

Multi-trait Analysis of GWAS (MTAG) of Substance Use Traits Identifies Novel Genetic Loci and Phenomic Associations

Heng Xu,¹ Sylvanus Toikumo,^{1,2} Richard C. Crist,^{1,2} Klaudia Glogowska,¹ Joseph D. Deak,^{3,4} Joel Gelernter,^{3,4} Emma C. Johnson,⁵ Henry R. Kranzler,^{1,2*} Rachel L. Kember^{1,2*}

¹Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104

²Mental Illness Research, Education and Clinical Center, Veterans Integrated Service Network 4, Crescenzo Veterans Affairs Medical Center, Philadelphia, PA 19104

³Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

⁴Veterans Affairs Connecticut Healthcare Center, West Haven, CT, USA

⁵Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA

*Correspondence to Rachel L. Kember, Ph.D. or Henry R. Kranzler, M.D., Center for Studies of Addiction, 3535 Market Street, Suite 500, Philadelphia, PA 19104; email: rkember@penncare.upenn.edu, telephone: 215-746-1956 or email: kranzler@penncare.upenn.edu; telephone: 215-746-1943

Abstract

Introduction: A large majority of genome-wide significant (GWS) loci identified for substance use traits (SUTs) in genome-wide association studies (GWAS) have been for alcohol and smoking-related phenotypes. GWAS of opioid use disorder (OUD) and cannabis use disorder (CUD) have lagged those of the two historically legal substances. **Methods:** We applied multi-trait analysis of GWAS (MTAG) to 2,888,727 single nucleotide polymorphisms (SNPs) common to GWAS of four SUTs (OUD, CUD, alcohol use disorder [AUD], and smoking initiation [SMK]) in European-ancestry (EUR) subjects. We calculated polygenic risk scores (PRS) for the four traits in an independent sample (i.e., the Yale-Penn sample; N=5,692 EUR) and examined the power increment for each set of MTAG-GWAS summary statistics relative to those of the input GWAS. **Results:** MTAG increased the effective sample size for all four SUTs, which showed high pairwise genetic correlations. After clumping, MTAG identified independent GWS SNPs for all 4 traits: 41 SNPs in 36 loci (including 5 novel loci not previously associated with any SUT) for OUD; 74 SNPs in 60 loci (including 4 novel loci) for CUD; 63 SNPs in 52 loci (including 10 novel loci) for AUD; and 183 SNPs in 144 loci (including 8 novel loci) for SMK. In PRS analyses in the Yale-Penn sample, the MTAG-derived PRS consistently yielded more significant associations with both the corresponding substance use disorder diagnosis and multiple related phenotypes than each of the 4 GWAS-derived PRS. **Conclusions:** MTAG boosted the number of GWS loci for the 4 SUTs, including identifying genes not previously linked to any SUT. MTAG-derived PRS also showed stronger associations with expected phenotypes than PRS for the input GWAS. MTAG can be used to identify novel associations for SUTs, especially those with sample sizes smaller than for historically legal substances.

Introduction

Genome-wide association studies (GWAS) of substance use traits (SUTs) have successfully identified genome-wide significant (GWS) risk variants. However, despite high estimates of heritability (~50%) for many SUTs, difficulty in recruiting large study samples has limited gene discovery efforts¹. Because common genetic variants have small effects on trait susceptibility, large samples are usually needed to provide adequate statistical power to identify GWS variant associations. Thus, variant discovery for some traits [e.g., opioid use disorder (OUD) and cannabis use disorder (CUD)] has lagged that of more common SUTs, including alcohol and smoking-related phenotypes.

Recent GWAS of OUD, CUD, alcohol use disorder (AUD), and smoking initiation (SMK) have identified more than 300 substance-associated loci, some of which are shared between multiple SUTs. SUTs have also shown high degrees of genetic correlation in both GWAS and twin and family studies of the respective traits². A GWAS of OUD in the Million Veteran Program (MVP), with 31,473 OUD cases, identified 10 loci in the cross-ancestry meta-analysis³ including *OPRM1*, *FURIN*, *NCAM1*, and 7 other novel genes. An initial GWAS of cannabis use disorder (CUD) identified one GWS locus (on chromosome 8, near *CHRNA2* and *EPHX2*)⁴, which was subsequently confirmed in a meta-analysis that included 20,916 CUD cases and identified a second GWS locus on chromosome 7 (*FOXP2*)⁵. A cross-ancestry GWAS of AUD in the MVP sample found 26 associated loci, of which 4 (*ZNF804A*, *MLN*, *NICN1*, *MIR5694*), though previously associated with psychiatric or other SUTs, were novel for alcohol-related phenotypes⁶. Finally, a meta-analysis of smoking initiation (SMK) in 1.2 million individuals identified 278 loci (including *NCAM1*)⁷.

Multi-trait analysis of GWAS (MTAG), developed to boost the statistical power of GWAS by incorporating information from effect estimates across traits⁸, enables the joint analysis of multiple, genetically correlated traits. MTAG is also valuable because it can be applied to summary statistics of GWAS rather than requiring individual genotypes, addressing the sample

overlap often present across GWAS discovery samples for different traits using linkage disequilibrium (LD) score regression⁸. Further, it generates trait-specific estimates of the effects of each single nucleotide polymorphism (SNP), yielding a lower mean-squared error than standard, single-trait GWAS estimates when a key assumption of the procedure—that all SNPs share a variance-covariance matrix of effect sizes across traits—is not satisfied⁸.

MTAG has been employed to boost genetic findings for SUTs, both for traits involving the same substance and for cross-substance traits. In a meta-analysis of problematic alcohol use (PAU)⁹, MTAG analysis of PAU and a measure of weekly alcohol consumption⁷ increased the number of independent loci for PAU from 29 to 76. MTAG analysis of OUD with AUD and CUD¹⁰ increased the number of GWS loci for OUD to 18 from 3 in the initial GWAS meta-analysis. Because MTAG for multiple, genetically correlated SUTs yields greater statistical and interpretive power than individual trait GWAS, the availability of large GWAS of alcohol- and smoking-related traits could enhance findings from GWAS of traits with smaller accumulated samples (e.g., CUD and OUD).

Here we conduct an MTAG analysis of the largest available GWAS for four SUTs: OUD, CUD, AUD, and SMK. We integrated information from the four sets of GWAS summary statistics to identify novel loci for each SUT. We also conducted gene prioritization, gene-set, and protein-protein interaction analyses to characterize the underlying biology of the novel genes in the context of SUTs. Finally, we generated polygenic risk scores (PRS) to examine the power increment for each set of MTAG-GWAS summary statistics in an independent sample.

Methods

GWAS Summary Statistics

We examine four substance use traits (SUTs) in European ancestry (EUR) samples: OUD³ (Effective N = 74,635), CUD⁵ (Effective N = 48,900), AUD⁶ (Effective N = 171,601), and SMK⁷ (Effective N = 632,802). Both the OUD and AUD GWAS were conducted in the Million Veteran Program (MVP), a large genomic dataset linked to electronic health records data. The CUD GWAS was a meta-analysis of three datasets (Psychiatric Genomics Consortium, iPSYCH, deCODE). The SMK GWAS was a meta-analysis of multiple datasets conducted by the GWAS consortium of alcohol and nicotine use (GSCAN). Because our downstream PRS analysis used the Yale-Penn dataset, and the CUD GWAS included genotype data from Yale-Penn subjects, we avoided sample overlap by generating summary statistics for CUD using a “leave-one-out” meta-analysis that excluded Yale-Penn subjects. All other GWAS were independent of the Yale-Penn dataset.

Genetic Correlations

We calculated pairwise genetic correlations (r_g) for the four SUTs using linkage disequilibrium score regression (LDSC)¹¹ and HapMap 3 SNPs. Pre-computed European LD scores and weights were downloaded from the LDSC GitHub website (<https://github.com/bulik/ldsc>).

Multi-trait Analysis of GWAS Summary Statistics (MTAG)

Given the high pairwise genetic correlations for the four SUTs, which ranged from 0.45 to 0.80, we conducted a joint analysis of these traits using MTAG, a generalized meta-analysis method that outputs trait-specific SNP associations. It uses bivariate LD score regression to account for potential sample overlap between two or more input summary statistics. We used trait-specific effective sample sizes and transformed Z-scores. SNPs present in all four sets of

SUT summary statistics were included in the MTAG calculation. Default MTAG parameters were used (i.e., each SNP's sample size must be larger than two-thirds of the 90th percentile of all SNPs' sample sizes). We used PLINK1.9 to perform clumping procedures on the four MTAG-GWAS results across a range of 3000 kb and $r^2 > 0.1$. GWS variants located within 1 Mb were merged into a single locus. Loci were annotated with the nearest protein-coding gene (within 1 Mb) using SNPsnap¹². We calculated maxFDR⁸ as the upper bound of FDR for each MTAG-GWAS result. Low maxFDR values support the robustness of the MTAG-GWAS results. Of note, SNP rs1229984, located in the alcohol dehydrogenase gene *ADH1B*, did not pass the default MTAG quality control parameter as it was not present in at least two-thirds of the sample. Because of the well-known strong association of rs1229984 with alcohol phenotypes [6,9], we conducted a separate analysis with a less stringent filter to include that SNP and report the results in Supplemental Table 18.

Identification of novel variants for each SUT

We systematically evaluated whether variants identified in the MTAG analysis were previously associated with either the primary SUT or other SUTs. First, for each lead variant, we determined whether any nearby variant (within 1 Mb) was GWS in the initial GWAS or in the other three contributing GWAS. We then determined whether the locus had prior SUT associations using the GWAS catalog¹³ implemented in FUMA¹⁴, annotating all lead variants with prior associations with any SUT (not limited to the GWAS included in this study). For completeness, we also labelled the variant with the closest protein-coding gene (within 1 Mb) and searched the GWAS catalog¹³ for prior associations of that gene with SUTs. Loci that were not previously GWS for any SUT were labeled as “novel”.

Gene-Set Analysis

To determine whether the identified genes are involved in important biological processes, GWS SNPs were mapped to genes by ANNOVAR¹⁵ implemented in FUMA¹⁴. Next, we curated gene-set enrichment and Gene Ontology (GO) terms using the GO annotation database^{16,17}, with gene-set enrichment p-values adjusted using a Bonferroni correction for each test.

Protein-Protein Interaction

We used STRING database v11.5¹⁸ to conduct protein-protein interaction (PPI) analyses. For each SUT MTAG result, we used annotated GWS genes as input to query the PPIs in the database. PPI enrichment p-value and pairwise interaction scores were reported by the STRING database. We used a cut-off interaction score >0.4 to identify PPIs with high confidence.

Yale-Penn Dataset

The Yale-Penn sample was recruited for genetic studies of substance use disorders (SUDs)¹⁹. It was deeply phenotyped using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA), a comprehensive psychiatric interview schedule that assesses the physical, psychosocial, and psychiatric manifestations of SUDs and co-occurring psychiatric disorders^{20,21}. Using the SSADDA, trained interviewers elicit information on demographics, substance use history, psychosocial history, medical history, and lifetime diagnostic criteria for DSM-IV²² and DSM-5²³ substance use and DSM-IV psychiatric disorders. Diagnoses and individual criteria for SUDs and psychiatric disorders obtained with the SSADDA show acceptable reliability^{20,21}. The Yale-Penn PheWAS dataset includes 5,692 unrelated European-ancestry genotyped individuals and over 650 summarized phenotypes categorized into 20 substance, medical, demographic, and psychiatric sections¹⁹.

Polygenic Risk Score (PRS) and Phenotype Association Test

We used PRS-continuous shrinkage (PRS-CS)²⁴ to generate PRS in the Yale-Penn dataset¹⁹, pre-computed LD reference for HapMap3 SNPs in EUR samples to account for local LD, and an optimal global shrinkage parameter learned from the data. We generated four GWAS-based PRS and four MTAG-based PRS, i.e., one of each for each of the four SUTs. To ensure comparability between GWAS-based and MTAG-based PRS, only the SNPs used in the MTAG calculation were included in PRS calculation. Effective sample size was used to generate GWAS- and MTAG-based PRS.

To compare the power of GWAS-based PRS with MTAG-based PRS, we tested the associations of each PRS in the Yale-Penn dataset with phenotypes in the corresponding substance section (e.g., AUD PRS were tested with alcohol-related phenotypes). We used linear regression or logistic regression models as appropriate, covarying for age, sex, and 10 genetic principal components. Each substance section includes DSM-IV and DSM-5 diagnoses for the corresponding substance use disorder. For diagnoses, we calculated the incremental pseudo R^2 value after adding the polygenic score to the logistic regression model.

Results

Genetic Correlation Between SUTs

Pairwise r_g calculated using LDSC were significant and were moderate or high across all pairs of SUTs (Figure 1, Supplementary Table 1). The strongest r_g was between OUD and AUD (0.80), followed by OUD and CUD (0.67), CUD and SMK (0.65), CUD and AUD (0.63), AUD and SMK (0.52) and OUD and SMK (0.45).

MTAG SUTs and Locus Discovery

We retained 2,888,727 SNPs for the MTAG analysis after extracting common SNPs from all four input sets of GWAS summary statistics and filtering them with the criteria described in the MTAG methods description. The mean χ^2 statistics for the MTAG-GWAS results are: $\chi^2_{MTAG- OUD} = 1.38$, $\chi^2_{MTAG-CUD} = 1.46$, $\chi^2_{MTAG-AUD} = 1.46$ and $\chi^2_{MTAG-SMK} = 1.87$. The maxFDR values are 6.7×10^{-2} for MTAG-OUD, 1.1×10^{-1} for MTAG-CUD, 8.9×10^{-3} for MTAG-AUD and 1.2×10^{-4} for MTAG-SMK (Table 1).

Among the four sets of MTAG results (Figure 2, Supplementary Tables 2-5), 20 loci were significantly associated with all four SUTs, the most significant of which was the locus containing *NCAM1*, supported by two intronic variants in complete LD (rs1940701 for OUD and AUD and rs4479020 for CUD and SMK). Other loci associated with all four SUTs included intronic variants in *DPP4* (rs6432708) and *CADM2* (rs62250713) and an intergenic variant near *ZNF184* (rs35984974).

Opioid use disorder (MTAG-OUD)

MTAG increased the effective sample size for OUD from 74,635 in the input GWAS to 176,876 in the MTAG analysis. After clumping, we identified 41 independent GWS SNPs in 36 genomic risk loci (Supplementary Table 2). Of these, 3 loci were GWS in the OUD GWAS, 17 were GWS in at least one of the other of the three SUT GWAS used for MTAG, and 30 were

GWS in other GWAS of SUTs, not limited to the GWAS included in this study. Variants including exonic SNPs in *OPRM1* (rs1799971, $p = 7.52 \times 10^{-9}$) and *FURIN* (rs4702, $p = 3.29 \times 10^{-12}$) have been shown to be GWS in previous OUD GWAS^{3,10}. Five GWS loci are novel with no prior associations found with OUD or any SUT. These include a variant in the 3' UTR of *POR* (rs17685, $p = 1.42 \times 10^{-8}$), intronic variants in *CNOT4* (rs2696880, $p = 3.97 \times 10^{-9}$) and *MTMR2* (rs7110786, $p = 1.38 \times 10^{-8}$), and intergenic variants near *TMEM170B* (rs112126124, $p = 8.35 \times 10^{-9}$) and *SNAI1* (rs73274724, $p = 9.92 \times 10^{-9}$).

Cannabis use disorder (MTAG-CUD)

MTAG increased the effective sample size from 48,900 to 223,956, yielding 74 independent GWS SNPs in 60 genomic risk loci (Supplementary Table 3). Of the 60 loci, 3 were GWS in the CUD GWAS, 43 were GWS in at least one of the other three SUT GWAS used for MTAG, and 51 were GWS in other GWAS of SUTs. We replicated previously associated loci for CUD⁵, including an intergenic variant near *CLU* in the same locus identified in the input CUD GWAS, an intronic variant in *GBF1* (rs4919626, $p = 1.38 \times 10^{-8}$), and an exonic variant in *FURIN* (rs4702, $p = 4.69 \times 10^{-8}$). Four novel GWS loci with no prior associations with any SUT were identified, including three intronic variants - one each in *CNOT4* (rs2696880, $p = 4.44 \times 10^{-8}$), *TMEM245* (rs11794648, $p = 4.26 \times 10^{-8}$), and *MTMR2* (rs7110786, $p = 8.21 \times 10^{-9}$) - and an intergenic variant near *TMEM170B* (rs112126124, $p = 4.47 \times 10^{-9}$).

Alcohol Use Disorder (MTAG-AUD)

With an effective sample size that increased from 171,601 in the input GWAS to 282,208 in the MTAG analysis, we identified 63 independent GWS SNPs in 52 genomic risk loci for AUD (Supplementary Table 4). Of these, 9 loci were GWS in the AUD GWAS, 23 were GWS in at least one of the other three SUT GWAS used for MTAG, and 40 were GWS in other GWAS of SUTs. These included multiple previously reported AUD-associated variants – exonic variants in

GCKR (rs1260326, $p = 1.53 \times 10^{-11}$) and *SLC39A8* (rs13107325, $p = 1.47 \times 10^{-18}$), and intronic variants in *ANKK1* (rs12360992, $p = 3.29 \times 10^{-8}$) and *FTO* (rs7206122, $p = 2.03 \times 10^{-9}$)⁶. We also identified 10 novel GWS loci, including exonic variants in *POR* (rs17685, $p = 3.44 \times 10^{-8}$) and *SYNGAP1* (rs411136, $p = 9.10 \times 10^{-9}$); intronic variants in *DNM3* (rs742510, $p = 9.63 \times 10^{-9}$), *CSMD3* (rs6469450, $p = 8.91 \times 10^{-9}$), *CNOT4* (rs2686880, $p = 8.26 \times 10^{-9}$), *LRFN5* (rs58734839, $p = 4.55 \times 10^{-9}$), *ZNF804A* (rs1366839, $p = 1.53 \times 10^{-8}$), and *TCF20* (rs9306356, $p = 1.30 \times 10^{-8}$); and two intergenic variants near *TMEM170B* (rs112126124, $p = 5.62 \times 10^{-11}$) and *SORCS3* (rs73274724, $p = 1.27 \times 10^{-8}$).

Smoking Initiation (MTAG-SMK)

With an effective sample size increase from 637,082 to 709,603, MTAG identified 183 independent GWS SNPs in 144 genomic risk loci (Supplementary Table 5). Of these, 86 were GWS in the SMK GWAS, 7 were GWS in at least one of the other three SUT GWAS used for MTAG, and 130 were GWS in other GWAS of SUTs. Eight GWS loci were novel, including a variant within *TNRC6B* (rs5750911, $p = 2.38 \times 10^{-9}$) and seven intergenic variants: one each near *WDR12* (rs4675308, $p = 2.50 \times 10^{-8}$), *PCDH7* (rs7680926, $p = 8.81 \times 10^{-9}$), *ITGA1* (rs7680926, $p = 4.05 \times 10^{-8}$), *TMEM170B* (rs12526369, $p = 7.41 \times 10^{-9}$), *SP4* (rs6974377, $p = 3.37 \times 10^{-9}$), *CTDP1* (rs56235016, $p = 7.76 \times 10^{-9}$), and *SORCS3* (rs10884186, $p = 4.01 \times 10^{-8}$).

Gene-Set Analysis

After Bonferroni correction, we identified enriched gene sets for all studied traits including 28 for MTAG-OD, 73 for MTAG-CUD, 51 for MTAG-AUD, and 70 for MTAG-SMK, as shown in Supplementary Tables 6-9. The gene-set “regulation of cell differentiation”, which contains the novel gene *POR* and the opioid-specific gene *OPRM1*, showed significant enrichment for OD ($P_{\text{Bon}} = 1.0 \times 10^{-3}$). “Protein dimerization activity”, one of the significantly

enriched gene-sets for CUD ($P_{\text{Bon}} = 3.7 \times 10^{-9}$), harbors the novel genes *MTMR2* and *FOXP2*. For AUD, the enriched gene set “regulation of synapse structure or activity” ($P_{\text{Bon}} = 1.2 \times 10^{-2}$) contained the novel genes *DNM3*, *LRFN5*, and *SYNGAP1* along with *DRD2*. For SMK, 70 gene sets showed significant enrichment, with the novel gene *TNRC6B* mapping to 12 of them. Of note, the gene set “neuron differentiation”, which includes *NCAM1*, is significantly enriched ($P_{\text{Bon}} = 3.1 \times 10^{-2}$).

Protein-protein Interaction

Using the STRING database, we observed significant PPI enrichment ($p < 0.05$) for genes that were GWS for MTAG-AUD and MTAG-SMK, whereas PPI enrichment was non-significant for GWS genes in MTAG-ODD ($p = 0.175$) and MTAG-CUD ($p = 0.175$) (Supplementary Tables 10-13). For GWS genes in MTAG-AUD, we identified high PPI for *NF1* and *SYNGAP1* (interaction score = 0.697), *NF1* and *CSMD3* (interaction score = 0.42), and *NCAM1* and *SEMA6D* (interaction score = 0.467). For GWS genes in MTAG-SMK, *TNRC6B* and *FOXO3*, two smoking-associated genes, showed high PPI (interaction score = 0.9).

PRS Associations with Primary Phenotypes

As shown in Figure 3, all four MTAG-based PRS (PRS_{MTAG}) showed stronger associations with the primary diagnosis than the individual-trait GWAS-based PRS (PRS_{GWAS}) (Supplementary Tables 14-17). This difference was most evident for DSM-IV cannabis dependence, where the Bonferroni-corrected association was non-significant with $\text{PRS}_{\text{GWAS-CUD}}$ (OR = 1.16, $p = 1.94 \times 10^{-3}$) but significant with $\text{PRS}_{\text{MTAG-CUD}}$ (OR = 1.30, $p = 2.18 \times 10^{-9}$). Associations with the corresponding primary diagnoses for the other three SUTs were all orders of magnitude less significant for the PRS_{GWAS} than the PRS_{MTAG} : $\text{PRS}_{\text{GWAS-ODD}}$: OR = 1.30, $p = 2.64 \times 10^{-16}$ vs. $\text{PRS}_{\text{MTAG-ODD}}$: OR = 1.50, $p = 3.34 \times 10^{-33}$; $\text{PRS}_{\text{GWAS-AUD}}$: OR = 1.27, $p = 5.44 \times 10^{-12}$

vs. $PRS_{MTAG-AUD}$: OR = 1.40, $p = 9.94 \times 10^{-21}$; $PRS_{GWAS-SMK}$: OR = 1.61, $p = 1.57 \times 10^{-44}$ vs.

$PRS_{MTAG-SMK}$: OR = 1.71, $p = 5.14 \times 10^{-53}$.

For each of the SUTs, there were more Bonferroni-corrected significant associations with PRS_{MTAG} than PRS_{GWAS} (Supplementary Tables 14-17). Notably, the number of significantly associated phenotypes for $PRS_{GWAS-CUD}$ was six, while for $PRS_{MTAG-CUD}$ it was 25. Phenotypes that became significantly associated when using $PRS_{MTAG-CUD}$ included age of first use of marijuana, the DSM-IV cannabis dependence diagnosis, and the DSM-5 CUD criterion count.

In addition, the incremental R^2 values both for diagnoses and related phenotypes were higher for PRS_{MTAG} than PRS_{GWAS} (Table 2). The greatest improvement was between $PRS_{MTAG-CUD}$ and $PRS_{GWAS-CUD}$, where the incremental R^2 for “ever used cannabis” was 1.62% and 0.38%, respectively. Moderate improvement was also observed for $PRS_{MTAG-SMK}$ compared with the well-powered $PRS_{GWAS-SMK}$.

Discussion

We performed a joint analysis of four SUTs using MTAG, which yielded an effective sample size that was up to four-fold that of the four individual GWAS. This led to our identifying novel associated variants in loci not previously linked to any SUT. Among the most significant risk loci for substance-use phenotypes are those that encode proteins with clear connections to the substance involved. This includes the mu-opioid receptor for OUD²⁵, various nicotinic cholinergic receptor subunits for smoking initiation⁷, and alcohol metabolism enzymes for alcohol consumption and AUD²⁶.

However, beyond the substance-specific proteins that directly interact with the drug are a wide range of biological mechanisms common to all addictive behavior, which includes reward pathways, learning and memory, withdrawal, and other functions²⁷. We therefore expected that the greater statistical power of MTAG would reveal novel genes with associations to multiple substances based on common mechanisms of risk (i.e., an addiction factor^{28,29}). In fact, five of the 19 novel loci identified in our analysis were significantly associated with two or more substances. Variants in *POR* were significant for OUD and AUD, in *MTMR2* was significant for OUD and CUD, near *SORCS3* was significant for AUD and SMK, in *CNOT4* was significant for OUD, CUD and AUD, and near *TMEM170B* was significant for all four SUTs. Interestingly, *CNOT4* was recently identified in a GWAS of maximum alcohol use³⁰, which supports our discovery here. The use of multiple substances is common and has been associated with poorer treatment outcomes in individuals with SUDs³¹. These shared genes may represent targets for therapies aimed at treating co-occurring SUDs.

POR encodes a reductase that contributes electrons to cytochrome P450 enzymes³², which are essential components of drug metabolism. Variation in *POR* can affect the enzymatic activities of multiple CYP450 enzyme family members³³⁻³⁵, including CYP2D6 and CYP3A4³³, enzymes responsible for the primary metabolism of many opioids. Our MTAG results indicate that the A-allele of the *POR* variant rs17685 is associated with decreased risk of OUD and AUD.

This SNP was also found to be an expression quantitative trait locus (eQTL) for *POR*, with the A-allele predicting higher expression in cerebellum³⁶. If increased *POR* correlates with increased CYP450 activity, the association of rs17685 with OUD could reflect altered opioid pharmacokinetics. There may be a similar connection between *POR* and AUD. CYP2E1, which accounts for ~20% of alcohol metabolism^{37,38}, has been shown to interact with *POR*³⁹. CYP2E1 expression is also induced by ethanol itself³⁸. Thus, there is a potential synergistic effect wherein rs17685-A allele carriers who consume alcohol may simultaneously have an increased amount and activity of the enzyme.

We also identified novel loci associated with individual SUTs. For OUD, we detected a novel variant near *SNAI1* and for CUD, a novel intronic variant in *TMEM245*. For AUD, six variants were identified that had not previously been associated with any SUT (*SYNGAP1*, *DNM3*, *CSMD3*, *LRFN5*, *ZNF804A*, *TCF20*). For SMK, six novel variants not associated with any other SUTs were identified (*TNRC6B*, *WDR12*, *PCDH7*, *ITGA1*, *SP4*, *CTDP1*).

In addition to polysubstance use, patients with SUDs also have higher rates of comorbid psychiatric disorders than the general population⁴⁰. Phenome-wide association studies (PheWAS) using data from electronic health records (EHR) have shown associations between PRS for SUTs and non-substance use psychiatric diagnostic codes⁹, suggesting genetic overlap between them. Consistent with this hypothesis, nine of the 19 novel genes identified in our MTAG analysis were significant in GWAS of depression or schizophrenia⁴¹⁻⁴⁵.

A prior MTAG analysis of PAU⁹, which leveraged information from a GWAS of drinks per week, identified 119 GWS variants. Of the variants in common between that MTAG analysis and the present analysis, we also observed association with 14 of the 45 SNPs in our MTAG-AUD results. In our MTAG-OUD analysis, we also replicated six of the nine variants in common with a recent OUD MTAG analysis of OUD, CUD and AUD¹⁰. Although *OPRM1* was not significantly associated in the prior MTAG, our MTAG-OUD analyses identified the *OPRM1* SNP rs1799971

as a lead variant, potentially due to our inclusion of the smoking initiation GWAS or the larger input OUD GWAS sample in our analysis.

We functionally annotated the GWS loci in all four SUTs to explore plausible underlying biological processes. We found that associations for AUD were enriched in genes (including *CNOT4*) involved in regulating synapse structure and activity, an enrichment not previously observed in the AUD GWAS⁶. In addition, we find evidence for neuronal differentiation in smoking initiation, in line with findings in the original SMK GWAS⁷. In the PPI analysis, we observed significantly enriched protein interaction networks for MTAG-AUD and MTAG-SMK. In addition, we observed multiple interactions between novel genes identified by MTAG and previously identified SUT-associated genes, which provide biological support for the MTAG results.

In PRS analyses, the PRS_{MTAG} outperformed the PRS_{GWAS} for all SUTs. This was of particular importance for the CUD PRS, where the PRS_{MTAG} was significantly associated with the diagnosis, whereas the PRS_{GWAS} was not. As PRS for many traits are being considered as potential biomarkers for disorders⁴⁶, the use of MTAG may yield more powerful PRS without having to recruit larger samples for GWAS. However, careful evaluation is required to assess the broader impact of MTAG-based PRS on phenotypic associations beyond the primary phenotype. A loss of specificity may result from the inclusion of multiple genetically correlated traits, potentially confounding PheWAS of MTAG-PRS.

In summary, in an MTAG analysis of four SUTs we identified 19 novel loci and, in an independent dataset, found that the associations of PRS_{MTAG} with relevant traits were more significant than with PRS_{GWAS} . As the size of GWAS samples continues to increase, MTAG analyses could provide a complementary method that leverages more powerful GWAS to boost the findings of risk variants for genetically correlated traits in which case ascertainment is more challenging, thereby enhancing our understanding of the biology underlying these phenotypes.

Figure Legends

Figure 1: Overview of the analysis. The four SUTs that were used in MTAG-analysis have pairwise genetic-correlations between 0.45 to 0.80. For each SUT MTAG-GWAS result, we identified novel loci, performed gene-set analysis, protein-protein interaction analysis, and examined the increased predictive power of the corresponding polygenic risk score.

Figure 2: Manhattan plots of MTAG-ODD ($N_{\text{Effective}} = 176,876$), MTAG-CUD ($N_{\text{Effective}} = 223,956$), MTAG-AUD ($N_{\text{Effective}} = 282,208$) and MTAG-SMK ($N_{\text{Effective}} = 709,603$). Dashed-lines indicate genome-wide significance ($P < 5 \times 10^{-8}$) and yellow dots indicate genome-wide significance SNPs.

Figure 3: Comparison of GWAS-based PRS and MTAG-based PRS. Red and blue dots in each plot represent the GWAS-based polygenic risk score and MTAG-based polygenic risk score respectively. Vertical dashed lines indicate significance threshold after Bonferroni correction.

References

1. Palotie, A., Widén, E. & Ripatti, S. From genetic discovery to future personalized health research. *New Biotechnol.* **30**, 291–295 (2013).
2. Deak, J. D. & Johnson, E. C. Genetics of substance use disorders: a review. *Psychol. Med.* **51**, 2189–2200.
3. Kember, R. L. *et al.* Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects on brain. 2021.12.13.21267480 (2021) doi:10.1101/2021.12.13.21267480.
4. Demontis, D. *et al.* Genome-wide association study implicates CHRNA2 in cannabis use disorder. *Nat. Neurosci.* **22**, 1066–1074 (2019).
5. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *Lancet Psychiatry* **7**, 1032–1045 (2020).
6. Kember, R. L. *et al.* Genetic underpinnings of the transition from alcohol consumption to alcohol use disorder: shared and unique genetic architectures in a cross-ancestry sample. 2021.09.08.21263302 (2021) doi:10.1101/2021.09.08.21263302.
7. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics* vol. 51 237–244 (2019).
8. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
9. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.* **23**, 809–818 (2020).

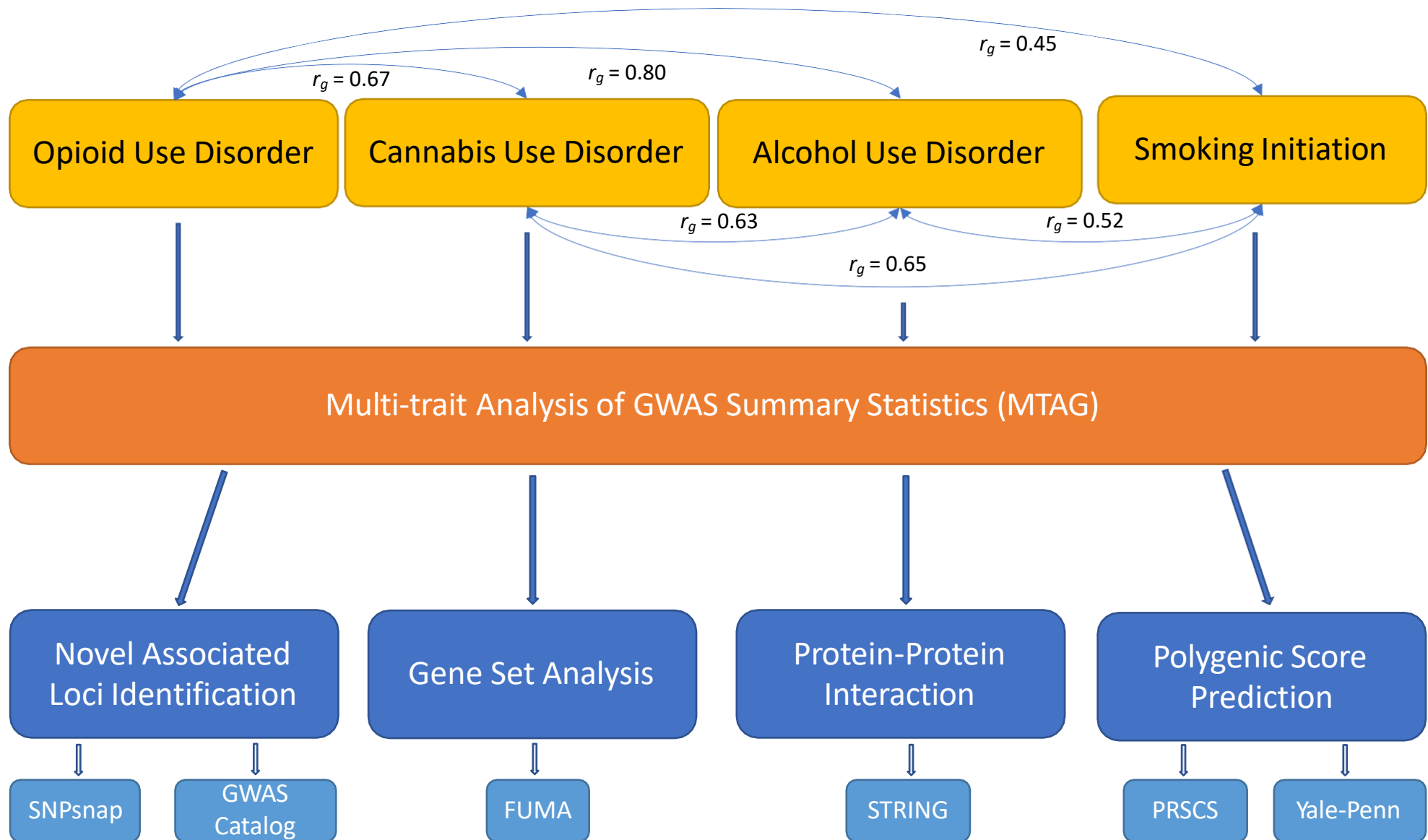
10. Deak, J. D. *et al.* Genome-wide association study and multi-trait analysis of opioid use disorder identifies novel associations in 639,709 individuals of European and African ancestry. 2021.12.04.21267094 (2021) doi:10.1101/2021.12.04.21267094.
11. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
12. Pers, T. H., Timshel, P. & Hirschhorn, J. N. SNPsnap: a Web-based tool for identification and annotation of matched SNPs. *Bioinformatics* **31**, 418–420 (2015).
13. Buniello, A. *et al.* The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012 (2019).
14. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
15. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
16. Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
17. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
18. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

19. Kember, R. L. *et al.* Phenome-wide association analysis of substance use disorders in a deeply phenotyped sample. 2022.02.09.22270737 (2022) doi:10.1101/2022.02.09.22270737.
20. Pierucci-Lagha, A. *et al.* Diagnostic reliability of the Semi-structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend.* **80**, 303–312 (2005).
21. Pierucci-Lagha, A. *et al.* Reliability of DSM-IV Diagnostic Criteria Using the Semi-Structured Assessment for Drug Dependence and Alcoholism (SSADDA). *Drug Alcohol Depend.* **91**, 85–90 (2007).
22. Diagnostic and Statistical Manual of Mental Disorders, 4th ed. (DSM-IV). *Am. J. Psychiatry* **152**, 1228–1228 (1995).
23. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (<https://doi.org/10.1176/appi.books.9780890425596.dsm05>).
24. Ge, T., Chen, C.-Y., Ni, Y., Feng, Y.-C. A. & Smoller, J. W. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat. Commun.* **10**, 1776 (2019).
25. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* **77**, 1072–1080 (2020).
26. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, 1499 (2019).
27. Koob, G. F. & Volkow, N. D. Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry* **3**, 760–773 (2016).

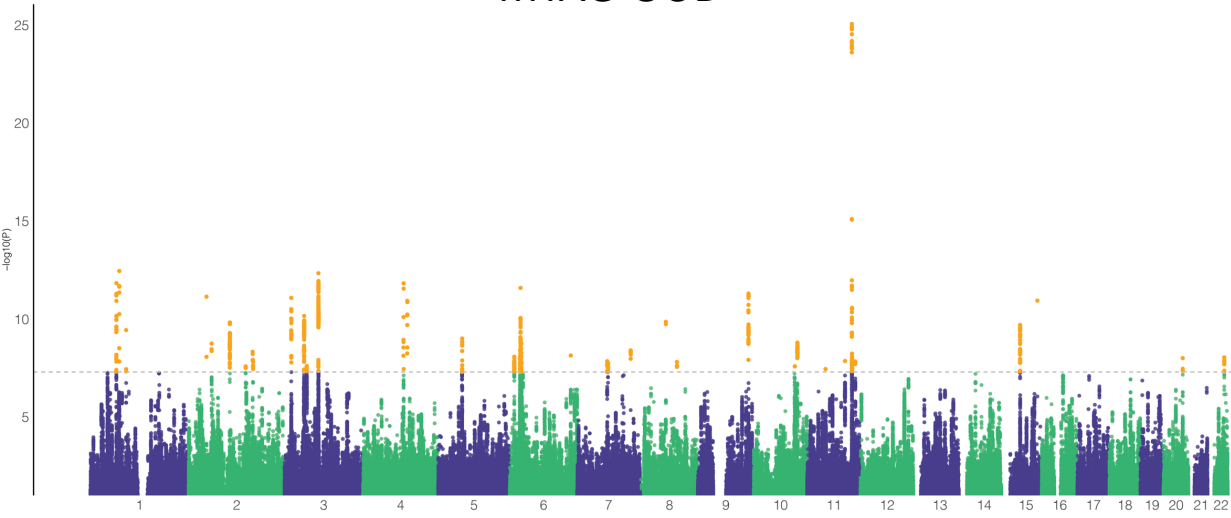
28. Hatoum, A. S. *et al.* Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. 2022.01.06.22268753 (2022) doi:10.1101/2022.01.06.22268753.
29. Hatoum, A. S. *et al.* The addiction risk factor: A unitary genetic vulnerability characterizes substance use disorders and their associations with common correlates. *Neuropsychopharmacology* 1–7 (2021) doi:10.1038/s41386-021-01209-w.
30. Deak, J. D. *et al.* Genome-wide investigation of maximum habitual alcohol intake (MaxAlc) in 247,755 European and African Ancestry U.S. Veterans informs the relationship between habitual alcohol consumption and alcohol use disorder. 2022.05.02.22274580 (2022) doi:10.1101/2022.05.02.22274580.
31. Wang, L. *et al.* Polydrug use and its association with drug treatment outcomes among primary heroin, methamphetamine, and cocaine users. *Int. J. Drug Policy* **49**, 32–40 (2017).
32. Pandey, A. V. & Sproll, P. Pharmacogenomics of human P450 oxidoreductase. *Front. Pharmacol.* **5**, 103 (2014).
33. Tomalik-Scharte, D. *et al.* Impaired hepatic drug and steroid metabolism in congenital adrenal hyperplasia due to P450 oxidoreductase deficiency. *Eur. J. Endocrinol.* **163**, 919–924 (2010).
34. Oneda, B. *et al.* The P450 oxidoreductase genotype is associated with CYP3A activity in vivo as measured by the midazolam phenotyping test. *Pharmacogenet. Genomics* **19**, 877–883 (2009).

35. Yang, G. *et al.* Effects of the CYP Oxidoreductase Ala503Val Polymorphism on CYP3A Activity In Vivo: A Randomized, Open-Label, Crossover Study in Healthy Chinese Men. *Clin. Ther.* **33**, 2060–2070 (2011).
36. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
37. Vincent, D., Cohadon, S., Loiseau, P. & Fontanges, X. The electroencephalogram in acute necrotizing encephalitis. *Rev. Neurol. (Paris)* **120**, 466–475 (1969).
38. Stahl, H. D. *et al.* Differential antibody screening of cloned *Plasmodium falciparum* sequences expressed in *Escherichia coli*: procedure for isolation of defined antigens and analysis of human antisera. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 2456–2460 (1984).
39. Jamakhandi, A. P., Kuzmic, P., Sanders, D. E. & Miller, G. P. Global analysis of protein-protein interactions reveals multiple CYP2E1-reductase complexes. *Biochemistry* **46**, 10192–10201 (2007).
40. Langås, A.-M., Malt, U. F. & Opjordsmoen, S. Comorbid mental disorders in substance users from a single catchment area - a clinical study. *BMC Psychiatry* **11**, 25 (2011).
41. Baselmans, B. M. L. *et al.* Multivariate genome-wide analyses of the well-being spectrum. *Nat. Genet.* **51**, 445–451 (2019).
42. Levey, D. F. *et al.* Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. *Nat. Neurosci.* **24**, 954–963 (2021).
43. Xu, C. *et al.* BCL9 and C9orf5 are associated with negative symptoms in schizophrenia: meta-analysis of two genome-wide association studies. *PloS One* **8**, e51674 (2013).

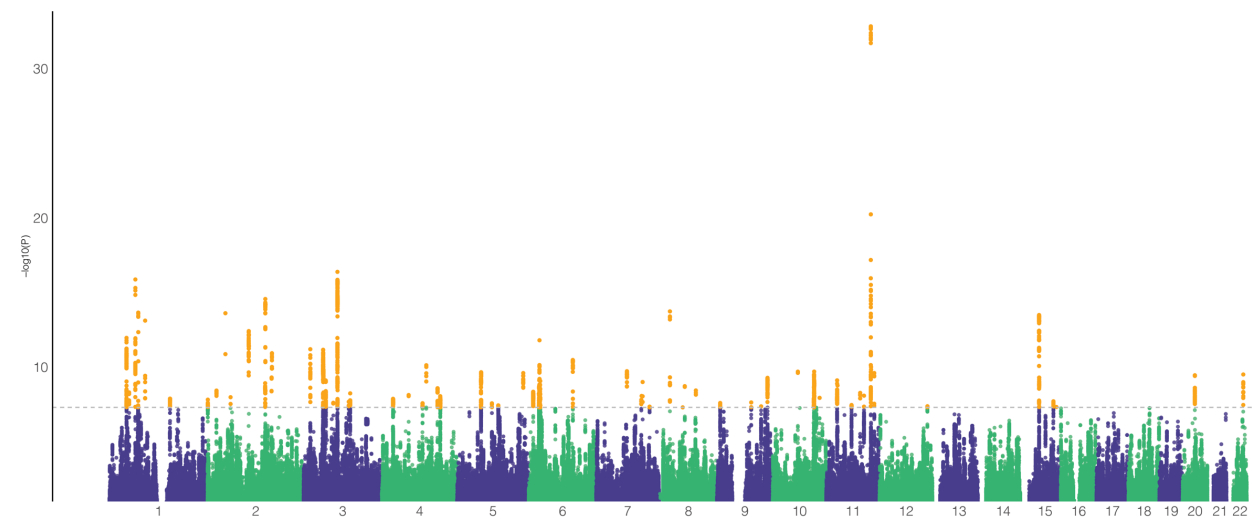
44. Ikeda, M. *et al.* Genome-Wide Association Study Detected Novel Susceptibility Genes for Schizophrenia and Shared Trans-Populations/Diseases Genetic Effect. *Schizophr. Bull.* **45**, 824–834 (2019).
45. Yao, X. *et al.* Integrative analysis of genome-wide association studies identifies novel loci associated with neuropsychiatric disorders. *Transl. Psychiatry* **11**, 69 (2021).
46. Fullerton, J. M. & Nurnberger, J. I. Polygenic risk scores in psychiatry: Will they be useful for clinicians? *F1000Research* **8**, F1000 Faculty Rev-1293 (2019).



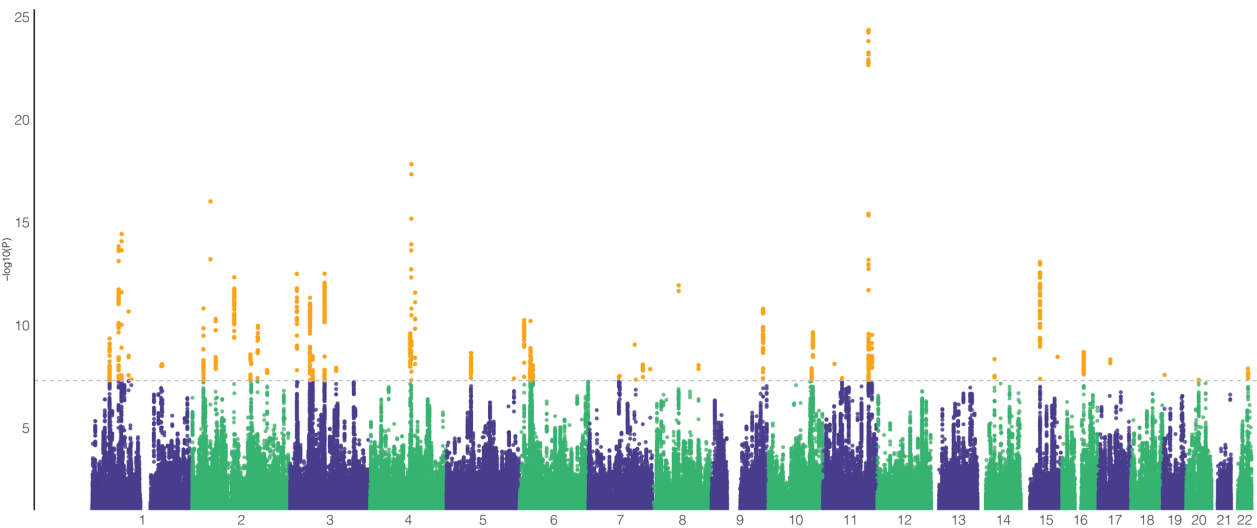
MTAG-ODU



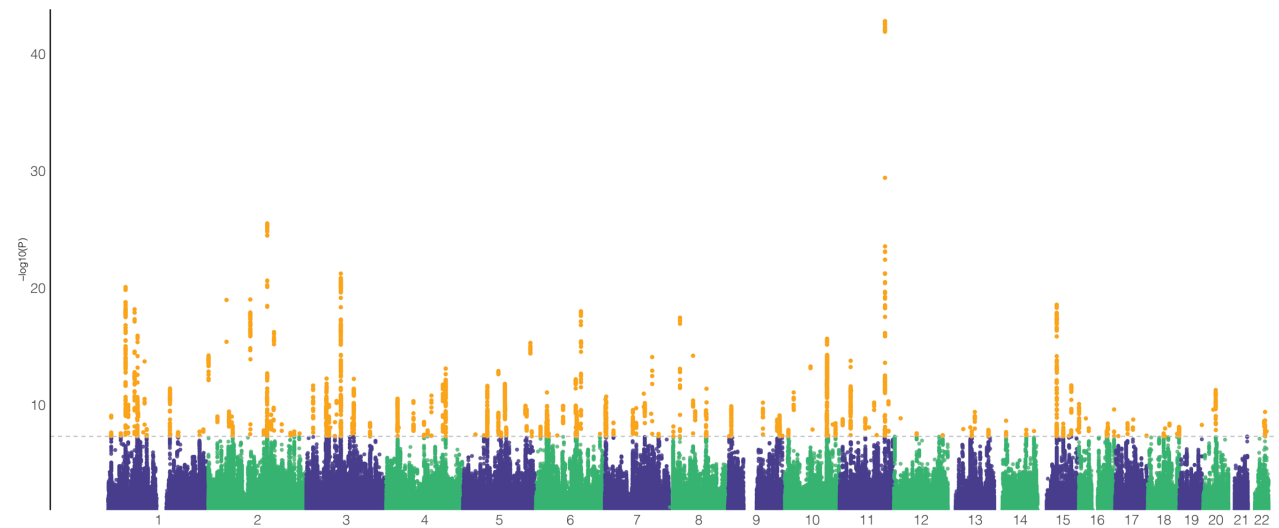
MTAG-CUD



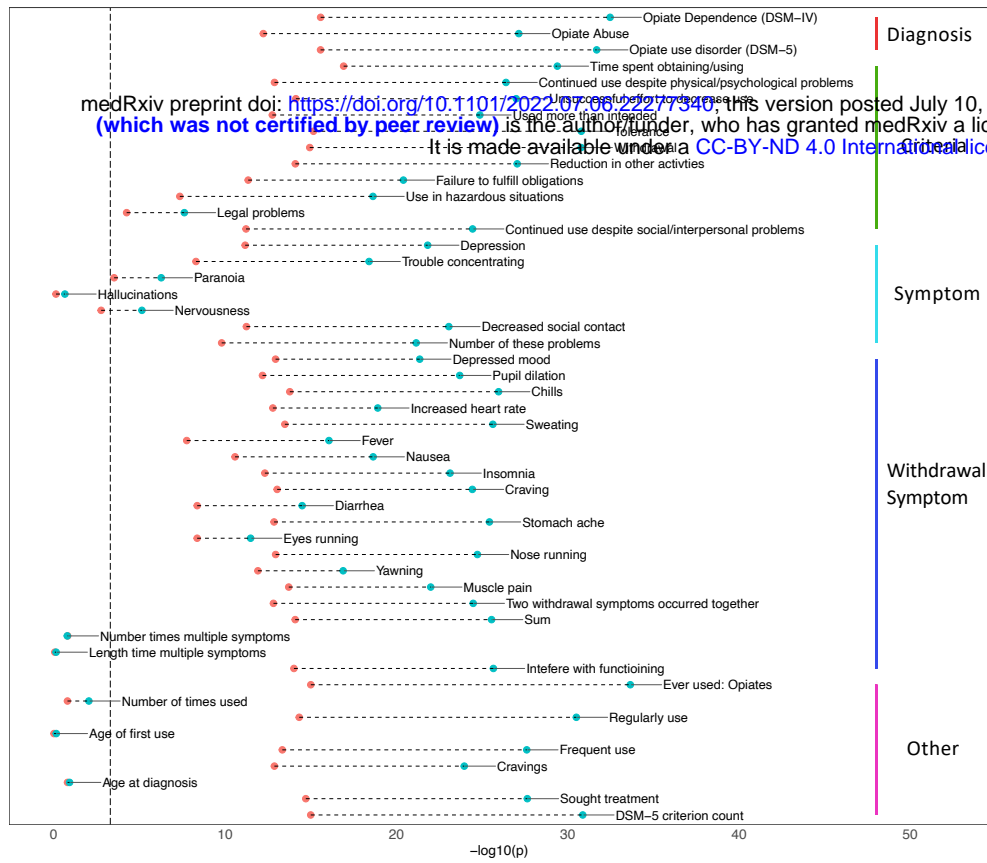
MTAG-AUD



MTAG-SMK



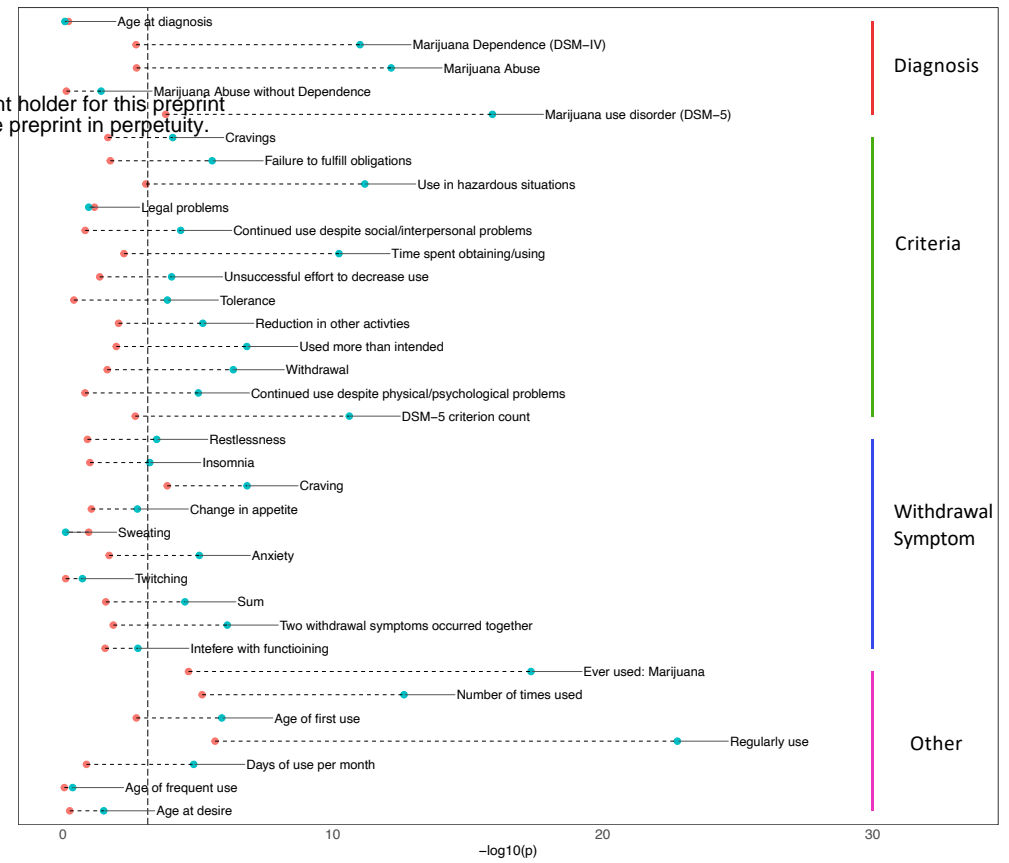
ODU PRS



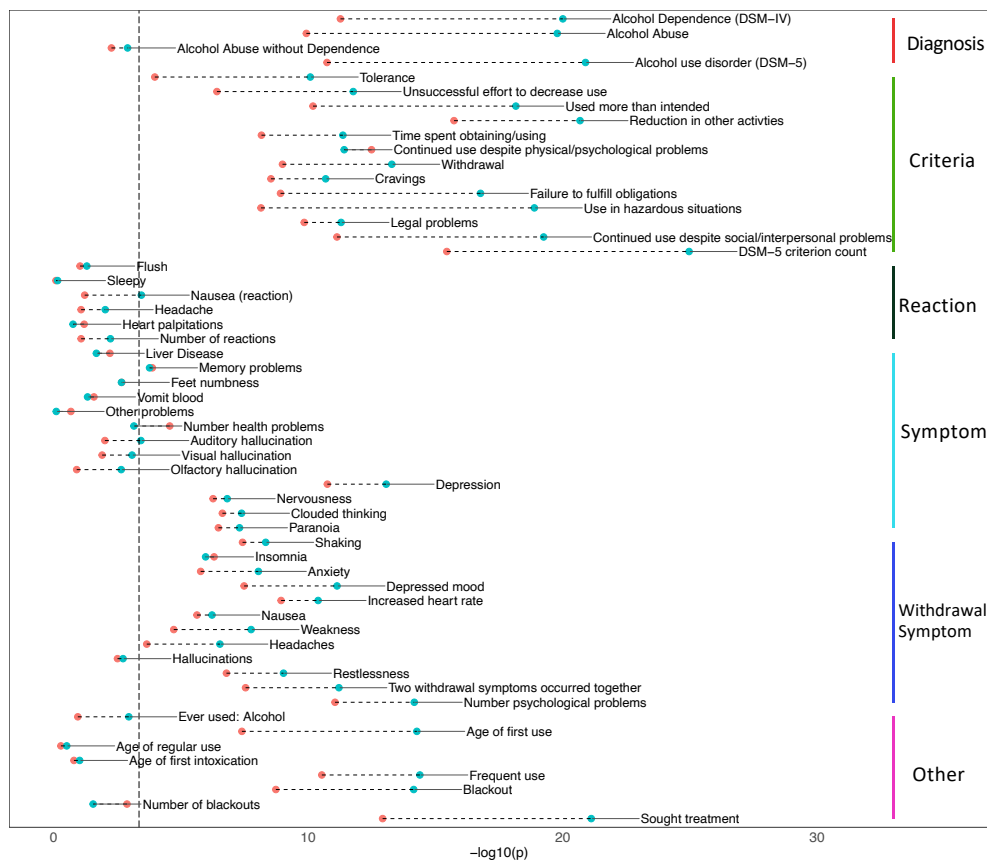
● GWAS-based PRS

● MTAG-based PRS

CUD PRS



AUD PRS



SMK PRS

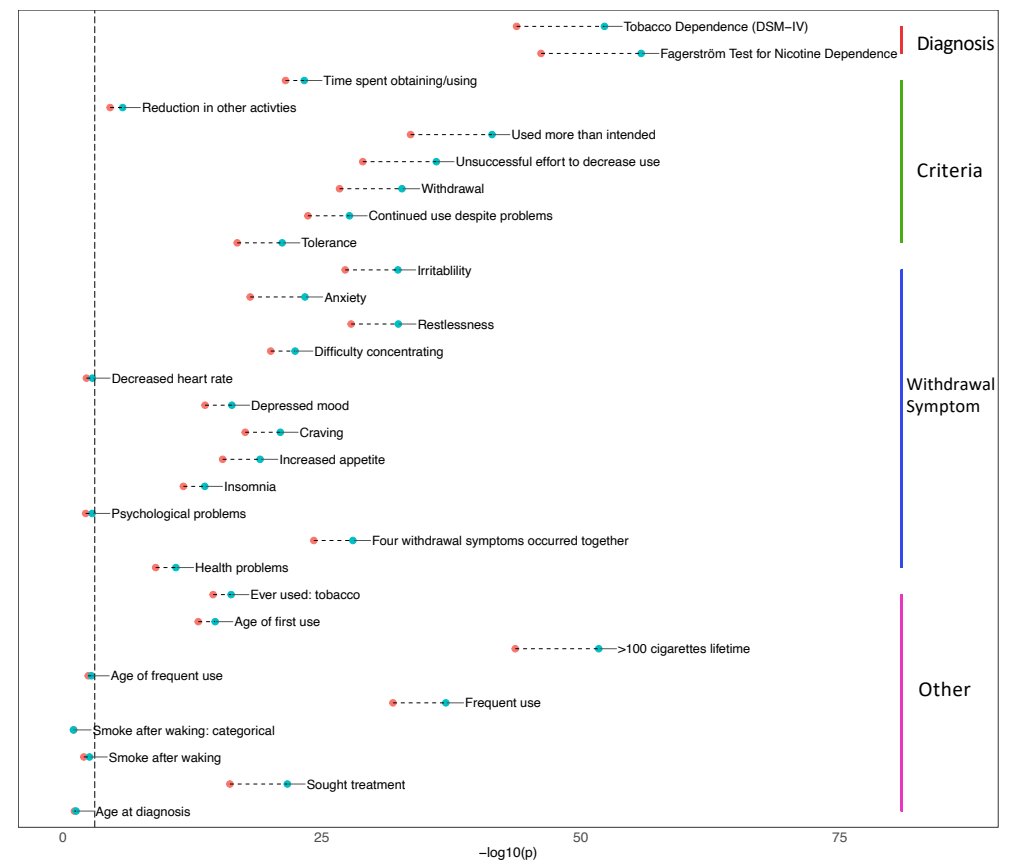


Table 1: Summary of MTAG Results

	OUD		CUD		AUD		SMK	
	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG	GWAS	MTAG
<u>SigSNPs</u>	18	1,465	7	2,547	363	1,854	3,049	5,380
<u>SigSNPs (After Clumping)</u>	4	41	2	74	20	63	105	183
Independent Locus	4	36	2	60	12	52	74	144
χ^2	1.159	1.376	1.099	1.455	1.281	1.463	1.771	1.865
Effective sample size	74,635	176,876	48,900	223,956	171,601	282,208	632,802	709,603

Table 2: PRS Increment R²

		GWAS-PRS	MTAG-PRS
Opioid	Ever Used Opioid	0.97%	2.31%
	DSM-IV Diagnosis	1.03%	2.26%
	DSM-5 Diagnosis	1.01%	2.16%
Cannabis	Ever Used Cannabis	0.38%	1.62%
	DSM-IV Diagnosis	0.19%	0.91%
	DSM-5 Diagnosis	0.23%	1.13%
Alcohol	Ever Used Alcohol	0.26%	1.11%
	DSM-IV Diagnosis	0.91%	1.68%
	DSM-5 Diagnosis	0.84%	1.72%
Smoking	Ever Used Smoking	1.81%	2.04%
	DSM-IV Diagnosis	3.43%	4.16%