iPSYCH Cannabis use disorder GWAS Results, June 2019 Release

The file "CUD_GWAS_iPSYCH_June2019" contains results from the GWAS of cannabis use disorder by the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) released in June 2019.

Citatio for studies using these data:
Ditte Demontis, Veera Manikandan Rajagopal, Thorgeir E. Thorgeirsson, Thomas D. Als, Jakob Grove, Kalle Leppälä, Daniel F. Gudbjartsson, Jonatan Pallesen, Carsten Hjorthøj, Gunnar W. Reginsson, Thorarinn Tyrfingsson, Valgerdur Runarsdottir, Per Qvist, Jane Hvarregaard Christensen, Jonas Bybjerg-Grauholm, Marie Bækvad-Hansen, Laura M. Huckins, Eli A. Stahl, Allan Timmermann, Esben Agerbo, David M. Hougaard, Thomas Werge, Ole Mors, Preben Bo Mortensen, Merete Nordentoft, Mark J. Daly, Hreinn Stefansson, Kari Stefansson, Mette Nyegaard, Anders D. Børglum. ***Genome-wide association study implicates CHRNA2 in cannabis use disorder.*** Nature Neuroscience, 2019, http://dx.doi.org/10.1038/s41593-019-0416-1

Disclaimer

Methods

See the article for full details. Briefly:

Statistical analyses

Genotyping, quality control and GWAS
DNA was extracted from dried blood spot samples and whole genome amplified in triplicates as described previously. 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 different calling algorithms (GenCall, Birdseed and Zcall). GenCall and Birdseed were used to call genotypes with minor allele frequency (maf) > 0.01.
Stringent quality control was applied and only samples with individual call rate (> 0.98) and genotypes with high call rate (>0.98), no strong deviation from Hardy-Weinberg equilibrium ($P > 1x10^{-6}$ in controls or $P > 1x10^{-10}$ 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) imputation reference panel and SHAPEIT and IMPUTE2. Relatedness and population stratification were evaluated using a set of high-quality markers (genotyped autosomal markers with minor allele frequency (maf) >0.05, HWE $P > 1x10^{-4}$ 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. were excluded). Genetic relatedness was estimated using PLINK v1.9 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 analysis (PCA) using EIGENSOFT. 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. After outlier exclusion PCA was redone and PCs from this analysis were included in the association analysis (see below).
Association analysis was done using logistic regression and the imputed marker dosages including 2,387 CUD cases and 48,985 controls. 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 (See paper: Supplementary Table 1). 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).

File Description

CUD_GWAS_iPSYCH_June2019.gz: GWAS of CUD (2,387 CUD cases and 48,985 controls)

CHR Chromosome (hg19)
SNP Marker name
BP Base pair location (hg19)
A1 Reference allele for OR (may or may not be minor allele)
A2 Alternative allele
INFO Imputation information score
OR Odds ratio for the effect of the A1 allele
SE Standard error of the log(OR)
P P-value for association test in the meta-analysis

Additional Notes

For long insertion/deletion variants, the A1/A2 alleles are truncated to the first 13 bases with a specification of the remaining length (e.g. AACACACACACAC+16)

Allele frequencies and case/control counts per variant are currently omitted from public release for data privacy. For inquiries about accessing this data, please contact Ditte Demontis (ditte@biomed.au.dk).

Data Use Agreement

If you download these data, you and your immediate collaborators (“investigators”) acknowledge and agree to all of the following conditions:

1. These data are provided on an "AS-IS" basis, without warranty of any type, expressed or implied, including but not limited to any warranty as to their performance, merchantability, or fitness for any particular purpose;
2. Investigators will use these results for scientific research and educational use only;
3. The downloaded results can be shared among collaborators but the reposting or public distribution of the result file is prohibited;
4. Investigators certify that they are in compliance with all applicable local, state, and federal laws or regulations and institutional policies regarding human subjects and genetics research;
5. Investigators will cite the appropriate publication in any communications or publications arising directly or indirectly from these data; and
6. Investigators will never attempt to identify any participant.

Experience has taught us that the appropriate use of these data requires considerable attention to detail, prior experience, and technical skill. Errors are easy to make. If investigators use these data, any and all consequences are entirely their responsibility.