Genome-wide association study implicates *CHRNA2* in cannabis use disorder

Ditte Demontis[1,2,3*], Veera Manikandan Rajagopal[1,2], Thomas D. Als[1,2,3], Jakob Grove[1,2,3], Jonatan Pallesen[1,2], Carsten Hjorthøj[2,4], Per Qvist[1,2], Jane Hvarregaard Christensen[1,2], Jonas Bybjerg-Grauholm[2,5], Marie Bækvad-Hansen[2,5], Laura M. Huckins[6,7], Eli A. Stahl[6,7], Allan Timmermann[8], Esben Agerbo[2,8,9], David M. Hougaard[2,5], Thomas Werge[2,10,11], Ole Mors[2,12], Preben Bo Mortensen[2,8,9], Merete Nordentoft[2,4,13], Mark Daly[14,15,16], Mette Nyegaard[1,2], Anders D. Børglum[1,2,3*]

1. Department of Biomedicine – Human Genetics, Aarhus University, Denmark
2. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Aarhus, Denmark
3. Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark
4. Copenhagen University Hospital, Mental Health Centre Copenhagen Mental Health Services in the Capital Region of Denmark, Hellerup, Denmark
5. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark
6. Division of Psychiatric Genomic, Icahn School of Medicine at Mount Sinai, NYC, NY;
7. Department of Genetics and Genomics, Icahn School of Medicine at Mount Sinai, NYC, NY;
8. National Centre for Register-based Research (NCRR) Aarhus University, Aarhus, Denmark
9. Centre for Integrated Register-Based Research, Aarhus University, CIRRAU
10. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark
11. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark
12. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark
13. Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark
14. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA
15. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
16. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

*Correspond with: Ditte Demontis (ditte@biomed.au.dk) Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark. Anders D. Børglum (anders@biomed.au.dk) Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark.

Introductory paragraph

Cannabis is the most frequently used illicit psychoactive substance worldwide[1]. Life time use has been reported among 35-40% of adults in Denmark[2] and the United States[3]. Cannabis use is increasing in the population[4-6] and among users around 9% become dependent[7]. The genetic risk component is high with heritability estimates of 51[8]–70%[9]. Here we report the first genome-wide significant risk locus for cannabis use disorder (CUD, $P=9.31x10^{-12}$) that replicates in an independent population ($P_{replication}=3.27x10^{-3}$, $P_{metaanalysis}=9.09x10^{-12}$). The finding is based on a genome-wide association study (GWAS) of 2,387 cases and 48,985 controls followed by replication in 5,501 cases and 301,041 controls. The index SNP (rs56372821) is a strong eQTL for *CHRNA2* and analyses of the genetic regulated gene expressions identified significant association of *CHRNA2* expression in cerebellum with CUD. This indicates a potential therapeutic use in CUD of compounds with agonistic effect on the neuronal acetylcholine receptor alpha-2 subunit encoded by *CHRNA2*. At the polygenic level analyses revealed a significant decrease in the risk of CUD with increased load of variants associated with cognitive performance.

Main

Overall the prevalence of diagnosed CUD in the population has been estimated to 1-1.5% among Europeans[10,11] and Americans[5]. CUD is associated with a range of adverse health problems[12,13] including risk of psychosis[14,15], bipolar disorder[16], anxiety disorder[17] and cognitive impairment with more persistent use associated with greater decline[18]. Estimates of the heritability for cannabis use initiation and life-time cannabis use, with respect to the amount of variance explained by common variants (i.e. the SNP heritability), has been estimated to 0.06[19] - 0.2[20]. Four GWASs related to cannabis use have been conducted without genome-wide significant findings: one study of cannabis dependence[21] and three studies of lifetime cannabis use[19,20,22]. In addition, two recent GWASs have reported genome-wide significant associations, albeit with negative or ambiguous replication

results: a study of DSM-IV cannabis dependence criterion counts in a combined sample of 14,754 European Americans and African Americans[23], which reported three genome-wide significant loci associated with cannabis use severity; and a GWAS of cannabis dependence of 2,080 European cases and 6,435 controls which identified one genome-wide significant locus[24].

Here we present results from a GWAS and subsequent replication based on analyses of the largest cohorts of diagnosed CUD reported so far. Individuals included in the discovery GWAS come from the Danish nation-wide population based cohort collected by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH)[25]. The iPSYCH cohort was ascertained to study six major psychiatric disorders (schizophrenia, bipolar disorder, major depressive disorder, attention-deficit hyperactivity disorder (ADHD), anorexia nervosa and autism spectrum disorder) and consists of 79,492 genotyped individuals. The present GWAS was based on 2,387 individuals with a diagnosis of CUD (ICD10 F12.1-12.9) and 48,985 individuals not diagnosed with CUD, all identified in the iPSYCH cohort. Data analysis was conducted using the Ricopili pipeline[26], including stringent quality control of genotyped variants and individuals (online methods). Information about non-genotyped markers was obtained by imputation[27,28] using the 1000 genomes phase 3 as reference panel[29]. GWAS was performed using imputed marker dosages and an additive logistic regression model with relevant principal components to correct for confounders such as population stratification, and the major psychiatric disorders studied in iPSYCH as covariates. Only markers with an imputation INFO score > 0.7, minor allele frequency (maf) > 0.01 and bi-allelic markers were retained, in total 8,971,679 genetic markers. We identified 26 genome-wide significant SNPs ($P< 5x10^{-8}$), located in a single locus on chromosome 8 (Figure 1 and 2). The index SNP, rs56372821, showed an odds ratio of 0.73 ($P= 9.31x10^{-12}$) with respect to the minor allele A (Table 1, Figure 1 and 2).

The genome-wide significant locus on chromosome 8 was replicated in an independent European sample consisting of 5,501 cases with CUD and 301,041 population controls (deCODE cohort). The cases were diagnosed with CUD while undergoing inpatient treatment at the SAA Treatment Centre, Vogur Hospital in Iceland (www.saa.is). We tested nine markers located in the risk locus in the Icelandic sample; the index SNP and eight correlated variants ($0.2 < r^2 < 0.7$) with P values less than $1x10^{-6}$ (four genome-wide significant). All variants demonstrated consistent direction of association and the most strongly associated variant (rs56372821) in the discovery GWAS had a P-value of $3.27x10^{-3}$ in the deCODE cohort. In the meta-analysis rs56372821, became slightly stronger associated with CUD ($P = 9.09x10^{-12}$), and additional two variants became genome-wide significant (Table 1).

There was no evidence of association of previously identified genome-wide significant cannabis risk varaints with CUD in our analyses[23,24] (Supplementary Table 1).). This might be due to different phenotype definitions among the studies as Sherva et al.[23] analysed association with cannabis criterion counts, and Agrawal et al.[24] used cannabis exposed (but not dependent) individuals as controls in their study. Additionally, the composition of the cohorts analysed also differ as the previous GWASs were based on cohorts established to study genetics of substance use disorders while the iPSYCH cohort is ascertained for major mental illnesses (Supplementary Table 2).

To assess the proportion of phenotypic variance explained by common variants we applied LD score regression[30] and the GREML method implemented in GCTA[31], assuming a population prevalence of 1% for cannabis use disorder (Online methods). Estimates of the liability-scale SNP heritability were h^2_{snp}= 0.09 (SE=0.03) and h^2_{snp}= 0.042 (SE=0.014) using LD score regression and

GCTA, respectively. The results imply a smaller contribution from common variants than was previously found for cannabis use[20]. The estimate would probably increase with larger sample size, as a result of a decrease in the size of the error terms of the SNP effect estimates[32].

We found no signs of contributions from confounding factors like population stratification and cryptic relatedness to the inflation in the distribution of the test statistics (see quantile-quantile plot, Figure 1.B.) using LD score regression, since the intercept was one (intercept = 0.996; SE=0.0079) (online methods).

In the GWAS we corrected for diagnoses of the major psychiatric disorders studied in iPSYCH (see Supplementary Table 2 for distribution of psychiatric disorders among CUD cases and the control group). In addition, we evaluated the impact on the odds ratio for the index SNP by leave-one-out analyses excluding psychiatric phenotypes one at a time in the association analysis (online methods). The odds ratio remained stable (Table 2), supporting that the association was independent of a diagnosis with one the psychiatric disorders evaluated.

The risk locus on chromosome 8 has also been found to be genome-wide significantly associated with schizophrenia[26]. In our leave-one-phenotype out analysis the locus remained genome-wide significant when individuals with schizophrenia were removed, excluding any potential confounding from schizophrenia (Table 2). The signal observed for the index SNP (rs73229090) in the GWAS meta-analysis of schizophrenia[26] is consistent with the direction of association observed in our analysis. Since individuals with schizophrenia often use cannabis (around 13% - 16% [33,34]) it could be speculated that the significant signal observed in schizophrenia, is driven by a sub-group of schizophrenia cases also having CUD. This hypothesis is supported by analysis of the association of rs56372821 with schizophrenia in the iPSYCH sample. We found a nominal significant association of the SNP with schizophrenia when including individulas with co-morbid CUD (2,281

schizophrenia cases and 23,134 controls; OR=0.9; P=0.036), while after exclusion of individuals with CUD (556 cases and 101 controls excluded) the association signal disappeared (1,727 cases, 23,033 controls; OR = 0.97, P=0.63). In order to evaluate further the impact of CUD comorbidity on the odds ratio of rs56372821 in schizophrenia a null distribution for the odds ratio was generated by performing 10,000 permutations randomly removing 556 and 101 individuals among the cases and controls, respectively (Supplementary Figure 1). The observed odds ratio for rs56372821, when excluding individuals with comorbid CUD was 0.97, and differed significantly from random removal of the same number of cases and controls ($P_{two\text{-}sided}$=0.0027, $P_{one\text{-}sided}$=0.0015). Thus the permutation test supports the hypothesis that the subgroup diagnosed with CUD among schizophrenia cases drives the nominal association observed for rs56372821 in the iPSYCH schizophrenia sample.

We performed a CUD-only analysis testing all SNPs in LD with the index SNP (19 SNPs; $r^2 > 0.7$) for association with age at first diagnosis. This analysis suggested the risk alleles to be associated with earlier age at first diagnosis (most significant SNP: rs35236974, P = 0.020). On average CUD cases homozygous for the protective allele got their diagnosis (mean age: 22.41; st.dev = 3.63) one year later than CUD cases having at least one risk allele (mean age: 21.01; st.dev = 3.56).

Among the brain tissues analysed in GTEx the index SNP rs56372821 was found to be a strong eQTL for *CHRNA2* in cerebellum (P-value in GTEx =2.1×10^{-7}), with the risk allele (G-allele) being associated with decreased expression of the gene (Supplementary Figure 2). In order to further evaluate the potential regulatory impact of the identified locus on chromosome 8 as well as gene expression differences between cases and controls genome-wide, we imputed the genetically regulated gene expression in 11 brain tissues using PrediXcan[35]. The SNP weights used were

derived from models trained on reference transcriptome data sets including 10 brain tissues from GTEx[36] and transcriptome data from dorsolateral prefrontal cortex generated by the CommonMind Consortium[37,(Huckins et al. submitted to Nature Genetics).] We tested for association of expression of 2,460 – 10,930 protein coding genes (depending on the tissue; see Supplementary Table 3) with CUD using logistic regression corrected by the same covariates as in the GWAS. One gene, *CHRNA2*, was significantly differently expressed between cases and controls ($P = 2.713 \times 10^{-6}$; beta = -0.21; SE = 0.045) in the cerebellum. The expression model for *CHRNA2* in cerebellum was based on 47 SNPs including four genome-wide significant SNPs rs59724122, rs73229093, rs7838316 and rs11783093 (Figure 4). *CHRNA2* expression was predicted with a valid model in two additional two brain tissues with nominal significant under-expression in cases compared to controls (dorsolateral prefrontal cortex, $P=5.19 \times 10^{-4}$; cerebellar hemisphere $P=5.30 \times 10^{-3}$). That the risk locus for CUD can be linked to *CHRNA2* expression is also supported by Won et al.[38] who generated high-resolution 3D maps of chromatin contacts (by Hi-C sequencing approach), in order to capture the functional relationships between regulatory and transcribed elements during human corticogenesis. They applied their map to the set of credible SNPs from the large PGC schizophrenia GWAS[26] and found physical interaction of the genome-wide significant risk locus on chromosome 8, with the regulatory region of *CHRNA2*.

CHRNA2 is expressed in the brain[39,40] and encodes the neuronal acetylcholine receptor (nAChR) alpha-2 subunit, which is incorporated in heteropentameric neuronal nAChRs mainly with the beta-2 or beta-4 subunits[41,42]. Candidate gene studies of common variants have linked this gene to e.g. substance abuse[43] and nicotine dependence[44,45] but no genome-wide significant findings have connected *CHRNA2* to any substance abuse or psychiatric disorder, besides the fuctional link

between schizophrenia risk variants and *CHRNA2* expression identified by Won et al[38] described above.

The result suggests *CHRNA2* underexpression in cerebellum (and potentially other brain regions, Supplementary Table 3), to be involved in CUD. Cerebellum may play a role in addiction with respect to reward and inhibitory control[46,47], and reward-anticipating responses[48]. The cerebellum has high density of cannabinoid receptor 1 (CB1), which mediates the effect of delta9-tetrahydrohydrocannabinol (THC)[49], the psychoactive compound in cannabis and has also been found to be affected by cannabis use in neuroanatomic studies[50-53].

There is no reported functional link between cannabis use and alpha-2 subunit containing nAChRs. We hypothesize three potential ways of the involvement of *CHRNA2* in CUD: 1) Substances in cannabis might interact directly with the alpha-2 subunit containing nAChRs as studies have found cannabidiol, a non-psychoactive component of cannabis, to inhibit the alpha-7 containing nAChRs[54]. 2) Cannabis could indirectly affect the alpha-2 subunit containing nAChRs. After binding of an agonist e.g. acetylcholine, the nAChR responds by opening of an ion-conducting channel across the plasma membrane. This causes depolarization of the membrane and can result in presynaptic neurotransmitter release[55] including dopamine[56,57]. Since the psychoactive active compound of cannabis THC has been found to affect the release of acetylcholine in various brain regions[58,59,60] it could be speculated that this, through alpha-2 subunit containing nAChRs, could affect dopamine release, a known neurotransmitter involved in addiction[61]. 3) There could be a strong biological link between expression of *CHRNA2* and the cannabinoid receptor 1 gene (*CNR1*). This hypothesis is based on evaluation of gene-expression correlations from genome-wide microarray gene expression profiles in the Allan Brain Atlas[62] (http://www.brain-map.org/). We found that of all genes evaluated (58692 probes analyzed), *CNR1* demonstrated the strongest negative correlation with *CHRNA2* expression (r_{max}= -0.498; Supplementary Figure 3). The signal

was driven by opposite expression patterns in a large number of brain tissues, e.g in cerebellum where *CNR1* had a relatively high expression in the cerebellar cortex and *CHRNA2* had a relatively low expression and the opposite was observed for cerebellar nuclei (Supplementary Figure 3). This suggests the existence of a currently uncharacterized biological interaction between the endocannabinoid system and alpha-2 subunit containing nAChR, by which the identified risk locus which is associated with decreased *CHRNA2* expression could be related to increased *CNR1* expression.

The observed association of *CHRNA2* unerexpression with CUD, also implies that the alpha-2 subunit can be a potential drug target in the treatment of CUD, by the use of an agonist selective for alpha-2 subunit containing nAChRs[63,64]. The impact of some compounds with an agonistic effect on the alpha-2 subunit containing nAChRs, like NS9283[65], ABT-418[66] and TC-1734[67], have already been studied primarily due to their potential impact on memory and cognitive processes[65,67-69]. Additionally TC-1734 has been tested in clinical phase I and II trials for its improvement on memory[67,69,70], and the favorable safty profile of this drug suggests a potential re-purposing for CUD treatment.

CHRNA2 is related to other nicotine receptor genes that have been identified as risk loci for nicotine dependence and smoking behavior[71,72]. It is therefore relevant to question if our finding is confounded by smoking. In the iPSYCH sample[25] there is no avaible information about smoking behavior. As much as 70-90% of cannabis users have been reported also to smoke cigarettes[73,74], which impose the risk of confounding. However, smoking is also highly comorbid with psychiatric disorders[75,76], which is prevalent among the control group (66,79% was diagnosed with at least one of the major psychiatric disorders studied in iPSYCH; supplementary Table 2). Since we expect smoking to be prevalent both among cases and the control group, it is unlikely that smoking alone

would generate the strong association we observe on chromosome 8. Additionally, SNPs located in or near *CHRNA3*, *CHRNA4*, *CHRNB3*, *CYP2A6* previously found to demonstrate extreme strong association with nicotine dependence and smoking behavior[71,72,77-79], was not associated in the present analysis (lowest P=0.31 for rs1051730; Supplementary Table 4), and the risk locus for CUD identified in our GWAS did not demonstrate strong association with smoking behaviour in the published large GWASs of smoking[71,72,77-79] (lowest P-value for life time smoking (P=0.003); Supplementary Table 5).

In order to evaluate the genetic overlap between CUD and a range of other phenotypes at the polygenic level, we conducted analyses of polygenic risk scores (PRS) for 22 phenotypes related to cognition, personality, psychiatric disorders, reproduction and smoking behaviour (online methods; list of phenotypes and results can be found in Supplementary Table 6). PRS for eight phenotypes (three measures of cognitive perfomance, age at first birth, life time smoking, ADHD, depressive symptoms and schizophrenia) demonstrated strong association with CUD ($4.33x10^{-4} < P < 7.44x10^{-15}$; Figure 4 and Supplementary Table 6). Strikingly, PRS for measures of educational attainment, SSGAC[80,81] (educational years: z-score= -7.78; $P=7.44x10^{-15}$ and college completion z-score=-5.10; $P=3.33x10^{-7}$), were strongly negatively associated with the risk of CUD. A finding which was reinforced by the significant negative association of PRS for human intelligence[82] with CUD (z-score = -3.51; $P=4.33x10^{-4}$). Numerous epidemiological studies have observed an association of cannabis use disorder and decreased educational performance[83-86]. The direction of causation is unclear and it is unknown if the relationship is caused by cannabis use and subsequent disengagement from education or wether and to what extent the two share overlapping genetic factors [87,88]. Our results suggest an overlap in genetic risk factors, with a decrease in the odds ratio for cannabis use disorder with increased number of educational years/cognitive performance

(Supplementary Figure 4.A-C.). The decreased risk of CUD with increased age at first birth (z-score =-7.41; P=1.26x10^{-13}) is supporting the relationship with educational attainment as number of educational years is known to correlate genetically with later birth of the first child[89]. In order to avoid confounding in the PRS analyses of psychiatric disorders we excluded individuals among CUD cases and the control group with a diagnosis of the psychiatric disorder being analysed. PRS for ADHD[90] (z-score =5.10; P=3.45x10^{-7}), depressive symptoms[91] (z-score 4.34; P=6.58x10^{-6}) and schizophrenia[26] (z-score=5.47; P=4.45x10^{-8}) were all associated with an increased risk for CUD (Figure 4). The overlap in common genetic risk factors for cannabis use disorder with depressive symptoms is in line with recent observations[92] and the association of PRS for ADHD with CUD suggest that the epidemiological observations of high comorbidity of ADHD with CUD[93,94] are influenced by shared genetic risk components. Likewise, the PRS analysis supports an overlap in common genetic risk factors between CUD and schizophrenia (in line with epidemiological observations[33,34]). However we cannot rule out if the overlap involves variants with pleiotropic effect on the two disorders, or if the overlap in general is caused by heterogeneity among schizophrenia cases due to a subgroup comorbid with CUD, in line with the CUD risk locus identified here, which does not seem to have a pleitropic effect across the two disorders.

In summary, we identified a genome-wide significant CUD locus on chromosome 8, which replicated in an independent sample. An impact of the locus on earlier age of first diagnosis was further suggested. The locus implicates *CHRNA2* in CUD through analyses of the imputed genetically regulated gene expression, showing decreased *CHRNA2* expression in cerebellum (and other brain regions) of individuals with cannabis use disorder. An intriguing result which suggests that CHRNA2 could be a potential therapeutic target for further testing in randomized controlled

trials. Analyses of PRS for other phenotypes revealed a significant decrease in the risk of CUD with increased measures of cognitive performance.

Online Methods

Sample

The individuals in this study is a part of the large Danish nation-wide population based sample collected by iPSYCH[25]. The iPSYCH sample was originally designed for studying major psychiatric disorders and includes genotypes from 79,492 individuals, including 54,249 cases diagnosed with at least one of six mental disorders (schizophrenia, bipolar disorder, major depressive disorder, ADHD, anorexia and autism spectrum disorder) and 25,243 individuals not having any of the six psychiatric disorders. The iPSYCH sample was selected from a birth cohort comprising individuals born in Denmark between May 1, 1981, and December 31, 2005, who were residents in Denmark on their first birthday and who have a known mother (N = 1,472,762). The iPSYCH cases were identified based on information in the Danish Psychiatric Central Research Register[95], and 30,000 randomly selected controls were identified from the same nationwide birth cohort in the Danish Civil Registration System[96]. Subseqeuntly blood spot samples (Guthrie cards) were identified in the in the Danish Newborn Screening Biobank (DNSB) from the included individuals[97]. Processing of DNA, genotyping and genotype calling as well as imputing of genotypes (see below) of the samples were done in 23 waves of approximately 3,500 individuals each.

Individuals with CUD were identified in the iPSYCH sample through information in the Danish Psychiatric Central Research Register[95] as well as in the Danish National Patient Register[98] using information up to 2013. All individuals with an ICD10 diagnosis of F12.1-12.9, were included.

Individuals with acute intoxication (diagnosis code F12.0) were not included. The F12.0 diagnosis is related to the acute pharmacological effects of cannabis use and resolve with time, and does not necessary reflect long term problematic cannabis use, which is the focus of this study. All individuals in the iPSYCH sample not having a diagnosis of CUD were used as controls and covariates for the major psychiatric disorders studied by iPSYCH were used in the analyses in order to correct for the presence of mental disorders. Only 54 individuals had a diagnosis of CUD and anorexia, and all individuals having a diagnosis of anorexia among cases and the control group were excluded. Supplementary Table 2 gives an overview of the distribution of individuals with a psychiatric disorder among the CUD cases and the control group.

Processing of samples, genotyping, QC and imputing was done in 23 waves. In this study individuals from the first four waves were excluded as only a few cannabis use disorder cases were present in these waves (range 0-15 cases). The waves represent approximate birth years and the four waves that were excluded represent to a large extend individuals, which in 2013 (the time of extraction of register diagnoses) were children.

Genotyping, quality control and GWAS

DNA was extracted from dried blood spot samples and whole genome amplified in triplicates as described previously[99,100]. Genotyping was performed at the Broad Institute of Harvard and MIT (Cambridge, MA, USA) using Illumina's Beadarrays (PsychChip; Illumina, CA, San Diego, USA) according to the manufacturer's protocols. Genotypes were a result of merging call sets from three different calling algorithms (GenCall, Birdseed and Zcall). GenCall[101] and Birdseed[102] was used to call genotypes with minor allele frequency (maf) > 0.01 and zCall[103] was used to call genotypes with maf < 0.01. The merging was done after pre-QC on individual call sets.

Stringent quality control was applied to the individual call rate (> 0.98) and only genotypes with high call rate (> 0.98), no strong deviation from Hardy-Weinberg equilibrium ($P > 10^{-6}$ in controls or $P > 10^{-10}$ in cases) and low heterozygosity rates ($| F_{het} | < 0.2$) were included. Genotypes were phased and imputed using the 1000 Genomes Project phase 3 (1KGP3)[29,104] imputation reference panel and SHAPEIT[27] and IMPUTE2[28]. Relatedness and population stratification were evaluated using a set of high quality markers (genotyped autosomal markers with minor allele frequency (maf) > 0.05, HWE P > 1x 10^{-4} and SNP call rate > 0.98), which were pruned for linkage disequilibrium (LD) ($r^2 < 0.075$) resulting in a set of 37,425 pruned markers (markers located in long-range LD regions defined by Price et al.[105] were excluded). Genetic relatedness was estimated using PLINK v1.9[106,107] to identify first and second-degree relatives ($\hat{\pi} > 0.2$) and one individual was excluded from each related pair (cases preferred kept over controls). Genetic outliers were identified for exclusion based on principal component analyses (PCA) using EIGENSOFT[108,109]. A genetic homogenous sample was defined based on a subsample of individuals being Danes for three generations (identified based on register information about birth country of the individuals, their parents and grandparents). The subsample of Danes was used to define the center based on the mean values of principal component (PC) 1 and PC2. Subsequently PC1 and PC2 were used to define a genetic homogenous population by excluding individuals outside an ellipsoid with axes greater than six standard deviations from the mean.

Association analysis was done using logistic regression and the imputed marker dosages. The following covariates were used: principal component 1-4 and principal components from the PCA associated with case-control status, the 19 data-processing waves and diagnosis of major psychiatric disorders studied by iPSYCH (supplementary Table 2). Results for 9,729,295 markers were generated, subsequently markers with imputation info score < 0.7 (n = 608.367), markers with maf < 0.01 (n=10.220) and multi-allelic markers (n = 143,083) were removed. In total after filtering

8,969,939 markers remained for further analysis. All analyses of the iPSYCH sample were performed at the secured national GenomeDK high performance-computing cluster in Denmark (https://genome.au.dk).

Replication

The genome-wide significant locus on chromosome 8 was replicated in an independent European cohort consisting of 5,501 cases with diagnosed CUD and 301,041 population controls collected by deCODE genetics. The characteristics of the SAA treatment sample have been described previously[110] and currently diagnoses at Vogur Hospital are made using DSM-V, but most of the diagnoses in this study are based on DSM-IV or DSM-IIIR. The genotypes were obtained based on SNP array data and whole genome sequences using long range phasing and two types of imputations[111]. For the replication study, markers were looked up in the results of a genome-wide association study performed by logistic regression treating disease status as the response and genotype counts as covariates. Other available individual characteristics that correlate with disease status were also included in the model as nuisance variables, using previously described methods[112]. The resulting P values were corrected for inflation by the method of genomic controls (correction factor=1.42). Nine genetic variants all representing the same association signal on chromosome 8 with P values less than $1x10^{-6}$ were looked up in the deCODE results. The variants in the locus were selected based on LD ($0.2 < r^2 < 0.7$). We included additional markers besides the index SNP in order to be able evaluate the consistency of direction of association in the replication cohort over a set of markers located in the associated risk locus. The data were meta-analyzed using summary statistics and inverse variance weighted fixed effects model, implemented in the software METAL[113]. All tested variants demonstrated consistent direction of association in the replication cohort.

Heritability

SNP heritability (h^2_{SNP}) was estimated by LD score regression[30] using summary statistics from the GWAS of CUD and pre-computed LD scores (available on https://github.com/bulik/ldsc). The SNP heritability was calculated on the liability scale using a prevalence of 1% in the population. The h^2_{SNP} estimated by LD score regression was inaccurate as the ratio between the estimated h^2_{SNP} and the standard error was less than four (h^2_{SNP}/SE = 2,86). Therefore, the h^2_{SNP} was also estimated using GCTA[31]. A genomic relationship matrix (GRM) between all pairwise combinations of individuals was estimated for each autosome separately and subsequently merged using the GCTA software. The strict best-guess-genotypes (i.e. 4,299,887 SNPs with info score > 0.8, missing rate < 0.01 and MAF > 0.01) from imputation with phase 3 of 1000 genomes were used for GRM estimation. Univariate GREML analyses in GCTA was used to estimate h^2_{SNP} on the liability scale, using the combined GRM and the same covariates as in the GWAS.

Leave-one phenotype out GWAS

In the GWAS of CUD we corrected, by covariates, for five major psychiatric disorders analysed in iPSYCH (Supplementary Table 2). However, in order to further evaluate the potential impact on the association signal of the index SNP from the psychiatric phenotypes, sensitivity analyses were performed where one phenotype at a time was excluded from the association analysis. The phenotypes evaluated were schizophrenia, bipolar disorder, major depressive disorder, attention-deficit hyper activity disorder, autism and individuals not having any of the five diagnoses. GWASs were performed as previously described excluding one of the six phenotypes one at a time (sample size for number of cases and controls is shown in Table 2).

Permutation based evaluation of the odds ratio of rs56372821 for schizophrenia with and without CUD

We applied a permutation based approach in order to evaluate how the inclusion and exclusion of individuals with CUD affected the association of our index SNP rs56372821 with schizophrenia. The analyses were based on the iPSYCH sample consisting of 2,281 schizophrenia cases and 23,134 population based controls (genotyping, QC and imputation were done using the same procedures as explained above). Odds ratios were obtained by the glm() function in R using dosage data from the imputed SNP rs56372821, and including principle components PC1-PC4 + PCs signicantly associated with SZ and/or CUD and wave number as covariates.

In the association analysis of schizophrenia the odds ratio of 56372821(CI95%) was OR=0.9 (0.81-0.99). When excluding all individuals with CUD, among cases and controls (1,727 cases and 23,033 controls) the odds ratio of rs56372821(CI95%) was OR = 0.97 (0.87-1.09). To evaluate whether this change in the odds ratio is due to removal of individuals with CUD (554 cases and 101 control with CUD removed) or due to reduction in sample size, a null distribution of OR's was generated by randomly removing 554 individuals among schizophrenia cases and 101 individuals among controls (removing a total of 655 individuals). This permutation of sample exlcusion was done 9,999 times to generate the null distribution of the odds ratio for rs56372821. To obtain a two-sided P-value, the observed $\ln(OR_{rs56372821}) = -0.03$ was mirrored around the mean $\ln(OR_{rs56372821})$ of the null distribution, as shown in Supplementary Figure 1. Removal of individuals with CUD (554 SZ cases and 101 controls) changed the OR of rs56372821 significantly, compared to random removal of the same number of schizophrina cases and controls ($P_{two\text{-}sided}=0.0027$), producing a significant increase of $OR_{rs56372821}$ compared to random removal ($P_{one\text{-}sided}=0.0015$).

Association with age at first diagnosis

In order to test for a potential impact of the risk locus on age at first diagnosis, a case only study was performed testing for association of the index SNP (rs56372821) and SNPs in LD with this (r^2 > 0.7; 19 variants) with age at first diagnosis. Date of diagnosis was identified from register information in the Danish Psychiatric Central Research Register[95]. Analysis for association with age at first diagnosis (natural logarithm (age at first diagnosis)) was done using linear regression and the same covariates as used in the GWAS (PCs, wave and the psychiatric disorders). A dominant model with respect to the risk allele, was applied.

PrediXcan

The genetically regulated gene expression was imputed in 11 brain tissues using PrediXcan[35] (models downloaded from: https://github.com/hakyimlab/PrediXcan; version 6 data). PrediXcan was used to impute the transcriptome for cases and the control group using SNP weights derived from models trained on reference transcriptome data sets including 10 brain tissues from GTEx (https://www.gtexportal.org/home/) and transcriptome data from dorsolateral prefrontal cortex generated by the CommonMind Consortium[37,(Huckins et al. submitted to Nature Genetics)]. The models were trained on 1000 genome SNPs and contained only gene expression prediction models with a false discovery rate less than 5%. Gene expression levels in the iPSYCH data were imputed wave-wise and subseqeuntly the imputed data were merged. We tested for association of 2,459 – 10,930 protein coding genes (depending on the tissue; see Supplementary Table 3), with CUD using logistic regression corrected by the relevant principal components from PCA and the psychiatric disorder as described for the GWAS. Since gene expression among the different brain tissues is highly correlated, evaluated based on GTEx data[114], we corrected the P-value for each gene by the

total number genes tested in all tissues with a valid model available for the gene, with respect to *CHRNA2* we corrected for 13,166 genes tested in three tissues.

PRS risk score analyses

PRS analyses were done using GWAS summary statistics from 22 GWASs (Supplementary Table 6). The summary files were downloaded from public databases and processed using the munge script which is a part of the LDscore regression software[30]. All variants with INFO < 0.9, MAF < 0.01, missing values, out of bounds P-values, ambiguous strand alleles and duplicated rs-ids were removed using the munge script. In addition, mult-allelic variants and insertion and deletion (indels) were removed. The processed summary files were then LD-clumped using Plink, with the following parameter settings: --clump-p1 1 --clump-p2 1 --clump-r2 0.1 --clump-kb 500. The clumped file were used as the training dataset. Genetic risk scores were estimated at different P-value thresholds for SNP inclusion: $5x10^{-8}$, $1x10^{-6}$, $1x10^{-4}$, $1x10^{-3}$, 0.01, 0.05, 0.1, 0.2, 0.5 and 1.0 for all individuals in the target sample (CUD cases and the control group) from the genotype dosages using Plink's '--score' method, with default arguments. However the PRS scores for ADHD, were generated using the approach described Demontis et al.[90]. For each P-value threshold the variance in the phenotype explained by PRS was estimated using Nagelkerke's R^2 (R package 'BaylorEdPsych'), and association of PRS with CUD was estimated using logistic regression including the same covariates used in the GWAS analysis (PCs from PCA and the psychiatric disorders listed in Supplementary Table 2). In PRS analyses of psychiatric disorders (ADHD, schizophrenia and depression related phenotypes) individuals with a diagnosis of the disorder being analysed were excluded. The number of individuals excluded with ADHD, schizophrenia and major depressive disorder are listed in Supplementary Table 2.

References

1 Degenhardt, L. & Hall, W. Extent of illicit drug use and dependence, and their contribution to the global burden of disease. *Lancet* **379**, 55-70, doi:10.1016/S0140-6736(11)61138-0 (2012).

2 Mounteney, J. *et al.* The drug situation in Europe: an overview of data available on illicit drugs and new psychoactive substances from European monitoring in 2015. *Addiction* **111**, 34-48, doi:10.1111/add.13056 (2016).

3 Administration, S. A. a. M. H. S. Results from the 2013 National Survey on drugs Use and Health: Summary of National Findings, NSDUH Series H-48, HHS Publication No. (SMA) 14-4863. (Rockville, MD, 2014).

4 Charilaou, P. *et al.* Trends of Cannabis Use Disorder in the Inpatient: 2002 - 2011. *Am J Med*, doi:10.1016/j.amjmed.2016.12.035 (2017).

5 Compton, W. M., Han, B., Jones, C. M., Blanco, C. & Hughes, A. Marijuana use and use disorders in adults in the USA, 2002-14: analysis of annual cross-sectional surveys. *Lancet Psychiatry* **3**, 954-964, doi:10.1016/S2215-0366(16)30208-5 (2016).

6 Azofeifa, A. *et al.* National Estimates of Marijuana Use and Related Indicators - National Survey on Drug Use and Health, United States, 2002-2014. *MMWR Surveill Summ* **65**, 1-28, doi:10.15585/mmwr.ss6511a1 (2016).

7 Lopez-Quintero, C. *et al.* Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC). *Drug and alcohol dependence* **115**, 120-130, doi:10.1016/j.drugalcdep.2010.11.004 (2011).

8 Verweij, K. J. *et al.* Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies. *Addiction* **105**, 417-430, doi:10.1111/j.1360-0443.2009.02831.x (2010).

9 Kendler, K. S. *et al.* A population-based Swedish Twin and Sibling Study of cannabis, stimulant and sedative abuse in men. *Drug and alcohol dependence* **149**, 49-54, doi:10.1016/j.drugalcdep.2015.01.016 (2015).

10 Addiction, E. M. C. f. D. a. D. Treatment of cannabis-related disorders in Europe. (2015).

11 Pedersen, C. B. *et al.* A comprehensive nationwide study of the incidence rate and lifetime risk for treated mental disorders. *JAMA Psychiatry* **71**, 573-581, doi:10.1001/jamapsychiatry.2014.16 (2014).

12 Arria, A. M., Caldeira, K. M., Bugbee, B. A., Vincent, K. B. & O'Grady, K. E. Marijuana use trajectories during college predict health outcomes nine years post-matriculation. *Drug and alcohol dependence* **159**, 158-165, doi:10.1016/j.drugalcdep.2015.12.009 (2016).

13 Volkow, N. D., Baler, R. D., Compton, W. M. & Weiss, S. R. Adverse health effects of marijuana use. *N Engl J Med* **370**, 2219-2227, doi:10.1056/NEJMra1402309 (2014).

14 Nielsen, S. M., Toftdahl, N. G., Nordentoft, M. & Hjorthoj, C. Association between alcohol, cannabis, and other illicit substance abuse and risk of developing schizophrenia: a nationwide population based register study. *Psychological medicine* **47**, 1668-1677, doi:10.1017/S0033291717000162 (2017).

15 Marconi, A., Di Forti, M., Lewis, C. M., Murray, R. M. & Vassos, E. Meta-analysis of the Association Between the Level of Cannabis Use and Risk of Psychosis. *Schizophr Bull* **42**, 1262-1269, doi:10.1093/schbul/sbw003 (2016).

16 Cougle, J. R., Hakes, J. K., Macatee, R. J., Chavarria, J. & Zvolensky, M. J. Quality of life and risk of psychiatric disorders among regular users of alcohol, nicotine, and cannabis: An analysis of the National Epidemiological Survey on Alcohol and Related Conditions (NESARC). *J Psychiatr Res* **66-67**, 135-141, doi:10.1016/j.jpsychires.2015.05.004 (2015).
17 Kedzior, K. K. & Laeber, L. T. A positive association between anxiety disorders and cannabis use or cannabis use disorders in the general population--a meta-analysis of 31 studies. *BMC psychiatry* **14**, 136, doi:10.1186/1471-244X-14-136 (2014).
18 Meier, M. H. *et al.* Persistent cannabis users show neuropsychological decline from childhood to midlife. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E2657-2664, doi:10.1073/pnas.1206820109 (2012).
19 Verweij, K. J. *et al.* The genetic aetiology of cannabis use initiation: a meta-analysis of genome-wide association studies and a SNP-based heritability estimation. *Addict Biol* **18**, 846-850, doi:10.1111/j.1369-1600.2012.00478.x (2013).
20 Stringer, S. *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. *Transl Psychiatry* **6**, e769, doi:10.1038/tp.2016.36 (2016).
21 Agrawal, A. *et al.* A genome-wide association study of DSM-IV cannabis dependence. *Addict Biol* **16**, 514-518, doi:10.1111/j.1369-1600.2010.00255.x (2011).
22 Minica, C. C. *et al.* Heritability, SNP- and Gene-Based Analyses of Cannabis Use Initiation and Age at Onset. *Behav Genet* **45**, 503-513, doi:10.1007/s10519-015-9723-9 (2015).
23 Sherva, R. *et al.* Genome-wide Association Study of Cannabis Dependence Severity, Novel Risk Variants, and Shared Genetic Risks. *JAMA Psychiatry* **73**, 472-480, doi:10.1001/jamapsychiatry.2016.0036 (2016).
24 Agrawal, A. *et al.* Genome-wide association study identifies a novel locus for cannabis dependence. *Molecular psychiatry*, doi:10.1038/mp.2017.200 (2017).
25 Pedersen, C. B. *et al.* The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Molecular psychiatry*, doi:10.1038/mp.2017.196 (2017).
26 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).
27 Delaneau, O., Marchini, J. & Zagury, J. F. A linear complexity phasing method for thousands of genomes. *Nature methods* **9**, 179-181, doi:10.1038/nmeth.1785 (2011).
28 Howie, B., Marchini, J. & Stephens, M. Genotype imputation with thousands of genomes. *G3* **1**, 457-470, doi:10.1534/g3.111.001198 (2011).
29 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
30 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211 (2015).
31 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *American journal of human genetics* **88**, 76-82, doi:10.1016/j.ajhg.2010.11.011 (2011).
32 Wray, N. R. *et al.* Pitfalls of predicting complex traits from SNPs. *Nat Rev Genet* **14**, 507-515, doi:10.1038/nrg3457 (2013).

33 Koskinen, J., Lohonen, J., Koponen, H., Isohanni, M. & Miettunen, J. Rate of cannabis use disorders in clinical samples of patients with schizophrenia: a meta-analysis. *Schizophr Bull* **36**, 1115-1130, doi:10.1093/schbul/sbp031 (2010).
34 Toftdahl, N. G., Nordentoft, M. & Hjorthoj, C. Prevalence of substance use disorders in psychiatric patients: a nationwide Danish population-based study. *Social psychiatry and psychiatric epidemiology* **51**, 129-140, doi:10.1007/s00127-015-1104-4 (2016).
35 Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics* **47**, 1091-1098, doi:10.1038/ng.3367 (2015).
36 Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics* **45**, 580-585, doi:10.1038/ng.2653 (2013).
37 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).
38 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
39 Quik, M. *et al.* Subunit composition of nicotinic receptors in monkey striatum: effect of treatments with 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine or L-DOPA. *Mol Pharmacol* **67**, 32-41, doi:10.1124/mol.104.006015 (2005).
40 Ishii, K., Wong, J. K. & Sumikawa, K. Comparison of alpha2 nicotinic acetylcholine receptor subunit mRNA expression in the central nervous system of rats and mice. *J Comp Neurol* **493**, 241-260, doi:10.1002/cne.20762 (2005).
41 Whiteaker, P. *et al.* Pharmacological and immunochemical characterization of alpha2* nicotinic acetylcholine receptors (nAChRs) in mouse brain. *Acta Pharmacol Sin* **30**, 795-804, doi:10.1038/aps.2009.68 (2009).
42 Chavez-Noriega, L. E. *et al.* Pharmacological characterization of recombinant human neuronal nicotinic acetylcholine receptors h alpha 2 beta 2, h alpha 2 beta 4, h alpha 3 beta 2, h alpha 3 beta 4, h alpha 4 beta 2, h alpha 4 beta 4 and h alpha 7 expressed in Xenopus oocytes. *J Pharmacol Exp Ther* **280**, 346-356 (1997).
43 Corley, R. P. *et al.* Association of candidate genes with antisocial drug dependence in adolescents. *Drug and alcohol dependence* **96**, 90-98, doi:10.1016/j.drugalcdep.2008.02.004 (2008).
44 Philibert, R. A. *et al.* Examination of the nicotine dependence (NICSNP) consortium findings in the Iowa adoption studies population. *Nicotine Tob Res* **11**, 286-292, doi:10.1093/ntr/ntn034 (2009).
45 Wang, S. *et al.* Significant associations of CHRNA2 and CHRNA6 with nicotine dependence in European American and African American populations. *Hum Genet* **133**, 575-586, doi:10.1007/s00439-013-1398-9 (2014).
46 Miquel, M. *et al.* Have we been ignoring the elephant in the room? Seven arguments for considering the cerebellum as part of addiction circuitry. *Neurosci Biobehav Rev* **60**, 1-11, doi:10.1016/j.neubiorev.2015.11.005 (2016).
47 Moulton, E. A., Elman, I., Becerra, L. R., Goldstein, R. Z. & Borsook, D. The cerebellum and addiction: insights gained from neuroimaging research. *Addict Biol* **19**, 317-331 (2014).
48 Wagner, M. J., Kim, T. H., Savall, J., Schnitzer, M. J. & Luo, L. Cerebellar granule cells encode the expectation of reward. *Nature* **544**, 96-100, doi:10.1038/nature21726 (2017).
49 Sim-Selley, L. J. Regulation of cannabinoid CB1 receptors in the central nervous system by chronic cannabinoids. *Crit Rev Neurobiol* **15**, 91-119 (2003).

50 Lorenzetti, V., Solowij, N. & Yucel, M. The Role of Cannabinoids in Neuroanatomic Alterations in Cannabis Users. *Biological psychiatry* **79**, e17-31, doi:10.1016/j.biopsych.2015.11.013 (2016).
51 Block, R. I. *et al.* Cerebellar hypoactivity in frequent marijuana users. *Neuroreport* **11**, 749-753 (2000).
52 Stella, N. Chronic THC intake modifies fundamental cerebellar functions. *J Clin Invest* **123**, 3208-3210, doi:10.1172/JCI70226 (2013).
53 Cutando, L. *et al.* Microglial activation underlies cerebellar deficits produced by repeated cannabis exposure. *J Clin Invest* **123**, 2816-2831, doi:10.1172/JCI67569 (2013).
54 Mahgoub, M. *et al.* Effects of cannabidiol on the function of alpha7-nicotinic acetylcholine receptors. *Eur J Pharmacol* **720**, 310-319, doi:10.1016/j.ejphar.2013.10.011 (2013).
55 Itier, V. & Bertrand, D. Neuronal nicotinic receptors: from protein structure to function. *FEBS Lett* **504**, 118-125 (2001).
56 Threlfell, S. *et al.* Striatal dopamine release is triggered by synchronized activity in cholinergic interneurons. *Neuron* **75**, 58-64, doi:10.1016/j.neuron.2012.04.038 (2012).
57 Cachope, R. *et al.* Selective activation of cholinergic interneurons enhances accumbal phasic dopamine release: setting the tone for reward processing. *Cell reports* **2**, 33-41, doi:10.1016/j.celrep.2012.05.011 (2012).
58 Nava, F., Carta, G., Colombo, G. & Gessa, G. L. Effects of chronic Delta(9)-tetrahydrocannabinol treatment on hippocampal extracellular acetylcholine concentration and alternation performance in the T-maze. *Neuropharmacology* **41**, 392-399 (2001).
59 Lopez-Moreno, J. A., Gonzalez-Cuevas, G., Moreno, G. & Navarro, M. The pharmacology of the endocannabinoid system: functional and structural interactions with other neurotransmitter systems and their repercussions in behavioral addiction. *Addict Biol* **13**, 160-187, doi:10.1111/j.1369-1600.2008.00105.x (2008).
60 Schlicker, E. & Kathmann, M. Modulation of transmitter release via presynaptic cannabinoid receptors. *Trends Pharmacol Sci* **22**, 565-572 (2001).
61 Nutt, D. J., Lingford-Hughes, A., Erritzoe, D. & Stokes, P. R. The dopamine theory of addiction: 40 years of highs and lows. *Nature reviews. Neuroscience* **16**, 305-312, doi:10.1038/nrn3939 (2015).
62 Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391-399, doi:10.1038/nature11405 (2012).
63 Posadas, I., Lopez-Hernandez, B. & Cena, V. Nicotinic receptors in neurodegeneration. *Curr Neuropharmacol* **11**, 298-314, doi:10.2174/1570159X11311030005 (2013).
64 Jensen, A. A., Frolund, B., Liljefors, T. & Krogsgaard-Larsen, P. Neuronal nicotinic acetylcholine receptors: structural revelations, target identifications, and therapeutic inspirations. *J Med Chem* **48**, 4705-4745, doi:10.1021/jm040219e (2005).
65 Timmermann, D. B. *et al.* Augmentation of cognitive function by NS9283, a stoichiometry-dependent positive allosteric modulator of alpha2- and alpha4-containing nicotinic acetylcholine receptors. *Br J Pharmacol* **167**, 164-182, doi:10.1111/j.1476-5381.2012.01989.x (2012).
66 Papke, R. L., Thinschmidt, J. S., Moulton, B. A., Meyer, E. M. & Poirier, A. Activation and inhibition of rat neuronal nicotinic receptors by ABT-418. *Br J Pharmacol* **120**, 429-438, doi:10.1038/sj.bjp.0700930 (1997).

67 Frolich, L., Ashwood, T., Nilsson, J., Eckerwall, G. & Sirocco, I. Effects of AZD3480 on cognition in patients with mild-to-moderate Alzheimer's disease: a phase IIb dose-finding study. *J Alzheimers Dis* **24**, 363-374, doi:10.3233/JAD-2011-101554 (2011).
68 Potter, A. *et al.* Acute effects of the selective cholinergic channel activator (nicotinic agonist) ABT-418 in Alzheimer's disease. *Psychopharmacology (Berl)* **142**, 334-342 (1999).
69 Dunbar, G. *et al.* Effects of TC-1734 (AZD3480), a selective neuronal nicotinic receptor agonist, on cognitive performance and the EEG of young healthy male volunteers. *Psychopharmacology (Berl)* **191**, 919-929, doi:10.1007/s00213-006-0675-x (2007).
70 Gatto, G. J. *et al.* TC-1734: an orally active neuronal nicotinic acetylcholine receptor modulator with antidepressant, neuroprotective and long-lasting cognitive effects. *CNS Drug Rev* **10**, 147-166 (2004).
71 Thorgeirsson, T. E. *et al.* Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. *Nature genetics* **42**, 448-453, doi:10.1038/ng.573 (2010).
72 Tobacco & Genetics, C. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nature genetics* **42**, 441-447, doi:10.1038/ng.571 (2010).
73 Schauer, G. L., Berg, C. J., Kegler, M. C., Donovan, D. M. & Windle, M. Assessing the overlap between tobacco and marijuana: Trends in patterns of co-use of tobacco and marijuana in adults from 2003-2012. *Addict Behav* **49**, 26-32, doi:10.1016/j.addbeh.2015.05.012 (2015).
74 Agrawal, A., Budney, A. J. & Lynskey, M. T. The co-occurring use and misuse of cannabis and tobacco: a review. *Addiction* **107**, 1221-1233, doi:10.1111/j.1360-0443.2012.03837.x (2012).
75 Matcham, F. *et al.* Smoking and common mental disorders in patients with chronic conditions: An analysis of data collected via a web-based screening system. *General hospital psychiatry* **45**, 12-18, doi:10.1016/j.genhosppsych.2016.11.006 (2017).
76 Kutlu, M. G., Parikh, V. & Gould, T. J. Nicotine Addiction and Psychiatric Disorders. *Int Rev Neurobiol* **124**, 171-208, doi:10.1016/bs.irn.2015.08.004 (2015).
77 Hancock, D. B. *et al.* Genome-wide meta-analysis reveals common splice site acceptor variant in CHRNA4 associated with nicotine dependence. *Transl Psychiatry* **5**, e651, doi:10.1038/tp.2015.149 (2015).
78 Rice, J. P. *et al.* CHRNB3 is more strongly associated with Fagerstrom test for cigarette dependence-based nicotine dependence than cigarettes per day: phenotype definition changes genome-wide association studies results. *Addiction* **107**, 2019-2028, doi:10.1111/j.1360-0443.2012.03922.x (2012).
79 Kumasaka, N. *et al.* Haplotypes with copy number and single nucleotide polymorphisms in CYP2A6 locus are associated with smoking quantity in a Japanese population. *PloS one* **7**, e44507, doi:10.1371/journal.pone.0044507 (2012).
80 Okbay, A. *et al.* Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539-542, doi:10.1038/nature17671 (2016).
81 Rietveld, C. A. *et al.* GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467-1471, doi:10.1126/science.1235488 (2013).
82 Sniekers, S. *et al.* Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature genetics* **49**, 1107-1112, doi:10.1038/ng.3869 (2017).
83 Bray, J. W., Zarkin, G. A., Ringwalt, C. & Qi, J. The relationship between marijuana initiation and dropping out of high school. *Health Econ* **9**, 9-18 (2000).

84 Fergusson, D. M., Horwood, L. J. & Beautrais, A. L. Cannabis and educational achievement. *Addiction* **98**, 1681-1692 (2003).
85 Silins, E. *et al.* Adolescent substance use and educational attainment: An integrative data analysis comparing cannabis and alcohol from three Australasian cohorts. *Drug and alcohol dependence* **156**, 90-96, doi:10.1016/j.drugalcdep.2015.08.034 (2015).
86 Horwood, L. J. *et al.* Cannabis use and educational achievement: findings from three Australasian cohort studies. *Drug and alcohol dependence* **110**, 247-253, doi:10.1016/j.drugalcdep.2010.03.008 (2010).
87 Lynskey, M. & Hall, W. The effects of adolescent cannabis use on educational attainment: a review. *Addiction* **95**, 1621-1630 (2000).
88 Verweij, K. J., Huizink, A. C., Agrawal, A., Martin, N. G. & Lynskey, M. T. Is the relationship between early-onset cannabis use and educational attainment causal or due to common liability? *Drug and alcohol dependence* **133**, 580-586, doi:10.1016/j.drugalcdep.2013.07.034 (2013).
89 Barban, N. *et al.* Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nature genetics* **48**, 1462-1472, doi:10.1038/ng.3698 (2016).
90 Demontis, D. *et al.* Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*, doi:10.1101/145581 (2017).
91 Okbay, A. *et al.* Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics* **48**, 624-633, doi:10.1038/ng.3552 (2016).
92 Carey, C. E. *et al.* Associations between Polygenic Risk for Psychiatric Disorders and Substance Involvement. *Front Genet* **7**, 149, doi:10.3389/fgene.2016.00149 (2016).
93 De Alwis, D., Lynskey, M. T., Reiersen, A. M. & Agrawal, A. Attention-deficit/hyperactivity disorder subtypes and substance use and use disorders in NESARC. *Addict Behav* **39**, 1278-1285, doi:10.1016/j.addbeh.2014.04.003 (2014).
94 Kolla, N. J. *et al.* Adult attention deficit hyperactivity disorder symptom profiles and concurrent problems with alcohol and cannabis: sex differences in a representative, population survey. *BMC psychiatry* **16**, 50, doi:10.1186/s12888-016-0746-4 (2016).
95 Mors, O., Perto, G. P. & Mortensen, P. B. The Danish Psychiatric Central Research Register. *Scand J Public Health* **39**, 54-57, doi:10.1177/1403494810395825 (2011).
96 Pedersen, C. B. The Danish Civil Registration System. *Scand J Public Health* **39**, 22-25, doi:10.1177/1403494810387965 (2011).
97 Nørgaard-Pedersen, B. & Hougaard, D. M. Storage policies and use of the Danish Newborn Screening Biobank. *Journal of Inherited Metabolic Disease* **30**, 530-536 (2007).
98 Lynge, E., Sandegaard, J. L. & Rebolj, M. The Danish National Patient Register. *Scand J Public Health* **39**, 30-33, doi:10.1177/1403494811401482 (2011).
99 Borglum, A. D. *et al.* Genome-wide study of association and interaction with maternal cytomegalovirus infection suggests new schizophrenia loci. *Molecular psychiatry* **19**, 325-333, doi:10.1038/mp.2013.2 (2014).
100 Hollegaard, M. V. *et al.* Robustness of genome-wide scanning using archived dried blood spot samples as a DNA source. *BMC Genet* **12**, 58, doi:10.1186/1471-2156-12-58 (2011).
101 Illumina. illumina GenCall Data Analysis Software. *Illumina Tech Note* (2005).

102 Korn, J. M. *et al.* Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. *Nature genetics* **40**, 1253-1260, doi:10.1038/ng.237 (2008).
103 Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543-2545, doi:10.1093/bioinformatics/bts479 (2012).
104 Sudmant, P. H. *et al.* An integrated map of structural variation in 2,504 human genomes. *Nature* **526**, 75-81, doi:10.1038/nature15394 (2015).
105 Price, A. L. *et al.* Long-range LD can confound genome scans in admixed populations. *American journal of human genetics* **83**, 132-135; author reply 135-139, doi:10.1016/j.ajhg.2008.06.005 (2008).
106 Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *American journal of human genetics* **81**, 559-575, doi:10.1086/519795 (2007).
107 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
108 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics* **38**, 904-909, doi:10.1038/ng1847 (2006).
109 Galinsky, K. J. *et al.* Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. *American journal of human genetics* **98**, 456-472, doi:10.1016/j.ajhg.2015.12.022 (2016).
110 Tyrfingsson, T. *et al.* Addictions and their familiality in Iceland. *Ann N Y Acad Sci* **1187**, 208-217, doi:10.1111/j.1749-6632.2009.05151.x (2010).
111 Gudbjartsson, D. F. *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nature genetics* **47**, 435-444, doi:10.1038/ng.3247 (2015).
112 Sveinbjornsson, G. *et al.* Rare mutations associating with serum creatinine and chronic kidney disease. *Hum Mol Genet* **23**, 6935-6943, doi:10.1093/hmg/ddu399 (2014).
113 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).
114 Park, Y., Sarkar, A. K., Bhutani, K. & Kellis, M. Multi-tissue polygenic models for transcriptome-wide association studies. *bioRxiv*, doi:10.1101/107623 (2017).

Acknowledgements

We are grateful to the following researchers for providing the results for the replication: Thorgeir E. Thorgeirsson[1], Gunnar W. Reginsson[1], Thorarinn Tyrfingsson[2], Valgerdur Runarsdottir[2], Hreinn Stefansson[1], Kari Stefansson[1] ([1]deCODE genetics / Amgen, Reykjavik, Iceland; [2]National Center of Addiction Medicine, SAA, Vogur Hospital, Reykjavik, Iceland)

The iPSYCH project is funded by the Lundbeck Foundation (grant numbers R102-A9118 and R155-2014-1724) and the universities and university hospitals of Aarhus and Copenhagen. Genotyping of iPSYCH samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789 to MJD), and NIMH (5U01MH094432-02 to MJD). The Danish National Biobank resource was supported by the Novo Nordisk Foundation. Data handling and analysis on the GenomeDK HPC facility was supported by NIMH (1U01MH109514-01 to Michael O'Donovan and ADB). High-performance computer capacity for handling and statistical analysis of iPSYCH data on the GenomeDK HPC facility was provided by the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to ADB).

We gratefully acknowledge all the studies and databases that made GWAS summary data available: GPC (Genetics of Personality Consortium), PGC (Psychiatric Genomics Consortium), SSGAC (Social Science Genetics Association Consortium), TAG (Tobacco and Genetics Consortium), UK Biobank, CTGLAB (Complex Traits Genetics Lab).

We gratefully acknowledge the data made available by the the Genotype-Tissue Expression (GTEx) Project (supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS). The data used for the analyses described in this manuscript were obtained from: https://www.gtexportal.org/home/ on 01/11/2017.

Disclosures

T. Werge has been a lecturer and advisor to H. Lundbeck A/S.
T. E. Thorgeirsson, G. W. Reginsson, H. Stefansson and K. Stefansson are employees of deCODE genetics/Amgen.

Author contributions

This section will be filled during the review process

Correspondence and requests for materials should be addressed to:

Anders D. Børglum, anders@biomed.au.dk
Ditte Demontis, ditte@biomed.au.dk

Table 1. Genome-wide significant SNPs associated with cannabis use disorder

Results for genome-wide significant SNPs in the associated locus at chromosome 8 in the primary GWAS and/or in the meta-analysis with results from an independent cohort collected by deCODE. Results for the index SNP (Index SNP) is shown together with results from supporting correlated SNPs in the locus ($0.2 < r^2 < 0.7$) (LD SNPs). Alleles for the variants (A1 and A2), frequency of A1 in controls (Frq A1 con), the odds ratio (OR) for the effect of A1, and P-values (P) are given.

				iPSYCH					deCODE					meta-analysis			
Index SNP	LD SNPs	CHR	BP (hg19)	A1	A2	Frq A1 con	OR	P	A1	A2	Frq A1 con	OR	P	A1	A2	OR	P
rs56372821		8	27436500	A	G	0.163	0.728	**9.31×10^{-12}**	A	G	0.144	0.878	3.27×10^{-03}	A	G	0.803	**9.09×10^{-12}**
	rs4732724	8	27423062	C	G	0.324	0.820	**1.34×10^{-08}**	C	G	0.367	0.905	1.60×10^{-03}	C	G	0.866	**7.74×10^{-10}**
	rs6558008	8	27438306	A	C	0.755	1.237	**2.54×10^{-08}**	C	A	0.228	0.915	1.45×10^{-02}	A	C	1.159	**1.97×10^{-08}**
	rs73229090	8	27442127	A	C	0.110	0.702	**7.41×10^{-10}**	A	C	0.105	0.881	1.30×10^{-02}	A	C	0.797	**2.82×10^{-09}**
	rs937220	8	27429730	A	G	0.259	0.8258	2.46×10^{-07}	A	G	0.269	0.912	7.56×10^{-03}	A	G	0.871	**4.53×10^{-08}**
	rs17466060	8	27422740	A	G	0.418	0.8594	3.94×10^{-06}	A	G	0.430	0.900	5.11×10^{-04}	A	G	0.881	**1.29×10^{-08}**

Table 2. The effect on the odds ratio of rs56372821 of leave-one-phenotype out association analysis

In the analyses individuals with a psychiatric diagnosis are excluded one disorder at a time (without ADHD (woADHD), without autism spectrum disorder (woASD), without bipolar disorder (woBP), without the iPSYCH control cohort (woCON), without major depressive disorder (woMDD), without schizophrenia (woSZ). The P-value for association with CUD (P-value), number of cases (cases), frequency of the minor allele in cases (Freq cases), number of controls (con), frequency of the minor allele in controls (Freq con), odds ratio (OR) and standard error (SE), are given. For illustration ln(OR), and the corresponding standard error.

	P-value	cases	Freq cases	con	Freq con	OR	SE
Cannabis GWAS	9.31×10^{-12}	2387	0.124	48985	0.163	0.727	0.047
woADHD	2.68×10^{-10}	1461	0.118	39284	0.165	0.683	0.060
woASD	7.19×10^{-10}	2234	0.126	40287	0.164	0.744	0.048
woBP	5.61×10^{-10}	2199	0.126	47809	0.163	0.742	0.048
woCON	2.35×10^{-11}	2290	0.124	32716	0.162	0.726	0.048
woMDD	2.81×10^{-07}	1253	0.121	32592	0.162	0.711	0.066
woSZ	7.15×10^{-10}	1818	0.125	47196	0.163	0.725	0.052

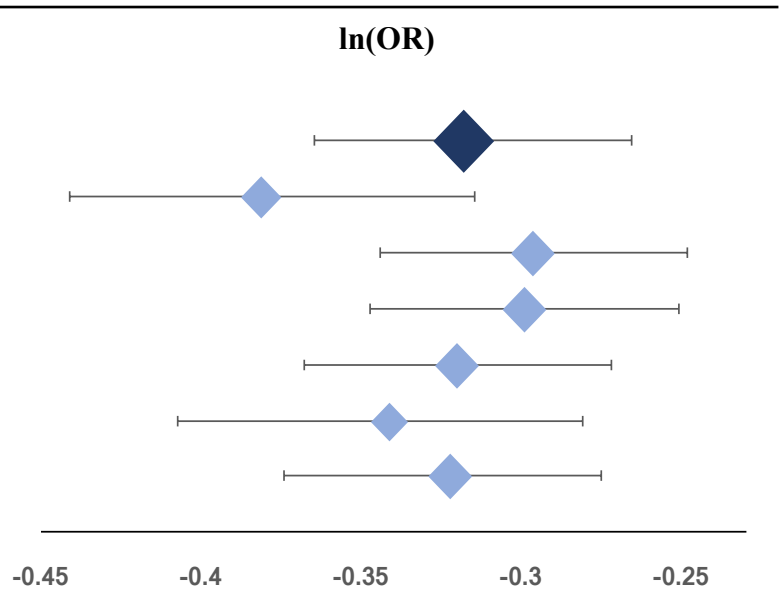


Figure legends

Figure 1. A) Manhattan plot Manhattan plot of the results from the GWAS of CUD. The index is highlighted as a green diamond and SNPs in high LD with the index SNP are marked in green. **B)** Quantile-quantile plot of the expected and observed P-values from GWAS of CUD.

Figure 2. Regional association plot of the local association results for the risk locus at chromosome 8. The index SNP (rs56372821) and additional three correlated genome-wide significant SNPs (LD with index SNP: $0.2 < r^2 < 0.7$) are marked with letters (a-d), the triangle represents the P-value from meta-analysis with the replication cohort from deCODE. The location and orientation of the genes in the region and the local estimates of recombination rate is shown. The association P-value (p), odds ratio (or), minor allele frequency (maf) and imputation info-score (info) is presented in upper right corner.

Figure 3. Association of the imputed genetically regulated expression of *CHRNA2* in three brain tissues with a valid model (cerebellar hemisphere, dorsolateret prefrontal contex and cerebellum). The P-value for association of expression with CUD (-log10(PrediXcan:P-value) and P-value from CUD GWAS (-log10(GWAS:P-value)) is given on the y-axis, with a red dotted line indicating statistical significance. Chromosome position is given on the x-axis and the light grey lines indicate which SNPs that are included in the models used to predict gene expression. The thin red lines indicate genetic predictors that are genome-wide significantly associated with CUD.

Figure 4. Association of PRS with CUD. PRSs was generated for phenotypes related to cognition, personality, psychiatric disorders, reproduction and smoking behavior based on summary statistics from 22 published GWASs. The variance explained by the scores (Nagelkerke-R^2) is given on the x-axis and the P-value for association of the PRS with CUD on the y-axis.

Figure 1.A.

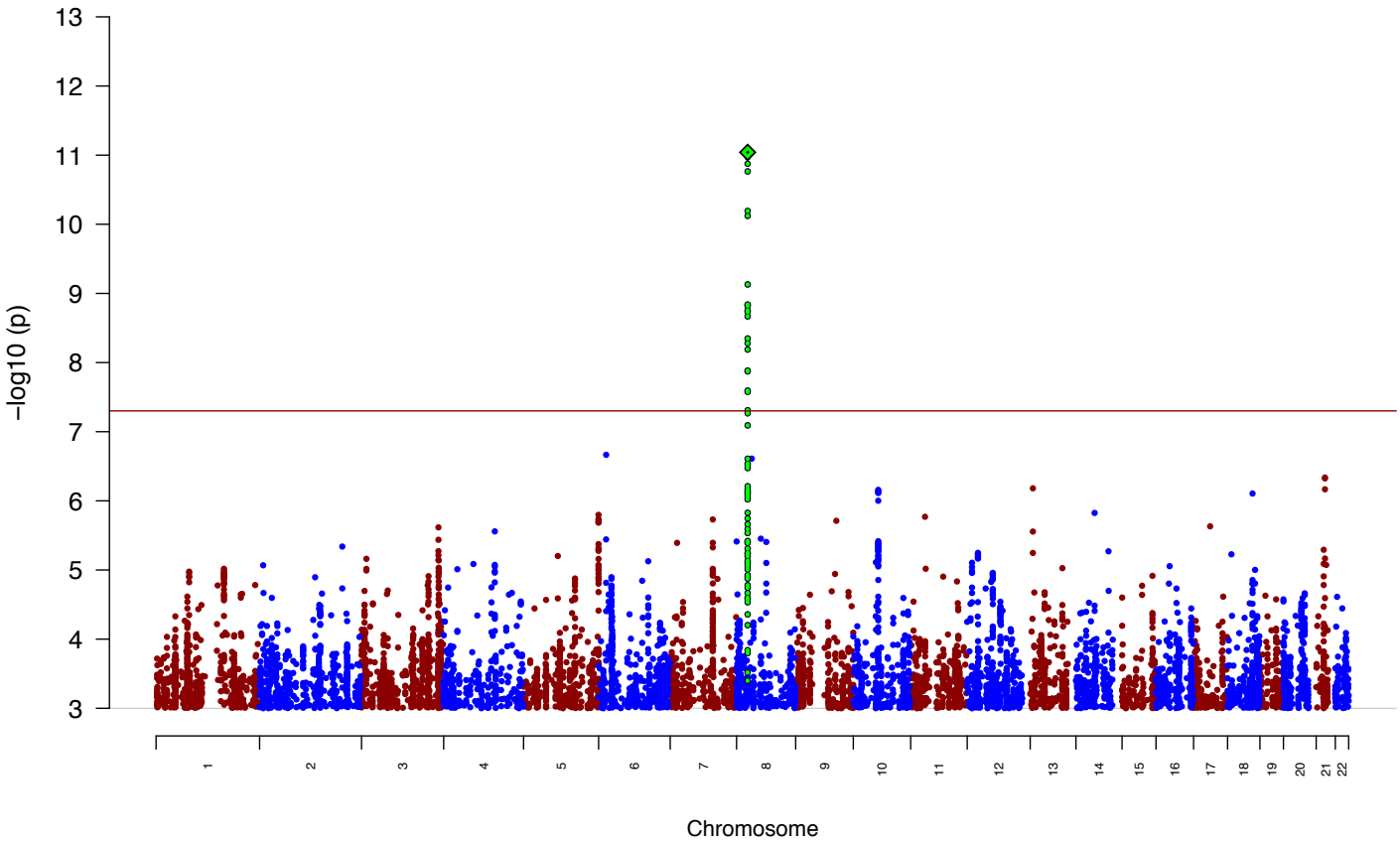


Figure 1.B.

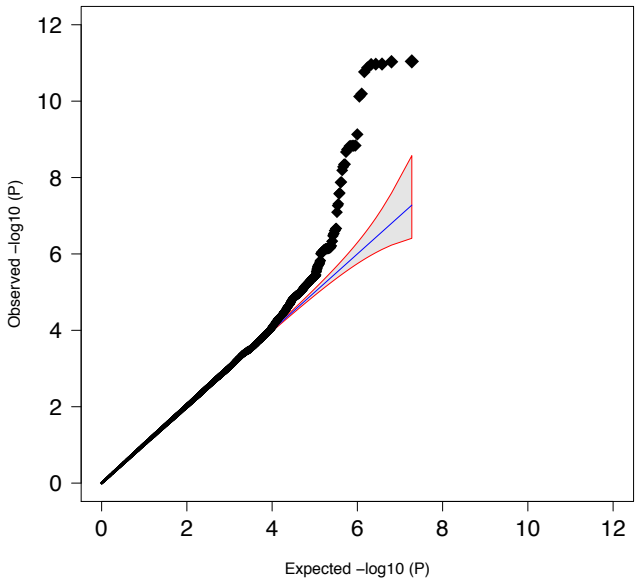


Figure 2.

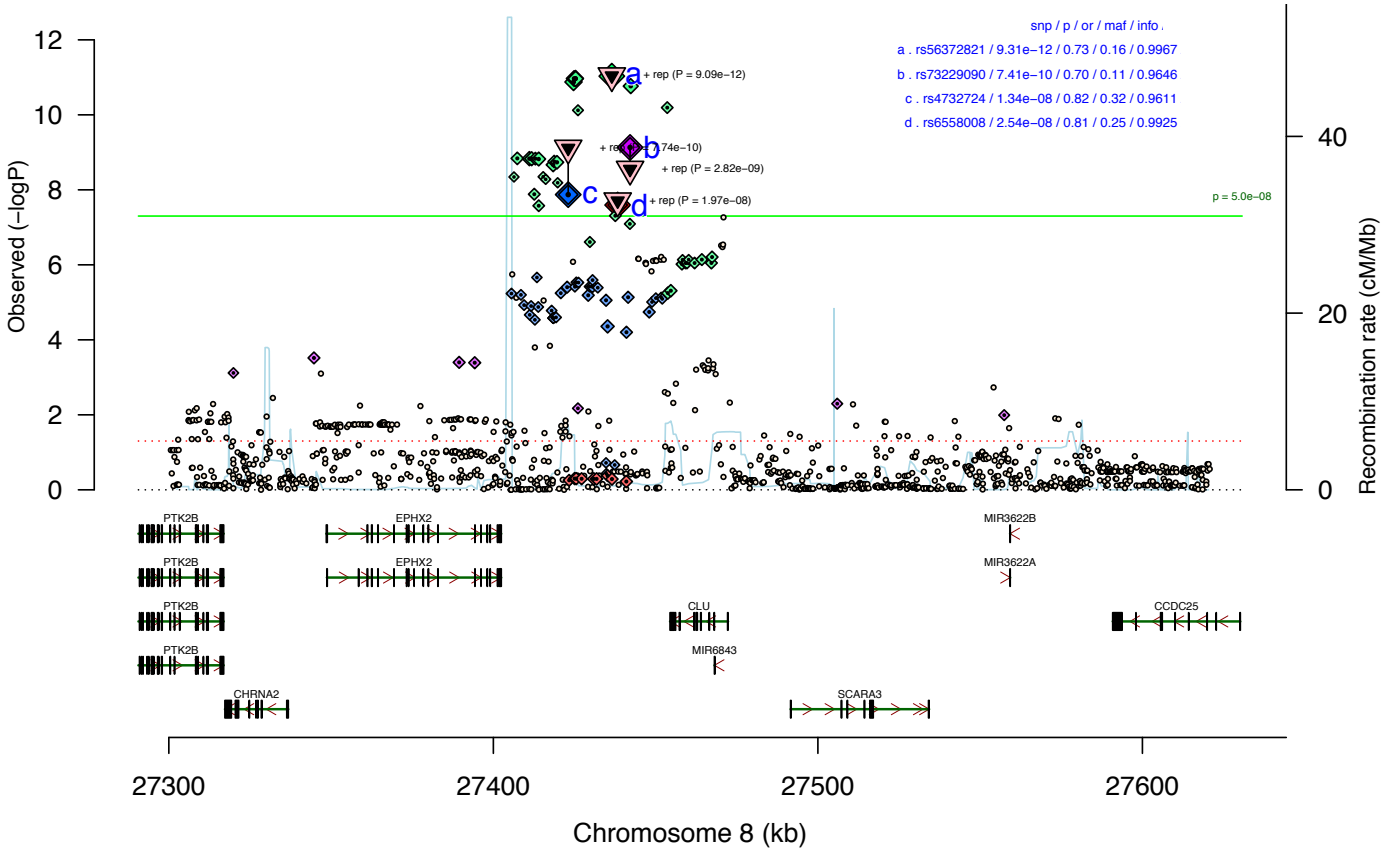

snp / p / or / maf / info .
a . rs56372821 / 9.31e−12 / 0.73 / 0.16 / 0.9967
b . rs73229090 / 7.41e−10 / 0.70 / 0.11 / 0.9646
c . rs4732724 / 1.34e−08 / 0.82 / 0.32 / 0.9611
d . rs6558008 / 2.54e−08 / 0.81 / 0.25 / 0.9925
+ rep (P = 9.09e−12)
+ rep (P = 2.82e−09)
+ rep (P = 1.97e−08)
p = 5.0e−08
a
b
c
d
Observed (−logP)
Recombination rate (cM/Mb)
PTK2B
EPHX2
MIR3622B
MIR3622A
CLU
CCDC25
MIR6843
CHRNA2
SCARA3
27300
27400
27500
27600
Chromosome 8 (kb)

Figure 3.

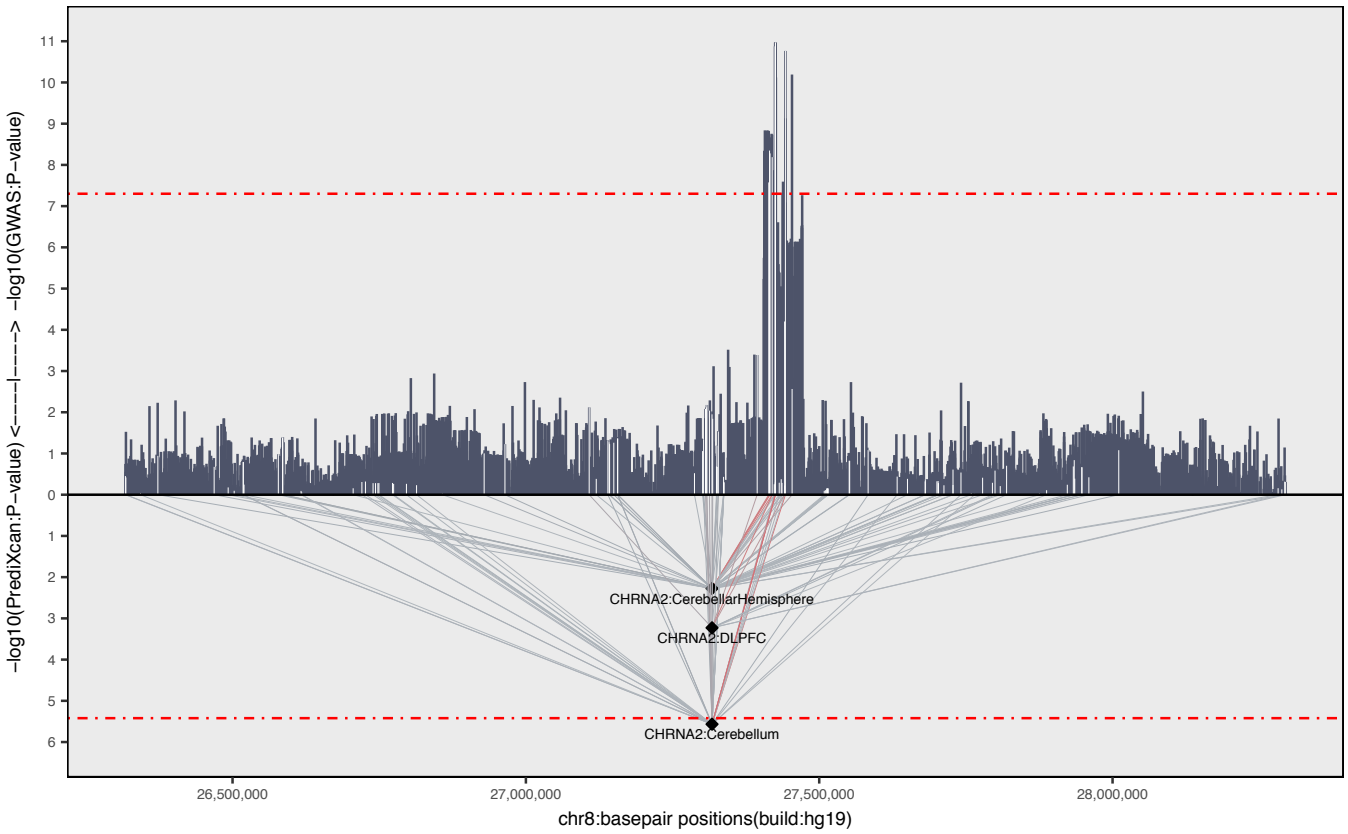
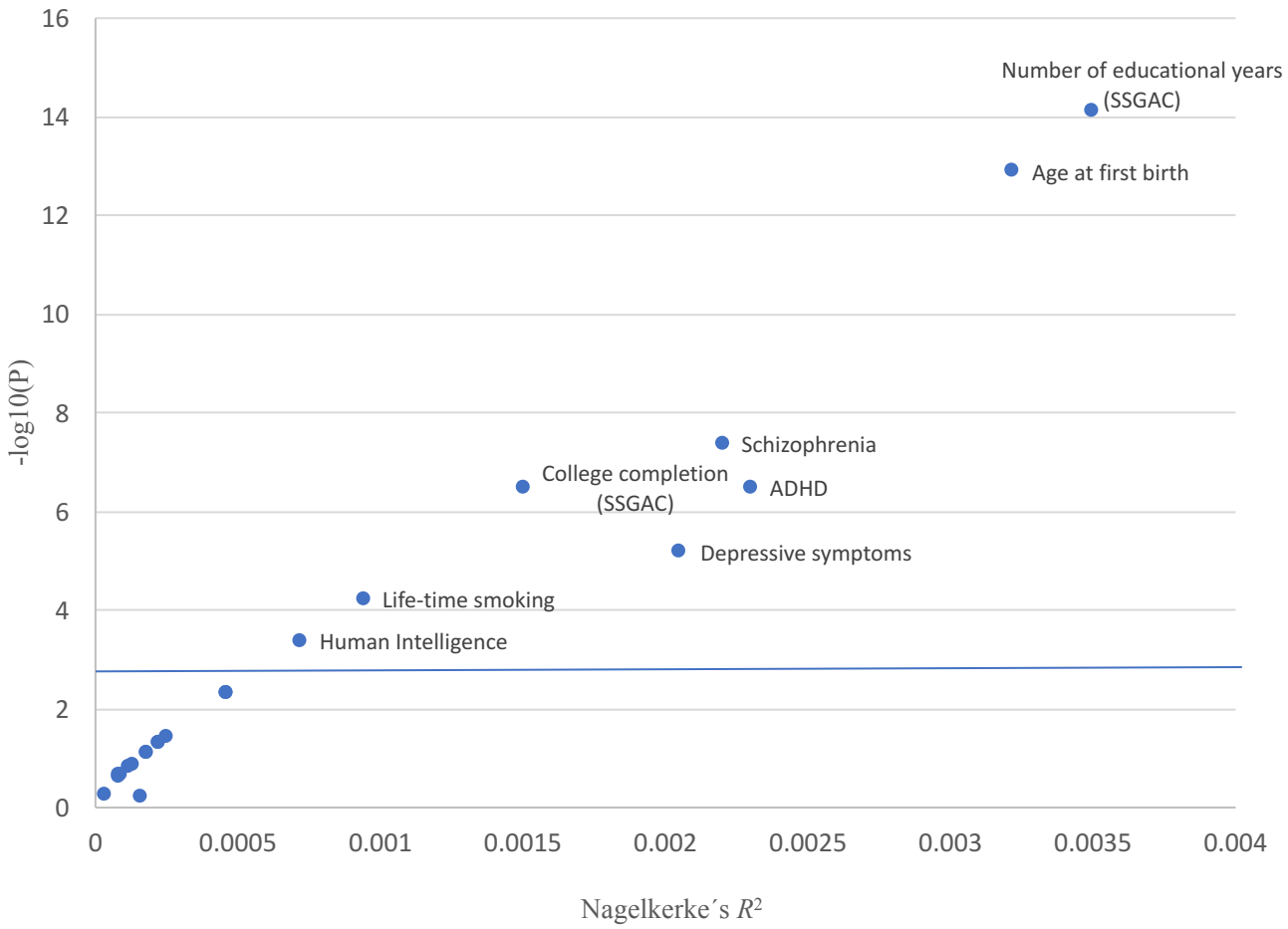

−log10(PrediXcan:P−value) <−−−−I−−−−> −log10(GWAS:P−value)
11
10
9
8
7
6
5
4
3
2
1
0
1
2
3
4
5
6
CHRNA2:CerebellarHemisphere
CHRNA2:DLPFC
CHRNA2:Cerebellum
26,500,000
27,000,000
27,500,000
28,000,000
chr8:basepair positions(build:hg19)

Figure 4.


16
14
12
10
8
6
4
2
0
-log10(P)
Number of educational years (SSGAC)
Age at first birth
Schizophrenia
College completion (SSGAC)
ADHD
Depressive symptoms
Life-time smoking
Human Intelligence
0
0.0005
0.001
0.0015
0.002
0.0025
0.003
0.0035
0.004
Nagelkerke´s R^2