Submission Type: Original Article
Abstract: 248/250
Text Word Count: 3950/4000
Display Items: 4/4

The Addiction Genetic Factor $a_{(g)}$: A Unitary Genetic Vulnerability Characterizes Substance Use Disorders and Their Associations with Common Correlates

Alexander S. Hatoum[1#], Emma C. Johnson[1], Renato Polimanti[2,3], Hang Zhou[2,3], Raymond Walters[4,5], Substance Use Disorders Working Group of the Psychiatric Genomics Consortium, Joel Gelernter[2,3,6,7], Howard J. Edenberg[8,9], Ryan Bogdan[10]* & Arpana Agrawal[1]*

#Please send all correspondence to: Washington University School of Medicine, Department of Psychiatry, 660 S. Euclid, CB 8134, Saint Louis, MO 63110, USA; email: ashatoum@wustl.edu;

[1]Washington University School of Medicine, Department of Psychiatry, Saint Louis, USA
[2]Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA
[3]Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA
[4]Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA
[5]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA
[6]Department of Genetics, Yale School of Medicine, New Haven, CT, USA
[7]Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA
[8]Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
[9]Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA
[10]Department of Psychological & Brain Sciences, Washington University in St. Louis
*These authors contributed equally to the work

Abstract

Background. Substance use disorders (SUDs) commonly co-occur with one another and with other psychiatric disorders. They share common features including high impulsivity, negative affect, and lower executive function. We estimate the shared genetic architecture across distinct SUDs, its independence from genetic liability to substance use, and its relation to genetic liability to impulsivity, negative affect, and executive function as well as non-substance psychopathology.

Methods. We tested whether a common genetic factor undergirds liability to problematic alcohol use (PAU), problematic tobacco use (PTU), cannabis use disorder (CUD), and opioid use disorder (OUD) by applying Genomic structural equation modelling to genome-wide association study statistics (Total N = **1,019,521**; substance specific Ns range: **82,707-435,563** of European ancestry), while adjusting for the genetics of substance use (Ns = **184,765-632,802**). We tested whether shared liability across SUDs is associated with behavioral constructs (risk taking, executive function, neuroticism; Ns = **328,339-427,037**) and non-substance use psychopathology (mood/psychotic, compulsive, and early neurodevelopmental disorders).

Results. Shared genetic liability to PAU, PTU, CUD, and OUD was characterized by a unidimensional addiction factor (termed $a_{(g)}$), independent of substance use. OUD and CUD demonstrated the largest loadings. $a_{(g)}$ was associated with risk taking, neuroticism, executive function, and non-substance psychopathology, but retained specific variance (standardized residual= .579).

Interpretation. A common genetic addictions factor partly explains susceptibility for alcohol, tobacco, cannabis and opioid use disorder. $a_{(g)}$ has unique pathways that are not shared with substance use or non-substance psychopathology, suggesting that addiction is not the linear combination of substance use and psychopathology.

Disclosures/COI: JG is named as an inventor on PCT patent application #15/878,640 entitled: "Genotype-guided dosing of opioid agonists," filed January 24, 2018.

Funding. This research was supported by MH109532 (AA, JG, HJE, ECJ) and T32DA007261 (ASH). AA acknowledges K02DA32573. ECJ was supported by F32AA027435. RP acknowledges R21DA047527. RB acknowledges R21-AA027827.
The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium (PGC-SUD) is supported by funds from NIDA and NIMH to MH109532.

Acknowledgments: We gratefully acknowledge our contributing studies and the participants in those studies without whom this effort would not be possible.

The MVP summary statistics were obtained via an approved dbGaP application (phs001672.v4.p1). The authors thank Million Veteran Program (MVP) staff, researchers, and volunteers, who have contributed to MVP, and especially participants who previously served their country in the military and now generously agreed to enroll in the study. (For details, see https://www.research.va.gov/mvp/ and Gaziano, J.M. et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. J Clin Epidemiol 70, 214-23 (2016)). This research is based on data from the Million Veteran Program, Office of Research and Development, Veterans Health Administration, and was supported by the Veterans Administration (VA) Cooperative Studies Program (CSP) award #G002.

This study included summary statistics of a genetic study on cannabis use (Pasman et al, 2018 Nature Neuroscience). We would like to acknowledge all participating groups of the International Cannabis Consortium, and in particular the members of the working group including Joelle Pasman, Karin Verweij, Nathan Gillespie, Eske Derks, and Jacqueline Vink. Pasman et al, (2018) included data from the UK Biobank resource under application numbers 9905, 16406 and 25331.

INTRODUCTION

Substance use and use disorders (SUDs) represent large and growing international public health problems that account for nearly 6% of global disease burden[1]. SUDs, both licit and illicit, commonly co-occur and also co-occur with non-substance psychopathology; comorbidity is associated with increased symptom severity[2] and worse outcomes (e.g., less responsive to treatment, greater socioeconomic costs[3]). However, the etiology underlying shared risk across these disorders is poorly understood.

Shared Genetic Liability

According to twin studies, the moderate-large heritability (50-60%) of distinct SUDs (i.e., alcohol, nicotine, cannabis, and other illicit drugs) is partly attributable to a shared latent genetic vulnerability.[4] Similarly, genetic correlations estimated from genome-wide association study (GWAS) summary statistics support a shared genetic vulnerability across distinct SUDs (e.g., SNP-rG = .73 between problematic alcohol use and opioid use disorder),[7] between SUDs and substance use (e.g., SNP-rG = .78 between problematic alcohol use and drinks per week),[8] and between SUDs and other psychopathology (e.g., SNP-rG = .33 between cannabis use disorder and major depressive disorder[9]). What remains unclear is the extent to which genetic liability across substance use *disorders* is shared with *and distinct* from substance use and non-substance psychopathology, and what putative intermediate phenotypes may link shared genetic liability between SUDs and non-substance psychopathology.

Substance Use and Use Disorder. Substance use and SUDs have substantial genetic overlap; however, there remain genetic mechanisms that relate to SUD liability beyond normative or frequently occurring substance use behaviors. For opioids[14], alcohol[15,16,17], and cannabis[9], the

use and use disorder dimensions show differing associations with psychopathology (e.g. schizophrenia) and life outcomes (e.g. educational attainment). The genetic contributions to SUDs can be confounded by genetic influences on earlier stages of substance use[18], since the definition of an SUD requires the individual to have initiated and escalated use. Explorations of shared genetic variation across substance use disorders can potentially also address vulnerability to initiation and use of substances.

Substance use and Psychopathology. Recently, Lee and colleagues[10] identified 3 broad clusters (mood/psychotic, compulsive, and early neurodevelopmental) representing shared and distinct genetic liability to 8 non-substance psychiatric disorders. While there has been limited integration of substance use phenotypes into psychopathology models, polygenic liability to cross-psychiatric disorders is associated with general substance use and SUDs[6]. Further, emerging evidence suggests partial overlap: tobacco *use* shares variance with ADHD[11], alcohol and cannabis dependence load with antisocial behavior[12], and alcohol use and use disorder load together onto an Externalizing factor[13]. Collectively, these data suggest that substance use and use disorders share a common genetic liability with psychopathology.

***Stage-Based Addiction Constructs and Substance Use Disorder Genetics*.** SUD vulnerability has been conceptualized within a 3-stage neurobiological model consisting of binge/intoxication, preoccupation/anticipation, and withdrawal/negative affect[19]. According to this model, initial positive reinforcement associated with substance use is derived from stimulation of neural reward circuitry that drives impulsive and risk-taking behaviors in the context of under-developed physiological tolerance. With continued use and progression towards SUD, the reinforcing properties of substances shift from positive to negative reinforcement; as use becomes compulsive, it functions to return the body to drug-present homeostasis and alleviate low mood, a

predisposition to which is broadly indexed by neuroticism[20]. Following repeated drug-reward and drug-homeostasis pairings, cognitive preoccupation with the drug in expectation of reward/relief emerges in the context of impaired executive functioning[21]. While GWAS support genetic correlations between SUDs and risk-taking[7,22], cognition[23], and negative affect[7,22], the extent to which common genetic liability across SUDs relates to these stage-based constructs has yet to be examined. Given that risk-taking, neuroticism and executive functioning also share genetic liability with non-SUD psychopathology, it is possible that these constructs may account for shared genetic liability between SUDs and non-substance psychiatric disorders.

The Current Study

Given accumulating evidence of shared liability to SUDs and other forms of psychopathology, understanding the shared and unique genetic contributions to SUDs and how these relate to stage-based addiction constructs, non-substance psychopathology, and substance use may generate etiologic insights that improve psychiatric nosology, prevention, and treatment. To this end, we *first* estimate the shared and genetic structure across SUDs by applying genomic structural equation modelling (gSEM)[24] to summary statistics generated by the largest GWAS of problematic alcohol use (PAU)[22], problematic tobacco use (PTU)[25,26], cannabis use disorder (CUD)[9], and opioid use disorder (OUD)[7]. We name the shared variance across SUDs the Addiction genetic factor, $a_{(g)}$. *Second*, we relate $a_{(g)}$ to genetics of behavioral constructs underlying the stage-based model of SUDs. We estimate the extent to which genetic liability to risk-taking (N = 328,339), executive function (N = 427,037), and neuroticism (N = 390,278) are related to a common genetic liability to SUDs. *Third*, we examine whether $a_{(g)}$ is associated with the 3 factors representing genetic liability to non-substance psychopathology[10] (i.e., mood/psychotic, compulsive, and early

neurodevelopment) and whether stage-based addiction constructs (i.e., risk-taking, executive function, neuroticism) indirectly link $a_{(g)}$ to psychopathology. *Finally*, given that genetic liability to substance use (e.g., ever using, quantity-frequency) and later stages of SUDs may be partially distinct (e.g., risk-taking more closely related to use)[8], we repeat all analyses while incorporating genetic liability to substance use (i.e., alcohol drinks/week: N = 537,349[26]; tobacco ever regularly use: N = 632,802[26], cannabis ever use N = 162,082[27]) as covariates at the indicator level.

We hypothesized that, similar to general intelligence (*g-factor*) and psychopathology (*p-factor*), SUDs and problem substance use would be largely characterized by a common genetic vulnerability (i.e., an addiction genetic factor [$a_{(g)}$]) with evidence of small but potentially important substance-specific liability (e.g., metabolic and signaling pathways for a specific drug such as *ADH1B* variants with alcohol[28]). We hypothesized that (i) $a_{(g)}$ would be associated with all 3 non-substance psychiatric disorder clusters while retaining variance unique to itself, (ii) genetic liability to behavioral phenotypes representing vulnerability to stage-based addiction constructs (i.e., risk-taking, executive function, and neuroticism) would be associated with $a_{(g)}$ and account for a proportion of the association between $a_{(g)}$, but also show specific associations with $a_{(g)}$ beyond psychopathology, and (iii) after accounting for genetics of substance use, $a_{(g)}$ would retain unique variance (i.e., we expect significant genetic correlations among SUDs) maintain similar patterns with non-substance psychopathology and stage-based constructs.

METHODS

Samples

Summary statistics from the largest available discovery GWASs were used to represent genetic risk for each construct (details and Ns are in **Supplemental Table 1**). These include: **i) 4 *SUDs*** (problematic alcohol use[22], problematic tobacco use[25], cannabis use disorder[9], opioid use disorder[7]); **ii)** 3 ***substance use phenotypes*** (alcohol drinks/week[26], lifetime ever smoking[26], lifetime cannabis use[27]); **iii)** 3 ***traits representing the stage-based model of SUDs*** (risk-taking, executive function, neuroticism); and **iv) 9 *non-substance psychiatric disorders***. Unfortunately, analyses were restricted to data from individuals of European ancestry because GWAS on these constructs in other ancestral origins are not available or are underpowered, and cross-ancestry analysis can confound genetic correlation estimates[29]. All GWAS summary statistics were filtered to retain variants with minor allele frequencies > 0.01 and INFO score > 0.90 for PGC and INFO score > 0.70 for the MVP.

Problematic Substance Use/Substance Use Disorder Summary Statistics

Problematic Alcohol Use: Summary statistics for problematic alcohol use (PAU) were derived from a meta-analysis of GWASs of DSM-IV alcohol dependence (AD) from the Psychiatric Genomics Consortium (PGC; n = 11,569 case, 34,999 controls), ICD-9/10 based diagnoses of alcohol use disorders from the Million Veteran Program phase 1 and 2 data (MVP; n = 45,995 cases; 221,396 controls) and the Problem subscale score from the Alcohol Use Disorders Identification Test (AUDIT-P) that was administered via an online health questionnaire to a subset of participants in the UK Biobank (n = 121,604)[22]. The final GWAS summary statistics included data on 435,563 participants.

Problematic Tobacco Use: We used summary statistics from the GWAS of the Fagerström Test for Nicotine Dependence[25] (FTND). Because the genetic correlation between FTND and cigarettes per day is high (calculated rG = 0.95[30]), we applied Multi-Trait Analysis of Genome-wide association study summary statistics (MTAG[31]) to summary statistics generated from the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN) GWAS of cigarettes per day to create a combined phenotype[26]. As this is the only observed genetic correlation between a use and use disorder category that is sufficiently high to allow for MTAG analysis, we used this procedure only for tobacco. The final GWAS summary statistics had an effective sample size of n = 270,120 individuals.

Cannabis Use Disorder: Summary statistics were derived from a GWAS meta-analysis of DSM-IV and DSM-III-R cannabis abuse and dependence from the Psychiatric Genomics Consortium (n = 5,289 cases; n = 10,004 controls), ICD 10 cannabis use disorder from the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) (n = 2,758 cases; n = 53,326 controls), and hospital-based diagnoses from deCODE (n = 6,033 cases; n = 280,396 controls)[9]. The final sample included 14,080 cases with CUD and 343,726 controls.

Opioid Use Disorder: Opioid use disorder (OUD) summary statistics were derived from an meta-analytic GWAS of ICD-9/10-based opioid abuse or dependence from Yale-Penn, Study of Addiction: Genetics and Environment, and the Million Veteran Program (n = 10,544 cases; n = 72,163 opioid-exposed controls)[7].

Substance Use Summary Statistics

*Alcohol Drinks/Week***.** Alcohol use summary statistics were derived from the GSCAN GWAS for current reported drinks/week (n = 537,349). There was a moderate correlation with PAU (SNP-rG between drinks/week and PAU = 0.77±0.02)[22].

*Lifetime Tobacco Use***.** Summary statistics came from the GSCAN GWAS of reported ever/never regular cigarette smoking (ever n = 301,524, never n = 331,278). There was a moderate correlation with PTU (SNP-rG = 0.28±0.03).

*Lifetime Cannabis Use***.** We used summary statistics from a meta-analysis of lifetime cannabis ever-use from the International Cannabis Consortium and UK Biobank (ever n = 43,380; never n = 118,702). There was a moderate correlation with CUD (SNP-rG = 0.47±0.05)[9].

Stage-based Behavioral Constructs

The behavioral three-stage model of addiction focuses on "state" changes in substance use behaviors. As GWAS measure individual differences in traits, we selected behaviors that (1) are known to convey vulnerability to each stage, and (2) are heritable.

*Risk-taking and sensitivity to reward***.** A GWAS of risk-taking derived from a single item in the UK Biobank ("Would you describe yourself as someone who takes risks?"; data field #2040; risk taker n = 83,677; non-risk taker n = 244,662)[32] was used as a trait representing this stage.

*Executive Function***.** The **"**preoccupation/anticipation" stage is characterized by maladaptive reward valuation and future planning. Recent work argues that this vulnerability is captured by executive functioning[33]. Summary statistics from a GWAS of a latent factor representing general executive functioning were used (N = 427,037)[23].

*Negative Emotionality and Sensitivity to Stress***.** The stage of withdrawal/negative affect represents substance use functioning to mitigate aversive withdrawal symptoms, such as negative affect. Neuroticism has been found to modify stress sensitivity and neural reward processing[34]. Neuroticism was chosen as a trait-based measure representing liability to negative affect as opposed to depression because depression was included in the non-substance psychiatric disorder factor generation and because neuroticism includes trans-diagnostic constructs such as negative urgency (i.e., impulsive attempts to cope with negative affect) that may place individuals at risk for the negative reinforcing aspects of SUDs. Thus, we selected the largest GWAS of neuroticism to represent this stage (n = 390,278)[20].

Non-Substance Summary Statistics

Summary statistics from the PGC Cross-disorder GWAS on the 8 disorders that were previously shown to fit a 3 factor confirmatory model were used[10]. These disorders included Schizophrenia[35], Bipolar disorder[36], Major depressive disorder[37], Attention Deficit Hyperactivity disorder[38], Obsessive Compulsive Disorder[39], Anorexia Nervosa[40], Tourette Syndrome[41], and Autism Spectrum Disorder[42] (See **Supplemental Table 1 for details**).

Statistical analysis

First, we estimated the pairwise genetic correlation between PAU, PTU, CUD and OUD using Linkage Disequilibrium Score Regression (LDSR)[29]. After confirming that the four SUDs were significantly genetically correlated (see **Results**), we applied confirmatory factor analysis to the covariance matrix generated by LDSR using gSEM[43]; a single latent factor (i.e., $a_{(g)}$-factor) was allowed to load freely on PAU, PTU, OUD, and CUD. Variance of this common latent factor was scaled to 1.0. A residual correlation between PAU and OUD was also estimated

to account for measurement overlap, because the Million Veterans Project sample was contained in both PAU and OUD GWAS.

Second, we used a series of structural regression models to estimate the extent to which genetic liability to stage-based constructs of addiction (i.e., risk-taking, executive function and neuroticism) are related to $a_{(g)}$. Here, the $a_{(g)}$ variance was freed, and the OUD loading was set to 1.0 to scale the model. Intercorrelations were estimated between risk-taking, executive function and neuroticism.

Third, we recapitulated the three factors from Lee et al[10]. (i.e., psychotic/mood disorders, compulsive disorders, and early neuro-developmental disorders) and estimated their relationship with $a_{(g)}$ while allowing for inter-factor correlations. This allowed us to estimate the unique association between each of the 3 psychopathology factors and $a_{(g)}$ and to estimate variance that was residual to $a_{(g)}$. We then examined whether stage-based addiction constructs (i.e., risk-taking, executive function and neuroticism) indirectly linked $a_{(g)}$ to the 3 non-substance psychopathology factors using a multiple mediator model. To estimate residual associations (i.e., direct paths) between the stage-based constructs and $a_{(g)}$, we re-structured the mediation model to one in which the 3 non-SUD psychopathology factors served as "mediators" of the relationship between risk-taking, executive functioning, neuroticism, and $a_{(g)}$.

To separate the genetics of SUD from the genetics of substance use, we ran a second gSEM model where substance *use* GWAS summary statistics were endogenous predictors of all measured variables in the model. For example, in the model estimating the association between $a_{(g)}$ and psychiatric factors, the 8 psychiatric disorders and the 4 SUD disorder variables were regressed on the 3 substance use variables. In this way, covariate effects were estimated robustly and simultaneously to our associations of interest.

RESULTS

The $a_{(g)}$-Factor

Genetic correlations between problematic alcohol use (PAU[22]), problematic tobacco use (PTU[25,26]), cannabis use disorder (CUD[9]), and opioid use disorder (OUD[7]) ranged from 0.19(S.E. = .04) to 0.78(.09) (**Supplemental Figure 1**). PTU showed the lowest SNP-rG with other SUD phenotypes [i.e., PAU = 0.19(.04), CUD = 0.31(.05), OUD = 0.26(.08)] while OUD showed the highest [PAU = 0.69(.07), CUD = 0.78(.09)]. A confirmatory factor model specifying a unidimensional $a_{(g)}$-factor underlying the genetic covariance between PAU, PTU, CUD and OUD fit the data well [$X^2(1) = .017$, p = .895, CFI = 1, SRMR = .002; residual r = .51, p = 0.016; **Figure 1A**]. Loadings were uniformly high except for PTU, where there was stronger evidence for tobacco-specific genetic influences.

The inclusion of genetic liability to typical substance use did not modify the single factor structure of $a_{(g)}$ (**Figure 1B**); all SUDs continued to load significantly on the factor. However, factor loadings were lower for all substances, especially for PAU, which may be attributable to the high genetic correlation between Drinks/Week and PAU.

Shared Liability to Stage-based Behavioral Phenotypes

Genetic liability to stage-based addiction constructs was shared with $a_{(g)}$ (**Figures 2**). As expected, $a_{(g)}$ was positively associated with genetic liability to risk-taking (β = 0.45) and neuroticism (β = 0.25), and negatively associated with executive function (β = -0.165; **Figure 2A**). Despite significant genetic overlap between $a_{(g)}$ and stage-based behavioral phenotypes, $a_{(g)}$ retained unique variance ($a_{(g)}$ residual = 0.68). When conditioning for genetic liability for substance use, $a_{(g)}$ remained significantly associated with increased genetic liability to risk-taking

(β = 0.22) and neuroticism (β = 0.18) and decreased genetic liability to executive function (β = -0.28; **Figure 2B**). Accounting for genetic liability for substance use substantially reduced the association between $a_{(g)}$ and risk-taking from 0.45 to 0.22 ($p_{df = 1}$ = 4e-09) and accentuated the negative association with executive function from β = -0.17 to -0.28 ($p_{(df = 1)}$ = 0.013); there was a smaller effect on the association with neuroticism (from β = 0.25 to 0.18, $p_{(df = 1)}$ = 0.012).

Shared Liability to Non-substance Psychopathology

Genetic liability to non-substance psychopathology (i.e., compulsive disorders, psychotic and mood disorders, and neurodevelopmental disorders) was shared with $a_{(g)}$ (**Figure 3**). Psychotic/mood disorders (β = 0.45) and neurodevelopmental disorders (β = 0.74) were positively associated with $a_{(g)}$ while compulsive disorders showed a negative association (β = -0.32; **Figure 3A**). Due to the strong correlation between $a_{(g)}$ and early-onset neurodevelopmental disorders (which includes ADHD) we allowed ADHD to load on $a_{(g)}$; here, a residual association between $a_{(g)}$ and early-onset neurodevelopmental disorders remained, but was significantly attenuated (from β = 0.74 to 0.43, $p_{(df = 1)}$ = 5e-5). When conditioning $a_{(g)}$ for substance *use*, only the psychotic/mood and early neurodevelopmental disorder factors remained significantly associated with $a_{(g)}$ (**Figure 3B**). Despite the significant genetic overlap with other psychiatric disorder domains, $a_{(g)}$ retained unique variance representing genetic liability specific to SUDs ($a_{(g)}$ residual = 0.30, p = 4.54e-3). This unique variance was potentiated by modeling genetic liability to substance use ($a_{(g)}$ residual = 0.58, p = 0.015).

The specifications for the mediation models is shown **in Supplemental Figure 2**. Genetic liability to risk taking accounted for a proportion of the associations between all non-substance psychopathology domains and $a_{(g)}$ (**Table 1**). Executive function uniquely indexed an

indirect effect between mood/psychotic disorders and $a_{(g)}$ (**Table 1**). When conditioning $a_{(g)}$ for genetic liability to substance use, risk-taking no longer accounted for a portion of the association between any non-substance psychopathology domain and $a_{(g)}$, but executive function continued to account for a proportion of the overlap (indirect effect of 0.048) between psychotic/mood disorders and $a_{(g)}$ (**Table 1**). *Post-hoc* analyses revealed that executive function retained a unique association with $a_{(g)}$ after accounting for genetic liability to both substance use and non-substance psychopathology (**Supplementary Table 2**).

DISCUSSION

We applied genomic structural equation modeling (gSEM)[24] to GWAS summary statistics to characterize the genetic influences shared across SUDs and estimated how common genetic liability is related to trait conceptualizations of a theoretical stage-based SUD model as well as to non-substance psychopathology. Three primary findings emerged. *First*, genetic risk for specific SUD phenotypes (i.e., PAU[22], PTU[25,26], CUD[9], and OUD[7]) was largely attributable to a single addiction genetics factor, $a_{(g)}$ (**Figure 1**). *Second*, $a_{(g)}$ was associated with genetic liability to trait representations of stage-based facets of addiction (risk taking [binge/intoxication], executive function [preoccupation/anticipation], neuroticism [negative affect][19]; **Figure 2**). It was also associated with non-substance psychopathology factors (compulsive disorders, psychotic and mood disorders, neurodevelopmental disorders; **Figure 3**). Trait representations of stage-based facets of addiction partially accounted for the shared genetic liability between non-substance psychopathology and $a_{(g)}$. *Third*, associations between $a_{(g)}$ and stage-based constructs and non-substance psychopathology were largely independent of genetic liability to substance *use* phenotypes (i.e., tobacco use, cannabis use, alcoholic drinks/week). However, consistent with the stage-based model of addiction, accounting for substance use attenuated associations between risk taking and $a_{(g)}$ while potentiating associations with executive functioning. Collectively, our findings suggest that SUDs are characterized by a common genetic factor, $a_{(g)}$. This shared genetic etiology may be leveraged to refine nosology and increase statistical power for research in treatment and prevention.

$a_{(g)}$ retains variance that is not shared with other psychopathology

Even after accounting for genetic liability to substance use as well as the commonality between $a_{(g)}$ and non-substance psychopathology, $a_{(g)}$ retained a substantial proportion of specific genetic variance. These data suggest that $a_{(g)}$ may be characterized by unique pathways that are not shared with substance use or non-substance psychopathology, i.e., addiction is not the linear combination of substance use and psychopathology.

A single latent factor, $a_{(g)}$, fit these data well, but specific SUDs showed varying degrees of association. The illicit SUDs (CUD and OUD; **Figure 1**) were almost entirely captured by the common latent factor, unlike PAU and PTU, suggesting that heightened $a_{(g)}$ may confer liability for potentially illicit substance use disorder. Notably, the loading for PTU on $a_{(g)}$ was lower than other substances. One potential contributor to the residual genetic variance of PTU may be the use of FTND and cigarettes/day as indices of PTU. Unlike the Diagnostic and Statistical Manual (DSM) criteria which index psychological and physiological aspects of tobacco use disorder, the FTND, which includes amount smoked, is an index of biochemical dependence and shows only moderate agreement with DSM-defined nicotine dependence (r = 0.50; kappa = 0.3)[44].

Stage-based Behavioral Constructs and $a_{(g)}$

Behavioral stage-based models of SUD posit a cyclical relationship between positive reinforcement, negative reinforcement, and incentive salience[19] that we hypothesized can be (partially) captured by individual differences in genetic liability to risk-taking, executive functioning, and negative emotionality. The strongest association with $a_{(g)}$ was for risk-taking. Consistent with this, when substance use was included in the model the shared genetic loading between $a_{(g)}$ and both risk-taking and neuroticism was attenuated while the association with executive function increased. The reduction in the association with neuroticism is counter to

expectations from the stage-based model which posits a more prominent role of negative affect for SUD relative to substance use. We speculate that neuroticism, which represents an amalgam of negative affect traits, may be too broad a construct when considering SUD-specific negative affect; large-scale studies of domains of negative affectivity (e.g., negative urgency) are needed.

Non-substance Psychopathology and $a_{(g)}$

We found that the 3 non-substance psychopathology clusters, derived from 8 psychiatric disorders[10], were genetically associated with $a_{(g)}$. The association with early neurodevelopmental disorders, which include ADHD, was the strongest. Cross-loading ADHD on $a_{(g)}$ attenuated the loading but it remained high, possibly because Major Depressive Disorder also loaded on this factor. Associations between $a_{(g)}$ and the psychopathology clusters were greater in magnitude than associations with trait representations of behavioral stages of addiction (with the possible exception of risk-taking). For instance, the genetic association between $a_{(g)}$ and the two disorder clusters that included Major Depressive Disorder (i.e., mood/psychotic disorders and early neurodevelopmental disorders) was greater in magnitude than the $a_{(g)}$-neuroticism association. These findings reinforce the idea that from a genetic perspective, SUDs align closely with pathophysiological states.

Of the 3 behavioral correlates, risk-taking was the most prominent contributor to the association between $a_{(g)}$ and all non-substance psychopathology factors. Executive function also contributed to the association between $a_{(g)}$ and psychotic/mood disorders. After accounting for substance use, risk-taking and executive function only mediated $a_{(g)}$ associations with the psychotic/mood disorder factor. Executive function maintained the only direct association with $a_{(g)}$ after accounting for genetics of substance use and genetics of non-substance

psychopathology. Thus, we speculate that while risk-taking may characterize the genetic overlap between substance use and other psychopathology, executive function impairment is a risk factor that not only shapes the overlap between addiction and non-substance psychopathology but also explains variance in addiction above and beyond that overlap. Thus, our analyses highlight the prominence of frontal brain executive function impairments in SUDs.

Limitations

There are several limitations of note. First, we had to restrict our analyses to those of European descent due to the lack of well-powered discovery GWAS informative of other ancestry groups. Second, we focused on common genetic variants, so it is unclear if our identified factor structure will generalize to rare genetic variation; because our findings align with evidence from twin studies, which aggregate effects of common and rare variants, we suspect it may. Third, to maximize sample size of discovery GWASs, our alcohol and tobacco use GWAS incorporated measures of "problematic" use that, while genetically highly correlated with AUD and ND, may include behavioral patterns that are less severe than those represented by use disorder. Finally, the analyses contain an over-representation of men, in part because the MVP samples contributed most of OUD and half of PAU and this sample is ~90% male. Studies with larger numbers of women would allow stratified analyses to explore the differences between sexes observed in epidemiological studies.

Conclusions

Common genetic liability, the $a_{(g)}$-factor largely undergirds distinct SUDs and shares variance with putative behavioral intermediary phenotypes/SUD risk factors and non-substance

psychopathology. This genetic factor is more than a linear combination of substance use and psychopathology, but represents a unique addiction dimension that is partially captured by executive functions.

Reference Cited

1. Degenhardt, L. *et al.* The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Psychiatry* **5**, 987–1012 (2018).
2. Bhalla, I. P., Stefanovics, E. A. & Rosenheck, R. A. Clinical Epidemiology of Single Versus Multiple Substance Use Disorders. *Med. Care* **55**, S24–S32 (2017).
3. Merikangas, K. R. & Kalaydjian, A. Magnitude and impact of comorbidity of mental disorders from epidemiologic surveys. *Current Opinion in Psychiatry* vol. 20 353–358 (2007).
4. Palmer, R. H. C. *et al.* Genetic etiology of the common liability to drug dependence: evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug Alcohol Depend.* **123 Suppl**, S24-32 (2012).
5. Fu, Q. *et al.* Shared Genetic Risk of Major Depression, Alcohol Dependence, and Marijuana Dependence. *Arch. Gen. Psychiatry* **59**, 1125 (2002).
6. Carey, C. E. *et al.* Associations between polygenic risk for psychiatric disorders and substance involvement. *Front. Genet.* **7**, (2016).
7. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant with Opioid Use Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* (2020) doi:10.1001/jamapsychiatry.2020.1206.
8. Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent Efforts to Dissect the Genetic Basis of Alcohol Use and Abuse. *Biological Psychiatry* vol. 87 609–618 (2020).
9. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of cannabis use disorder. *The Lancet Psychiatry* **0**, (2020).
10. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469-1482.e11 (2019).
11. Jang, S. K. *et al.* Genetic correlation, pleiotropy, and causal associations between substance use and psychiatric disorder. *Psychol. Med.* 1–11 (2020) doi:10.1017/S003329172000272X.
12. Waldman, I. D., Poore, H. E., Luningham, J. M. & Yang, J. Testing structural models of psychopathology at the genomic level. *World Psychiatry* **19**, 350–359 (2020).
13. Karlsson Linnér, R. *et al.* Multivariate genomic analysis of 1.5 million people identifies genes related to addiction, antisocial behavior, and health. *Henry R. Kranzler* **17**, 20 (2020).
14. Polimanti, R. *et al.* Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium. *Mol. Psychiatry* **25**, 1673–1687 (2020).
15. 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**, (2019).
16. Sanchez-Roige, S. *et al.* Genome-wide association study meta-analysis of the alcohol use disorders identification test (AUDIT) in two population-based cohorts. *Am. J. Psychiatry* **176**, 107–118 (2019).
17. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
18. Neale, M. C. & Kendler, K. S. Models of comorbidity for multifactorial disorders. *Am. J. Hum. Genet.* (1995).
19. Goldstein, R. Z. & Volkow, N. D. Drug addiction and its underlying neurobiological

basis: Neuroimaging evidence for the involvement of the frontal cortex. *American Journal of Psychiatry* vol. 159 1642–1652 (2002).

20. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* **9**, 905 (2018).
21. Koob, G. F. & Volkow, N. D. Neurobiology of addiction: a neurocircuitry analysis. *The Lancet Psychiatry* vol. 3 760–773 (2016).
22. 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.* 1–10 (2020) doi:10.1038/s41593-020-0643-5.
23. Hatoum, A. *et al.* GWAS of Over 427,000 Individuals Establishes GABAergic and Synaptic Molecular Pathways as Key for Cognitive Executive Functions. *GWAS Over 427,000 Individ. Establ. GABAergic Synaptic Mol. Pathways as Key Cogn. Exec. Funct.* 674515 (2019) doi:10.1101/674515.
24. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
25. Hancock, D. B. *et al.* Genome-wide association study across European and African American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence. *Mol. Psychiatry* **23**, 1–9 (2018).
26. 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).
27. Pasman, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–1170 (2018).
28. Edenberg, H. J. & McClintick, J. N. Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental Research* vol. 42 2281–2297 (2018).
29. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
30. Quach, B. C. *et al.* Expanding the genetic architecture of nicotine dependence and its shared genetics with multiple traits. *Nat. Commun.* **11**, 1–13 (2020).
31. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat. Genet.* **50**, 229–237 (2018).
32. Strawbridge, R. J. *et al.* Genetics of self-reported risk-taking behaviour, trans-ethnic consistency and relevance to brain gene expression. *Transl. Psychiatry* **8**, (2018).
33. Volkow, N. D., Koob, G. F. & McLellan, A. T. Neurobiologic Advances from the Brain Disease Model of Addiction. *N. Engl. J. Med.* **374**, 363–371 (2016).
34. Bondy, E. *et al.* Neuroticism and reward-related ventral striatum activity: Probing vulnerability to stress-related depression. (2020) doi:10.31234/osf.io/5wd3k.
35. Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the P. G. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–7 (2014).
36. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
37. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).

38. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
39. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
40. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214 (2019).
41. Yu, D. *et al.* Interrogating the genetic determinants of Tourette's syndrome and other tiC disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227 (2019).
42. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum disorder. *Nat. Genet.* **51**, 431–444 (2019).
43. Grotzinger, A. D. *et al.* Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *bioRxiv* 305029 (2018) doi:10.1101/305029.
44. Agrawal, A. *et al.* A latent class analysis of DSM-IV and fagerström (FTND) criteria for nicotine dependence. *Nicotine Tob. Res.* **13**, 972–981 (2011).

Table 1. Behavioral Liabilities Mediate the Association Between $a_{(g)}$ and Psychiatric Factors

	F1 Indirect %	**F1 Indirect P**	**F2 Indirect %**	**F2 Indirect P**	**F3 Indirect %**	**F3 Indirect P**
No covariates						
executive function	0.007	0.463	**0.051***	**0.001***	0.003	0.486
Neuroticism	0.005	0.842	0.003	0.836	0.013	0.777
Risk-taking	**-0.090***	**0.001***	**0.152***	**1.47E-08***	**0.043***	**0.013***
Controlling for Genetics of Substance Use						
executive function	0.013	0.233	**0.048***	**0.047***	-0.007	0.491
Neuroticism	0.011	0.687	-0.022	0.693	-0.040	0.689
Risk-taking	0.001	0.509	0.019	0.446	-0.002	0.741

Table 2. Behavioral Liabilities Mediate the Association Between $a_{(g)}$ and Psychiatric Factors
Indirect associations from a mediation model (see Figure 4) where stage-based constructs link non-substance psychopathology (3 factors from Lee et al.,) and $a_{(g)}$. F1 = compulsive disorders, F2 = Psychotic disorders, F3 = Neurodevelopmental disorders. The proportion accounted for by the indirect association (%) and the significance of the indirect association are shown. * P < .05 for the indirect association pathway.

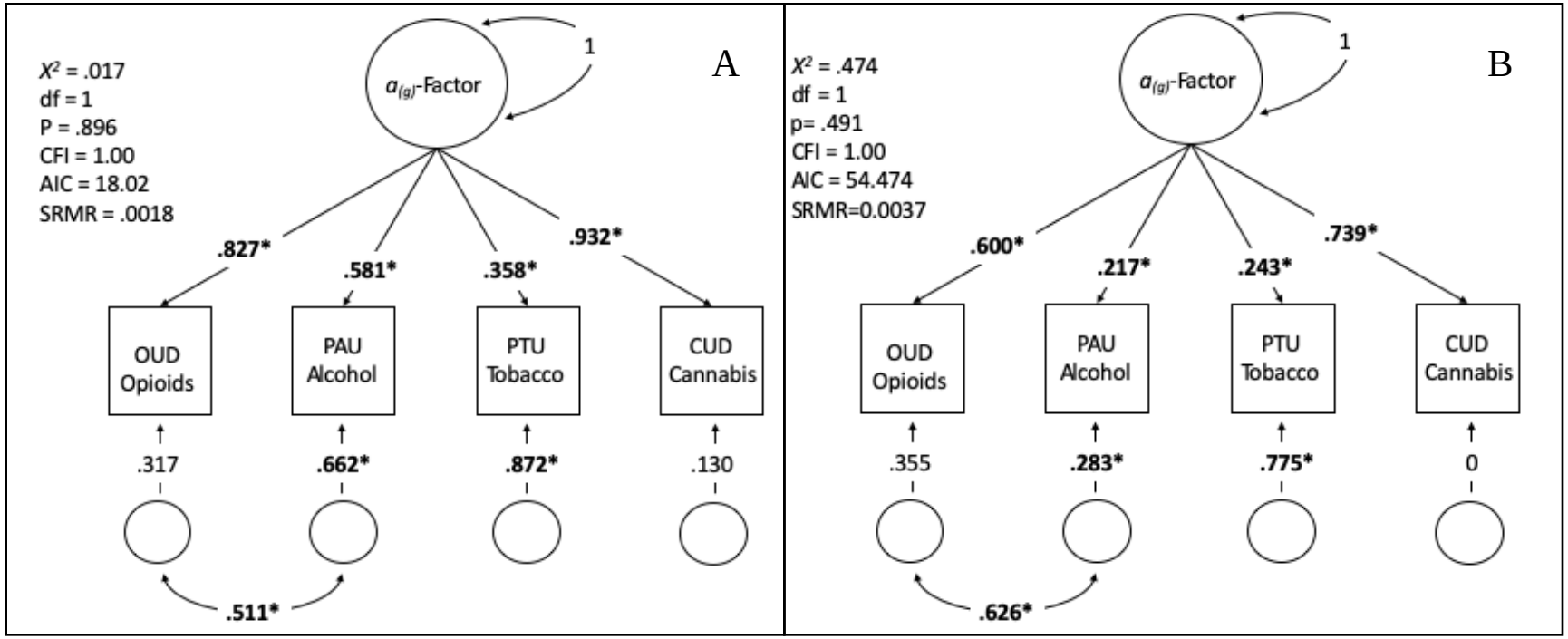

Figure 1. Factor Structure of 4 SUD GWAS. Panel A: the model, loadings, and fit for a model that allowed all 4 SUD categories to load on a latent factor. A residual correlation was added between OUD and PAU to account for large sample overlap (both used the Million Veterans Project data; models without residual correlations also fit well: Supplemental Figure 1). Panel B: the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level, i.e. the three substance use measures are exogenous to all indicators in this model and the model represents the residual associations after accounting for substance use. Both models provided excellent fit to the data. **Bold*** represents significance at $p < .05$.

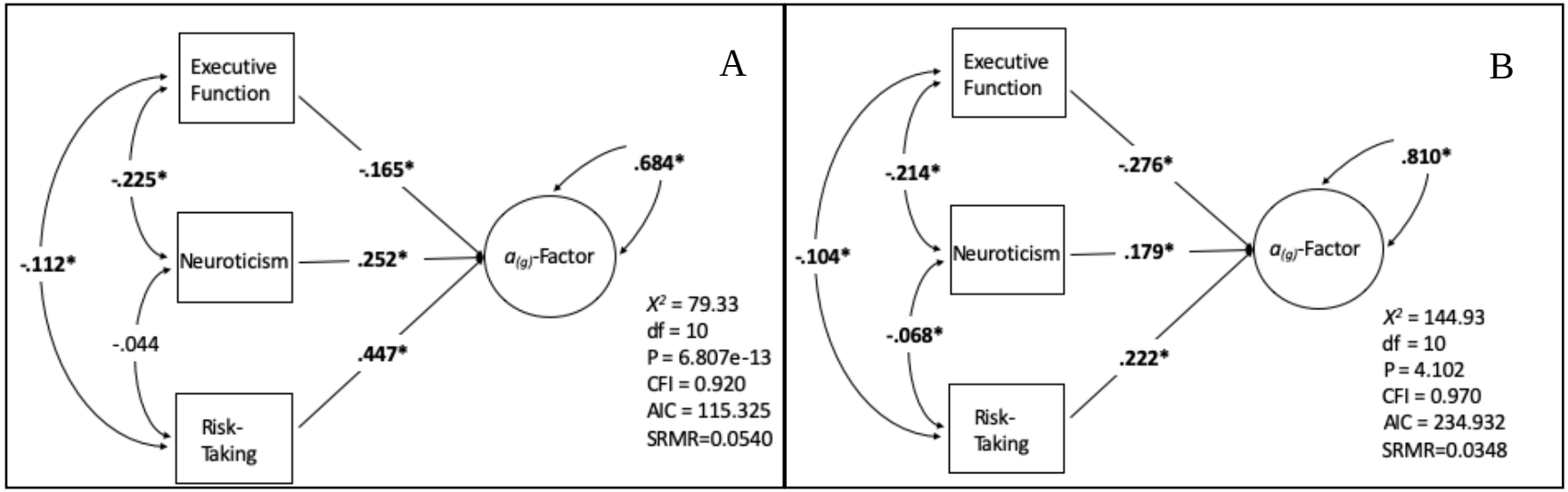

Figure 2. Genetic associations between $a_{(g)}$ and behavioral traits from the behavioral model of Addiction: Executive Function, Neuroticism and Risk-taking. Panel A the model, fit, and regression pathways without accounting for common substance use i.e. the three substance use measures are exogenous to all indicators in this model and the model is the residual associations after accounting for substance use (model was scaled by setting the Opioid Use Disorder loading to 1). Panel B is the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level (regressed on all measured variables/GWAS). **Bold*** represents significance at $p < .05$.

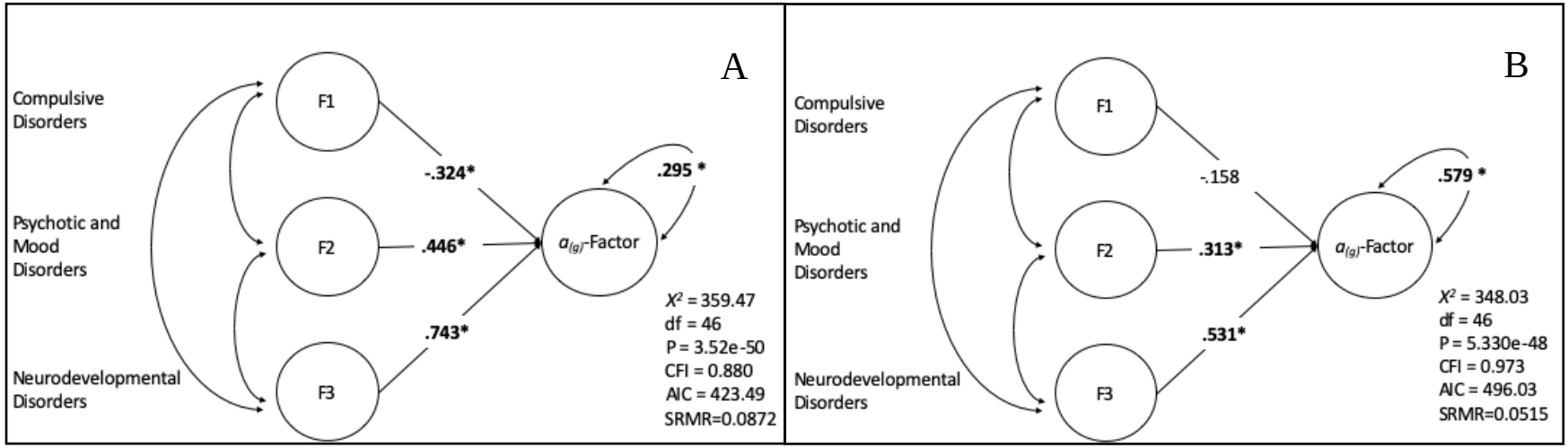

Figure 3. Genetic associations between $a_{(g)}$ and Latent Psychopathology Factors: Compulsive disorders (F1; Tourette's syndrome, Obsessive compulsive disorder, and Eating Disorders), Psychotic Disorders (F2; Major Depressive Disorder, Schizophrenia, and Bipolar Disorder) and neurodevelopmental dysfunction (F3; ADHD, Autism, and Major Depressive Disorder). Panel A the model, fit, and regression pathways without accounting for common substance use (model was scaled by setting the Opioid Use Disorder loading to 1). Panel B is the same model, but accounting for common substance use (ever smoke, ever use marijuana, and drinks per week) as covariates at the indicator level (regressed on all measured variables/GWAS), i.e. the three substance use measures are exogenous to all indicators in this model and the model is the residual associations after accounting for substance use. **Bold*** represents significance at $p < .05$.