Alcohol Dependence GWAS Results, August 2018 Release

Introduction

These are the GWAS results files from the meta-analysis of alcohol dependence by the Psychiatric Genomics Consortium (PGC) Substance Use Disorder (SUD) workgroup released in August 2018. Note that most of the meta-analyses include here are not conventional case/control GWAS; **please be careful** to understand the difference between these analyses before applying these results.

Current citation for studies using these data (preprint to be replaced with final paper upon publication):

Walters, R.K., Adams, M.J., Adkins, A.E., Aliev, F., Bacanu, S., Batzler, A., Bertelsen, S., Biernacka, J.M., Bigdeli, T.B., Chen, L., Clarke, T., Chou, Y., Degenhardt, F., Docherty, A.R., Fontanillas, P., Foo, J., Fox, L., Frank, J., Giegling, I., Gordon, S., Hack, L.M., Hartmann, A.M., Hartz, S.M., Heilmann-Heimbach, S., Herms, S., Hodgkinson, C., Hoffmann, P., Hottenga, J.J., Kennedy, M.A., Alanne-Kinnunen, M., Konte, B., Lahti, J., Lahti-Pulkkinen, M., Ligthart, L., Loukola, A., Maher, B.S., Mbarek, H., McIntosh, A.M., McQueen, M.B., Milaneschi, Y., Palviainen, T., Pearson, J.F., Peterson, R.E., Polimanti, R., Ripatti, S., Ryu, E., Saccone, N.L., Salvatore, J.E., Sanchez-Roige, S., Schwandt, M., Sherva, R., Streit, F., Strohmaier, J., Thomas, N., Wang, J., Webb, B.T., Wedow, R., Wetherill, L., Wills, A.G., 23andMe Research Team, Boardman, J.D., Chen, D., Choi, D., Copeland, W.E., Culverhouse, R.C., Dahmen, N., Degenhardt, L., Domingue, B.W., Elson, S.L., Frye, M., Gäbel, W., Ising, M., Johnson, E.C., Keyes, M., Kiefer, F., Kramer, J., Kuperman, S., Lucae, S., Lynskey, M.T., Maier, W., Mann, K., Männistö, S., McClintick, J.N., Meyers, J.L., Müller-Myhsok, B., Nurnberger, J.I., Palotie, A., Preuss, U., Räikkönen, K., Reynolds, M.D., Ridinger, M., Scherbaum, N., Schuckit, M., Soyka, M., Treutlein, J., Witt, S., Wodarz, N., Zill, P., Adkins, D.E., Boden, J.M., Boomsma, D., Bierut, L.J., Brown, S.A., Bucholz, K.K., Cichon, S., Costello, E.J., de Wit, H., Diazgranados, N., Dick, D.M., Eriksson, J.G., Farrer, L.A., Foroud, T.M., Gillespie, N.A., Goate, A.M., Goldman, D., Grucza, R.A., Hancock, D.B., Harris, K.M., Heath, A.C., Hesselbrock, V., Hewitt, J.K., Hopfer, C., Horwood, J., Iacono, W., Johnson, E.O., Kaprio, J.A., Karpyak, V., Kendler, K.S., Kranzler, H.R., Krauter, K., Lichtenstein, P., Lind, P.A., McGue, M., MacKillop, J., Madden, P.A.F., Maes, H., Magnusson, P., Martin, N.G., Medland, S.E., Montgomery, G.W., Nelson, E.C., Nöthen, M.M., Palmer, A.A., Pedersen, N.L., Penninx, B., Porjesz, B., Rice, J.P., Rietschel, M., Riley, B.P., Rose, R., Rujescu, D., Shen, P., Silberg, J., Stallings, M.C., Tarter, R.E., Vanyukov, M.M., Vrieze, S., Wall, T.L., Whitfield, J.B., Zhao, H., Neale, B.M., Gelernter, J., Edenberg, H.J., & Agrawal, A. Trans-ancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. Preprint at https://www.biorxiv.org/content/early/2018/03/10/257311 (2018).

Please also include the following **acknowledgements statement**:

The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium (PGC-SUD) is supported by funds from NIDA and NIMH to MH109532 and, previously, with analyst support from NIAAA to U01AA008401 (COGA). We gratefully acknowledge our contributing studies and the participants in those studies without whom this effort would not be possible.

Disclaimer

These data are provided "as is", and without warranty, for scientific and educational use only. See the full data use agreement at the end of this document.

Methods

We collected individual genotypic data from 14 case/control studies and 9 family-based studies and summary statistics from GWAS of alcohol dependence (AD) from 5 additional cohorts. AD was defined as meeting criteria for a DSM-IV (or DSM-IIIR in one instance) diagnosis of AD.

QC and imputation of the 14 case/control studies was performed using ricopili (https://github.com/Nealelab/ricopili). For 9 family-based cohorts, an equivalent pipeline, picopili (https://github.com/Nealelab/picopili), was developed for QC, imputation, and analysis appropriate for diverse family structures. After initial sample and variant QC, PCA was used to identify population outliers for exclusion and to stratify European (EU) and African (AA) ancestry samples in each study. Samples were filtered for cryptic relatedness within and between cohorts. Each cohort was imputed using SHAPEIT and IMPUTE2, using the 1000 Genomes Phase 3 reference panel. Imputed SNPs were then filtered for INFO score > 0.8 and allele frequency > 0.01 prior to analysis.

A GWAS for AD status was performed within each ancestry stratum of each sample using an association model appropriate for the study design. For case/control studies, GWAS was performed using logistic regression with imputed dosages. For family-based studies of small, simple pedigrees (e.g., sibships), association with imputed genotypes was tested using generalized estimating equations (GEE). For more complex pedigrees, imputed genotypes were tested using logistic mixed models. Sex was included as a covariate, along with principal components to control for population structure. In addition, subsets of genetically unrelated individuals were selected from each family-based cohort (i.e. taking one individual per family) and used to perform a conventional case/control GWAS using logistic regression.

The primary discovery meta-analysis of all ancestry-stratified GWAS was performed using effective sample size-based weights to account for the different study designs (family vs. case-control). Separate ancestry-specific discovery meta-analyses of EU and AA cohorts, respectively, were also performed.

In addition to the discovery meta-analyses, we conducted meta-analyses for two design subsets. First, we performed sample size weighted meta-analysis of the subset of genetically unrelated individuals in EU and AA cohorts for use in LD score regression analysis. Second, we performed inverse-variance weighted meta-analysis of genetically unrelated individuals in genotyped cohorts to estimate within-ancestry effect sizes for EU (N = 28,757) and AA (N = 5,799). These effect sizes were then used to compare trans-ancestral fine mapping results using inverse-variance weighted fixed effects, random effects[30], and Bayesian[31] models.

File Description

`pgc_alcdep.discovery.aug2018_release.txt.gz`

- Full discovery GWAS (all samples); 14,904 cases, 37,944 controls

`pgc_alcdep.afr_discovery.aug2018_release.txt.gz`

- African ancestry (AA) discovery GWAS; 3,335 cases, 2,945 controls

`pgc_alcdep.eur_discovery.aug2018_release.txt.gz`

- European ancestry (EA) discovery GWAS; 11,569 cases, 34,999 controls

`pgc_alcdep.afr_unrelated.aug2018_release.txt.gz`

- AA GWAS of unrelated individuals (including summary statistic cohorts); 2,991 cases, 2,808 controls

`pgc_alcdep.eur_unrelated.aug2018_release.txt.gz`

- EU GWAS of unrelated individuals (including summary statistic cohorts); 10,206 cases, 28,480 controls

`pgc_alcdep.afr_unrel_genotyped.aug2018_release.txt.gz`

- AA GWAS of unrelated genotyped individuals (excluding summary statistic cohorts); 2,991 cases, 2,808 controls

`pgc_alcdep.eur_unrel_genotyped.aug2018_release.txt.gz`

- EU GWAS of unrelated genotyped individuals (excluding summary statistic cohorts) 8,485 cases, 20,272 controls

`pgc_alcdep.trans_fe_unrel_geno.aug2018_release.txt.gz`

- Fixed effects trans-ancestral GWAS (using unrelated genotyped individuals); 11,476 cases, 23,080 controls

`pgc_alcdep.trans_re2_unrel_geno.aug2018_release.txt.gz`

- Random effects (RE2) trans-ancestral GWAS (using unrelated genotyped individuals); 11,476 cases, 23,080 controls

`pgc_alcdep.trans_mantra_unrel_geno.aug2018_release.txt.gz`

- Bayesian (MANTRA) trans-ancestral GWAS (using unrelated genotyped individuals); 11,476 cases, 23,080 controls

See adapted version of Supplementary Table S2 at end of this readme for more details on the study design for each of these meta-analyses.

File Contents

Each file includes all of the following basic fields:

CHR	Chromosome (hg19)
SNP	Marker name
BP	Base pair location (hg19)
A1	Effect allele (corresponds to the effect size's sign; may not be the minor allele)
A2	Non-effect allele

The coding of the alleles and marker names differs slightly between the files:

- *For "Unrelated Genotyped" and "Trans-ancestral" meta-analyses:* Marker names are given as rsIDs where available (taken from the 1000 Genomes Phase 3 reference panel), otherwise "chromosome:position". For multiallelic variants, "m" is appended to the marker name for different alternative alleles in order to insure that the marker name is unique. 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). These marker names and allele codings can be expected to match PGC studies and other GWAS performed using ricopili with the 1000 Genomes Phase 3 reference panel.

- *For "Discovery" and "Unrelated" meta-analyses:* Marker names are given as rsid:position:ref:alt, or "chr:position:ref:alt" if rsid is unavailable (rsIDs taken from the 1000 Genomes Phase 3 reference panel). Multiallelic variants do not have an appended "m" in the marker name, and indel alleles are not truncated. Note that this means that the marker names and alleles for multiallelic and indel variants likely *will not match* previous PGC studies unless these changes are accounted for. They should match other GWAS using the locus representation from 1000 Genomes Phase 3.

The remaining fields in each file differ between the analyses. For the "Discovery" and "Unrelated" meta-analyses, the remaining fields are:

Z	Z score for association; sign corresponds to the effect of the A1 allele
P	P-value for the test of association

Weight	Total estimated effective sample size

For the “Unrelated Genotyped” meta-analyses (i.e. the meta-analyses excluding summary statistic cohorts), the remaining fields are:

INFO	Imputation information score (weighted average across non-missing cohorts)
OR	Odds ratio for the effect of the A1 allele
SE	Standard error of the log(OR)
P	P-value for the test of association
Neff	Total effective case/control sample size

The fields for the remaining trans-ancestral meta-analyses are analysis specific. In the inverse-variance weighted fixed effects analysis:

INFO	Imputation information score (weighted average across non-missing cohorts)
BETA	Log odds ratio for the effect of the A1 allele
SE	Standard error of the estimated beta
P	P-value for the test of association
Neff	Total effective case/control sample size

In the Hand & Eskin random effects (RE2) meta-analysis:

INFO	Imputation information score (weighted average across non-missing cohorts)
STAT1_RE2	Test statistic for mean effect in the RE2 model
STAT2_RE2	Test statistic for heterogeneous effects in the RE2 model
P	P-value for the overall test of association
Neff	Total effective case/control sample size

Lastly, in the Bayesian (MANTRA) meta-analysis:

INFO	Imputation information score (weighted average across non-missing cohorts)
Log10BF	Log10(Bayes Factor) for association
Neff	Total effective case/control sample size

For all files, allele frequencies and raw case/control counts per variant are currently omitted from public release for data privacy. For inquiries about accessing this data, please contact the SUD Data Access Committee (DAC) representative (pgc.dac.sud@gmail.com).

Data Use Agreement

The PGC has made the full results from all published PGC studies available for download. 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. Downloaded PGC results can be shared among collaborators but the reposting or public distribution of PGC results files 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 PGC publication(s) 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.

Filename	Name	Model	Ancestry	N Cases	N Controls	Family Cohorts	Sum Stats	Notes
pgc_alcdep.discovery.aug2018_release.txt.gz	Discovery	Neff	Both	14,904	37,944	All	Yes	Primary analysis
pgc_alcdep.afr_discovery.aug2018_release.txt.gz	AA Discovery	Neff	Afr.	3,335	2,945	All	Yes	
pgc_alcdep.eur_discovery.aug2018_release.txt.gz	EU Discovery	Neff	Eur.	11,569	34,999	All	Yes	
pgc_alcdep.afr_unrelated.aug2018_release.txt.gz	AA Unrelated	Neff	Afr.	2,991	2,808	Unrelateds	Yes	LDSC, MAGMA
pgc_alcdep.eur_unrelated.aug2018_release.txt.gz	EU Unrelated	Neff	Eur.	10,206	28,480	Unrelateds	Yes	LDSC, MAGMA
pgc_alcdep.afr_unrel_genotyped.aug2018_release.txt.gz	AA Unrel. Genotyped	IVW	Afr.	2,991	2,808	Unrelateds	No	Effect sizes, GRS
pgc_alcdep.eur_unrel_genotyped.aug2018_release.txt.gz	EU Unrel. Genotyped	IVW	Eur.	8,485	20,272	Unrelateds	No	Effect sizes, GRS
pgc_alcdep.trans_fe_unrel_geno.aug2018_release.txt.gz	Trans-ethnic Fixed	IVW	Both	11,476	23,080	Unrelateds	No	
pgc_alcdep.trans_re2_unrel_geno.aug2018_release.txt.gz	Trans-ethnic Random	Han & Eskin	Both	11,476	23,080	Unrelateds	No	
pgc_alcdep.trans_mantra_unrel_geno.aug2018_release.txt.gz	Trans-ethnic Bayes	MANTRA	Both	11,476	23,080	Unrelateds	No	

Adapted version of Supplementary Table S2, summarizing the meta-analysis study design for each of the provided sets of GWAS summary statistics. Neff: effective sample size weighted meta-analysis. IVW: inverse-variance weighted meta-analysis.

MD5 checksums:

```
8a27a895364a2fb3498402de6421f655  pgc_alcdep.afr_discovery.aug2018_release.txt.gz
14ba33f46d90567de98c92771f3e987c  pgc_alcdep.afr_unrelated.aug2018_release.txt.gz
52e2c2a01511f1b4a838f510176113fc  pgc_alcdep.afr_unrel_genotyped.aug2018_release.txt.gz
c5182417c874329cbec967a64f904097  pgc_alcdep.discovery.aug2018_release.txt.gz
26309f918d1aaae65a118eed22b8dc1b  pgc_alcdep.eur_discovery.aug2018_release.txt.gz
5e65ffdf4074a676a83c17d3e68550a2  pgc_alcdep.eur_unrelated.aug2018_release.txt.gz
edb4356c8e9755fc55b622dcd07d4e99  pgc_alcdep.eur_unrel_genotyped.aug2018_release.txt.gz
279b02e792a55eff7d49fdcd806cc780  pgc_alcdep.trans_fe_unrel_geno.aug2018_release.txt.gz
8c091c14726e8e713743dd6e5d93f0de  pgc_alcdep.trans_mantra_unrel_geno.aug2018_release.txt.gz
26661d2ce5b0429a7d189216df63dc77  pgc_alcdep.trans_re2_unrel_geno.aug2018_release.txt.gz
```