Supporting Information

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

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

					Overall sample (N=10,091)					Australian (N=7175)		UK (N=2916)	
Chr	**SNP**	**BP location**	**Closest Gene**	**Location**	**Ref Allele**	**Allele Freq**	**Effect size**	**P-value**	**P-value GDT**	**Effect size**	**P-value**	**Effect size**	**P-value**
13	rs1417205	47898879	GNG5P5	Intergenic	A	.95	-.064	$8.41*10^{-7}$	$1.62*10^{-7}$	-.062	$7.48*10^{-4}$	-.067	$5.07*10^{-4}$
13	rs10507554	47901802	GNG5P5	Intergenic	T	.05	.064	$8.41*10^{-7}$	$2.04*10^{-7}$	.062	$7.17*10^{-4}$	.067	$5.14*10^{-4}$
13	rs1417202	47896427	GNG5P5	Intergenic	C	.05	.064	$1.03*10^{-6}$	$2.11*10^{-7}$	.061	$8.78*10^{-4}$	.067	$4.96*10^{-4}$
13	rs1538803	47896895	GNG5P5	Intergenic	T	.05	.064	$1.03*10^{-6}$	$2.11*10^{-7}$	.061	$8.64*10^{-4}$	.067	$4.98*10^{-4}$
13	rs9316288	47862283	GNG5P5	Intergenic	A	.95	-.065	$1.31*10^{-6}$	$4.14*10^{-7}$	-.060	$1.43*10^{-3}$	-.070	$2.98*10^{-4}$
13	rs13378547	47877760	GNG5P5	5kb upstream	A	.95	-.065	$1.31*10^{-6}$	$4.14*10^{-7}$	-.060	$1.35*10^{-3}$	-.070	$3.05*10^{-4}$
13	rs9316290	47878309	GNG5P5	5kb upstream	A	.05	.065	$1.31*10^{-6}$	$4.14*10^{-7}$	.060	$1.35*10^{-3}$	.070	$3.05*10^{-4}$
13	rs7324154	47891364	GNG5P5	Intergenic	T	.05	.065	$1.58*10^{-6}$	$3.05*10^{-7}$	.061	$1.01*10^{-3}$	.068	$4.30*10^{-4}$
17	rs9900808	13062795		Intergenic	A	.04	.082	$2.02*10^{-6}$	$8.82*10^{-5}$	.050	$4.61*10^{-2}$	.112	$4.30*10^{-6}$
13	rs9316284	47849528	GNG5P5	Intergenic	T	.95	-.065	$2.20*10^{-6}$	$3.76*10^{-7}$	-.060	$1.30*10^{-3}$	-.071	$2.97*10^{-4}$
13	rs9595635	47849915	GNG5P5	Intergenic	A	.05	.065	$2.20*10^{-6}$	$3.76*10^{-7}$	.060	$1.31\text{-}10^{-3}$	.071	$2.88*10^{-4}$
13	rs17069824	47854082	GNG5P5	Intergenic	T	.95	-.065	$2.20*10^{-6}$	$4.14*10^{-7}$	-.060	$1.39*10^{-3}$	-.071	$2.88*10^{-4}$
13	rs9595637	47854425	GNG5P5	Intergenic	A	.05	.065	$2.20*10^{-6}$	$4.14*10^{-7}$	.060	$1.39\text{-}10^{-3}$	.071	$2.81*10^{-4}$
13	rs9595638	47854508	GNG5P5	Intergenic	T	.05	.065	$2.20*10^{-6}$	$4.14*10^{-7}$	.060	$1.39\text{-}10^{-3}$	.071	$2.84*10^{-4}$

13	rs9591025	47846583	GNG5P5	Intergenic	T	.05	.065	$2.60*10^{-6}$	$3.76*10^{-7}$	.060	$1.30*10^{-3}$	.070	$3.48*10^{-4}$
13	rs9591024	47846508	GNG5P5	Intergenic	T	.95	-.064	$3.64*10^{-6}$	$3.76*10^{-7}$	-0.60	$1.29*10^{-3}$	-.068	$5.43*10^{-4}$
11	rs1573535	21620948	NELL1	Intergenic	A	.56	.026	$4.30*10^{-6}$	$8.83*10^{-6}$	.030	$2.00*10^{-4}$	.022	$6.45*10^{-3}$
11	rs1599203	21624588	NELL1	Intergenic	T	.45	-.026	$4.30*10^{-6}$	$8.83*10^{-6}$	-.031	$1.28*10^{-4}$	-.021	$1.12*10^{-2}$
11	rs7123021	21631528	NELL1	Intergenic	T	.55	.026	$4.30*10^{-6}$	$6.12*10^{-6}$	.031	$1.29*10^{-4}$	.021	$1.13*10^{-2}$
17	rs12603221	13075556		Intergenic	T	.05	.068	$4.66*10^{-6}$	$4.54*10^{-4}$	.039	6.00*10-2	.097	$2.93*10^{-6}$
17	rs4789400	74997897	MGAT5B	Intergenic	A	.08	-.069	$5.23*10^{-6}$	$5.67*10^{-3}$	-.061	$3.00*10^{-3}$	-.079	$5.19*10^{-4}$
6	rs10455657	68984930	RP11-406O16.1	Intronic	T	.81	-.032	$6.03*10^{-6}$	$7.91*10^{-5}$	-.033	$1.72*10^{-3}$	-.031	$3.20*10^{-3}$
6	rs12207424	68985325	RP11-406O16.1	Intronic	T	.81	-.032	$6.03*10^{-6}$	$6.94*10^{-5}$	-.033	$1.72*10^{-3}$	-0.31	$3.20*10^{-3}$
6	rs9458975	158389741	SYNJ2	Intergenic	T	.45	.026	$6.55*10^{-6}$	$1.22*10^{-4}$	.025	$2.53*10^{-3}$	.026	$2.02*10^{-3}$
6	rs9456922	158390147	SYNJ2	Intergenic	A	.45	.026	$6.55*10^{-6}$	$1.03*10^{-4}$	.025	$2.62*10^{-3}$	.026	$2.06*10^{-3}$
6	rs12662277	158393824	SYNJ2	Intergenic	A	.45	.026	$6.55*10^{-6}$	$1.03*10^{-4}$	.025	$2.73*10^{-3}$	.026	$2.14*10^{-3}$
13	rs9316275	47834216	GNG5P5	Intergenic	T	.95	-.062	$6.89*10^{-6}$	$1.32*10^{-6}$	-.061	$1.24*10^{-3}$	-.063	$1.55*10^{-3}$
13	rs9316276	47834285	GNG5P5	Intergenic	A	.05	.062	$6.89*10^{-6}$	$1.32*10^{-6}$	.061	$1.24*10^{-3}$	.063	$1.55*10^{-3}$
13	rs9595632	47840718	GNG5P5	Intergenic	A	.05	.062	$6.89*10^{-6}$	$1.32*10^{-6}$	.061	$1.22*10^{-3}$	.063	$1.57*10^{-3}$
13	rs9316277	47841322	GNG5P5	Intergenic	A	.05	.062	$6.89*10^{-6}$	$1.32*10^{-6}$	.061	$1.22*10^{-3}$	.063	$1.57*10^{-3}$

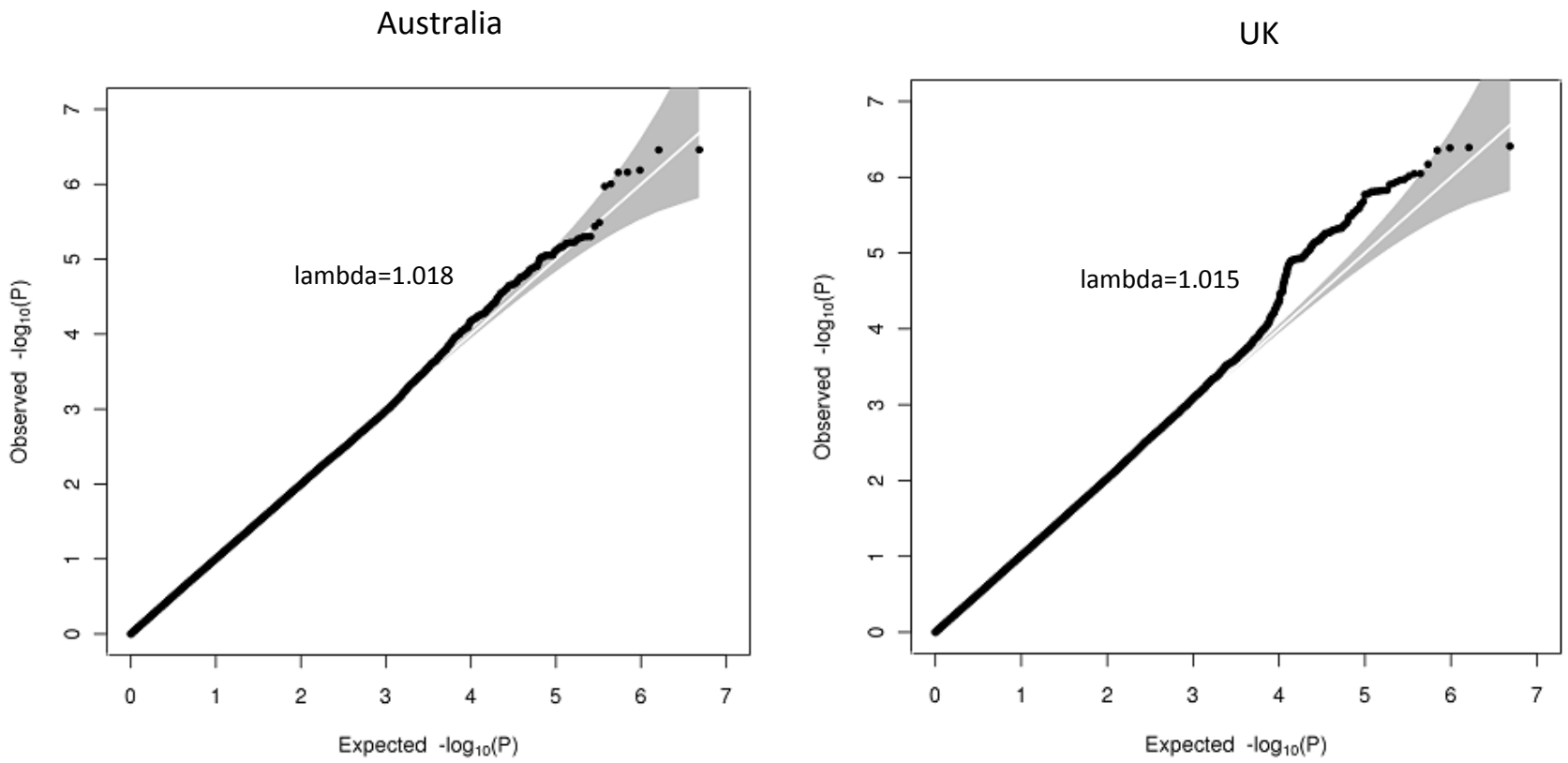
Chr = Chromosome; BP = base pair; Ref allele = reference allele; Allele Freq = allele frequency of the reference allele. Closest gene: name of gene if the SNP is located in a known gene or within 50 kb distance from a gene. P-value GDT = the P-values obtained from the meta-analysis on the genome-wide association results from the analyses preformed in GDT. The base pair locations in this table were obtained from the HapMap I + II (b36r22) CEU legend files, the genes closest to the SNP were obtained from WGA Viewer.

Supplementary Information Table S2. Ten genes showing strongest association with cannabis use initiation.

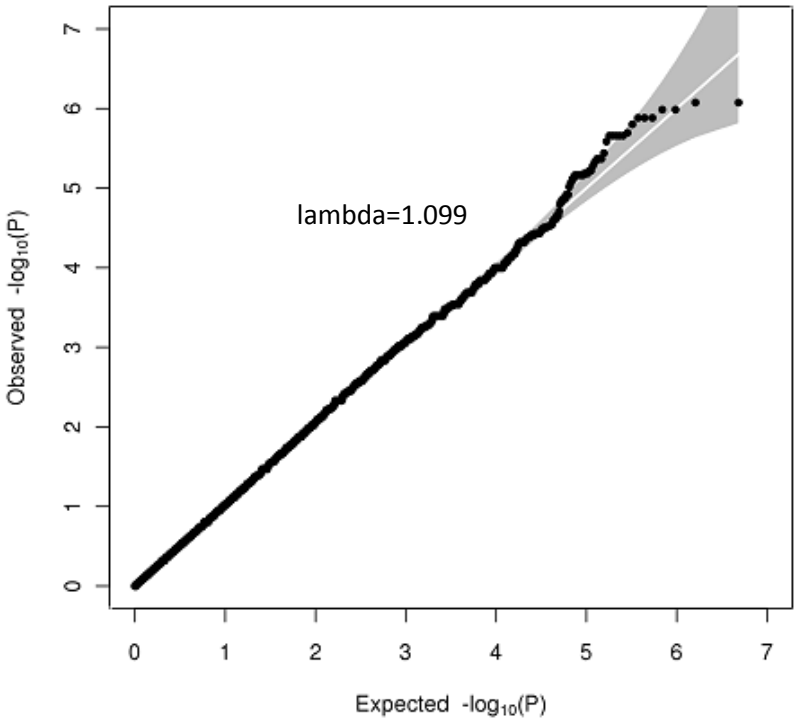
Chr	Gene	Start position gene	End position gene	Number of SNPs in gene (±50 kb)	P-value	Best SNP	SNP P-value
10	ADD3	111755715	111885313	108	$3.03*10^{-4}$	rs17126936	$1.01*10^{-4}$
2	ARHGEF4	131390693	131521306	136	$3.68*10^{-4}$	rs11690726	$5.28*10^{-5}$
5	ISOC1	128458340	128477618	134	$6.28*10^{-4}$	rs10068088	$3.71*10^{-5}$
3	IFT57	109362348	109424107	177	$6.66*10^{-4}$	rs9881619	$1.40*10^{-4}$
1	UCHL5	191251521	191295144	40	$1.03*10^{-3}$	rs7542466	$6.89*10^{-4}$
15	ZFP106	40492312	40537022	67	$1.03*10^{-3}$	rs12442040	$4.93*10^{-5}$
1	C1orf162*	111818126	111822657	100	$1.11*10^{-3}$	rs1361133	$2.18*10^{-3}$
1	ATP5F1*	111793265	111806048	111	$1.22*10^{-3}$	rs6664998	$5.67*10^{-4}$
1	OVGP1*	111758459	111771922	116	$1.41*10^{-3}$	rs6664998	$5.67*10^{-4}$
1	WDR77*	111784034	111793353	99	$1.42*10^{-3}$	rs6664998	$5.67*10^{-4}$

* Boundaries of these genes (±50 kb) are overlapping - partly the same SNPs are driving the signal in the different genes.

Supporting Information 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.



Meta-analysis



Supporting Information – Appendix S1

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

Using a recently developed method (Genome-wide Complex Trait Analysis (GCTA); see http://gump.qimr.edu.au/gcta/; Yang, Lee, Goddard, et al., 2011), we estimated the proportion of variation explained by the aggregate effect of all SNPs. This is achieved by determining to what extent genetic similarity (at the SNP level) between all individuals corresponds to their phenotypic similarity. Technically, the SNP effects are treated as random effects in a mixed linear model and the total trait variance explained by all the SNPs is estimated by restricted maximum likelihood (REML) analysis, as implemented in the freely available GCTA program (Yang, et al., 2011). The method is described in detail in Yang et al (2010), Yang et al (2011), and Visscher et al. (2010). Recently, this program has been further extended for the use of case-control data by Lee et al. (2011).

Because genotyping artefacts can severely bias the estimate, the SNP data were subjected to very restrictive quality control: SNPs with minor allele frequency <0.01, p-values for Hardy-Weinberg equilibrium test $<10^{-3}$, and SNPs with missing rates greater than 5% were removed.

After QC we estimated the genetic similarity matrix between all individuals using the hard-call imputed autosomal SNPs that passed quality control. We excluded one of each pair of individuals with an estimated genetic similarity of >0.025 (approximately closer than third cousins), in order to reduce the possibility that the phenotypic resemblance between close relatives could be caused by shared environmental effects and/or causal variants not correlated with SNPs but captured by pedigree (Visscher, et al., 2010; Yang, et al., 2010). This led to an exclusion of 5479 individuals, resulting in a retained dataset of 4612 individuals.

Population structure can inflate the genetic variance estimates. We controlled for population structure by including the first 20 principal components (eigenvectors of the genetic relatedness matrix) and cohort status as covariates in the analysis. As in the genome-wide association analyses, we also controlled for the effects of sex, age, age^2, and sex*age.

Note that in order to correct the SNP-based heritability estimate for sample ascertainment (in the Australian sample), we used a population prevalence for initiation of cannabis use of 28%, lower than the prevalence in our combined samples.

References

Lee SH, Wray NR, Goddard ME, Visscher PM (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. *American Journal of Human Genetics, 88*(3): 294-305.

Visscher PM, Yang JA, Goddard ME (2010). A Commentary on 'Common SNPs Explain a Large Proportion of the Heritability for Human Height' by Yang et al. (2010). *Twin Research and Human Genetics, 13*(6): 517-524.

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. *Nature Genetics, 42*(7): 565-U131.

Yang JA, Lee SH, Goddard ME, Visscher PM (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics, 88*(1): 76-82.