S5a-k.

----Figure 2 and Table 2 about here----

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 GC was applied or not (P = 1.08e-08). This SNP has a MAF ranging from 0.105 to 0.185 across the discovery samples (commensurate with MAFs reported for European ancestry populations by Ensemble), and an imputation quality ≥ 0.89 (see Supplementary Table S4a for more details on this SNP).

The I^2 statistic for the top SNP was 32.6% ($\chi^2(7)=10.38$, 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 (Figure 2b).

Gene-based tests of association

Figure 3 provides an overview of the gene-based results. The quantile-quantile plot (Supplementary Figure 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 noncoding regions, were enriched for SNPs that yielded strong association signals in the single variant analysis (Supplementary Figure S6).

-- Figure 3 about here--

As shown in the Manhattan plot in Figure 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, corrected P=0.034). See Supplementary Table S5 for the top 100 genes identified in the discovery meta-analysis and Figure 3b for the zoom plot of the significant gene.

ATP2C2 is located at 16q24.1 (Figure 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 ATP-ase activity, calcium ion transmembrane transport, ATP binding and metal ion binding.

SNP-based heritability analyses

The selected SNPs did not significantly contribute 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).

Replication analyses

The power to replicate the top 8 SNPs was low, ranging from 0.04 to 0.10 (see Supplemental file S5Table 2-S5). We refer to Supplemental File S5 for results of the replication analyses.

DISCUSSION

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 8,055 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 intronic 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 likely caused 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

impairement (e.g. 61). *ATP2C2* has also been linked to cocaine dependence. Gelernter et al. (62) found that the highest ranked gene networks significantly associated 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 (65). 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. 66, 67). 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 behaviors 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. 68) and twin studies have shown that part of the covariation is due to overlapping genetic influences (69-71). It is therefore plausible that genes for age at first cannabis use also play a role in the broader spectrum of externalizing behavior. Unfortunately, existing GWASs of conduct and antisocial behavior have not been sufficiently powered to identify genes robustly associated with these behaviours (72, 73). However, using the combined effect of all SNPs, Tielbeek et al.(73) showed a significant genetic correlation between antisocial behavior 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 7 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 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 (74). 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-analytic 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 rather modest in the context of standard GWAS of highly polygenic traits (75), 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 most likely due to lack of statistical power. Second, we imposed stringent selection criteria on the SNPs comprising the polygenic scores by selecting only variants present in at least 7 discovery samples and genotyped in the NTR2/RADAR replication sample (i.e. we removed imputed SNPs). Although this was done to maximize the prediction accuracy of the polygenic scores, it is possible SNPs in imperfect linkage disequilibrium with the causal variants were retained, as SNPs GWASs do not perfectly tag all causal variants, in particular, those with low frequency and rare variants, see (76). Rare genetic variants have been shown to explain part of the variation in addiction phenotypes (77). However, sequencing of much larger samples is required to reliably locate rare variants. 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. Third, 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 of 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, likely due 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 across the samples and there was no evidence of significant between-cohort heterogeneity in the estimated effects (Figure 2b, Supplemental Figures S5 and Supplementary Table S3 for I^2 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. Fourth, the sample was limited to individuals of European ancestry. Whether our

conclusions generalize to populations of other ethnicities remains subject to further investigation. Fifth, 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 likely attributable to lack of statistical power. Further investigation of these signals in larger samples is warranted and may yield valuable insights into the genetic etiology of substance use initiation.

Acknowledgements

JMV, CCM and HM are supported by the European Research Council [Beyond the Genetics of Addiction ERC-284167, PI JM Vink]. EMD is supported by the Foundation Volksbond Rotterdam. KJHV is supported by a 2014 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. NAG is supported by US National Institutes of Health, National Institute on Drug Abuse R00DA023549. CCM and MCN are supported by NIDA grant DA-018673. 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 *SURF*sara 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.

Study site acknowledgements

ALSPAC 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 ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corportation of America) using support from 23andMe.

JJW is supported by a Postdoctoral Research Fellowship from the Oak Foundation. JJW and MRM 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. MH is a member of NIHR School of Public Health Research and NIHR Health Protection Research Unit in Evaluation. JJW and MRM are members of UK Centre for Tobacco and Alcohol Studies, and MH is a member of DeCIPHER (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 (R00DA023549) awarded to Nathan Gillespie, by the Australian Research Council to Margie Wright (Nos. DP0343921, DP0664638, and DP1093900), and by Australian National Health and Medical Research Council Australia Fellowships awarded to Ian Hickie (No. 464914) and Grant Montgomery (No. 619667). We acknowledge and thank the following project staff: Anjali Henders, Leanne Wallace and Lisa Bowdler 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.

CADD The Center on Antisocial Drug Dependence (CADD) data reported here were funded by grants from the National Institute on Drug Abuse (P60 DA011015, R01 DA012845, R01 DA021913, R01 DA021905, 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, U Kulmala-Gråhn, M Mantere, K Saanakorpi, M Saarinen, R Sipilä, 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 J. Kaprio), National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-

00145, and AA-09203 to R. J. Rose and AA15416 and K02AA018755 to D. M. Dick), Sigrid Juselius Foundation (to J. Kaprio), and the Welcome Trust Sanger Institute, UK. Antti-Pekka 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.

HUVH We are grateful to patients and controls who kindly participated in this research. Financial support was received "Instituto de Salud Carlos III-FIS" (PI12/01139, PI14/01700, PI15/01789, PI16/01505), and cofinanced by the European Regional Development Fund (ERDF), Agència de Gestió d'Ajuts Universitaris i de Recerca-AGAUR, Generalitat de Catalunya (2014SGR1357), Departament de Salut, Government of Catalonia, Spain , the European College of Neuropsychopharmacology (ECNP network: 'ADHD across the lifespan'), and a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation. This project also received funding from the European Community's Seventh Framework Program (under grant agreement number 602805, Aggressotype) and from the European Community's H2020 Program (under grant agreement number 667302, CoCA).

Marta Ribasés is a recipient of a Miguel de Servet contract from the Instituto de Salud Carlos III, Ministerio de Economía, Industria y Competitividad, Spain (CP09/00119 and CPII15/00023)Iris Garcia-Martínez is a recipient of a contract from the 7th Framework Programme for Research, Technological Development and Demonstration, European Commission (AGGRESSOTYPE_FP7HEALTH2013/602805). Cristina Sánchez-Mora is a recipient of a Sara Borrell contract from the Spanish Ministerio de Economía y Competitividad (CD15/00199) and a mobility grant from the Spanish Ministerio de Economía y Competitividad, Instituto de Salud Carlos III (MV16/00039).

NTR & NTR2 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 480-04-004; 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/2008.024. 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 (NIH R01 HD042157-01A1, MH081802, Grand Opportunity grants 1RC2 MH089951 and 1RC2 MH089995) 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 Supported by National Institutes of Health Grants AA07535, AA0758O, AA07728, AA10249, AA13320, AA13321, AA14041, AA11998, AA17688, DA012854, DA018267, DA018660, DA23668 and DA019951; by Grants from the Australian National Health and Medical Research Council (241944, 339462, 389927, 389875, 389891, 389892, 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 (CRESP - Australia).

We thank Anjali Henders, Richard Parker, Soad Hancock, Judith Moir, Sally Rodda, Pieta-Maree Shertock, Heather Park, Jill Wood, Pam Barton, Fran Husband, Adele Somerville, Ann Eldridge, Marlene Grace, Kerrie McAloney, Lisa Bowdler, Alexandre Todorov, Steven Crooks, David Smyth, Harry Beeby, and Daniel Park. Last, 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 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 & Cognition, 056-21-010). RADAR has been financially supported 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 (CID; 024.001.003), and various other grants from the Netherlands Organisation for Scientific Research, the VU University Amsterdam, and Utrecht University. AJH is supported by the Netherlands Organization for Health Research and Development, ZonMW 31160212.

Saguenay 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 University of Utrecht, the Radboud Medical Center Nijmegen, and the Parnassia Bavo group, all in the Netherlands. TRAILS has been financially supported by grants from the Netherlands Organization 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 60-60600-97-118; 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 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 financially supported 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 Willemijn van Gastel and numerous students for their work in the study. Foremost we like to thank our study participants. This study was financially supported by a grant of the NWO (Netherlands Organization for Scientific Research), grant no. 91207039. The study was performed at the University Medical Centre 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 fully funded 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, R01 DA12849, R01 DA18432, R01 AA11330, R01 AA017535, and the VA Connecticut and Philadelphia VA MIRECCs.

Conflict of interest:

HRK 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 three years by AbbVie, Alkermes, Ethypharm, Indivior, Lilly, Lundbeck, Otsuka, Pfizer, and XenoPort. **The other co-authors do not have a conflict of interest.**

REFERENCES

1. Substance Abuse and Mental Health Services Administration. Results from the 2013 National Survey on Drug Use and Health: Summary of National Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014.
2. World Health Organisation. Facts & Figures 2016 [cited 2016 November 24]. Available from: http://www.who.int/substance_abuse/facts/cannabis/en/.
3. Degenhardt L, Lynskey M, Hall W. Cohort trends in the age of initiation of drug use in Australia. Australian and New Zealand journal of public health. 2000;24(4):421-6.
4. Monshouwer K, Smit F, De Graaf R, Van Os J, Vollebergh W. First cannabis use: does onset shift to younger ages? Findings from 1988 to 2003 from the Dutch National School Survey on Substance Use. Addiction. 2005;100(7):963-70.
5. Kokkevi A, Nic Gabhainn S, Spyropoulou M. Early initiation of cannabis use: a cross-national European perspective. The Journal of adolescent health : official publication of the Society for Adolescent Medicine. 2006;39(5):712-9.
6. Degenhardt L, Chiu WT, Sampson N, Kessler RC, Anthony JC, Angermeyer M, et al. Toward a global view of alcohol, tobacco, cannabis, and cocaine use: findings from the WHO World Mental Health Surveys. PLoS medicine. 2008;5(7):e141.
7. Butterworth P, Slade T, Degenhardt L. Factors associated with the timing and onset of cannabis use and cannabis use disorder: results from the 2007 Australian National Survey of Mental Health and Well-Being. Drug and alcohol review. 2014;33(5):555-64.
8. UNODC. World Drug Report 2010: United Nations Publication; 2010.
9. Verweij KJH, Huizink AC, Agrawal A, Martin NG, Lynskey MT. Is the relationship between early-onset cannabis use and educational attainment causal or due to common liability? Drug and Alcohol Dependence. 2013;133(2):580-6.
10. Stiby AI, Hickman M, Munafo MR, Heron J, Yip VL, Macleod J. Adolescent cannabis and tobacco use and educational outcomes at age 16: birth cohort study. Addiction. 2015;110(4):658-68.
11. Tamm L, Epstein JN, Lisdahl KM, Molina B, Tapert S, Hinshaw SP, et al. Impact of ADHD and cannabis use on executive functioning in young adults. Drug Alcohol Depend. 2013;133(2):607-14.
12. Hall W, Degenhardt L. Adverse health effects of non-medical cannabis use. Lancet (London, England). 2009;374(9698):1383-91.
13. van der Pol P, Liebregts N, de Graaf R, Korf DJ, van den Brink W, van Laar M. Predicting the transition from frequent cannabis use to cannabis dependence: a three-year prospective study. Drug Alcohol Depend. 2013;133(2):352-9.
14. French L, Gray C, Leonard G, Perron M, Pike GB, Richer L, et al. Early Cannabis Use, Polygenic Risk Score for Schizophrenia and Brain Maturation in Adolescence. JAMA psychiatry. 2015;72(10):1002-11.
15. Pedersen W, Mastekaasa A, Wichstrøm L. Conduct problems and early cannabis initiation: a longitudinal study of gender differences. Addiction. 2001;96(3):415-31.
16. DuRant RH, Smith JA, Kreiter SR, Krowchuk DP. The relationship between early age of onset of initial substance use and engaging in multiple health risk behaviors among young adolescents. Archives of pediatrics & adolescent medicine. 1999;153(3):286-91.
17. Arseneault L, Cannon M, Poulton R, Murray R, Caspi A, Moffitt TE. Cannabis use in adolescence and risk for adult psychosis: longitudinal prospective study. BMJ (Clinical research ed). 2002;325(7374):1212-3.
18. Fergusson DM, Lynskey MT, Horwood LJ. The short-term consequences of early onset cannabis use. Journal of Abnormal Child Psychology. 1996;24(4):499-512.
19. Gage SH, Hickman M, Zammit S. Association Between Cannabis and Psychosis: Epidemiologic Evidence. Biol Psychiatry. 2016;79(7):549-56.

20. Schubart CD, van Gastel WA, Breetvelt EJ, Beetz SL, Ophoff RA, Sommer IE, et al. Cannabis use at a young age is associated with psychotic experiences. Psychol Med. 2011;41(6):1301-10.
21. Agrawal A, Grant JD, Waldron M, Duncan AE, Scherrer JF, Lynskey MT, et al. Risk for initiation of substance use as a function of age of onset of cigarette, alcohol and cannabis use: Findings in a Midwestern female twin cohort. Preventive Medicine. 2006;43(2):125-8.
22. Chen CY, Storr CL, Anthony JC. Early-onset drug use and risk for drug dependence problems. Addictive behaviors. 2009;34(3):319-22.
23. 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. Addictive behaviors. 2010;35(1):35-41.
24. King KM, Chassin L. A prospective study of the effects of age of initiation of alcohol and drug use on young adult substance dependence. Journal of studies on alcohol and drugs. 2007;68(2):256-65.
25. Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, et al. Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction. 2010;105(3):417-30.
26. Richmond-Rakerd LS, Slutske WS, Lynskey MT, Agrawal A, Madden PA, Bucholz KK, et al. Age at first use and later substance use disorder: Shared genetic and environmental pathways for nicotine, alcohol, and cannabis. Journal of abnormal psychology. 2016;125(7):946-59.
27. Lynskey MT, Agrawal A, Henders A, Nelson EC, Madden PA, Martin NG. An Australian twin study of cannabis and other illicit drug use and misuse, and other psychopathology. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2012;15(5):631-41.
28. Sartor CE, Agrawal A, Lynskey MT, Bucholz KK, Madden PA, Heath AC. Common genetic influences on the timing of first use for alcohol, cigarettes, and cannabis in young African-American women. Drug Alcohol Depend. 2009;102(1-3):49-55.
29. Vink JM, Willemsen G, Boomsma DI. Heritability of smoking initiation and nicotine dependence. Behav Genet. 2005;35(4):397-406.
30. Minica CC, Dolan CV, Hottenga JJ, Pool R, Fedko IO, Mbarek H, et al. Heritability, SNP- and Gene-Based Analyses of Cannabis Use Initiation and Age at Onset. Behav Genet. 2015.
31. Kiefer AK, Tung JY, Do CB, Hinds DA, Mountain JL, Francke U, et al. Genome-wide analysis points to roles for extracellular matrix remodeling, the visual cycle, and neuronal development in myopia. PLoS Genet. 2013;9(2):e1003299.
32. van der Net JB, Janssens AC, Eijkemans MJ, Kastelein JJ, Sijbrands EJ, Steyerberg EW. Cox proportional hazards models have more statistical power than logistic regression models in cross-sectional genetic association studies. European journal of human genetics : EJHG. 2008;16(9):1111-6.
33. Stringer S, Denys D, Kahn RS, Derks EM. What Cure Models Can Teach us About Genome-Wide Survival Analysis. Behav Genet. 2016;46(2):269-80.
34. Stringer S, Minica CC, Verweij KJ, Mbarek H, Bernard M, Derringer J, et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Translational psychiatry. 2016;6:e769.
35. Clarke T-K, Adams MJ, Davies G, Howard DM, Hall LS, Padmanabhan S, et al. Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N= 112,117). bioRxiv. 2017:116707.
36. Boutwell B, Hinds D, Tielbeek J, Ong K, Day F, Perry J, et al. Replication and characterization of CADM2 and MSRA genes on human behavior. bioRxiv. 2017:110395.
37. Day FR, Helgason H, Chasman DI, Rose LM, Loh PR, Scott RA, et al. Physical and neurobehavioral determinants of reproductive onset and success. 2016;48(6):617-23.
38. Gillespie NA, Henders AK, Davenport TA, Hermens DF, Wright MJ, Martin NG, et al. The Brisbane Longitudinal Twin Study: Pathways to cannabis use, abuse, and dependence project-current status, preliminary results, and future directions. Twin research and human genetics : the official journal of the International Society for Twin Studies. 2013;16(1):21-33.

39. Neale MC, Cardon LR. Methodology for Genetic Studies of Twins and Families. series NA, editor. Dordrecht: Kluwer; 1992.
40. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.
41. Therneau T. A Package for Survival Analysis in S. R package version 2.37-7. 2015.
42. Minică CC, Dolan CV, Kampert MM, Boomsma DI, Vink JM. Sandwich corrected standard errors in family-based genome-wide association studies. European Journal of Human Genetics. 2015;23(3):388-94.
43. 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(3):559-75.
44. Urbanek S. Rserve: Binary R server, R package version 1.7-3. 2013.
45. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics (Oxford, England). 2010;26(17):2190-1.
46. de Bakker PI, Neale BM, Daly MJ. Meta-analysis of genome-wide association studies. Cold Spring Harbor protocols. 2010;2010(6):pdb.top81.
47. Furberg H, Kim Y, Dackor J, Boerwinkle E, Franceschini N, Ardissino D, et al. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. Nature genetics. 2010;42(5):441-U134.
48. Allen HL, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature. 2010;467(7317):832-8.
49. Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association test using extended Simes procedure. American journal of human genetics. 2011;88(3):283-93.
50. Li MX, Kwan JS, Sham PC. HYST: a hybrid set-based test for genome-wide association studies, with application to protein-protein interaction-based association analysis. American journal of human genetics. 2012;91(3):478-88.
51. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society B. 1995;57:289-300.
52. So HC, Li M, Sham PC. Uncovering the total heritability explained by all true susceptibility variants in a genome-wide association study. Genetic epidemiology. 2011;35(6):447-56.
53. van Beek JH, de Moor MH, Geels LM, Willemsen G, Boomsma DI. Explaining individual differences in alcohol intake in adults: evidence for genetic and cultural transmission? Journal of studies on alcohol and drugs. 2014;75(2):201-10.
54. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291-5.
55. Vilhjalmsson BJ, Yang J, Finucane HK, Gusev A, Lindstrom S, Ripke S, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. American journal of human genetics. 2015;97(4):576-92.
56. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nature genetics. 2000;25(1):25-9.
57. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic acids research. 2015;43(Database issue):D1049-56.
58. Xiang M, Mohamalawari D, Rao R. A novel isoform of the secretory pathway Ca2+,Mn(2+)-ATPase, hSPCA2, has unusual properties and is expressed in the brain. The Journal of biological chemistry. 2005;280(12):11608-14.
59. Newbury DF, Winchester L, Addis L, Paracchini S, Buckingham LL, Clark A, et al. CMIP and ATP2C2 modulate phonological short-term memory in language impairment. American journal of human genetics. 2009;85(2):264-72.

60. Zheng JQ, Poo MM. Calcium signaling in neuronal motility. Annual review of cell and developmental biology. 2007;23:375-404.
61. Graham SA, Fisher SE. Decoding the genetics of speech and language. Current opinion in neurobiology. 2013;23(1):43-51.
62. Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR, et al. Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. Molecular psychiatry. 2014;19(6):717-23.
63. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H, et al. Genome-wide association study of opioid dependence: multiple associations mapped to calcium and potassium pathways. Biol Psychiatry. 2014;76(1):66-74.
64. Lynskey MT, Vink JM, Boomsma DI. Early onset cannabis use and progression to other drug use in a sample of Dutch twins. Behavior Genetics. 2006;36(2):195-200.
65. Lynskey MT, Agrawal A, Henders A, Nelson EC, Madden PA, Martin NG. An Australian twin study of cannabis and other illicit drug use and misuse, and other psychopathology. Twin Research and Human Genetics. 2012;15(05):631-41.
66. Gillespie NA, Neale MC, Kendler KS. Pathways to cannabis abuse: a multi-stage model from cannabis availability, cannabis initiation and progression to abuse. Addiction. 2009;104(3):430-8.
67. Agrawal A, Neale MC, Jacobson KC, Prescott CA, Kendler KS. Illicit drug use and abuse/dependence: modeling of two-stage variables using the CCC approach. Addictive behaviors. 2005;30(5):1043-8.
68. Fergusson DM, Horwood LJ, Ridder EM. Conduct and attentional problems in childhood and adolescence and later substance use, abuse and dependence: results of a 25-year longitudinal study. Drug Alcohol Depend. 2007;88 Suppl 1:S14-26.
69. Miles DR, van den Bree MB, Pickens RW. Sex differences in shared genetic and environmental influences between conduct disorder symptoms and marijuana use in adolescents. American journal of medical genetics. 2002;114(2):159-68.
70. Verweij KJH, Creemers HE, Korhonen T, Latvala A, Dick DM, Rose RJ, et al. Role of overlapping genetic and environmental factors in the relationship between early adolescent conduct problems and substance use in young adulthood. 2016.
71. Shelton K, Lifford K, Fowler T, Rice F, Neale M, Harold G, et al. The association between conduct problems and the initiation and progression of marijuana use during adolescence: A genetic analysis across time. Behavior Genetics. 2007;37(2):314-25.
72. Pappa I, St Pourcain B, Benke K, Cavadino A, Hakulinen C, Nivard MG, et al. A genome-wide approach to children's aggressive behavior: The EAGLE consortium. American journal of medical genetics Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics. 2016;171(5):562-72.
73. Tielbeek JJ, Vink JM, Polderman TJC, Popma A, Posthuma D, Verweij KJH. Genetic correlation of antisocial behaviour with alcohol, nicotine, and cannabis use. Drug Alcohol Depend. 2018;187:296-9.
74. Uher R. Gene–environment interactions in common mental disorders: an update and strategy for a genome-wide search. Social Psychiatry and Psychiatric Epidemiology. 2014;49(1):3-14.
75. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. The American Journal of Human Genetics. 2017;101(1):5-22.
76. Yang J, Zeng J, Goddard ME, Wray NR, Visscher PM. Concepts, estimation and interpretation of SNP-based heritability. Nat Genet. 2017;49(9):1304-10.
77. Vrieze SI, Feng S, Miller MB, Hicks BM, Pankratz N, Abecasis GR, et al. Rare Non-Synonymous Exonic Variants in Addiction and Behavioral Disinhibition. Biological psychiatry. 2014;75(10):783-9.

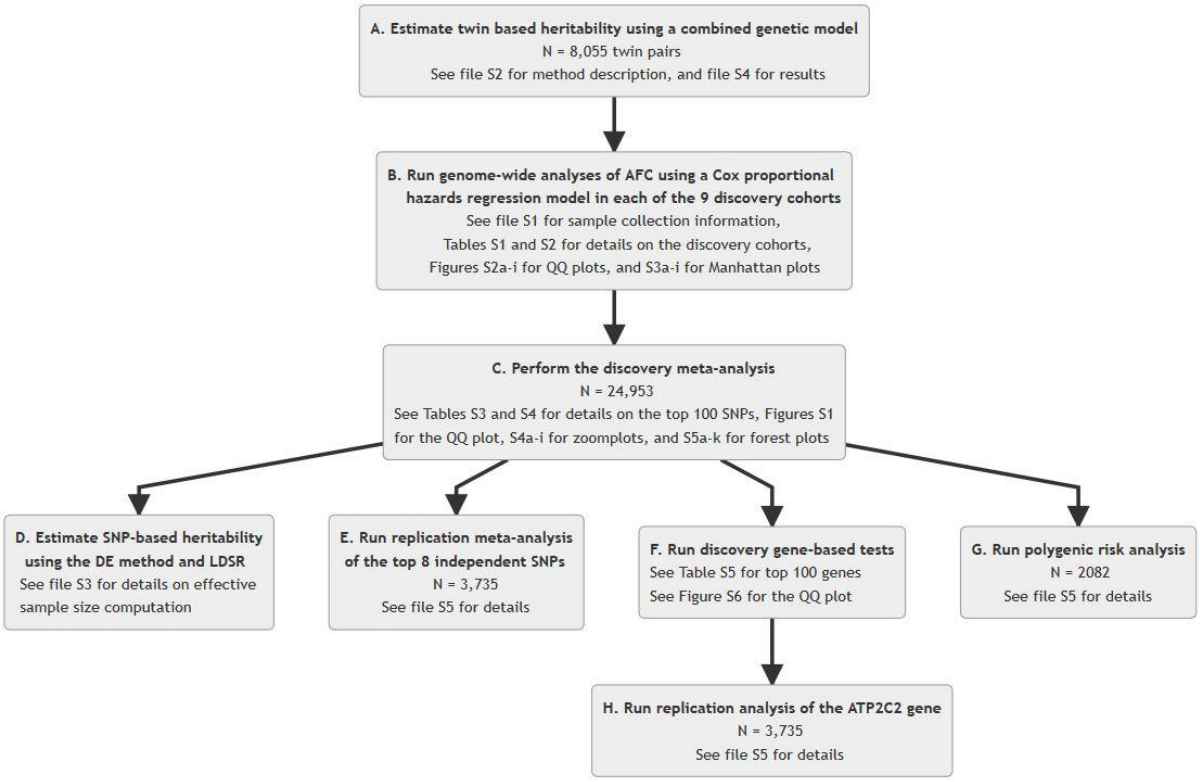


Figure 1: The outline of the analysis steps, and references to the Supplementary Material relevant to each step. Abbreviations: AFC – age at first cannabis use; DE – density estimation; LDSR – linkage disequilibrium score regression.

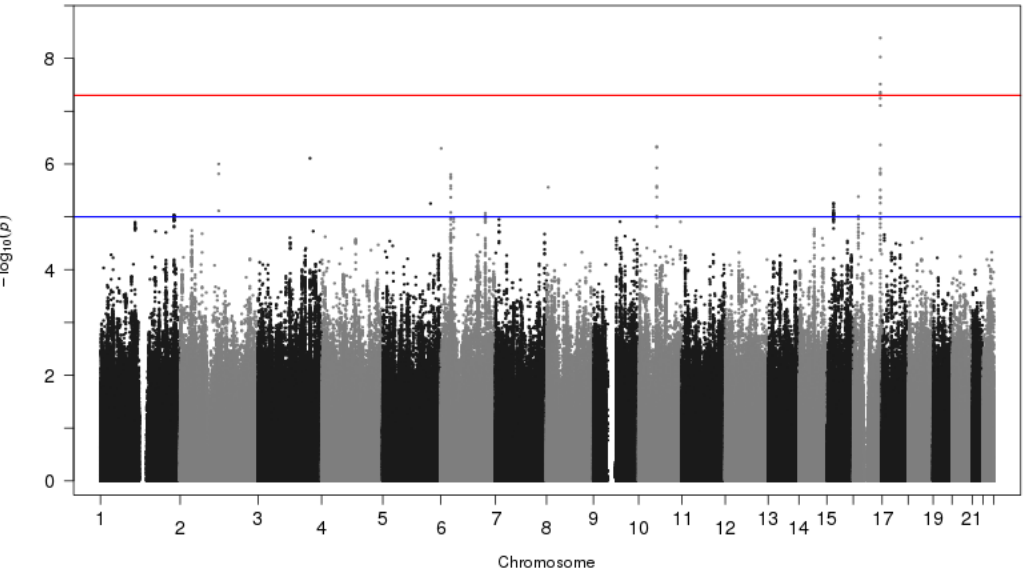
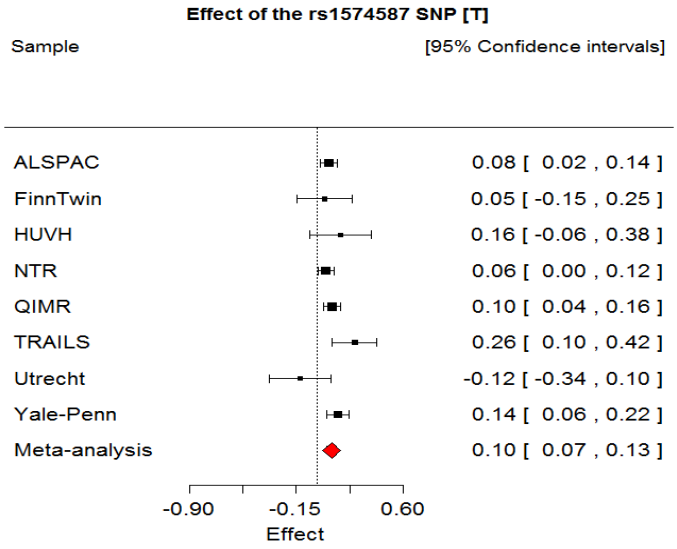


Figure 2a



Note: rs1574587 did not meet quality control criteria in the BLTS sample

Figure 2b

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 (-log10(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.

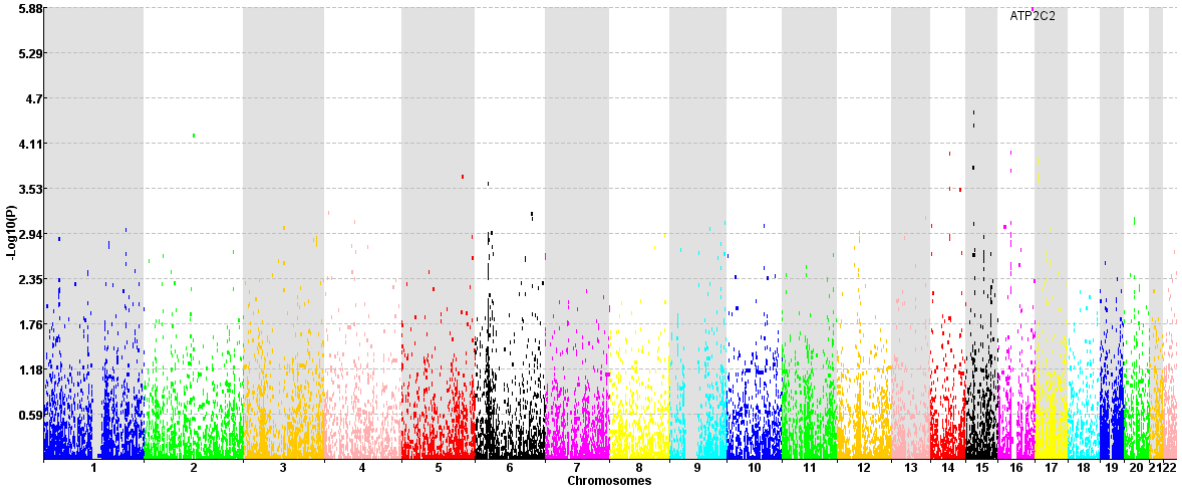


Figure 3a

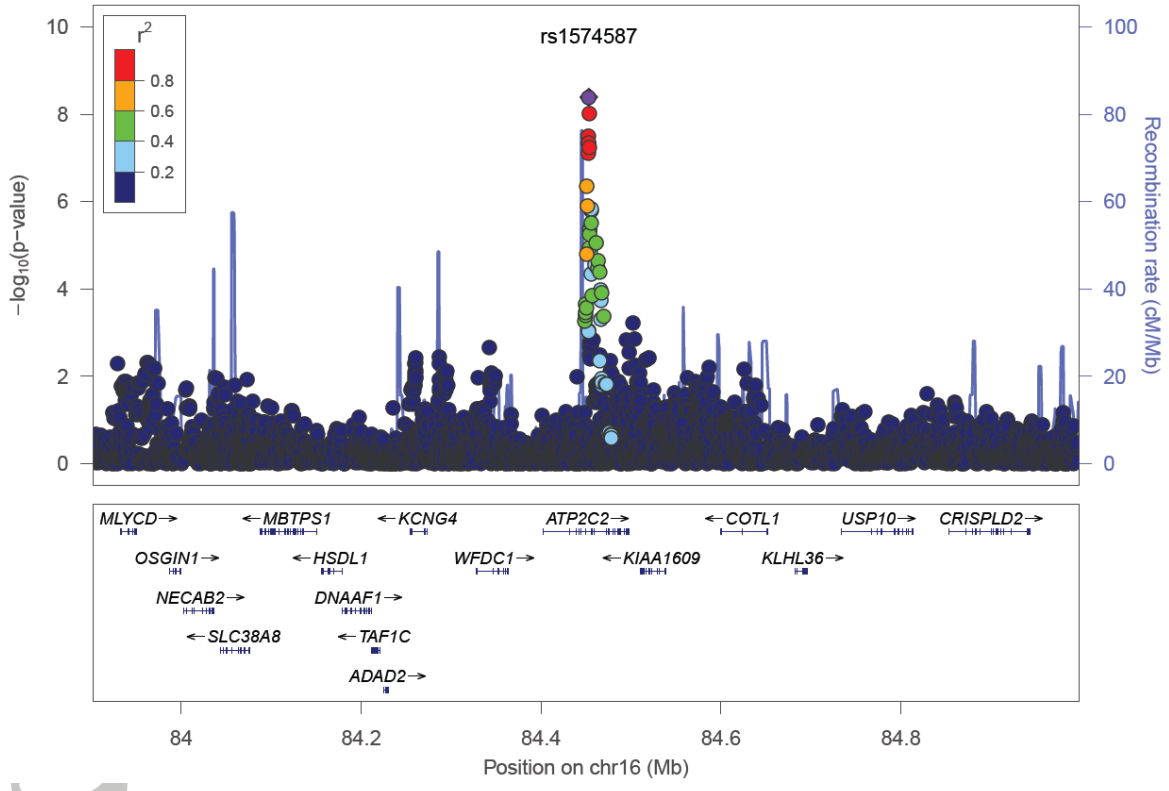


Figure 3b

Figure 3: Results of the gene-based tests: (a) Manhattan plot for the gene-based tests; and (b) Regional plot around the significantly associated gene

Table 1: Descriptive information on the participating discovery cohorts.

Cohort	N (or range)	% Females	%Uncensored Observations	Mean age (SD)	Mean age at first use (sd) (in users)	Number of SNPs
ALSPAC	6147	51.9	38.4	17.3 (1.7)	14.8 (1.6)	6,284,747
BLTS	721	57.1	59.5	26.2 (3.3)	18.8(2.8)	4,093,835
FinnTwin	1029	51.7	27.5	22.8 (1.3)	18.0 (2.5)	4,362,100
HUVH	581	31.3	30.3	28.7 (12.5)	16.0 (3.0)	4,319,651
NTR	5148	62.3	16.6	46.9 (17.5)	18.9 (5.1)	4,773,834
QIMR	6758	53.8	51.3	45.2 (10.9)	19.9 (5.8)	5,953,917
TRAILS	1249	53.8	61.7	20.0 (1.6)	16.3 (2.0)	4,819,504
Utrecht	958	51.3	59	17.4 (3.2)	15.5 (2.1)	4,139,839
Yale-Penn	2362	41.2	92.6	38.2 (10.6)	17.0 (9.4)	5,732,659

N = sample size (or range if sample size varied across SNPs), % uncensored observations (i.e., individuals who have initiated cannabis use). Mean age: age when completing survey or interview. Mean age at first use: mean age at first cannabis use.

***Table 2**. Top 8 independent 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 8).*

SNP	**Chr**	**BP (hg19)**	**A1**	**A2**	**Freq A1**	**beta (s.e.)**	**P**	***Direction****
rs1574587	16	84453056	T	C	0.1415	0.09 (0.016)	$4.0x10^{-9}$	+?+++++-+
rs4935127	10	56654986	C	G	0.7741	-0.06 (0.013)	$4.6x10^{-7}$	---+---+-
rs2249437	6	1595216	T	C	0.4595	0.07 (0.014)	$5.1x10^{-7}$	++++?+?++
rs9266245	6	31325702	A	G	0.2655	-0.07 (0.015)	$1.6x10^{-6}$	----?--?-
rs28622199	8	5392103	T	C	0.8012	0.07 (0.015)	$2.7x10^{-6}$	+++-+++++
rs215069	16	16091237	T	C	0.0685	-0.11 (0.025)	$3.8x10^{-6}$	-?-?--??-
rs4924506	15	41129467	A	C	0.7318	0.06 (0.013)	$5.5x10^{-6}$	++++++--+
rs7773177	6	139143088	A	G	0.7383	-0.06 (0.013)	$8.5x10^{-6}$	-------+-

** 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 2 samples were included in the meta-analysis. Order of samples in the discovery: ALSPAC, BLTS, FinnTwin, HUVH, NTR, QIMR, TRAILS, Utrecht, Yale Penn EA. Sample information can be found in Table 1.*

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, s.e. = standard error, P = p-value.