Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

Citation
Stringer, S., C. C. Minică, K. J. H. Verweij, H. Mbarek, M. Bernard, J. Derringer, K. R. van Eijk, et al. 2016. “Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium.” Translational Psychiatry 6 (3): e769. doi:10.1038/tp.2016.36. http://dx.doi.org/10.1038/tp.2016.36.

Published Version
doi:10.1038/tp.2016.36

Citable link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:27320365

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA
Cannabis is the most widely produced and consumed illicit psychoactive substance worldwide. Occasional cannabis use can progress to frequent use, abuse and dependence with all known adverse physical, psychological and social consequences. Individual differences in cannabis initiation are heritable (40–48%). The International Cannabis Consortium was established with the aim to identify genetic risk variants of cannabis use. We conducted a meta-analysis of genome-wide association data of 13 cohorts (N = 32 330) and four replication samples (N = 5627). In addition, we performed a gene-based test of association, estimated single-nucleotide polymorphism (SNP)-based heritability and explored the genetic correlation between lifetime cannabis use and cigarette use using LD score regression. No individual SNPs reached genome-wide significance. Nonetheless, gene-based tests identified four genes significantly associated with lifetime cannabis use: NCAM1, CADM2, SCOC and KNT2. Previous studies reported associations of NCAM1 with cigarette smoking and other substance use, and those of CADM2 with body mass index, processing speed and autism disorders, which are phenotypes previously reported to be associated with cannabis use.

1Department of Complex Trait Genetics, VU Amsterdam, Center for Neurogenomics and Cognitive Research, Amsterdam, The Netherlands; 2Department of Psychiatry, Academic Medical Centre, Amsterdam, The Netherlands; 3Department of Biological Psychology/Netherlands Twin Register, VU University, Amsterdam, The Netherlands; 4Neuroscience Campus Amsterdam, Amsterdam, The Netherlands; 5Department of Developmental Psychology and EMGO Institute for Health and Care Research, VU University, Amsterdam, The Netherlands; 6The Hospital for Sick Children Research Institute, Toronto, Canada; 7Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL, USA; 8Department of Human Neurogenetics, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands; 9Department of Psychology, University of Minnesota, Minneapolis, MN, USA; 10Department of Public Health, Hietl Institute, University of Helsinki, Helsinki, Finland; 11Estonian Genome Center, University of Tartu, Tartu, Estonia; 12Department of Epidemiology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 13Psychiatric Genetics Unit, VU University Medical Center, Amsterdam, The Netherlands; 14Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; 15Biomedical Network Research Centre on Mental Health (CIBERSAM), Barcelona, Spain; 16Department of Psychiatry, Hospital Universitari Vall d'Hebron, Barcelona, Spain; 17Biomedical Research Institute (HVB), Universitat Autònoma de Barcelona, Barcelona, Spain; 18Department of Psychiatry, Hospital Universitari Vall d’Hebron, Barcelona, Spain; 19Department of Medicine, Harvard Medical School, Boston, MA, USA; 20Department of Social and Community Medicine, University of Bristol, Bristol, UK; 21School of Social and Developmental Genetics, University of Toronto, Toronto, ON, Canada; 22MRC Integrative Epidemiology Unit (IEU), University of Bristol, Bristol, UK; 23Department of Psychiatry, Virginia Institute for Psychiatric and Behavior Genetics, Virginia Commonwealth University, Richmond, VA, USA; 24Research Centre Adolescent Development, Utrecht University, Utrecht, The Netherlands; 25Department of Psychiatry and Psychology, University of California San Diego, La Jolla, CA, USA; 26The LifeLines Cohort Study, University of Groningen, Groningen, The Netherlands; 27Department of Psychiatry and Legal Medicine, University Autònoma de Barcelona, Barcelona, Spain; 28Department of Psychiatry, University of Toronto, Toronto, ON, Canada; 29Department of Medicine, Harvard Medical School, Boston, MA, USA; 30Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA; 31Brain and Mind Research Institute, University of Sydney, Sydney, Australia; 32Department of Psychiatry, University of Colorado Denver, Aurora, CO, USA; 33Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland; 34Department of Mental Health and Substance Abuse Services, National Institute for Health and Welfare, Helsinki, Finland; 35Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, USA; 36Department of Molecular, Cellular and Developmental Biology, University of Colorado Boulder, Boulder, CO, USA; 37Department of Psychology, University of Notre Dame, Notre Dame, IN, USA; 38Developmental Psychology, Tilburg University, Tilburg, The Netherlands; 39Interdisciplinary Center for Pathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands; 40Department of Physiology and Nutritional Sciences, University of Toronto, Toronto, ON, Canada; 41Department of Psychological and Brain Sciences, Indiana University Bloomington, Bloomington, IN, USA; 42Department of Psychology and Neuroscience, Institute for Behavioral Genetics, University of Colorado Boulder, Boulder, CO, USA; 43Department of Psychiatry, University of California San Diego, La Jolla, CA, USA; 44Rotman Research Institute, Baycrest, Toronto, ON, Canada; 45Department of Psychology and Psychiatry, University of Toronto, Toronto, ON, Canada; 46Center for the Developing Brain, Child Mind Institute, New York, NY, USA; 47Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; 48UK Centre for Tobacco and Alcohol Studies and School of Experimental Psychology, University of Bristol, Bristol, UK; 49Department of Psychiatry, Genetics, and Neurobiology, Yale University School of Medicine and VA CT, West Haven, CT, USA; 50Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands. Correspondence: Professor JM Vink, Behavioural Science Institute, Radboud University, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands. E-mail: j.vink@bsi.ru.nl

Received 11 December 2015; accepted 21 December 2015
INTRODUCTION

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

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

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

Candidate gene studies, including reports examining the CNR1, GABRA2, FAAH and ABO1 genes have detected some significant associations with cannabis use but again, replication has been inconsistent. On the basis of a sample of 7452 Caucasian individuals, Verweij et al. found no gene-based associations between the frequency of cannabis use and 10 candidate genes identified by Agrawal and Lynskey. Overall, the results of candidate-gene studies are inconclusive; some associations have been replicated a few times, but failed to replicate in other studies. Moreover, findings may be further distorted due to publication bias favouring significant results.

As an alternative to the candidate-gene approach, the genome-wide association study (GWAS) is a hypothesis-free method that aims to detect novel genetic variants involved in complex traits. To date, three GWASs of cannabis use phenotypes have been published: one GWAS of cannabis dependence in 708 cannabis-dependent individuals and 2346 controls; a GWAS meta-analysis of lifetime cannabis use based on two studies with a combined sample size of 10,091 individuals (40.7% users); and a recent GWAS of lifetime cannabis use and age of cannabis use onset based on a sample of 6744 individuals (of whom 20% were users). None of the studies identified any genome-wide significant associations. This was likely due to the small effect sizes typical of common variants underpinning highly polygenic traits, thereby indicating a need for larger sample sizes. In this context, the success of larger GWASs and international consortia examining a variety of complex traits is encouraging. For example, multiple large meta-analyses of GWAS results for number of cigarettes smoked per day have independently identified associations on chromosome 15q25 spanning the α5, α3 and β4 nicotinic receptor subunit gene clusters (CHRNA5, CHRNA3, CHRN4).

These and other recent GWA findings clearly illustrate the need for larger sample sizes. In response to this need, the International Cannabis Consortium was established to combine the results of multiple GWA studies to identify the genetic variants underlying individual differences in cannabis use phenotypes. Our rationale for focusing on lifetime cannabis use (yes/no) is because this phenotype is heritable and shares significant genetic risks with that risk for cannabis abuse or dependence. In contrast to frequency of use or abuse and dependence, which are not commonly assessed in large-scaled genetic studies, most general population studies have assessed lifetime cannabis use, thereby increasing our sample size and power to detect associations. Currently, the combined International Cannabis Consortium sample size for lifetime cannabis is 32,330 individuals from 13 cohorts from Europe, the United States and Australia, along with four independent replication samples comprising 5627 individuals. This sample size is considerably larger than the sample size of the previous GWAS investigating lifetime cannabis use in two samples from Australia and the UK, thereby providing substantially greater power to detect genetic variants of small effect size. The aim of the present study is to identify genetic variants associated with lifetime cannabis use by meta-analysis of the GWAS results from all contributing International Cannabis Consortium samples. The tests of association for individual genetic variants will be complemented with gene-based tests of association. In addition, we will investigate which proportion of the heritability inferred by twin studies is explained by common SNPs captured on GWAS arrays. Finally, we will estimate the genetic correlation between lifetime cannabis and smoking initiation based on the analysis of our summary statistics and those from the publicly available Tobacco Alcohol and Genetics consortium.

MATERIALS AND METHODS

Cohorts

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

Four additional independent samples with a total of 5627 subjects were used for replication. One sample (n = 2660) consisted of African American...
Genotype imputation was based on the 1000 Genomes phase 1 reference data. Although some groups did do imputation using a logistic regression model including covariates (see above). For family-based samples, familial relatedness was taken into account by using a sandwich correction as implemented in PLINK.35 The analyses plan can be found in Supplementary Information 1. The proportion of phenotypic variance that could be explained by the SNPs was estimated using the density estimation method developed by So et al.41 The density estimation method estimates the genome-wide distribution of effect sizes on the difference between the observed distribution of test statistics in the meta-analysis and the corresponding null distribution. Before estimation, the SNPs present in at least 25% of the meta-analysis samples were pruned for LD. We used the $r^2 = 0.15$ pruning level as the primary result for consistency with other applications of this method. Additional details are located in the Supplementary Information 2. The meta-analysis was performed using a fixed-effects meta-analysis based on the cohort’s effect sizes and standard errors using METAL.36 Our meta-analysis combined association summary statistics for 6,444,471 SNPs that passed all the filters. We applied the conventional threshold of $5 \times 10^{-8}$ as an indication of genome-wide significance (see ref. 37). Although the combined sample size of the meta-analysis based on the discovery samples is 32,330, the sample size per SNP varies due to missingness across subsamples.

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

Statistical analyses

**GWA analysis in each discovery cohort.** The GWA analyses were performed by each group separately. Associations between the binary phenotype and the genotypes were tested genome-wide using a logistic regression model including covariates (see above). For family-based samples, familial relatedness was taken into account by using a sandwich correction as implemented in PLINK.35 The analyses plan can be found in Supplementary Information 3. It should be noted that some groups did do the analyses in a slightly different manner based on the characteristics of their sample. The proportion of phenotypic variance that could be explained by the SNPs was estimated using the density estimation method developed by So et al.41 The density estimation method estimates the genome-wide distribution of effect sizes on the difference between the observed distribution of test statistics in the meta-analysis and the corresponding null distribution. Before estimation, the SNPs present in at least 25% of the meta-analysis samples were pruned for LD. We used the $r^2 = 0.15$ pruning level as the primary result for consistency with other applications of this method. Additional details are located in the Supplementary Information 2. The meta-analysis was performed using a fixed-effects meta-analysis based on the cohort’s effect sizes and standard errors using METAL.36 Our meta-analysis combined association summary statistics for 6,444,471 SNPs that passed all the filters. We applied the conventional threshold of $5 \times 10^{-8}$ as an indication of genome-wide significance (see ref. 37). Although the combined sample size of the meta-analysis based on the discovery samples is 32,330, the sample size per SNP varies due to missingness across subsamples.

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

**Estimation of SNP-based heritability and genetic overlap with lifetime cigarette smoking.** The proportion of phenotypic variance that could be explained by the SNPs was estimated using the density estimation method developed by So et al.41 The density estimation method estimates the genome-wide distribution of effect sizes on the difference between the observed distribution of test statistics in the meta-analysis and the corresponding null distribution. Before estimation, the SNPs present in at least 25% of the meta-analysis samples were pruned for LD. We used the $r^2 = 0.15$ pruning level as the primary result for consistency with other applications of this method. Additional details are located in the Supplementary Information 2. The meta-analysis was performed using a fixed-effects meta-analysis based on the cohort’s effect sizes and standard errors using METAL.36 Our meta-analysis combined association summary statistics for 6,444,471 SNPs that passed all the filters. We applied the conventional threshold of $5 \times 10^{-8}$ as an indication of genome-wide significance (see ref. 37). Although the combined sample size of the meta-analysis based on the discovery samples is 32,330, the sample size per SNP varies due to missingness across subsamples.
RESULTS

Meta-analysis

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

Gene-based tests

The gene-based tests of associations were run on 24 576 genes/genetic regions (see ‘Materials and Methods’ section for details). The Manhattan and QQ plot for this test are shown in Figures 1a and b. Results for the top 100 genes can be found in Supplementary Table 5. As shown in Table 3, four genes and one intergenic noncoding RNA region were significantly (false discovery rate-corrected $P < 0.05$) associated with lifetime cannabis use: (i) neural cell adhesion molecule 1 ($NCAM1$, on 11q23); (ii) cell adhesion molecule 2 ($CADM2$, on 3p12); (iii) short coiled-coil protein (SCOC) and (iv) SCOC antisense RNA1 (SCOC-AS1), both located on 4q31; and (v) potassium channel, subfamily T, member 2 ($KCNV2$, on 1q31). Regional plots of these top genes are located in Supplementary Figure 6.

The smallest gene-based $P$-value was found for the $NCAM1$ gene. Within this gene, rs4471463 had the lowest SNP $P$-value, and was also among the top 10 associated SNPs. The forest plot in Figure 2 illustrates the effect of this SNP on each sample. In most samples, the effect is in the same direction, such that the major (T) allele is associated with a decreased risk of lifetime cannabis use. The forest plot for two SNPs with lowest $P$-values in the other significant gene regions can be found in Supplementary Figure 5.

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

SNP-based heritability and genetic overlap with lifetime cigarette smoking

Using the density estimation method (see ‘Materials and Methods’ section for a description), all the SNPs available in at least 25% of the samples when combined explained 20% of the total variance in lifetime cannabis use ($P < 0.001$). Alternative estimation with LD score regression also yielded a significant heritable component of 13% ($h^2_{LD} = 0.13$, s.e. = 0.02, $P = 1.4 \times 10^{-5}$). These variance estimates were robust across pruned sets with similar $r^2$ thresholds (see Supplementary Table 6). Stricter LD pruning (that is, $r^2 = 0.05$), or restricting analyses to SNPs present in all studies substantially decreased the estimate of variance explained. Both SNP heritability estimates confirmed that lifetime cannabis use has a significant heritable component ($13–20\%$), indicating that GWAS should be able to identify these common SNPs (but effect sizes are small and large sample sizes are thus required). However, because these estimates are only based on common SNPs, the total heritability of lifetime cannabis use is likely to be higher.

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

DISCUSSION

To date, this is the largest GWA study of lifetime cannabis use. We performed meta-analysis of the GWA results based on a discovery sample comprising 32 330 subjects from 13 cohorts, and a replication sample comprising 5627 subjects from four cohorts (including one African American cohort). There were no genome-wide significant SNP associations. However, heritability analyses revealed that between 13 and 20% of the variation in lifetime cannabis use could be explained by common SNPs. Moreover,
gene-based tests of association identified four protein-coding genes and one intergenic region significantly associated with lifetime cannabis use including NCAM1, which has previously been linked to substance use.45–48 Finally, we revealed that the genetic liability to lifetime cannabis use correlated to a large extent (r = 0.83) with the genetic liability to lifetime cigarette smoking. Our results are consistent with the hypothesis that lifetime cannabis use is a highly polygenic trait, comprising many SNPs each with small effects contributing to lifetime risk. Moreover, portions of the genetic risk in lifetime cannabis use likely correlates with other substances including cigarette smoking.

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

| Gene     | Chr | Start position (hg19) | BP length | N SNPs | Nominal P-values discovery | Corrected P-values discovery | Nominal P-values discovery EU replication samples | Nominal P-values replication African Americans |
|----------|-----|-----------------------|-----------|--------|-----------------------------|------------------------------|-----------------------------------------------|-----------------------------------------------|
| NCAM1    | 11  | 112831968             | 303 952   | 400    | 6.26 × 10^{-7}              | 0.015                        | 0.381                                         | 0.302                                         |
| CADM2    | 3   | 85008132              | 1 115 448 | 978    | 2.13 × 10^{-6}              | 0.026                        | 0.744                                         | 0.112                                         |
| SCOC-AS1 | 4   | 141204879             | 89 668   | 81     | 5.76 × 10^{-6}              | 0.046                        | 0.681                                         | 0.044                                         |
| SCOC     | 4   | 141264614             | 39 097   | 111    | 7.85 × 10^{-6}              | 0.046                        | 0.636                                         | 0.027                                         |
| KCNT2    | 1   | 196194909             | 382 653  | 237    | 9.38 × 10^{-6}              | 0.046                        | 0.269                                         | 0.201                                         |

Abbreviations: BP length, base pair length; chr, chromosome; hg19, human genome version 19; N SNPs, number of SNPs used for the meta-analysis; SNP, single-nucleotide polymorphism.
has been associated with smoking behavior and nicotine dependence, alcohol dependence, heroin dependence, as well as other substance use disorders. Interestingly, these phenotypes were associated with lifetime cannabis use implying that these genes are appropriate targets for future functional studies of cannabis use. Unfortunately, our gene-based results were not replicated in the replication samples, probably due to low sample sizes and therefore low power. In the African American replication sample, we did find suggestive association with SCOC-AS1 and SCOC.

On the basis of twin studies, the heritability of lifetime cannabis use is estimated at 40–50%. In our study, all common SNPs combined explained 13–20% of the variance in the liability to use cannabis depending on the method used. Stricter LD pruning (that is, r² = 0.05) or restricting to SNPs observed (genotyped or imputed) in all the analyses, substantially reduces the estimate of variance explained. Speculatively, this may indicate that much of the variance explained comes from SNPs located in the regions of weak LD. Such effects are likely to be poorly tagged for the estimation of variance explained after strict LD pruning, and are likely to be more difficult to impute owing to a lack of strongly correlated genotyped SNPs (and thus missing from some studies). Our SNP-based heritability estimates lie in between two previous heritability estimates for lifetime cannabis use based on the Genome-wide Complex Trait Analysis software package. Verweij et al. estimated that 6% of the variance in lifetime cannabis use is explained by aggregated common SNPs (MAF > 0.05). Minică et al. found an estimate of 25%. Provided that the current sample is much larger than the samples used in the previous studies, we conclude that approximately one-third to half of the heritability is explained by common SNPs captured on a GWAS array. Other sources of variation may explain the discrepancy between SNP- and twin-based heritability estimates. For example, age-related genetic differences, non-additive genetic variance, interactions between genetic variants and environmental risk factors, epistasis and/or rare mutations may also have a role.

Our results indicate a very high genetic overlap (r² = 0.83) between our measure of lifetime cannabis use and lifetime cigarette use when based on the SNP panel. Twin studies have shown moderate to high genetic correlations of 0.59–0.74 between lifetime cannabis and nicotine use. Combining variants within larger units (that is, genes) did however reveal four significant genes associated with lifetime cannabis use implying that these genes are important for future functional studies of cannabis use. Several reasons. First, complex traits are known to be influenced by many variants, each with very small effect sizes. Although power calculations reveals suitable power (96%) to detect odds ratios of 1.15 based on common SNPs (MAF = 0.2), the power to detect smaller effect sizes remains lower. For example, there is only 28% power to detect effect sizes with odds ratio of 1.1 and MAF = 0.2. Therefore, our data suggest that the effect sizes of single variants contributing to lifetime cannabis use are likely to be smaller than 1.15. Combining variants within larger units (that is, genes) did however reveal four significant genes associated with lifetime cannabis use implying that these genes are appropriate targets for future functional studies of cannabis use. Unfortunately, our gene-based results were not replicated in the replication samples, probably due to low sample sizes and therefore low power. In the African American replication sample, we did find suggestive association with SCOC-AS1 and SCOC.

Figure 2. Forest plot for the top-SNP rs4471463 in the NCAM1 gene on chromosome 11. SNP, single-nucleotide polymorphism.
samples were representative of the same population of users. Finally, the average age of participants varied between 18 (ALSPAC) and 45 (QIMR) years. Consequently, some younger participants might have initiated cannabis use at a later age, but have been classified as ‘never users’ in the current study. This can decrease power, but does not invalidate our results. In addition, we note that the average age of each sample did not correlate with sample prevalences ($r = -0.04, P = 0.91$).

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

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

CONCLUSION

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

CONFLICT OF INTEREST

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

ACKNOWLEDGMENTS

JMV, CCM and HM are supported by the European Research Council (Beyond the Genetics of Addiction ERC-284167, PI JMW). SS and EMD are supported by the Foundation Volksbond Rotterdam. KJHV is supported in part by the Netherlands Organization for Health Research and Development (ZonMW 311600212) and in part by a 2014 NARSAD Young Investigator Grant from the Brain and Behavior Research Foundation. NAG is supported by US National Institutes of Health, National Institute on Drug Abuse R00DA023549. GH is supported by NIDA R37 DA-018673 and NSF BCS-1229450. RW is supported by NIH U01 MH094432 and NSF BCS-1229450. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org) which is supported by the Netherlands Organization for Scientific Research (NWO 480-05-003 PI: Posthumo) along with a supplement from the Dutch Brain Foundation and the VU University Amsterdam. The study site acknowledgments are as follows: 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, researchers, secretaries and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/13/2) and the University of Bristol provide core support for ALSPAC. GWAS data were generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. AIS is supported by a Wellcome Trust 4-year PhD studentship in molecular, genetic and lifestyle epidemiology (WT083431MA). JWW is supported by a Postdoctoral Research Fellowship from the Oak Foundation. JWW 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. JWW and MRM are also members of the UK Centre for Tobacco and Alcohol Studies, a UK Public Health Research Collaborative. 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 NAG, by the Australian Research Council to MJW (Nos. DP0343921, DP0664638 and DP1093900) and by Australian Medical Health and Medicines Research Councils. Research Councils awarded to Ian Hickie (No. 646914) and GWM (No. 619667). We acknowledge and thank the following project staff: AKH, Leanne Wallace and Lisa Bowdler for the laboratory processing, genotyping and QC; Soad Hacock as Project Coordinator; Lenore Sullivan as Research Editor; our research interviewers Pieta-Marie Shertock and Jill Wood; and David Smyth for IT. We also thank the twins and their siblings for their willing cooperation. SEM was supported by an ARC Future fellowship FT110100548. CADD—the Center on Antisocial Drug Dependence (CADD) data reported here were funded by grants from the National Institute on Drug Abuse (P60 DA011015, R01 DA012845, R01 DA021913, R01 DA021905, R01 DA35804). EGCT—received funding by FP7 grants (278913, 306031, 313010), Center of Excellence in Genomics (EXCEGEN) and University of Tartu (SP1GIVARENG). We acknowledge EGCT technical personnel, especially Mr V Soo and S Smit. Data analyses were carried out in part at the High Performance Computing Unit at the University of Tartu. FinnTwin—We warmly thank the participating twin pairs and their family members for their contribution. We express our appreciation to the skilled study interviewers A-M Livonen, K Karhu, H-M Kuha, U Kulmala-Grähn, M Mantere, K Saanakorpi, M Saarinen, R Sipila, L Viljanen and E Voipio. Anja Höppel and Kauko Heikkilä are acknowledged for their valuable contribution in recruitment, data collection and data management. Phenotyping and genotyping of the Finnish twin cohorts has been supported by the Academy of Finland Center of Excellence in Complex Disease Genetics (grants 213506, 129680), the Academy of Finland (grants 100499, 205585, 118555, 141054, 265240, 263278 and 264146 to JK), National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145, and AA-09203 to RRR and AA15416 and K02AA018755 to DM Dick), Sigrid Juselius Foundation (to JK), and the Welcome Trust Sanger Institute, UK. Anti-Pelka Sarin and Samuli Ripatti are acknowledged for genotype data quality controls and imputation. GWAS analyses were run at the ELIXIR Finland node hosted at CSC—IT Center for Science for ICT resources. HUVH—Financial support was received from ‘Instituto de Salud Carlos III’ (P11/00671, P11/01629, P11/01139), ‘Plan Nacional Sobre Drogas’ (PNSD2011-0080), ‘Agencia de Gestión de Atuados Universitarios i de Recerca-AAGUR, Generalitat de Catalunya’ (2014SGR1357) and ‘Departament de Salut’, Government of Catalonia, Spain. MR is a recipient of a Miguel de Servert contract from the ‘Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación’, Spain. MCFTR—This research was supported in part by USPHS Grants from the National Institute on Alcohol Abuse and Alcoholism (AA09367 and AA11886), the National Institute on Drug Abuse (DA05147, DA13240 and DA024417) and the National Institute on Mental Health (MH061640). NTR—We thank the Netherlands Twin Register participants whose data we analyzed in this study. This work was supported by grants from the Netherlands Organization for Scientific Research (ZonMW Addiction 31160008; ZonMW 940-37-024; NWO/SPI 56-464-14192; NWO-400-05-717; NWO-MW 904-61-19; NWO-MagW 480-40-004; NWO-Veni 016-015-035), 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/2008/02. We acknowledge the EMGO+ Institute for Health and Care Research, the Neuroscience Campus Amsterdam, BBMRI-NL (18.4.201:007; 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 (R03DA018673-06, RC2 MH089951), Rutgers University Cell and DNA Repository cooperative agreement (National Institute of Mental Health U42 MH68457-06), and the National Institutes of Health (NH R01 HD02157-01A1, McIntosh Center and Oppenheimer Foundation funds). Genotyping 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 Organization for Scientific Research (NWO 480-05-003), the Dutch Brain Foundation and the Department of Psychology and Education of the VU University Amsterdam. QIMR—This is supported by National Institutes of Health (AA07535, AA07530, AA08652 and AA03149), AA Research (R01 AA2449, AA13320, AA AA12187) and the National Institute of Mental Health (AA199720, DA00768, DA012854, DA018627, DA018660, DA23668 and DA019951); by grants from the Australian National Health and Medical Research Council (214994, 339462, 398927, 398875, 398891, 398938, 442915, 442981, 469739, 552485, 552498 and 628911); by
Grants from the Australian Research Council (A7960034, A7990658, 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 (CRES-P—Australia). We thank AKH, Richard Parker, Soad Hancock, Judith Moir, Sally Rodda, Piet-a-Maree Shertock, Heather Park, Jill Wood, Pam Barton, Fran Husband, Adele Somerville, Ann Eldridge, Mark Grant, Diana Spooner, Lisa Bowdler, Agnieszka Todorow, Steven Cooks, David Smyth, Harry Beeby and Daniel Park. Last, we thank the twins and their families for their participation. RADAR—We thank all the adolescents and their families and friends for their participation. Moreover, we thank the various assistants that helped in recruiting participants as well as collecting and cleaning the data. The research was funded partly by the Netherlands Organisation for Scientific Research (Brain and Cognition, 056-21-010). Data of the RADAR study were used. RADAR has been financially supported by environmental research funding from the Netherlands Organisation for Scientific Research (GB-MAGW 480-03-005), and Stichting Achmea Slachtoffer en Samenleving (SASS) and various other grants from the Netherlands Organisation for Scientific Research, the VU Amsterdam University and Utrecht University. ACH is supported by the Netherlands Organization for Health Research and Development, ZonMW 31160212. Sangueney Youth Study—The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the SYS (TP, ZP). TP is the Tenanbaum Chair in Population Neuroscience (University of Toronto) and the Dr John and Consuela Phelan Scholar (Child Mind Institute). TRAILS—TRAILS (Tracking Adolescents’ Individual Lives Survey) is a collaborative project involving various departments of the University Medical Center and University of Groningen, the Erasmus University Medical Center Rotterdam, the University of Utrecht, the Radboud Medical Center Nijmegen and the Pammasia 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 60-06060-97-118; ZonMw Culture and Health grant 261-98-710; Social Sciences Council medium-sized investment grants GB-MAGW 480-01-006 and GB-MAGW 480-07-001; Social Sciences Council project grants GB-Magw 452-04-324 and GB-Magw 452-06-004; NWO large-sized investment grant 175.010.2003.005; NWO Longitudinal Survey and Panel Funding 481-08-013; the Dutch Ministry of Justice (WODC), the European Science Foundation (EuroSTRESS project FP-006), Biobanking and Biomolecular Resources Research Infrastructure BBMRI-NL (CP 32), the participating universities and Accare Center for Child and Adolescent Psychiatry. We are grateful to all the adolescents, their parents and teachers who participated in this research and to everyone who worked on this project and made it possible. Statistical analyses were carried out on the Genetic Cluster Computer (http://www.geneticcluster.org), which is financially supported by the Netherlands Scientific Organization (NWO 480-05-003) along with a supplement from the Dutch Brain Foundation. TwinsUK—The study was funded by the Wellcome Trust; European Community’s Seventh Framework Program (FP7-2007-2013). The study was financially supported by grants from the National Institute of Diabetes and Dineet BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. SNP Genotyping was performed by The Wellcome Trust Sanger Institute and National Eye Institute via NIH/C. Utrecht—We are grateful to Chris Schubart and Wim Jennis and numerous students for their work in the study. Foremost, we thank the study participants. This study was financially supported by grants 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 the National Institutes of Health grants R2C DA028909, R01 DA12690, R01 DA12849, R01 DA18432, R01 AA13330, R01 AA017535, and the VA Connecticut and Philadelphia VA MIRECC.

REFERENCES
1 United Nations Office on Drugs and Crime. 2010 World Drug Report, New York, USA.
2 Hall W, Solovij N. Adverse effects of cannabis. Lancet 1998; 352: 161–166.
3 Hall W, Babor TF. Cannabis use and public health: assessing the burden. Addiction 2000; 95: 485–490.
4 Aggarwal SK, Carter GT, Sullivan MD, ZumBrunnen C, Morrill R, Mayer JD. Medicinal use of cannabis in the United States: historical perspectives, current trends, and future directions. J Opioid Manag 2009; 5: 153–168.
5 Hall W. What has research over the past two decades revealed about the adverse health effects of recreational cannabis use? Addiction 2014; 110: 19–35.
6 Volkow ND, Compton WM, Weiss SR. Adverse health effects of marijuana use. N Engl J Med 2014; 371: 879.
33 Kendler KS, Karkowski LM, Corey LA, Prescott CA, Neale MC. Genetic and environ-
mental risk factors in the etiology of illicit drug initiation and subsequent
misuse in women. Br J Psychiatry 1999; 175: 351–356.

34 Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE et al. 
An integrated map of genetic variation from 1092 human genomes. Nature 2012; 
491: 56–65.

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

36 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-
wide association scans. Bioinformatics 2010; 26: 2190–2191.

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

38 Li MX, Gui HS, Kwan JS, Sham PC. GATES: a rapid and powerful gene-based association 
test using extended Simes procedure. Am J Hum Genet 2011; 88: 283–293.

39 Li MX, Kwan JS, Sham PC. HYS'T: a hybrid set-based test for genome-wide association
studies, with application to protein-protein interaction-based association analysis. 
Am J Hum Genet 2012; 91: 478–488.

40 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and
powerful approach to multiple testing. J R Stat Soc Ser B 1995; 57: 289–300.

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

42 Bulik-Sullivan BK, Lah PR, Finucane HK, Ripke S, Yang J. LD Score regression
estimates confounding from polygenicity in genome-wide association studies. 
Nat Genet 2015; 47: 291–295.

43 Bulik-Sullivan BK, Finucane HK, Anttila V, Gusev A, Day FR, Lah PR et al. An atlas of
genetic correlations across human diseases and traits. Nat Genet 2015; 47: 1236–1241.

44 Viechtbauer W. Conducting meta-analyses in R with the metafor package. J Stat 
Softw 2010; 36: 1–48.

45 Gelernter J, Yu Y, Weiss R, Brady K, Panhuysen C, Yang BZ et al. Haploptype 
spanning TTC12 and ANKK1, flanked by the DRD2 and NCAM1 loci, is strongly
associated to nicotine dependence in two distinct American populations. Mol 
Hum Genet 2006; 15: 3498–3507.

46 Rubinek T, Yu R, Hadani M, Barkai G, Nass D, Melmed S et al. The cell adhesion 
molecules N-cadherin and neural cell adhesion molecule regulate human growth 
hormone: a novel mechanism for regulating pituitary hormone secretion. J Clin 
Endocrinol Metab 2008; 83: 3724–3730.

47 Ducci F, Kaakinen M, Pouta A, Hartikainen AL, Veijola J, Isomaki M et al. TTC12-
ANKK1-DRD2 and CHRNA5-CHRNA3-CHRNB4 influence different pathways leading
to smoking behavior from adolescence to mid-adulthood. Biol Psychiatry 2011; 
69: 650–660.

48 Bidwell LC, McGary JE, Gray JC, Palmer RH, Knopik VS, MacKillop J. NCAM1-
TTC12-ANKK1-DRD2 variants and smoking motives as intermediate phenotypes for 
nicotine dependence. Psychopharmacology 2015; 232: 1177–1186.

49 Gelernter J, Panhuysen C, Weiss R, Brady K, Poling J, Krauthammer M et al. 
Genomewide linkage scan for nicotine dependence: identification of a chromo-
some 5 risk locus. Biol Psychiatry 2007; 61: 119–126.

50 Laucht M, Becker K, Frank J, Schmidt MH, Esser G, Treutlein J et al. Genetic 
variation in dopamine pathways differentially associated with smoking progress-
in adolescence. J Am Acad Child Adolesc Psychiatry 2008; 47: 673–681.

51 Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PAF et al. 
Cholinergic nicotinic receptor genes implicated in a nicotine dependence asso-
cia tion study targeting 348 candidate genes with 3713 SNPs. Hum Mol Genet 
2007; 16: 36–49.

52 Munafò M, Clark T, Johnstone E, Murphy M, Walton R. The genetic basis for
smoking behavior: a systematic review and meta-analysis. Nicotine Tob Res 2004; 
6: 583–597.

53 Yang BZ, Kranzler HR, Zhao H, Gruen JR, Luo X, Gelernter J. Association of
haploptypic variants in DRD2, ANKK1, TTC12 and NCAM1 to alcohol dependence in
independent case-control and family samples. Hum Mol Genet 2007; 16: 
2844–2853.

54 Yang BZ, Kranzler HR, Zhao H, Gruen JR, Luo X, Gelernter J. Haploptypic variants in 
DRD2, ANKK1, TTC12, and NCAM1 are associated with comorbid alcohol and drug 
dependence. Alcohol Clin Exp Res 2008; 32: 2117–2127.

55 Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL et al. ANKK1, 
TTC12, and NCAM1 polymorphisms and heroin dependence: importance of 
considering drug exposure. JAMA Psychiatry 2013; 70: 325–333.

56 Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU et al. 
Association analyses of 249,796 individuals reveal 18 new loci associated with 
body mass index. Nat Genet 2010; 42: 937–948.

57 Ibrahim-Verbaas CA, Bressler J, Debette S, Schuur M, Smith AV, Bis JC et al. GWAS
for executive function and processing speed suggests involvement of the 
CADD2 gene. Mol Psychiatry 2015; 21: 189–197.

58 Casey JP, Magalhaes T, Conroy JM, Regan R, Shah N, Anney R et al. A novel 
approach of homoyzogous haplotype sharing identifies candidate genes in autism
spectrum disorder. Hum Genet 2012; 131: 565–579.

59 Hayatbakhsh MR, O’Callaghan NJ, Mamun AA, Williams GM, Clavarrino A, Najman
JM. Cannabis use and obesity and young adults. Am J Drug Alcohol Abuse 2010; 
36: 350–356.

60 Kelleher LM, Stough C, Sergejev AA, Rolfe T. The effects of cannabis on
information-processing speed. Addict Behav 2004; 29: 1213–1219.

61 De Alwis D, Agrawal A, Reiersen AM, Constantino JN, Henders A, Martin NG et al. 
ADHD symptoms, autistic traits, and substance use and misuse in adult
Australian twins. J Stud Alcohol Drugs 2014; 75: 211–221.

62 Mason JM, Arndt KM. Coiled coil domains: stability, specificity, and 
biological implications. Chembiochem 2004; 5: 170–176.

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: 64–74.

64 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 iden-
tified as a risk gene. Mol Psychiatry 2014; 19: 717–723.

65 Yang JA, Lee SH, Goddard ME, Visser PM. GCTA: a tool for genome-wide complex 
trait analysis. Am J Hum Genet 2011; 88: 76–82.

66 Agrawal A, Silberg JL, Lynskey MT, Maes HH, Eaves LJ. Mechanisms underlying the
life-time co-occurrence of tobacco and cannabis use in adolescent and young
adult twins. Drug Alcohol Depend 2010; 108: 49–55.

67 Kendler KS, Schmitt E, Aggen SH, Prescott CA. Genetic and environmental influ-
ences on alcohol, caffeine, cannabis, and nicotine use from early adolescence
to middle adulthood. Arch Gen Psychiatry 2008; 65: 674–682.

68 Manchia M, Cullis J, Turecki G, Rouleau GA, Uher R, Alda M. The impact of phe-
notypic and genetic heterogeneity on results of genome wide association studies of
complex diseases. PLoS One 2013; 8: e76295.

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

70 Agrawal A, Lynskey MT, Pergadia ML, Bucholz KK, Heath AC, Martin NG et al. Early
onset cannabis use and DSM-IV nicotine dependence: a twin study. Addiction 2008; 103:
1896–1904.

71 Lynskey MT, Vink JM, Boomsma DI. Early onset cannabis use and progression to
other drug use in a sample of Dutch twins. Behav Genet 2006; 36: 195–200.

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