Addiction Biology

BRIEF REPORT

doi:10.1111/j.1369-1600.2012.00478.x

The genetic aetiology of cannabis use initiation: a meta-analysis of genome-wide association studies and a SNP-based heritability estimation

Karin J. H. Verweij[1,2], Anna A. E. Vinkhuyzen[1,3], Beben Benyamin[1,3], Michael T. Lynskey[4], Lydia Quaye[5], Arpana Agrawal[4], Scott D. Gordon[1], Grant W. Montgomery[1], Pamela, A. F. Madden[4], Andrew C. Heath[4], Timothy D. Spector[5], Nicholas G. Martin[1] & Sarah E. Medland[1]

Queensland Institute of Medical Research, Australia[1], School of Psychology, University of Queensland, Australia[2], Queensland Brain Institute, University of Queensland, Australia[3], Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA[4], King's College London St.Thomas' Hospital, Twin Research and Genetic Epidemiology Department, UK[5]

ABSTRACT

While initiation of cannabis use is around 40% heritable, not much is known about the underlying genetic aetiology. Here, we meta-analysed two genome-wide association studies of initiation of cannabis use with > 10 000 individuals. None of the genetic variants reached genome-wide significance. We also performed a gene-based association test, which also revealed no significant effects of individual genes. Finally, we estimated that only approximately 6% of the variation in cannabis initiation is due to common genetic variants. Future genetic studies using larger sample sizes and different methodologies (including sequencing) might provide more insight in the complex genetic aetiology of cannabis use.

Keywords Association, cannabis, genetics, heritability.

Correspondence to: Karin J.H. Verweij, Genetic Epidemiology, Queensland Institute of Medical Research, 300 Herston Road, Brisbane, Queensland 4006, Australia. E-mail: karin.verweij@qimr.edu.au

Cannabis is the most widely used illicit drug worldwide, and numerous population-based family and twin studies indicate that the risk of cannabis use runs in families. A recent meta-analysis of existing twin studies reported heritability estimates of 45% for men and 39% for women for cannabis use initiation (Verweij *et al.* 2010). Not much is known about the genetic variants and biological mechanisms underlying this heritability. The few linkage studies that examined cannabis use phenotypes reported non-overlapping linkage peaks that did not meet genome-wide significance (see Agrawal & Lynskey 2009). Candidate gene association studies for cannabis use have mainly focussed on genes known to be involved in the endogenous cannabinoid system: the cannabinoid receptor 1 gene and the fatty acid amide hydrolase gene. But again, while some nominally significant associations were reported, others were unable to replicate these (see Agrawal & Lynskey 2009). A recent large-scale study testing for association of these two and eight other candidate genes with lifetime frequency of cannabis use did not support association with any of the genes (Verweij *et al.* 2012; NB: this study used the same Australian subsample as used in the present study). The lack of replication may point to potential publication bias and false-positive findings in candidate gene association studies and also highlights the limited understanding of the neurobiology of cannabis use. Other, hypothesis-free, approaches are therefore needed to further unravel the genetic aetiology of cannabis use.

Recently, Agrawal *et al.* (2011) performed the first genome-wide association study for cannabis use phenotypes, using a sample of 708 cannabis-dependent cases and 2346 controls. They did not identify any genetic variants significantly associated with cannabis dependence, potentially because of a lack of statistical power. The present study has two components. Firstly, we meta-analysed results from two genome-wide association analyses of initiation of cannabis use (ever versus never) in order to identify genetic variants underlying cannabis use. With a combined sample size of over 10 000

individuals (from 4622 independent families) and a substantially higher number of cases, this study provided more power to identify genetic variants of small effect size than the study by Agrawal *et al*. (2011). Secondly, we used a recently developed method utilizing genome-wide single-nucleotide polymorphism (SNP) data to estimate the overall percentage of variance in initiation of cannabis use that is due to common genetic variants, which will provide insight into the genetic aetiology underlying cannabis use.

Data used in this study come from Australian and UK (TwinsUK, http://www.twinsuk.ac.uk; Spector & Williams 2006) twin registries (see Table 1 for sample details). As part of larger questionnaires, individuals were asked whether they had ever used cannabis. Table 1 shows the prevalence of cannabis use initiation for individuals included in the present study. It should be noted that the prevalence is substantially higher in the Australian sample. Higher levels of cannabis use have generally been reported in Australia than in the United Kingdom (UNODC 2010), but the difference may be larger due to age differences between the two samples and because a subset of the Australian sample has been ascertained for familial alcohol and nicotine use.

Genotype data were obtained using different Illumina SNP platforms (317 K, HumanCNV370-Quadv3, Human CNV370v1 and Human610-Quad). Standard quality control procedures were applied as outlined previously (Medland *et al*. 2009), including checks for ancestry outliers, Hardy–Weinberg equilibrium, Mendelian errors, call rate and minor allele frequency. Subsequently, both datasets were imputed separately using Markov Chain Haplotyping software (Li *et al*. 2010) using reference data from the European HapMap I + II samples, (Release 22 Build 36). Only SNPs with an imputation quality score (r^2) greater than 0.3 were retained, and SNPs were filtered on allele frequency. In total, ~2.4 million SNPs were available for the association analyses.

For each sample separately, the dosage score at each SNP was tested for association with cannabis use initiation using the family-based association test implemented in Merlin-offline (Chen & Abecasis 2007), correcting for sex, age, age^2 and sex × age effects. We performed a total test of association, which employs a linear regression method while explicitly correcting for the relatedness between family members (including identical twins) within the variance/covariance model. The strongest associations were verified using the Generalized Disequilibrium Test (GDT, Chen, Manichaikul & Rich 2009), which is specifically designed for use with dichotomous data. However, unlike Merlin-offline, which can analyse dosage data, GDT can only be used with hard-call genotypes, which are less informative than dosage data thus, resulting in a less powerful test. A meta-analysis of the association results from both studies was conducted in METAL (Abecasis 2009) using the effect sizes of the SNPs along with their standard errors (SEs). Next, we conducted a gene-based test (Versatile Gene-based Association Study, Liu *et al*. 2010) in order to determine if particular genes harbour an excess of associated variants. This test summarizes evidence for association on a per gene basis by considering the *P* value of all SNPs within the gene, while accounting for linkage disequilibrium and number of SNPs in the gene.

Results from the meta-analysis are summarized in Figure 1 (Manhattan plot), Supporting Information Figure S1 (Q-Q plot), and Supporting Information Table S1 (strongest associations; note that the results were similar with Merlin-offline and GDT), and the strongest gene-based associations are shown in Supporting Information Table S2. We did not identify any SNP or gene that was significantly associated with cannabis use initiation, after correcting for multiple testing. An approximate power calculation (Purcell, Cherny & Sham 2003) indicates the combined samples provided 65 and 98% power to detect a genetic variant (with a minor allele frequency of 0.25) with a relative risk of 1.15 and 1.2, respectively. This indicates that individual common genetic variants of this size or greater do not contribute to individual differences in initiation of cannabis use. The power calculation

Table 1 Sample descriptives.

	Australia (n = 2538 independent families)			*United Kingdom (n = 2084 independent families)*		
	Males	*Females*	*Total*	*Males*	*Females*	*Total*
Sample size	3292	3883	7175	216	2700	2916
Age (mean ± standard deviation)	45.2 ± 10.8	44.5 ± 11.1	44.8 ± 11.0	58.7 ± 13.5	58.7 ± 12.2	58.7 ± 12.3
Percentage of individuals that have used cannabis	58.4%	47.4%	52.4%	18.5%	11.3%	11.8%

Note that the Australian sample includes 403 and the UK sample 224 identical twin pairs.

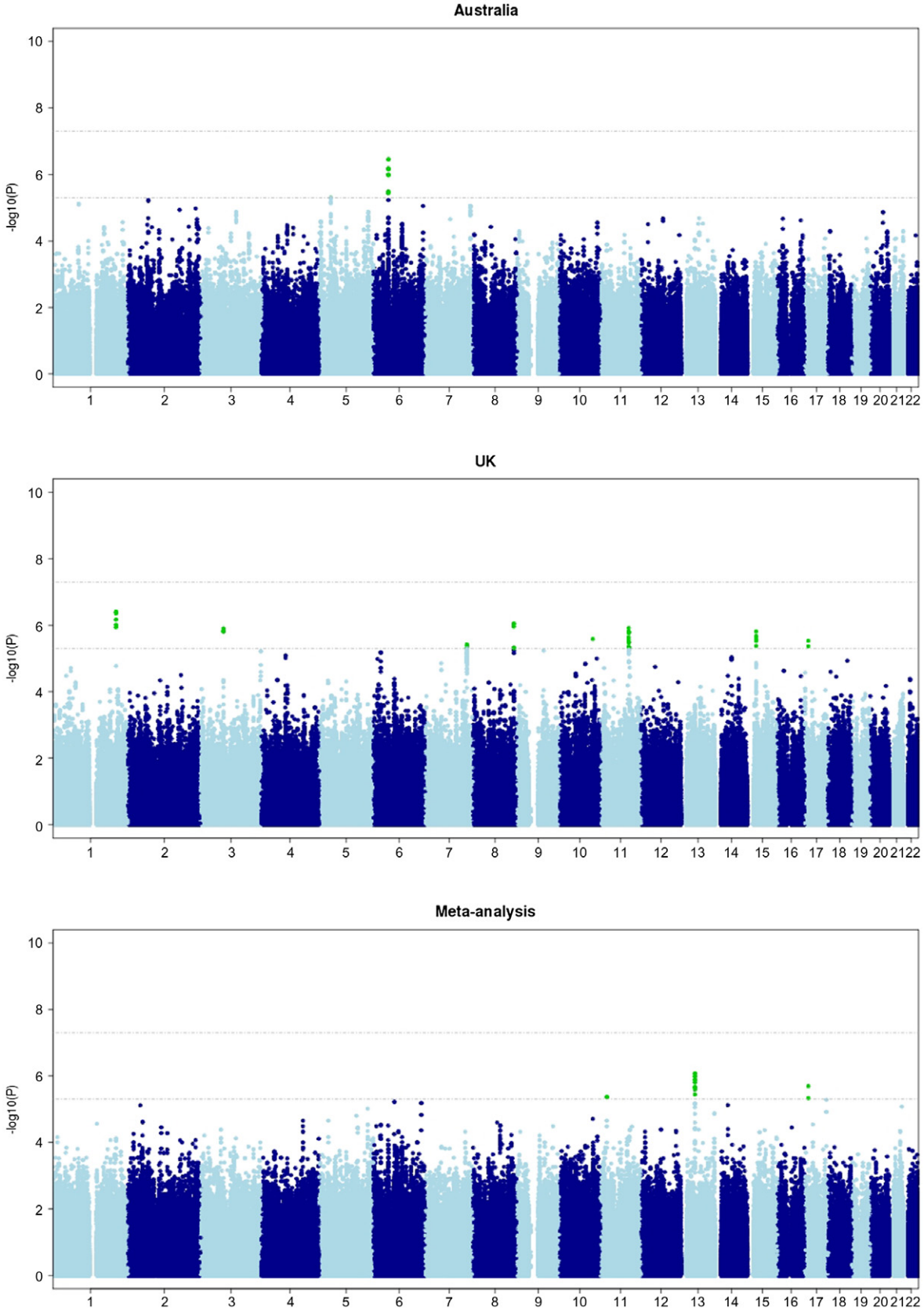

Figure 1 Results of the genome-wide association analyses for lifetime cannabis use. The *x*-axis shows the chromosome numbers and the *y*-axis the significance of the association signals [i.e. –log10(*P*) value]

shows that our sample is underpowered to detect common genetic variants of smaller effect sizes, indicating the need for larger sample sizes to enable the detection of these variants.

We then estimated the proportion of variance in initiation of cannabis use that could be explained by the aggregate effect of all SNPs using GCTA (Yang *et al.* 2011). This estimate is obtained by correlating the genetic similarity (i.e. the identity by state at the SNP level) between individuals (relatedness < 0.025) with their phenotypic similarity. The methodology is explained in more detail in the Supporting Information. We estimated that only 6.0% (SE = 10.2, $P = 0.28$) of the variance in cannabis use initiation is due to the aggregated effect of common variants. The relatively large SE indicates that this estimate is somewhat imprecise, mostly because of the limited power of the sample ($n = 4612$ unrelated individuals), but also by the measurement of the phenotype (binary, as opposed to continuous) and potentially by the heterogeneity of the phenotype between the two samples.

This SNP-based heritability estimate is substantially lower than the heritability estimate of ~40% obtained from twin studies (Verweij *et al.* 2010). The discrepancy between twin and SNP-based heritability estimates is larger than for some other phenotypes, such as height and intelligence (Yang *et al.* 2010; Davies *et al.* 2011), but similar to for example personality, where the SNP-based estimates for various personality traits are between 4 and 12% (Verweij *et al.* 2012; Vinkhuyzen *et al.* 2012), while the twin and family-based heritability estimates are between 30 and 60%. Note that heritability estimates from twin studies include the effects of all causal genetic variants, while the heritability estimated using GCTA includes only the effects of variants that are in linkage disequilibrium with the SNPs included in the analyses. These SNPs do not capture all genetic variants, especially not rare variants or variants with low minor allele frequencies. Although the SNP-based heritability estimate is somewhat imprecise because of limited power, the result raises the possibility that the role of common genetic variants in the heritability of initiation of cannabis use is low, which could help to explain why we were unable to find any genetic association. It may also suggest that non-additive genetic effects (dominance and/or epistasis), interactions with the environment, and/or rare mutations also play a role. Alternatively, twin studies may have overestimated the relative contribution of genetic influences (see Vinkhuyzen *et al.* 2012). As with other complex traits, future genetic studies using larger sample sizes and different methodologies (including sequencing) might provide more insight in the complex genetic aetiology of cannabis use.

Acknowledgements

We greatly thank the twins for their participation. This study was supported in part by grants AA013326, AA014041, AA13320, AA013321, DA23668 and DA12854 from the National Institutes of Health, Bethesda, Md. We would also like to thank the staff of the Twin Research Unit (KCL) for their help and support in undertaking this project. The Wellcome Trust provides core support for Department of Twin Research. KJHV is supported by an ANZ Trustees PhD scholarship in Medical Research. BB is the recipient of an NHMRC Biomedical Postdoctoral Fellowship (552498).

Conflict of interest

All authors declare that they have no conflict of interest.

Authors Contribution

KJHV and SEM were responsible for the study design. MTL, AA, GWM, PAFM, ACH, TDS and NGM contributed to the acquisition of the data. KJHV performed most analyses and was responsible for writing up the paper. AAEV, BB, SDG and SEM assisted in the data analysis. All authors provided critical revision of the manuscript and approved final version for publication.

References

Abecasis GR (2009) METAL. Available at: http://www.sph.umich.edu/csg/abecasis/Metal/ Accessed 19 December 2011.

Agrawal A, Lynskey MT (2009) Candidate genes for cannabis use disorders: findings, challenges and directions. Addiction 104:518–532.

Agrawal A, Lynskey MT, Hinrichs A, Grucza R, Saccone SF, Krueger R, Neuman R, Howells W, Fisher S, Fox L, Cloninger R, Dick DM, Doheny KF, Edenberg HJ, Goate AM, Hesselbrock V, Johnson E, Kramer J, Kuperman S, Nurnberger JI, Pugh E, Schuckit M, Tischfield J, Rice JP, Bucholz KK, Bierut LJ, Consortium G (2011) A genome-wide association study of DSM-IV cannabis dependence. Addict Biol 16:514–518.

Chen WM, Abecasis GR (2007) Family-based association tests for genomewide association scans. Am J Hum Genet 81:913–926.

Chen WM, Manichaikul A, Rich SS (2009) A generalized family-based association test for dichotomous traits. Am J Hum Genet 85:364–376.

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D, Ke X, Le Hellard S, Christoforou A, Luciano M, McGhee K, Lopez L, Gow AJ, Corley J, Redmond P, Fox HC, Haggarty P, Whalley LJ, McNeill G, Goddard ME, Espeseth T, Lundervold AJ, Reinvang I, Pickles A, Steen VM, Ollier W, Porteous DJ, Horan M, Starr JM, Pendleton N, Visscher PM, Deary IJ (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16:996–1005.

Li Y, Willer CJ, Ding J, Scheet P, Abecasis GR (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet Epidemiol 34:816–834.

Liu JZ, McRae AF, Nyholt DR, Medland SE, Wray NR, Brown KM, AMFS Investigators, Hayward NK, Montgomery GW, Visscher PM, Martin NG, Macgregor S (2010) A versatile gene-based test for genome-wide association studies. Am J Hum Genet 87:139–145.

Medland SE, Nyholt DR, Painter JN, McEvoy BP, McRae AF, Zhu G, Gordon SD, Ferreira MAR, Wright MJ, Henders AK, Campbell MJ, Duffy DL, Hansell NK, Macgregor S, Slutske WS, Heath AC, Montgomery GW, Martin NG (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85:750–755.

Purcell S, Cherny SS, Sham PC (2003) Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19:149–150.

Spector TD, Williams FMK (2006) The UK adult twin registry (TwinsUK). Twin Res Hum Genet 9:899–906.

UNODC (2010) World Drug Report 2010. Blue Ridge Summit, PA: United Nations Publication, Sales No. E.10.XI.13.

Verweij KJH, Yang J, Lahti J, Veijola J, Hintsanen M, Pulkki-Råback L, Heinonen K, Pouta A, Pesonen AK, Widen E, Taanila A, Isohanni M, Miettunen J, Palotie A, Penke L, Service SK, Heath AC, Montgomery GW, Raitakari O, Kähönen M, Viikari J, Räikkönen K, Eriksson JG, Keltikangas-Järvinen L, Lehtimäki T, Martin NG, Järvelin MR, Visscher PM, Keller MC, Zietsch BP (2012) Maintenance of genetic variation in human personality: testing evolutionary models by estimating heritability due to common causal variants and investigating the effect of distant inbreeding. Evolution. Epub ahead of print.

Verweij KJH, Zietsch BP, Liu JZ, Medland SE, Lynskey MT, Madden PAF, Agrawal A, Montgomery GW, Heath AC, Martin NG (2012) No association of candidate genes with cannabis use in a large sample of Australian twin families. Addict Biol 17:687–690.

Verweij KJH, Zietsch BP, Lynskey MT, Medland SE, Neale MC, Martin NG, Boomsma DI, Vink JM (2010) Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. Addiction 105:417–430.

Vinkhuyzen AAE, Pedersen NL, Yang J, Lee SH, Magnussen PKE, Iacono WG, McGue M, Madden PAF, Heath AC, Luciano M, Payton A, Horan M, Ollier W, Pendleton N, Deary IJ, Montgomery GW, Martin NG, Visscher PM, Wray NR (2012) Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. Transl Psychiatry. Epub ahead of print.

Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, Madden PA, Heath AC, Martin NG, Montgomery GW, Goddard ME, Visscher PM (2010) Common SNPs explain a large proportion of the heritability for human height. Nat Genet 42:565–U131.

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

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1 Q–Q plots of observed and expected –log10(P) of the associations between SNPs and cannabis use initiation. Grey areas represent 95% confidence intervals

Table S1 Association results of the top 30 SNPs for cannabis use initiation

Table S2 Ten genes showing strongest association with cannabis use initiation

Appendix S1 Estimation of the proportion of variance in cannabis use initiation explained by all common variants using GCTA. Methods description

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

