

**Submission Type:** Original Article

**Abstract:** 233/250

**Text Word Count:** 3992/4000

**Display Items:** 4/4

**The Addiction Risk Factor: A Unitary Genetic Vulnerability Characterizes Substance Use Disorders and Their Associations with Common Correlates**

Alexander S. Hatoum<sup>1#</sup>, Emma C. Johnson<sup>1</sup>, Sarah M.C. Colbert<sup>1</sup>, Renato Polimanti<sup>2,3</sup>, Hang Zhou<sup>2,3</sup>, Raymond Walters<sup>4,5</sup>, Joel Gelernter<sup>2,3,6,7</sup>, Howard J. Edenberg<sup>8,9</sup>, Ryan Bogdan<sup>10\*</sup> & Arpana Agrawal<sup>1\*</sup>

#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](mailto:ashatoum@wustl.edu); Phone: 903-539-7131

<sup>1</sup>Washington University School of Medicine, Department of Psychiatry, Saint Louis, USA

<sup>2</sup>Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA

<sup>3</sup>Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA

<sup>4</sup>Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

<sup>5</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>6</sup>Department of Genetics, Yale School of Medicine, New Haven, CT, USA

<sup>7</sup>Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA

<sup>8</sup>Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

<sup>9</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA

<sup>10</sup>Department of Psychological & Brain Sciences, Washington University in St. Louis

\*These authors contributed equally to the work

## Abstract

36  
37  
38 Substance use disorders commonly co-occur with one another and with other psychiatric  
39 disorders. They share common features including high impulsivity, negative affect, and lower  
40 executive function. We tested whether a common genetic factor undergirds liability to  
41 problematic alcohol use (PAU), problematic tobacco use (PTU), cannabis use disorder (CUD),  
42 and opioid use disorder (OUD) by applying genomic structural equation modelling to genome-  
43 wide association study summary statistics for individuals of European ancestry (Total N =  
44 **1,019,521**; substance specific Ns range: **82,707-435,563**), while adjusting for the genetics of  
45 substance use (Ns = **184,765-632,802**). We also tested whether shared liability across SUDs is  
46 associated with behavioral constructs (risk taking, executive function, neuroticism; Ns =  
47 **328,339-427,037**) and non-substance use psychopathology (psychotic, compulsive, and early  
48 neurodevelopmental disorders). Shared genetic liability to PAU, PTU, CUD, and OUD was  
49 characterized by a unidimensional addiction risk factor (termed *The Addiction-Risk-Factor*,  
50 independent of substance use. OUD and CUD demonstrated the largest loadings, while  
51 problematic tobacco use showed the lowest loading. *The Addiction-Risk-Factor* was associated  
52 with risk taking, neuroticism, executive function, and non-substance psychopathology, but  
53 retained specific variance before and after accounting for genetics of substance use. Thus, a  
54 common genetic factor partly explains susceptibility for alcohol, tobacco, cannabis, and opioid  
55 use disorder. *The Addiction-Risk-Factor* has a unique genetic architecture that is not shared with  
56 normative substance use or non-substance psychopathology, suggesting that addiction is not the  
57 linear combination of substance use and psychopathology.  
58

59 **INTRODUCTION**

60  
61 Substance use and use disorders (SUDs) represent large and growing public health problems that  
62 account for nearly 6% of global disease burden<sup>1</sup>. SUDs, both licit and illicit, commonly co-occur  
63 with each other and also with non-substance psychopathology; comorbidity is associated with  
64 increased symptom severity<sup>2</sup> and worse outcomes (e.g., less responsivity to treatment, greater  
65 socioeconomic costs<sup>3</sup>). However, the etiology underlying shared risk across these disorders is  
66 poorly understood.

67

68 **Shared Genetic Liability**

69 According to twin studies, the moderate-large heritability (50-60%) of distinct SUDs (i.e., alcohol,  
70 nicotine, cannabis, and other illicit drugs) is partly attributable to a shared genetic vulnerability.<sup>4</sup>  
71 Similarly, genetic correlations estimated from genome-wide association study (GWAS) data  
72 support a shared genetic vulnerability between SUDs (e.g., SNP-rG = .73 between alcohol use  
73 disorder and opioid use disorder),<sup>5</sup> between SUDs and substance use (e.g., SNP-rG = .78 between  
74 problematic alcohol use and drinks per week),<sup>6</sup> and between SUDs and psychopathology (e.g.,  
75 SNP-rG = .33 between cannabis use disorder and major depressive disorder<sup>7</sup>). What remains  
76 unclear is the extent to which genetic liability across substance use *disorders* is *shared with* and  
77 *distinct* from that of substance *use* (but not dependence) and non-substance psychopathology, and  
78 what putative intermediate phenotypes may link shared genetic liability between SUDs and non-  
79 substance psychopathology.

80 ***Substance Use and Use Disorder.*** Substance use and SUDs have substantial genetic  
81 overlap; however, genetic mechanisms that relate to SUD liability beyond normative or frequently  
82 occurring substance use remain. For opioids<sup>8</sup>, alcohol<sup>9,10,11</sup>, and cannabis<sup>7</sup>, the use and use disorder

83 dimensions show differing associations with psychopathology (e.g. schizophrenia) and life  
84 outcomes (e.g. educational attainment)<sup>7,8,12,13</sup>.

85 ***Substance use and Psychopathology.*** Recently, Lee and colleagues<sup>14</sup> identified 3 broad  
86 clusters (psychotic, compulsive, and early neurodevelopmental) representing shared and distinct  
87 genetic liability to 8 non-substance psychiatric disorders. While there has been limited integration  
88 of substance use phenotypes into psychopathology models, polygenic liability to cross-diagnostic  
89 vulnerability is associated with general substance use and SUDs<sup>15</sup>. Further, emerging evidence  
90 suggests partial overlap: tobacco *use* shares variance with ADHD<sup>16</sup>, alcohol and cannabis  
91 dependence load with antisocial behavior<sup>17</sup>, and alcohol use and use disorder load together onto  
92 an Externalizing factor<sup>18</sup>. Collectively, these data suggest that substance use and use disorders  
93 share some common genetic liability with psychopathology.

94 ***Stage-Based Addiction Individual Differences and Substance Use Disorders.*** SUD  
95 vulnerability has been conceptualized within a 3-stage neurobiological model consisting of  
96 binge/intoxication, preoccupation/anticipation, and withdrawal/negative affect<sup>19</sup>. In this model,  
97 initial positive reinforcement is derived from stimulation of neural reward circuitry that drives  
98 impulsive behaviors in the context of under-developed tolerance. With continued use and  
99 progression towards SUD, the reinforcing properties of substances shift from positive to negative  
100 reinforcement; as use becomes compulsive, it functions to return the body to drug-present  
101 homeostasis and alleviate low mood, a predisposition to which is broadly indexed by  
102 neuroticism<sup>20</sup>. Following repeated drug-reward and drug-homeostasis pairings, cognitive  
103 preoccupation with the drug in expectation of reward/relief emerges in the context of impaired  
104 executive functioning<sup>21</sup>. While GWASs support genetic correlations between SUDs and risk-

105 taking<sup>5,22</sup>, Executive Functioning<sup>23</sup>, and negative affect<sup>5,22</sup>, the extent to which common genetic  
106 liability across SUDs relates to these constructs has yet to be examined.

### 107 **The Current Study**

108 Given evidence of shared liability to SUDs and other forms of psychopathology, understanding  
109 the shared and unique genetic contributions to SUDs and how these relate to heritable proxies for  
110 stage-based addiction constructs, non-substance psychopathology, and substance use may generate  
111 etiologic insights that improve psychiatric nosology, prevention, and treatment. To this end, we  
112 *first* estimate the shared genetic structure across SUDs by applying genomic structural equation  
113 modelling (gSEM)<sup>24</sup> to summary statistics generated by the largest GWAS of problematic alcohol  
114 use (PAU)<sup>22</sup>, problematic tobacco use (PTU)<sup>25,26</sup>, cannabis use disorder (CUD)<sup>7</sup>, and opioid use  
115 disorder (OUD)<sup>5</sup>. We name the shared variance across SUDs the *Addiction-Risk-Factor*. *Second*,  
116 we relate the *Addiction-Risk-Factor* to genetics of behavioral constructs representing proxies of  
117 the stage-based model of SUDs. We estimate the extent to which genetic liability to risk-taking,  
118 executive function, and neuroticism are related to *The Addiction-Risk-Factor*. *Third*, we examine  
119 whether *The Addiction-Risk-Factor* is associated with the 3 factors representing genetic liability  
120 to non-substance psychopathology<sup>14</sup> (i.e., psychotic, compulsive, and neurodevelopment) and  
121 whether stage-based addiction constructs (i.e., risk-taking, executive function, neuroticism)  
122 indirectly link *The Addiction-Risk-Factor* to psychopathology. *Finally*, given that genetic liability  
123 to substance use (e.g., ever using, quantity-frequency) and later stages of SUDs are partially  
124 distinct<sup>7,12,13</sup>, we repeat all analyses while incorporating genetic liability to substance use (i.e.,  
125 alcohol drinks/week<sup>26</sup>; tobacco ever regularly use<sup>26</sup>, cannabis ever use<sup>27</sup>) as covariates.

126 We hypothesized that SUDs and problem substance use would be largely characterized by  
127 a common genetic vulnerability (i.e., *The Addiction-Risk-Factor*) with evidence of potentially

128 important substance-specific liability (e.g., metabolic and signaling pathways for a specific drug  
129 such as *ADH1B* variants with alcohol<sup>28</sup>). We hypothesized that (i) *The Addiction-Risk-Factor*  
130 would be associated with all 3 non-substance psychiatric clusters while retaining variance unique  
131 to itself, (ii) genetic liability to behavioral phenotypes representing vulnerability to stage-based  
132 addiction constructs (i.e., risk-taking, executive function, and neuroticism) would be associated  
133 with *The Addiction-Risk-Factor* and account for a proportion of the association between *The*  
134 *Addiction-Risk-Factor* and psychopathology factors, and (iii) after accounting for genetics of  
135 substance use, *The Addiction-Risk-Factor* would retain unique variance (i.e., we expect significant  
136 residual genetic correlations among SUDs) and maintain similar patterns with non-substance  
137 psychopathology and stage-based constructs.

138

139 **METHODS**

140 **Samples**

141 Summary statistics from the largest available discovery GWASs were used to represent genetic  
142 risk for each construct (more details are in **Supplemental Table 1**). These include: **i) 4 SUDs**  
143 (problematic alcohol use<sup>22</sup>, problematic tobacco use<sup>25</sup>, cannabis use disorder<sup>7</sup>, opioid use  
144 disorder<sup>5</sup>); **ii) 3 *substance use phenotypes*** (alcohol drinks/week<sup>26</sup>, lifetime ever smoking<sup>26</sup>,  
145 lifetime cannabis use<sup>27</sup>); **iii) 3 *traits proxying the stage-based model of SUDs*** (risk-taking,  
146 executive function, neuroticism); and **iv) 9 *non-substance psychiatric disorders***. Analyses were  
147 restricted to data from individuals of European ancestry because GWAS on these constructs in  
148 other ancestral origins are not available or are underpowered, and cross-ancestry analysis can  
149 confound genetic correlation estimates<sup>29</sup>. All GWAS summary statistics were filtered to retain  
150 variants with minor allele frequencies > 0.01 and INFO score > 0.90 for GSCAN and PGC<sup>7,26</sup> and  
151 INFO score > 0.70 for the MVP<sup>5,30</sup>.

152

153 **Problematic Substance Use/Substance Use Disorder Summary Statistics**

154 *Problematic Alcohol Use*: Summary statistics for problematic alcohol use (PAU) were derived  
155 from a meta-analysis of GWASs of DSM-IV alcohol dependence from the Psychiatric Genomics  
156 Consortium<sup>11</sup> (PGC-AD; n = 11,569 case, 34,999 controls), ICD-9/10 based diagnoses of alcohol  
157 use disorders from the Million Veteran Program phase 1 and 2 data (MVP; n = 45,995 cases;  
158 221,396 controls)<sup>9</sup> and the Problem subscale score from the Alcohol Use Disorders Identification  
159 Test (AUDIT-P)<sup>13</sup> from the UK Biobank (n = 121,604)<sup>22</sup>. The final GWAS summary statistics  
160 included data on 435,563 participants<sup>22</sup>. We also report on model fit with PGC-AD (instead of  
161 PAU) in the supplement (**Alternative Models, M1**).

162 *Problematic Tobacco Use (PTU)*: We used summary statistics from the GWAS of the  
163 Fagerström Test for Nicotine Dependence<sup>25</sup> (FTND). As cigarettes per day is an item within the  
164 FTND and the genetic correlation between FTND and cigarettes per day is high (calculated  $r_G =$   
165  $0.97$   $CI = .12$ )<sup>12</sup>, we combined CPD And FTND into a single indicator. We applied Multi-Trait  
166 Analysis of Genome-wide association study summary statistics (MTAG<sup>31</sup>) to summary statistics  
167 generated from the GWAS and Sequencing Consortium of Alcohol and Nicotine Use (GSCAN)  
168 GWAS of cigarettes per day to create the combined problematic tobacco use (PTU) phenotype<sup>26</sup>.  
169 The final GWAS summary statistics had an effective sample size of  $n = 270,120$  individuals. We  
170 also report on model fit with just FTND as an indicator in the supplement (**Alternative Models**  
171 **M2**).

172 *Cannabis Use Disorder (CUD)*: Summary statistics were derived from a GWAS meta-  
173 analysis<sup>7</sup> of DSM-IV and DSM-III-R cannabis abuse and dependence from the Psychiatric  
174 Genomics Consortium ( $n = 5,289$  cases;  $n = 10,004$  controls), ICD-10 cannabis use disorder  
175 from the Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) ( $n =$   
176  $2,758$  cases;  $n = 53,326$  controls), and hospital-based diagnoses from deCODE ( $n = 6,033$  cases;  
177  $n = 280,396$  controls). The final European-ancestry sample included  $14,080$  cases with CUD and  
178  $343,726$  controls.

179 *Opioid Use Disorder*: Opioid use disorder (OUD): Summary statistics were derived from  
180 a meta-analysis<sup>5</sup> of GWASs of DSM-IV opioid abuse or dependence from Yale-Penn, and the  
181 Study of Addiction: Genetics and Environment, and ICD-9/10 codes for opioid use disorder from  
182 the Million Veteran Program ( $n = 10,544$  cases;  $n = 72,163$  opioid-exposed controls).

183

184 **Substance Use Summary Statistics**

185 *Alcohol use.* Alcohol use summary statistics were derived from the GSCAN GWAS<sup>32</sup> for current  
186 (this past week or average in the past year) reported drinks/week (n = 537,349). There was a  
187 strong correlation with lifetime PAU (SNP-rG between drinks/week and PAU =  $0.77 \pm 0.02$ )<sup>22</sup>.

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

191 *Lifetime Cannabis Use.* We used summary statistics from a meta-analysis of lifetime  
192 cannabis ever-use from the International Cannabis Consortium and UK Biobank (ever n =  
193 43,380; never n = 118,702)<sup>27</sup>. There was a moderate correlation with CUD<sup>7</sup> (SNP-rG =  
194  $0.47 \pm 0.05$ ).

195

## 196 **Stage-based Behavioral Constructs**

197 The three-stage behavioral model of addiction focuses on “state” changes in substance use  
198 behaviors. Because GWASs measure individual differences in traits, we selected behaviors that  
199 (1) are known to convey vulnerability to each stage as proxies, and (2) are heritable.

200 *Risk-taking and sensitivity to reward.* A GWAS of risk-taking derived from a single item  
201 in the UK Biobank (“Would you describe yourself as someone who takes risks?”; data field  
202 #2040; risk taker n = 83,677; non-risk taker n = 244,662)<sup>33</sup>.

203 *Executive Function.* The “preoccupation/anticipation” stage is characterized by  
204 maladaptive reward valuation and future planning. Recent work argues that this vulnerability is  
205 captured by executive functioning<sup>34</sup>. Summary statistics from a GWAS of a latent factor  
206 representing common executive functioning were used (N = 427,037)<sup>23</sup>.

207 *Negative Emotionality and Sensitivity to Stress*. The stage of withdrawal/negative affect  
208 represents substance use functioning to mitigate aversive withdrawal symptoms, such as negative  
209 affect. Neuroticism has been found to modify stress sensitivity and neural reward processing<sup>35</sup>.  
210 Neuroticism was chosen as a trait-based measure representing liability to negative affect as  
211 opposed to depression because depression was included in the non-substance psychiatric disorder  
212 factor generation and because neuroticism includes trans-diagnostic constructs such as negative  
213 urgency (i.e., impulsive attempts to cope with negative affect) that may place individuals at risk  
214 for the negative reinforcing aspects of SUDs. We selected the largest GWAS of neuroticism as a  
215 heritable proxy (N = 390,278)<sup>20</sup>.

#### 216 **Non-Substance Summary Statistics**

217 Summary statistics from the PGC Cross-disorder GWAS on the 8 disorders that were  
218 previously shown to fit a 3 factor confirmatory model were used<sup>14</sup>. These disorders included  
219 Schizophrenia<sup>36</sup>, Bipolar disorder<sup>37</sup>, Major depressive disorder<sup>38</sup>, Attention Deficit Hyperactivity  
220 disorder<sup>39</sup>, Obsessive Compulsive Disorder<sup>40</sup>, Anorexia Nervosa<sup>41</sup>, Tourette Syndrome<sup>42</sup>, and  
221 Autism Spectrum Disorder<sup>43</sup> (See **Supplemental Table 1 for details**).

#### 223 **Statistical analysis**

224 *First*, we estimated the pairwise genetic correlations between PAU, PTU, CUD and OUD using  
225 Linkage Disequilibrium Score Regression (LDSR)<sup>29</sup>. After confirming that the four SUDs were  
226 significantly genetically correlated (see **Results**), we applied confirmatory factor analysis to the  
227 covariance matrix generated by LDSR using gSEM<sup>44</sup> with weighted least squares estimation;  
228 PAU, PTU, OUD, and CUD indicators were allowed to load freely on a single latent factor (i.e.,  
229 *The Addiction-Risk-Factor*). Variance of this common latent factor was scaled to 1.0. A residual

230 correlation between PAU and OUD was also estimated to account for measurement overlap,  
231 because the Million Veterans Project sample was contained in both PAU and OUD GWAS (but  
232 see model fit without this residual correlation in the supplement – **Alternative Models M1**). In  
233 supplemental analyses, we also examined alternative two factor models (**Alternative Models**  
234 **M3**).

235 *Second*, we used a series of structural regression models to estimate the extent to which  
236 genetic liability to stage-based constructs of addiction (i.e., risk-taking, executive function and  
237 neuroticism) are related to *The Addiction-Risk-Factor*. Here, the *Addiction-Risk-Factor* variance  
238 was freed, and the OUD loading was set to 1.0 to scale the model. Intercorrelations were  
239 estimated between risk-taking, executive function and neuroticism.

240 *Third*, we recreated the three factors from Lee et al<sup>14</sup>. (i.e., psychotic disorders,  
241 compulsive disorders, and early neuro-developmental disorders) and estimated their relationship  
242 with *The Addiction-Risk-Factor* while allowing for inter-factor correlations (the association  
243 between *The Addiction-Risk-Factor* and an alternative cross-disorder genetic model from a  
244 preprint<sup>45</sup> was also estimated; this alternative model is shown in **Supplemental Figure 5**). This  
245 allowed us to estimate the unique association between each of the 3 psychopathology factors and  
246 *The Addiction-Risk-Factor* and to estimate variance that was residual to *The Addiction-Risk-*  
247 *Factor*. We then examined whether proxies for stage-based addiction constructs (i.e., risk-taking,  
248 executive function and neuroticism) indirectly linked *The Addiction-Risk-Factor* to the 3 non-  
249 substance psychopathology factors using a multiple mediator model. We also conducted  
250 supplemental modified Q-Trait analyses<sup>45</sup> to examine the extent of the mediation (**Q-Trait**  
251 **Analysis**). To estimate residual associations (i.e., direct paths) between the stage-based  
252 constructs and *The Addiction-Risk-Factor*, we re-structured the mediation model to one in which

253 the 3 non-SUD psychopathology factors served as “mediators” of the relationship between risk-  
254 taking, executive functioning, neuroticism, and *The Addiction-Risk-Factor*.

255 To separate the genetics of SUD from the genetics of substance use, we estimated models  
256 where substance *use* GWAS summary statistics were endogenous predictors of all measured  
257 variables in the model. For example, in the model estimating the association between *The*  
258 *Addiction-Risk-Factor* and psychiatric factors, the 8 psychiatric disorders and the 4 SUD  
259 disorder variables were regressed on the 3 substance use variables. In this way, covariate effects  
260 were estimated simultaneously to our associations of interest.

261

262 **RESULTS**

263 **The Addiction Risk Factor**

264 Genetic correlations between problematic alcohol use (PAU<sup>22</sup>), problematic tobacco use  
265 (PTU<sup>25,26</sup>), cannabis use disorder (CUD<sup>7</sup>), and opioid use disorder (OUD<sup>5</sup>) ranged from 0.19  
266 (S.E. = .04) to 0.78 (.09) (**Supplemental Figure 1** and **Supplemental Figure 2**). PTU showed  
267 the lowest SNP-rG with other SUD phenotypes [i.e., PAU = 0.19 (.04), CUD = 0.31 (.05), OUD  
268 = 0.26 (.08)] while OUD showed the highest [PAU = 0.69 (.07), CUD = 0.78 (.09)]. A  
269 confirmatory factor model specifying a unidimensional *Addiction-Risk-Factor* underlying the  
270 genetic covariance among PAU, PTU, CUD and OUD fit the data well [ $X^2(1) = .017$ ,  $p = .895$ ,  
271 CFI = 1, SRMR = .002; residual  $r = .51$ ,  $p = 0.016$ ; **Figure 1A**]. Loadings were uniformly high  
272 except for PTU. Neither PAU nor PTU were impacted by the inclusion of non-diagnostic indices  
273 of addiction risk (**Supplemental Results Alternative Models M1-M2**). Alternative 2-factor  
274 models did not fit the data well (**Alternative Models M3**).

275 The inclusion of genetic liability to typical substance use did not modify the single factor  
276 structure of *The Addiction-Risk-Factor* (**Figure 1B**); all SUDs continued to load significantly on  
277 the factor. However, factor loadings were lower for all substances, especially for PAU, which  
278 may be attributable to the high genetic correlation between drinks/week and PAU. Alternative  
279 parameterization of substance use as covariates did not improve model fit (**Alternative Models**  
280 **M4**).

281 **Shared Liability to Stage-based Behavioral Phenotypes**

282 Genetic liability to stage-based addiction constructs was shared with *The Addiction-Risk-Factor*  
283 (**Figures 2, Q-Trait Analysis in Supplemental Methods**). As expected, *The Addiction-Risk-*  
284 *Factor* was positively associated with genetic liability to risk-taking ( $\beta = 0.45$ ) and neuroticism

285 ( $\beta = 0.25$ ), and negatively associated with executive function ( $\beta = -0.17$ ; **Figure 2A**). Despite  
286 significant genetic overlap between *The Addiction-Risk-Factor* and stage-based behavioral  
287 phenotypes, *The Addiction-Risk-Factor* retained unique variance (*Addiction-Risk-Factor* residual  
288 = 0.68). When conditioning for genetic liability for substance use, *The Addiction-Risk-Factor*  
289 remained significantly associated with increased genetic liability to risk-taking ( $\beta = 0.22$ ) and  
290 neuroticism ( $\beta = 0.18$ ) and decreased genetic liability to executive function ( $\beta = -0.28$ ; **Figure**  
291 **2B**). Accounting for genetic liability for substance use substantially reduced the association  
292 between *The Addiction-Risk-Factor* and risk-taking from 0.45 to 0.22 ( $p_{(df = 1)} = 4e-09$ ) and  
293 accentuated the negative association with executive function from  $\beta = -0.17$  to  $-0.28$  ( $p_{(df = 1)} =$   
294  $0.013$ ); there was a smaller effect on the association with neuroticism (from  $\beta = 0.25$  to  $0.18$ ,  $p_{(df$   
295  $= 1)} = 0.012$ ).

296

### 297 **Shared Liability to Non-substance Psychopathology**

298 Genetic liability to non-substance psychopathology (i.e., compulsive disorders, psychotic  
299 disorders, and neurodevelopmental disorders) was shared with *The Addiction-Risk-Factor*  
300 (**Figure 3**, full models in **Supplemental Figure 4**; **Supplemental Figure 5** shows results with  
301 an alternative cross-disorder model from a recent preprint<sup>45</sup>). Psychotic disorders ( $\beta = 0.45$ ) and  
302 neurodevelopmental disorders ( $\beta = 0.74$ ) were positively associated with *The Addiction-Risk-*  
303 *Factor* while compulsive disorders showed a negative association ( $\beta = -0.32$ ; **Figure 3A**). Due to  
304 the strong correlation between *The Addiction-Risk-Factor* and early-onset neurodevelopmental  
305 disorders (which includes ADHD) we allowed ADHD to load on *The Addiction-Risk-Factor* to  
306 control for ADHD; here, an association between *The Addiction-Risk-Factor* and early-onset  
307 neurodevelopmental disorders remained, but was significantly attenuated (from  $\beta = 0.74$  to  $0.43$ ,

308  $p_{(df = 1)} = 5e-5$ ). When conditioning *The Addiction-Risk-Factor* for substance use, the psychotic  
309 and early neurodevelopmental disorder factors remained significantly associated with *The*  
310 *Addiction-Risk-Factor* (**Figure 3B**). Despite the significant genetic overlap with other  
311 psychiatric disorder domains, *The Addiction-Risk-Factor* retained unique variance representing  
312 genetic liability specific to SUDs (*The Addiction-Risk-Factor* residual = 0.30,  $p = 4.54e-3$ ). This  
313 unique variance remained significant when accounting for genetic liability to substance use (*The*  
314 *Addiction-Risk-Factor* residual = 0.58,  $p = 0.015$ ).

315         The specifications for the mediation models are shown in **Supplemental Figure 6**.  
316 Genetic liability to risk taking accounted for a proportion of the associations between all non-  
317 substance psychopathology domains and *The Addiction-Risk-Factor* (**Table 1**). Executive  
318 function uniquely indexed an indirect effect between psychotic disorders and *The Addiction-*  
319 *Risk-Factor* (**Table 1**). When conditioning *The Addiction-Risk-Factor* for genetic liability to  
320 substance use, risk-taking no longer accounted for a portion of the association between any non-  
321 substance psychopathology domain and *The Addiction-Risk-Factor*, but executive function  
322 continued to account for a proportion of the overlap (indirect effect of 0.048) between psychotic  
323 disorders and *The Addiction-Risk-Factor* (**Table 1**). *Post-hoc* analyses revealed that executive  
324 function retained a unique association with *The Addiction-Risk-Factor* after accounting for  
325 genetic liability to both substance use and non-substance psychopathology (**Supplementary**  
326 **Table 2**).

327

328 **DISCUSSION**

329 We applied genomic structural equation modeling (gSEM)<sup>24</sup> to GWAS summary statistics to  
330 characterize the genetic influences shared across SUDs and estimate how common genetic  
331 liability is related to trait conceptualizations of a theoretical stage-based SUD model as well as to  
332 non-substance psychopathology. Three primary findings emerged. *First*, genetic risk for specific  
333 SUD phenotypes (i.e., PAU<sup>22</sup>, PTU<sup>25,26</sup>, CUD<sup>7</sup>, and OUD<sup>5</sup>) was largely attributable to a single  
334 Addiction risk factor, *The Addiction-Risk-Factor* (**Figure 1**). *Second*, *The Addiction-Risk-Factor*  
335 was associated with genetic liability to trait representations of stage-based facets of addiction  
336 (risk taking [binge/intoxication], executive function [preoccupation/anticipation], neuroticism  
337 [negative affect]<sup>19</sup>; **Figure 2**). It was also associated with non-substance psychopathology factors  
338 (compulsive disorders, psychotic disorders, neurodevelopmental disorders; **Figure 3**). Trait  
339 representations of stage-based facets of addiction partially accounted for the shared genetic  
340 liability between non-substance psychopathology and *The Addiction-Risk-Factor*. *Third*,  
341 associations between *The Addiction-Risk-Factor* and stage-based constructs and non-substance  
342 psychopathology were largely independent of genetic liability to substance *use* phenotypes (i.e.,  
343 tobacco use, cannabis use, alcoholic drinks/week). However, consistent with the stage-based  
344 model of addiction, accounting for substance use attenuated associations between risk taking and  
345 *The Addiction-Risk-Factor* while potentiating associations with executive functioning.  
346 Collectively, our findings suggest that SUDs are characterized by a common genetic factor,  
347 *Addiction-Risk-Factor*.  
348  
349 ***The Addiction-Risk-Factor* retains variance that is not shared with other psychopathology**

350 After accounting for genetic liability to substance use as well as the commonality between *The*  
351 *Addiction-Risk-Factor* and non-substance psychopathology, *The Addiction-Risk-Factor* retained  
352 significant variance. These data suggest that *The Addiction-Risk-Factor* may be characterized by  
353 unique pathways not shared with substance use or non-substance psychopathology, i.e., addiction  
354 is not the linear combination of substance use and psychopathology.

355 A single latent factor, fit these data well, but specific SUDs showed varying degrees of  
356 association. The illicit SUDs (CUD and OUD; **Figure 1**) were almost entirely captured by the  
357 common latent factor. Notably, the loading for PTU on *The Addiction-Risk-Factor* was the  
358 smallest. One potential contributor to the residual variance of PTU may be the use of FTND and  
359 cigarettes/day as indices of PTU. Unlike the Diagnostic and Statistical Manual (DSM) criteria  
360 which index psychological and physiological aspects of tobacco use disorder, the FTND is an  
361 index of biochemical dependence and although used widely by investigators to define addiction,  
362 phenotypically shows only moderate agreement with DSM-defined nicotine dependence ( $r =$   
363  $0.50$ ;  $\kappa = 0.3$ )<sup>46</sup>.

364

### 365 **Proxies of Stage-based Behavioral Constructs and *The Addiction-Risk-Factor***

366 Behavioral stage-based models of SUD posit a cyclical relationship between positive  
367 reinforcement, negative reinforcement, and incentive salience<sup>19</sup> that we found can be (partially)  
368 captured by genetic liability to risk-taking, executive functioning, and negative emotionality  
369 (neuroticism). The strongest association with *The Addiction-Risk-Factor* was for risk-taking.

370 When substance use was included as a covariate in the model, the shared genetic loading  
371 between *The Addiction-Risk-Factor* and both risk-taking and neuroticism was attenuated down  
372 while the association with executive function increased. The reduction in the association with

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

378

### 379 **Non-substance Psychopathology and *The Addiction-Risk-Factor***

380 We found that the 3 non-substance psychopathology clusters, derived from 8 psychiatric  
381 disorders<sup>14</sup>, were genetically associated with *The Addiction-Risk-Factor*. The association with  
382 early neurodevelopmental disorders, which include ADHD, was the strongest. Cross-loading  
383 ADHD on *The Addiction-Risk-Factor* to condition on ADHD attenuated the loading but it  
384 remained high. Associations between *The Addiction-Risk-Factor* and the psychopathology  
385 clusters were greater than associations with trait representations of behavioral stages of addiction  
386 (with the exception of risk-taking). For instance, the genetic association between *The Addiction-*  
387 *Risk-Factor* and the two disorder clusters that included Major Depressive Disorder (i.e.,  
388 psychotic disorders and early neurodevelopmental disorders) was greater in magnitude than the  
389 *Addiction-Risk-Factor*-neuroticism association. Interestingly the compulsive disorder factor did  
390 not show strong associations with *The Addiction-Risk-Factor*, suggesting that compulsive  
391 disorders and addiction-related compulsive behaviors have distinct etiologies.

392       Of the 3 behavioral correlates, risk-taking was the most prominent contributor to the  
393 association between *The Addiction-Risk-Factor* and all non-substance psychopathology factors.  
394 After accounting for substance use, only risk-taking and executive function mediated *The*  
395 *Addiction-Risk-Factor* associations with the psychotic disorder factor. Executive function

396 maintained the only direct association with *The Addiction-Risk-Factor* after accounting for  
397 genetics of substance use and genetics of non-substance psychopathology. Thus, we speculate  
398 that while risk-taking may characterize the genetic overlap between substance use and other  
399 psychopathology, executive function impairment is a risk factor that not only shapes the overlap  
400 between addiction and non-substance psychopathology but also explains variance in addiction  
401 above and beyond that overlap.

402

### 403 **Limitations**

404 There are several limitations. First, we had to restrict our analyses to individuals of European  
405 descent due to the lack of well-powered discovery GWAS informative for other ancestry groups.  
406 Second, to maximize sample size of discovery GWASs, our alcohol and tobacco use GWAS  
407 incorporated measures of “problematic” use that, while genetically highly correlated with AUD  
408 and ND, may include behavioral patterns that are less severe than those represented by use  
409 disorder and were not assessed based on clinical presentation. Third, the analyses contain an  
410 over-representation of men, in part because the MVP samples contributed most of OUD and half  
411 of PAU and the MVP is ~90% male. Studies with larger numbers of women would allow  
412 stratified analyses to explore the differences between sexes observed in epidemiological studies.  
413 Fourth, while it is unlikely that individuals completed assessments of risk-taking, neuroticism,  
414 and executive function while under the influence of substances, how substance use may have  
415 influenced these assessments cannot be determined. Fifth, though significant, mediation  
416 pathways were small in effect. Sixth, how these processes effect phenotypic patterns is unknown,  
417 however twin studies support a common factor model as well<sup>4</sup>.

418

419 **Conclusions**

420 Common genetic liability undergirds distinct SUDs and shares variance with putative behavioral  
421 intermediary phenotypes/SUD risk factors and non-substance psychopathology. This addiction  
422 genetic factor is more than a linear combination of substance use and psychopathology; it  
423 represents a unique addiction dimension that is partially captured by executive functions.  
424

425 **Funding.** This research was supported by MH109532 (AA, JG, HJE, ECJ) and T32DA007261  
426 (ASH). AA acknowledges K02DA32573. ECJ was supported by F32AA027435. RP  
427 acknowledges R21DA047527. RB acknowledges R21-AA027827.  
428 The Substance Use Disorders Working Group of the Psychiatric Genomics Consortium (PGC-  
429 SUD) is supported by funds from NIDA and NIMH to MH109532. Disclosures/COI: JG is  
430 named as an inventor on PCT patent application #15/878,640 entitled: "Genotype-guided dosing  
431 of opioid agonists," filed January 24, 2018. The authors have no conflicts of interest.

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

434 The MVP summary statistics were obtained via an approved dbGaP application  
435 ([phs001672.v4.p1](https://www.ncbi.nlm.nih.gov/bioproject/5001672)). The authors thank Million Veteran Program (MVP) staff, researchers, and  
436 volunteers, who have contributed to MVP, and especially participants who previously served  
437 their country in the military and now generously agreed to enroll in the study. (For details, see  
438 <https://www.research.va.gov/mvp/> and Gaziano, J.M. et al. Million Veteran Program: A mega-  
439 biobank to study genetic influences on health and disease. *J Clin Epidemiol* 70, 214-23 (2016)).  
440 This research is based on data from the Million Veteran Program, Office of Research and  
441 Development, Veterans Health Administration, and was supported by the Veterans  
442 Administration (VA) Cooperative Studies Program (CSP) award #G002.

443 This study included summary statistics of a genetic study on cannabis use (Pasman et al, 2018  
444 *Nature Neuroscience*). We would like to acknowledge all participating groups of the  
445 International Cannabis Consortium, and in particular the members of the working group  
446 including Joelle Pasman, Karin Verweij, Nathan Gillespie, Eske Derks, and Jacqueline Vink.  
447 Pasman et al, (2018) included data from the UK Biobank resource under application numbers  
448 9905, 16406 and 25331.

449

450

451

452 **Reference Cited**

- 453 1. Degenhardt, L. *et al.* The global burden of disease attributable to alcohol and drug use in  
454 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of  
455 Disease Study 2016. *The Lancet Psychiatry* **5**, 987–1012 (2018).
- 456 2. Bhalla, I. P., Stefanovics, E. A. & Rosenheck, R. A. Clinical Epidemiology of Single  
457 Versus Multiple Substance Use Disorders. *Med. Care* **55**, S24–S32 (2017).
- 458 3. Merikangas, K. R. & Kalaydjian, A. Magnitude and impact of comorbidity of mental  
459 disorders from epidemiologic surveys. *Current Opinion in Psychiatry* vol. 20 353–358  
460 (2007).
- 461 4. Palmer, R. H. C. *et al.* Genetic etiology of the common liability to drug dependence:  
462 evidence of common and specific mechanisms for DSM-IV dependence symptoms. *Drug*  
463 *Alcohol Depend.* **123 Suppl**, S24–32 (2012).
- 464 5. Zhou, H. *et al.* Association of OPRM1 Functional Coding Variant with Opioid Use  
465 Disorder: A Genome-Wide Association Study. *JAMA Psychiatry* (2020)  
466 doi:10.1001/jamapsychiatry.2020.1206.
- 467 6. Sanchez-Roige, S., Palmer, A. A. & Clarke, T. K. Recent Efforts to Dissect the Genetic  
468 Basis of Alcohol Use and Abuse. *Biological Psychiatry* vol. 87 609–618 (2020).
- 469 7. Johnson, E. C. *et al.* A large-scale genome-wide association study meta-analysis of  
470 cannabis use disorder. *The Lancet Psychiatry* **0**, (2020).
- 471 8. Polimanti, R. *et al.* Leveraging genome-wide data to investigate differences between  
472 opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics  
473 Consortium. *Mol. Psychiatry* **25**, 1673–1687 (2020).
- 474 9. Kranzler, H. R. *et al.* Genome-wide association study of alcohol consumption and use  
475 disorder in 274,424 individuals from multiple populations. *Nat. Commun.* **10**, (2019).
- 476 10. Sanchez-Roige, S. *et al.* Genome-wide association study meta-analysis of the alcohol use  
477 disorders identification test (AUDIT) in two population-based cohorts. *Am. J. Psychiatry*  
478 **176**, 107–118 (2019).
- 479 11. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common  
480 genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
- 481 12. Sanchez-Roige, S., Cox, N. J., Johnson, E. O., Hancock, D. B. & Davis, L. K. Alcohol and  
482 cigarette smoking consumption as genetic proxies for alcohol misuse and nicotine  
483 dependence. *Drug Alcohol Depend.* **221**, 108612 (2021).
- 484 13. Sanchez-Roige, S. *et al.* Genome-Wide Association Study Meta-Analysis of the Alcohol  
485 Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *Am. J.*  
486 *Psychiatry* **176**, 107–118 (2019).
- 487 14. Lee, P. H. *et al.* Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across  
488 Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.e11 (2019).
- 489 15. Carey, C. E. *et al.* Associations between polygenic risk for psychiatric disorders and  
490 substance involvement. *Front. Genet.* **7**, (2016).
- 491 16. Jang, S. K. *et al.* Genetic correlation, pleiotropy, and causal associations between  
492 substance use and psychiatric disorder. *Psychol. Med.* 1–11 (2020)  
493 doi:10.1017/S003329172000272X.
- 494 17. Waldman, I. D., Poore, H. E., Luningham, J. M. & Yang, J. Testing structural models of  
495 psychopathology at the genomic level. *World Psychiatry* **19**, 350–359 (2020).
- 496 18. Karlsson Linnér, R. *et al.* Multivariate genomic analysis of 1.5 million people identifies  
497 genes related to addiction, antisocial behavior, and health. *Henry R. Kranzler* **17**, 20

- 498 (2020).
- 499 19. Goldstein, R. Z. & Volkow, N. D. Drug addiction and its underlying neurobiological  
500 basis: Neuroimaging evidence for the involvement of the frontal cortex. *American Journal*  
501 *of Psychiatry* vol. 159 1642–1652 (2002).
- 502 20. Nagel, M., Watanabe, K., Stringer, S., Posthuma, D. & van der Sluis, S. Item-level  
503 analyses reveal genetic heterogeneity in neuroticism. *Nat. Commun.* **9**, 905 (2018).
- 504 21. Koob, G. F. & Volkow, N. D. Neurobiology of addiction: a neurocircuitry analysis. *The*  
505 *Lancet Psychiatry* vol. 3 760–773 (2016).
- 506 22. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563  
507 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.*  
508 1–10 (2020) doi:10.1038/s41593-020-0643-5.
- 509 23. Hatoum, A. *et al.* GWAS of Over 427,000 Individuals Establishes GABAergic and  
510 Synaptic Molecular Pathways as Key for Cognitive Executive Functions. *GWAS Over*  
511 *427,000 Individ. Establ. GABAergic Synaptic Mol. Pathways as Key Cogn. Exec. Funct.*  
512 674515 (2019) doi:10.1101/674515.
- 513 24. Grotzinger, A. D. *et al.* Genomic structural equation modelling provides insights into the  
514 multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
- 515 25. Hancock, D. B. *et al.* Genome-wide association study across European and African  
516 American ancestries identifies a SNP in DNMT3B contributing to nicotine dependence.  
517 *Mol. Psychiatry* **23**, 1–9 (2018).
- 518 26. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into  
519 the genetic etiology of tobacco and alcohol use. *Nature Genetics* vol. 51 237–244 (2019).
- 520 27. Pasmán, J. A. *et al.* GWAS of lifetime cannabis use reveals new risk loci, genetic overlap  
521 with psychiatric traits, and a causal influence of schizophrenia. *Nat. Neurosci.* **21**, 1161–  
522 1170 (2018).
- 523 28. Edenberg, H. J. & McClintick, J. N. Alcohol Dehydrogenases, Aldehyde Dehydrogenases,  
524 and Alcohol Use Disorders: A Critical Review. *Alcoholism: Clinical and Experimental*  
525 *Research* vol. 42 2281–2297 (2018).
- 526 29. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from  
527 polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 528 30. Zhou, H. *et al.* Genome-wide meta-analysis of problematic alcohol use in 435,563  
529 individuals yields insights into biology and relationships with other traits. *Nat. Neurosci.*  
530 **23**, 809–818 (2020).
- 531 31. Turley, P. *et al.* Multi-trait analysis of genome-wide association summary statistics using  
532 MTAG. *Nat. Genet.* **50**, 229–237 (2018).
- 533 32. Liu, M. *et al.* Association studies of up to 1.2 million individuals yield new insights into  
534 the genetic etiology of tobacco and alcohol use. *Nature Genetics* vol. 51 237–244 (2019).
- 535 33. Strawbridge, R. J. *et al.* Genetics of self-reported risk-taking behaviour, trans-ethnic  
536 consistency and relevance to brain gene expression. *Transl. Psychiatry* **8**, (2018).
- 537 34. Volkow, N. D., Koob, G. F. & McLellan, A. T. Neurobiologic Advances from the Brain  
538 Disease Model of Addiction. *N. Engl. J. Med.* **374**, 363–371 (2016).
- 539 35. Bondy, E. *et al.* Neuroticism and reward-related ventral striatum activity: Probing  
540 vulnerability to stress-related depression. (2020) doi:10.31234/osf.io/5wd3k.
- 541 36. Schizophrenia Working Group of the Psychiatric Genomics Consortium, S. W. G. of the  
542 P. G. *et al.* Biological insights from 108 schizophrenia-associated genetic loci. *Nature*  
543 **511**, 421–7 (2014).

- 544 37. Stahl, E. A. *et al.* Genome-wide association study identifies 30 loci associated with  
545 bipolar disorder. *Nat. Genet.* **51**, 793–803 (2019).
- 546 38. Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102  
547 independent variants and highlights the importance of the prefrontal brain regions. *Nat.*  
548 *Neurosci.* **22**, 343–352 (2019).
- 549 39. Demontis, D. *et al.* Discovery of the first genome-wide significant risk loci for attention  
550 deficit/hyperactivity disorder. *Nat. Genet.* **51**, 63–75 (2019).
- 551 40. Arnold, P. D. *et al.* Revealing the complex genetic architecture of obsessive-compulsive  
552 disorder using meta-analysis. *Mol. Psychiatry* **23**, 1181–1188 (2018).
- 553 41. Watson, H. J. *et al.* Genome-wide association study identifies eight risk loci and  
554 implicates metabo-psychiatric origins for anorexia nervosa. *Nat. Genet.* **51**, 1207–1214  
555 (2019).
- 556 42. Yu, D. *et al.* Interrogating the genetic determinants of Tourette’s syndrome and other tic  
557 disorders through genome-wide association studies. *Am. J. Psychiatry* **176**, 217–227  
558 (2019).
- 559 43. Grove, J. *et al.* Identification of common genetic risk variants for autism spectrum  
560 disorder. *Nat. Genet.* **51**, 431–444 (2019).
- 561 44. Grotzinger, A. D. *et al.* Genomic SEM Provides Insights into the Multivariate Genetic  
562 Architecture of Complex Traits. *bioRxiv* 305029 (2018) doi:10.1101/305029.
- 563 45. Grotzinger, A. D. *et al.* Genetic architecture of 11 major psychiatric disorders at  
564 biobehavioral, functional genomic, and molecular genetic levels of analysis. *medRxiv* vol.  
565 18 18–19 (2020).
- 566 46. Agrawal, A. *et al.* A latent class analysis of DSM-IV and fagerström (FTND) criteria for  
567 nicotine dependence. *Nicotine Tob. Res.* **13**, 972–981 (2011).
- 568



**Table 1. Behavioral Liabilities Mediate the Association Between *The Addiction-Risk-Factor* and Psychiatric Factors**

	<b>F1 Indirect %</b>	<b>F1 Indirect P</b>	<b>F2 Indirect %</b>	<b>F2 Indirect P</b>	<b>F3 Indirect %</b>	<b>F3 Indirect P</b>
<b>No covariates</b>						
executive function	0.007	0.463	<b>0.051*</b>	<b>0.001*</b>	0.003	0.486
Neuroticism	0.005	0.842	0.003	0.836	0.013	0.777
Risk-taking	<b>-0.090*</b>	<b>0.001*</b>	<b>0.152*</b>	<b>1.47E-08*</b>	<b>0.043*</b>	<b>0.013*</b>
<b>Controlling for Genetics of Substance Use</b>						
executive function	0.013	0.233	<b>0.048*</b>	<b>0.047*</b>	-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 1. Behavioral Liabilities Mediate the Association Between *The Addiction-Risk-Factor* 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 *The Addiction-Risk-Factor*. 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 3**). *Addiction-rf=The Addiction Risk-Factor*. **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  $p < .05$ . Note that in panel B, the residual of CUD is zero; this model constraint was necessary, as the model produced a negative residual without the constraint. Note: If you want to recreate the correlation matrix from both panels, the model with residual correlations cannot recover the implied correlation between PAU and OUD without taking the square root of the residual variance, rather than the value of the residual variance itself.

**Figure 2. Genetic associations between *The Addiction-Risk-Factor* and behavioral traits:** Executive Function, Neuroticism and Risk-taking. Panel A the model, fit, and regression pathways without accounting for common substance use. 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 *The Addiction-Risk-Factor* 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  $p < .05$ . *Addiction-rf* = *The Addiction-Risk-Factor*.

## **The Addiction Risk Factor: A Unitary Genetic Vulnerability Characterizes Substance Use Disorders and Their Associations with Common Correlates**

### **Supplemental Tables and Figures**

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

#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](mailto:ashatoum@wustl.edu);

<sup>1</sup>Washington University School of Medicine, Department of Psychiatry, Saint Louis, USA

<sup>2</sup>Department of Psychiatry, Division of Human Genetics, Yale School of Medicine, New Haven, CT, USA

<sup>3</sup>Veterans Affairs Connecticut Healthcare System, West Haven, CT, USA

<sup>4</sup>Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA

<sup>5</sup>Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA, USA

<sup>6</sup>Department of Genetics, Yale School of Medicine, New Haven, CT, USA

<sup>7</sup>Department of Neuroscience, Yale School of Medicine, New Haven, CT, USA

<sup>8</sup>Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA

<sup>9</sup>Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN, USA

<sup>10</sup>Department of Psychological & Brain Sciences, Washington University in St. Louis

\*These authors contributed equally to the work

## **Index**

**Alternative Models Pg. 2**

**Q-Trait Analysis pg. 2-3**

**Supplemental Table 1 Pg. 4-7**

**Supplemental Table 2 Pg. 8**

**Supplemental Figure 1 Pg. 9**

**Supplemental Figure 2 Pg. 10**

**Supplemental Figure 3 Pg. 11**

**Supplemental Figure 4 Pg. 12**

**Supplemental Figure 5 Pg. 13**

**Supplemental Figure 6 Pg. 14-15**

## Alternative Model specifications

To establish our factor model, we tried several different approaches. First, we confirmed that the problematic alcohol use (PAU) and problematic tobacco use (PTU) variables did not significantly alter factor structure, as these are more proxies of the underlying construct of interest. Samples that are about an order of magnitude smaller are available for this exploration.

First, we reran the model with alcohol dependence from the Psychiatric Genetics Consortium GWAS of European descent ( $N=38,686$ )<sup>1</sup> as the indicator for alcohol use disorder, instead of the larger problematic alcohol use (PAU) GWAS (Models M1). As the PGC alcohol dependence GWAS did not include data from the MVP, we also used this as an opportunity to test whether the residual genetic correlation between opioid use disorder (OUD) and alcohol dependence was present in the absence of measurement overlap. The model fit well ( $X^2 = 1.22$ ,  $p = .54$ ,  $df = 2$ ,  $CFI = 1$ ). When we tested the residual correlation, it was large although no longer significant ( $r=0.71$ ,  $P=4.46e-01$ ). The residual variance for alcohol dependence was also not significant ( $r=0.207$ ,  $p=.51$ ). As the alcohol dependence measure without MVP does not have a significant residual correlation, we posit that the need for the residual correlation when the MVP PAU summary data are used is to improve fit due to the larger sample or due to *measurement* overlap in the MVP.

Second, we tested whether inclusion of cigarettes per day with FTND created a poor proxy for tobacco dependence (Model M2). The model with FTND an indicator fit well, ( $X^2 = 0.027$ ,  $p = .987$ ,  $df = 2$ ,  $CFI = 1$ ) FTND indicator pathway was larger, but the confidence intervals overlapped substantially ( $r = .502$ ,  $CI = .121$ ).

For our next test of robustness, we conducted two follow-up analysis to examine whether parsing *Addiction-rf* into a two-factor structure improved fit (Models M3). The highest pairwise associations between SUD indicators were OUD with PAU. PTU with CUD also showed a high pairwise associations. Both factors were allowed to correlate. This model did not fit the data well (model fit:  $X^2 = 34.039$ ,  $p = 1.94e-07$ ,  $df = 3$ ,  $CFI = .924$ ,  $AIC = 47.039$ ). We also fit a model where PAU and PTU loaded on one factor (“licit drugs”) while CUD and OUD loaded on another correlated factor (“illicit drugs”). This model did not fit the data well either (model fit:  $X^2 = 293924.2$ ,  $p = <.001$ ,  $df = 5$ ,  $CFI = -721.5134$ ,  $AIC = 293934.2$ ).

Next, there are two alternative models we tested to see if there was a more appropriate analysis to account for substance use (Models M4). To be clear, substance use is a covariate, but its covariate effect may be better modeled via alternative models. First, we generated a substance use factor using ever smoking regularly, marijuana use, and drinks per week as indicators. We then regressed the use disorder indicators on this factor. This model fit the data poorly. ( $X^2=394.4224$ ,  $df=1$ ,  $p=5.739041e-77$   $AIC=426.4224$   $CFI=0.8833732$ ) and significantly worse than our original model ( $X^2\ diff=394.007$ ,  $df=9$ ,  $p=2e-64$ ).

Next, rather than have the substance use variables endogenous to the indicators of addiction factor, we tested whether an alternative model where the indicators predicted the addiction factor directly fit the data better via a nested comparison. A model with the substance use variables predicting addition-*rf* fit the data poorly ( $X^2=324.481$ ,  $df=10$ ,  $P=1.025754e-63$ ,  $AIC=360.481$ ,  $CFI=0.9040932$ ).

## Q-Trait Analysis

*Background. Background.* Q-Trait analysis<sup>1</sup> was developed to test whether a latent genetic factor estimated using gSEM mediates the genetic association between an external trait (i.e., one not loaded onto the latent genetic factor) and factor indicators (those loading on the latent genetic factor). We conducted Q-trait analysis to test whether *The Addiction-Risk-Factor* accounts for genetic associations between these traits and each individual SUD indicator loading onto it (i.e., OUD, PAU, PTU, CUD). Q-Trait analysis evaluates this by comparing one model in which only the latent genetic factor is associated with the external trait to another model in which each indicator (i.e., the trait loading onto the genetic factor) is associated with the external trait. A  $\chi^2$  difference between these models is calculated and if there is a significant difference, it suggests that the genetic association between the external trait and the factor indicators are not accounted for by the latent genetic factor. However, this analysis is conducted with all factor indicators in a single model independent of the latent factor, and does not model scenarios in which the latent factor may differentially mediate (e.g., fully, partially, or not at all) individual factor indicators. Because of this limitation, in cases where we found no evidence that the latent genetic factor completely mediated the genetic association between the external trait and factor indicators (i.e., when there was a significant  $\chi^2$  difference), we conducted a series of post-hoc models regressing the latent genetic factor and each individual genetic indicator on the external traits in an iterative processes (For example, a model in which the latent genetic *Addiction-Risk-Factor* and problematic tobacco use were regressed on executive function).

*Results.* Q-Trait analysis revealed that the latent genetic *Addiction-Risk-Factor* mediated associations between risk-taking and individual SUD indicators, as indicated by a lack of significance in model fit ( $X^2diff(3) = 6.478, P=0.091$ ). Although initial Q-trait modeling provided no evidence that the latent genetic *Addiction-Risk-Factor* accounted for associations between individual SUD indicators with executive function ( $X^2diff(3) = 25.7, P = 1.00e-5$ ) and neuroticism ( $X^2diff(3) = 625.7, P=1.00e-5$ ), post-hoc analyses revealed that the lack of evidence for mediation is attributable to differential mediation across indicators. Executive function is correlated with *The Addiction-Risk-Factor*, but shows stronger associations with problematic alcohol use, cannabis use disorder, and opioid use disorder than problematic tobacco use (see **Supplemental Table 1**). Because of this we tested an additional model wherein the latent genetic *Addiction-Risk-Factor* and problematic tobacco use were regressed on executive function simultaneously, which revealed no significant difference from the model in which all individual SUD indicators were included ( $X^2diff(2) = 4.635, P = .0985$ ), suggesting that the common latent *Addiction-Risk-Factor* mediates associations between executive function and PAU, OUD, and CUD, but not PTU. Further, additional regression path greatly bolstered upward the association between *The Addiction-Risk-Factor* and executive function ( $r=-.31, p = 1.00e-17$ ). For neuroticism, the addition of specific paths with PAU and CUD in addition to *The Addiction-Risk-Factor* revealed evidence that the latent genetic factor mediates associations with OUD and PTU ( $X^2diff(1) = 2.798, P=0.0944$ ). However, this did not change the association with the latent factor ( $r=.20, p=7.08e-3$ ); it is simply that neuroticism has associations with PAU and CUD that are independent of *The Addiction-Risk-Factor*.

*Interpretation for downstream analysis.* Downstream analysis did not retain these specific associations. For executive functioning, we still found significant associations between cEF and addiction risk, and mediating pathways despite the association being attenuated downward by our chosen model. Neuroticism did not maintain mediating pathways, and so

improving the model would not have led to additional discovery power for the neuroticism phenotype.

**Table 1. GWAS summary statistics used in this study.**

		N (EUR)	SNP-h2 (SE)	Link
Substance use disorders				
Problematic alcohol use (PAU)	Meta-analysis of DSM-IV alcohol dependence, ICD9/10 coded alcohol use disorder, and problem subscale scores from the AUDIT	435,563	0.068(0.004)	PMID: 32451486
Problematic Tobacco Use (PTU)	MTAG analysis combining FTND GWAS with GWAS of cigarettes per day	270,120	0.076(0.007)	<a href="#">PMC6358542</a>
Cannabis Use Disorder (CUD)	Meta-analysis of DSM-IV and ICD9/10 coded cannabis abuse and dependence	357,806	.120(0.010)	PMC5882602
Opioid Use Disorder (OUD)	Meta-analysis of ICD9/10 and DSM-IV	82,707	0.113(0.018)	<a href="#">PMC7270886</a>

	coded opioid use disorder			
Substance Use				
Alcohol drinks per week	Meta-analysis of self-reported drinks per week by GSCAN	537,349	0.042(0.002)	<a href="#">PMC6358542</a>
Ever smoking cigarettes	Meta-analysis of self-reported ever smoking by GSCAN	632,802	.078(0.002)	<a href="#">PMC6358542</a>
Cannabis use	Meta-analysis of any lifetime cannabis use	184,765	0.11(0.001)	<a href="#">PMC6386176</a>
Stage-based behavioral correlates				
Risk-taking	GWAS of self-reported risk-taking	328,339	0.055(0.003)	PMC6123450
Executive function	Common Factor of multiple executive functioning tasks	427,037	0.104(0.002)	<a href="https://www.biorxiv.org/content/10.1101/674515v2.full.pdf">https://www.biorxiv.org/content/10.1101/674515v2.full.pdf</a>
Neuroticism	GWAS meta-analysis of trait neuroticism	390,278	0.100(0.003)	PMID: 29942085
Other Psychiatric Disorders				

ADHD	GWAS meta-analysis of ADHD diagnosis from the PGC	53,293	0.222(0.014)	<a href="#">PMC6481311</a>
Anorexia Nervosa	GWAS meta-analysis of Anorexia Nervosa from the PGC	72,517	0.110(0.01)	<a href="#">PMC6779477</a>
Autism	GWAS meta-analysis of Autism from the PGC	46,350	.113(0.010)	<a href="#">PMC6454898</a>
Bipolar Disorder	GWAS meta-analysis of Bipolar disorder diagnosis from the PGC	51,710	0.182(0.011)	<a href="#">PMC6956732</a>
MDD	GWAS meta-analysis of self-reported Major Depression Diagnosis from the PGC	480,359	0.089(0.003)	<a href="#">PMC6522363</a>
OCD	GWAS meta-analysis of Obsessive Compulsive Disorder	9,725	0.280(0.041)	<a href="#">PMC6660151</a>

	Diagnosis from the PGC			
Schizophrenia	GWAS meta-analysis of Schizophrenia diagnosis from the PGC	77,096	0.222(0.012)	<a href="#">PMC4112379</a>
Tourette Syndrome	GWAS meta-analysis of Tourette Syndrome diagnosis from the PGC	13,340	0.200(0.026)	<a href="#">PMC6677250</a>

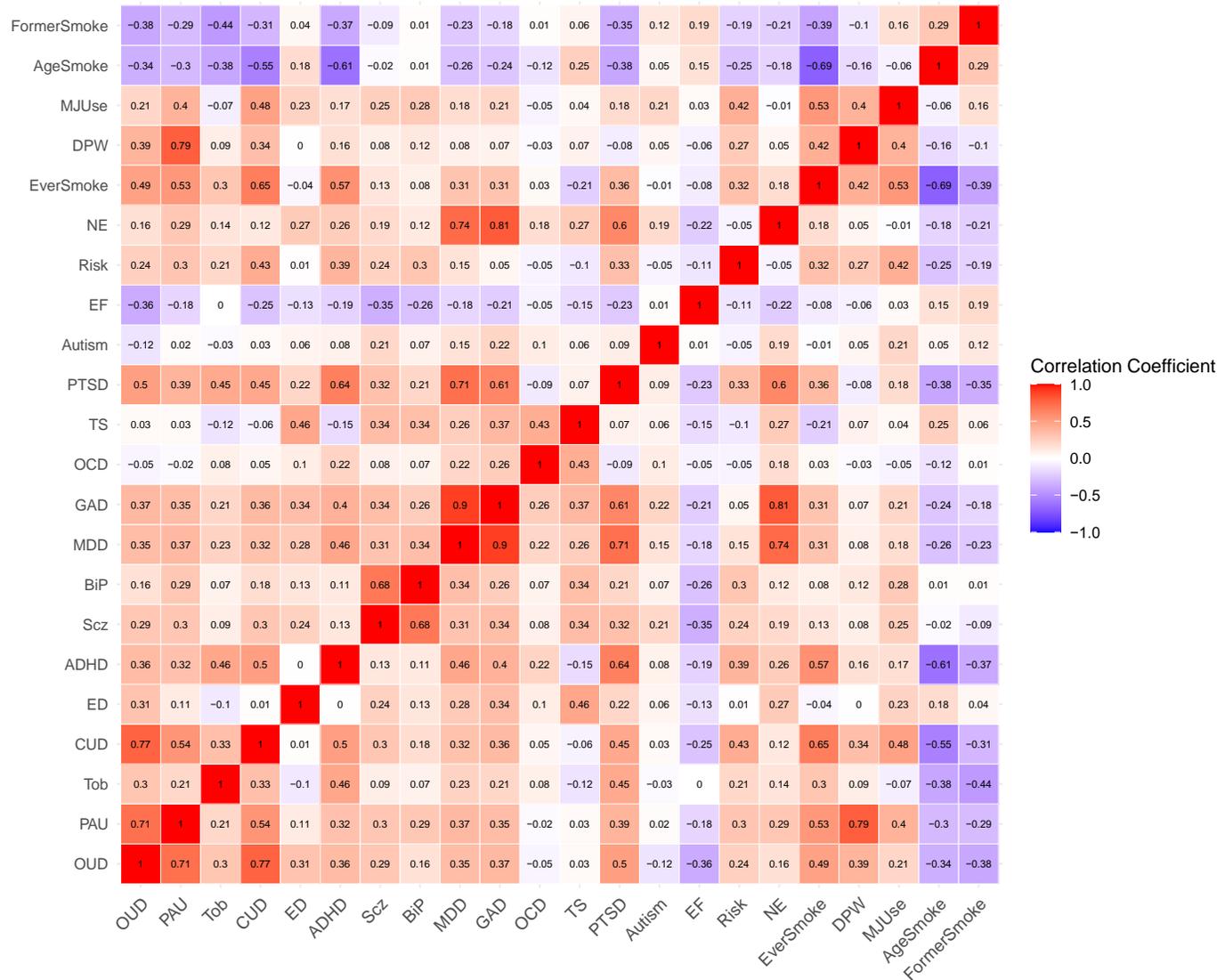
**Supplementary Table 2. Direct Associations between behavioral correlates and the Addiction Risk Factor – *Addiction-rf* - beyond non-substance psychopathology**

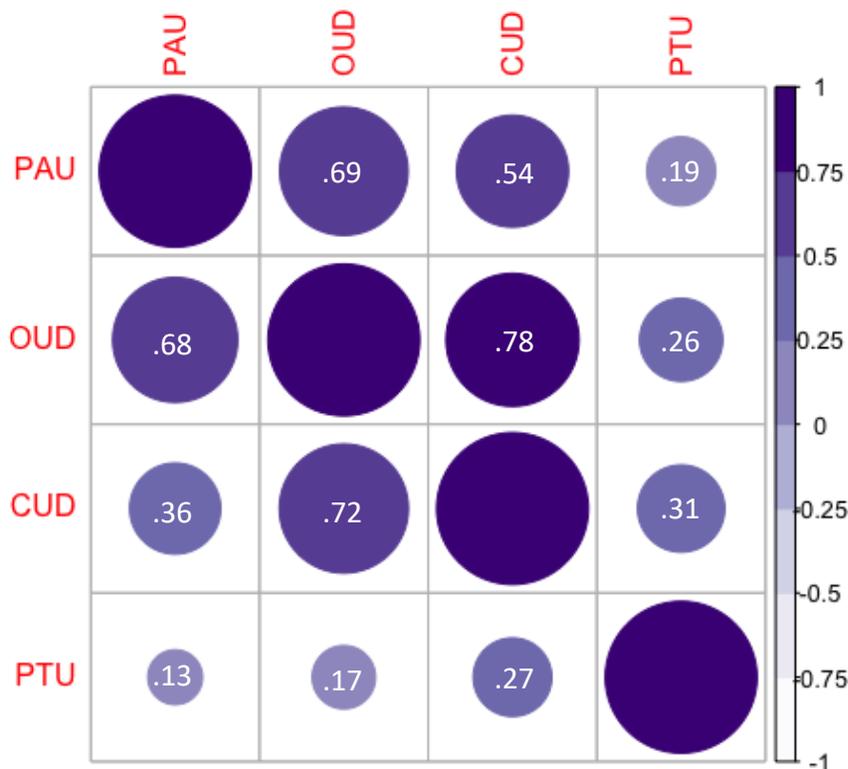
	<b>Direct Beta on <i>Addiction-rf</i></b>	<b>P-value</b>
<b>No covariates</b>		
Executive Function	<b>-0.101*</b>	0.004
Neuroticism	-0.009	0.874
Risk-taking	<b>0.261*</b>	<b>1.60e-8</b>
<b>Controlling for Genetics of Substance Use</b>		
Executive Function	<b>-0.122*</b>	<b>0.018</b>
Neuroticism	0.048	0.578
Risk-taking	0.057	0.307

**Supplementary Table 2. Behavioral Liabilities Mediate the Association Between -Factor and Psychiatric Factors**

Indirect associations from a mediation model (see Figure 4) where stage-based constructs link non-substance psychopathology (three factors from Lee et al.,) and *The Addiction-Risk Factor (Addiction-rf)*. Indirect association is represented by the proportion accounted for by the indirect association (%) and the significance of the indirect association. F1 = compulsive disorders, F2 = Psychotic disorders, F3 = Neurodevelopmental disorders.

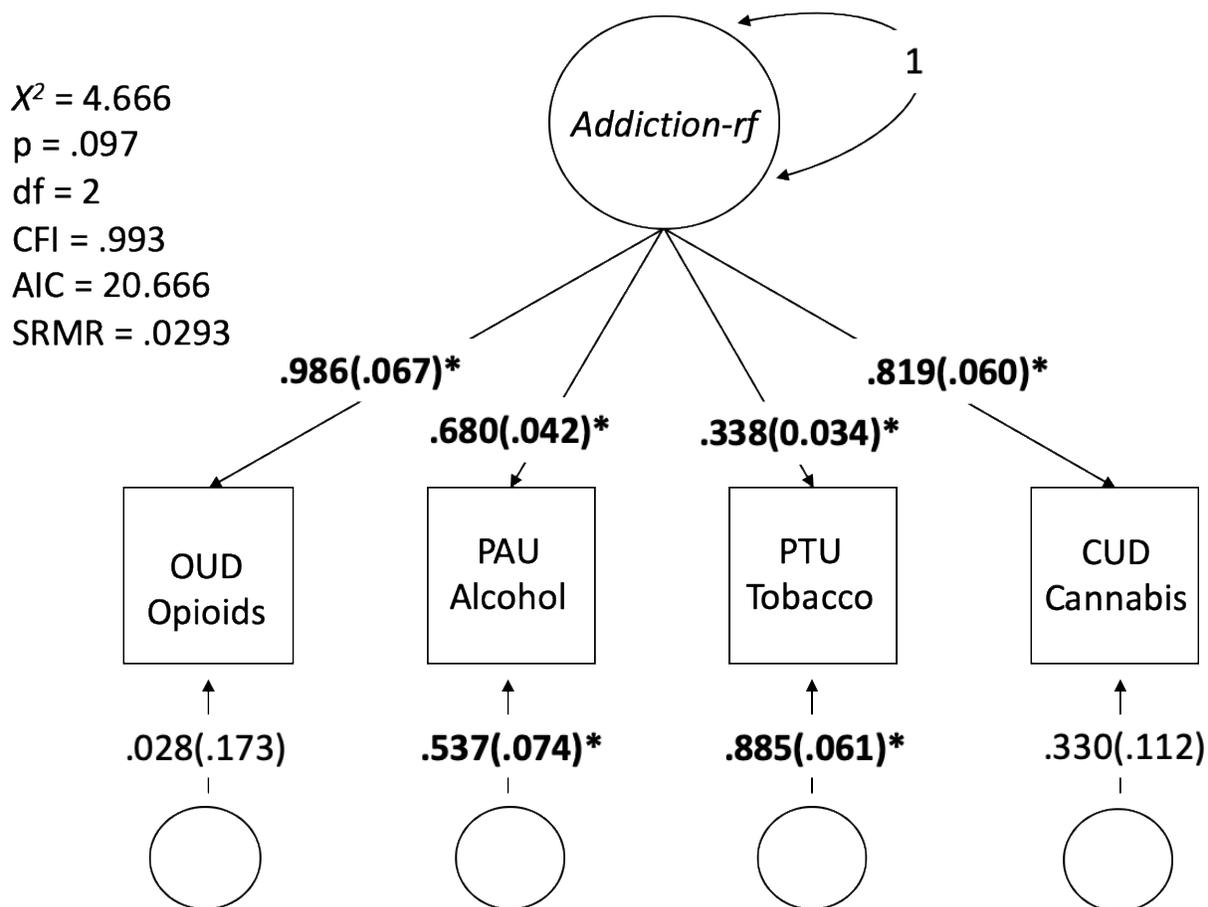
**Supplementary Figure 1.** Genetic correlations (SNP-rG) were estimated between all variables. The upper diagonal represents the bivariate associations from LD Score regression. PAU = Problematic Alcohol Use, OUD = Opioid Use Disorder, CUD = Cannabis Use Disorder, PTU = Problematic Tobacco Use, FormerSmoke = smoking cessation, AgeSmoke=Age first smoking regularly, MJUse = Marijuana Use, DPW = Drinks Per Week, EverSmoke = Ever Smoking regularly, NE = neuroticism, Risk = Risk-taking, EF = Executive Function, PTSD = Post-Traumatic Stress Disorder, TS = Tourette’s Syndrome, OCD = Obsessive Compulsive Disorder, GAD = General Anxiety Disorder, MDD = Major Depressive Disorder, BiP = Bipolar disorder, Scz = Schizophrenia, ADHD =



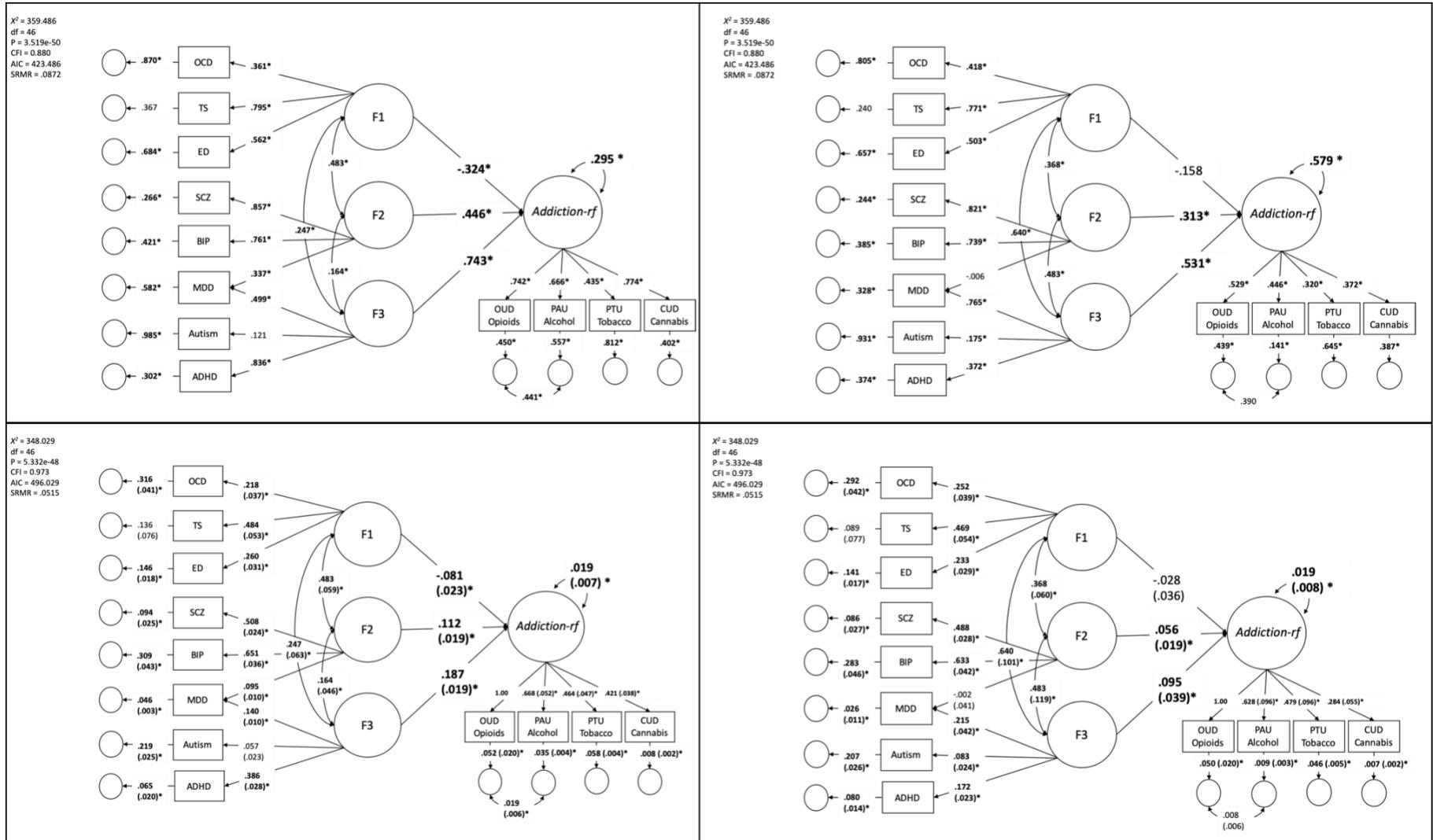


**Supplementary Figure 2. Genetic Correlation Matrix Between Substance Use Disorders.**

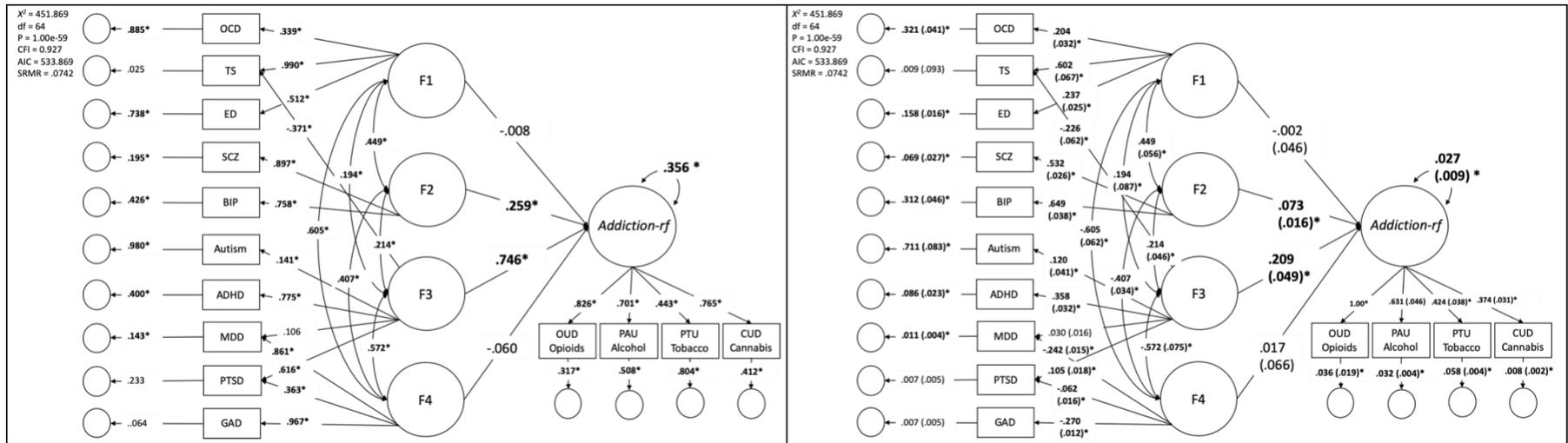
Genetic correlations (SNP-rG) were estimated between all substance use disorder categories. The upper diagonal represents the bivariate associations from LD Score regression. The lower diagonal are correlations estimated in Genomic Structural Equation Modeling when controlling for common substance use (drinks per week, ever smoking marijuana or tobacco). All genetic correlations were positive (shown in purple). PAU = Problematic Alcohol Use, OUD = Opioid Use Disorder, CUD = Cannabis Use Disorder, PTU = Problematic Tobacco Use. All rGs were significant (all  $p > .01$ ).



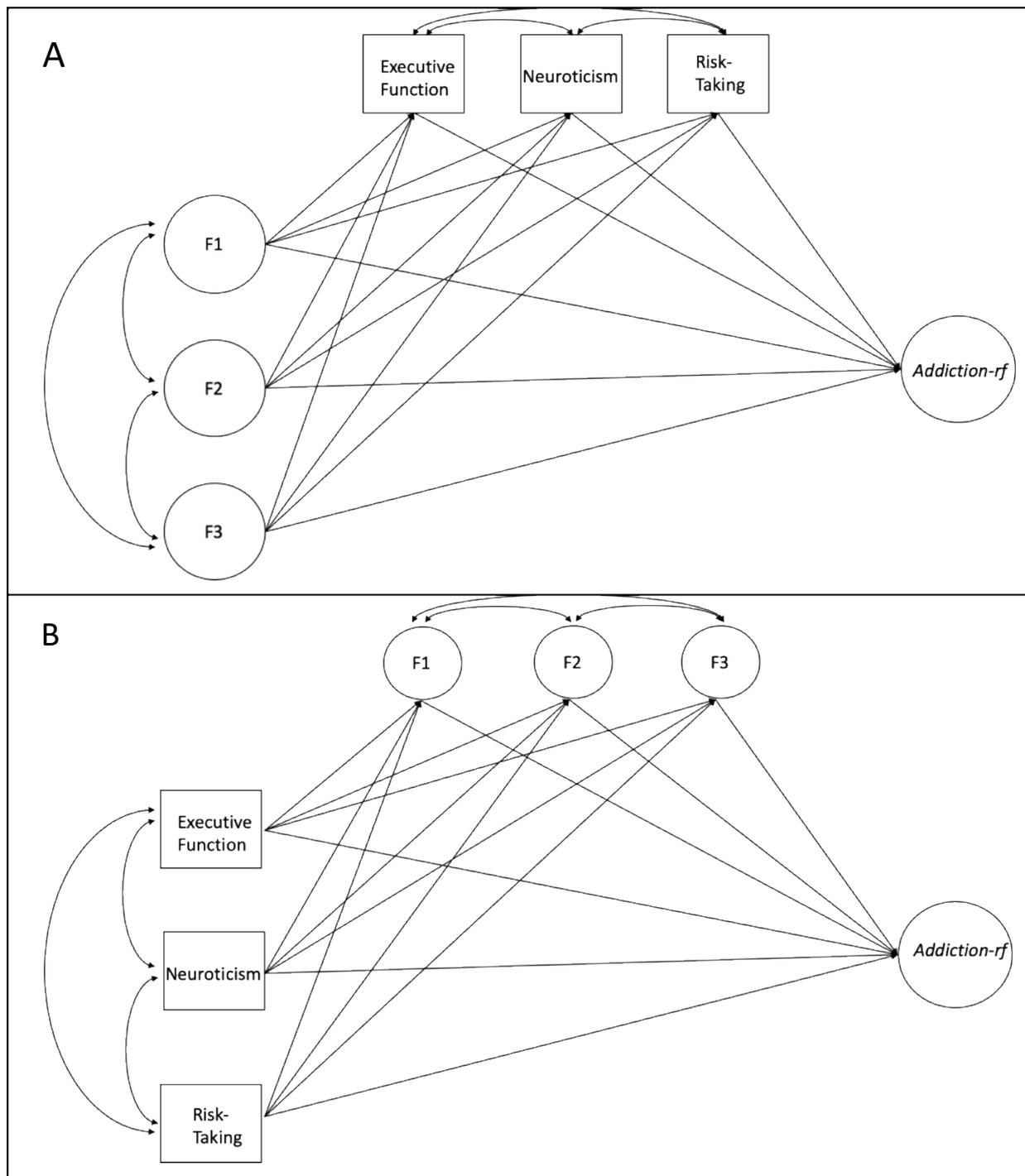
**Supplementary Figure 3.** Alternative specification of one factor model, without residual correlation between problematic alcohol use and opioid use disorder.



**Supplemental Figure 4. Full model specification for *Addiction-rf* association with non-substance psychopathology.** The top two panels represent the standardized estimates, the bottom two are unstandardized with standard errors. The left panels include the substance use variables as covariates. **Bold\***= p-value < .05.



**Supplementary Figure 5.** Alternative model of psychopathology, taken from Grotzinger et al.<sup>2</sup> the left panel represents the model with standardized estimates. The right panel represents the unstandardized estimates with standard errors. **Bold\***= p-value < .05.

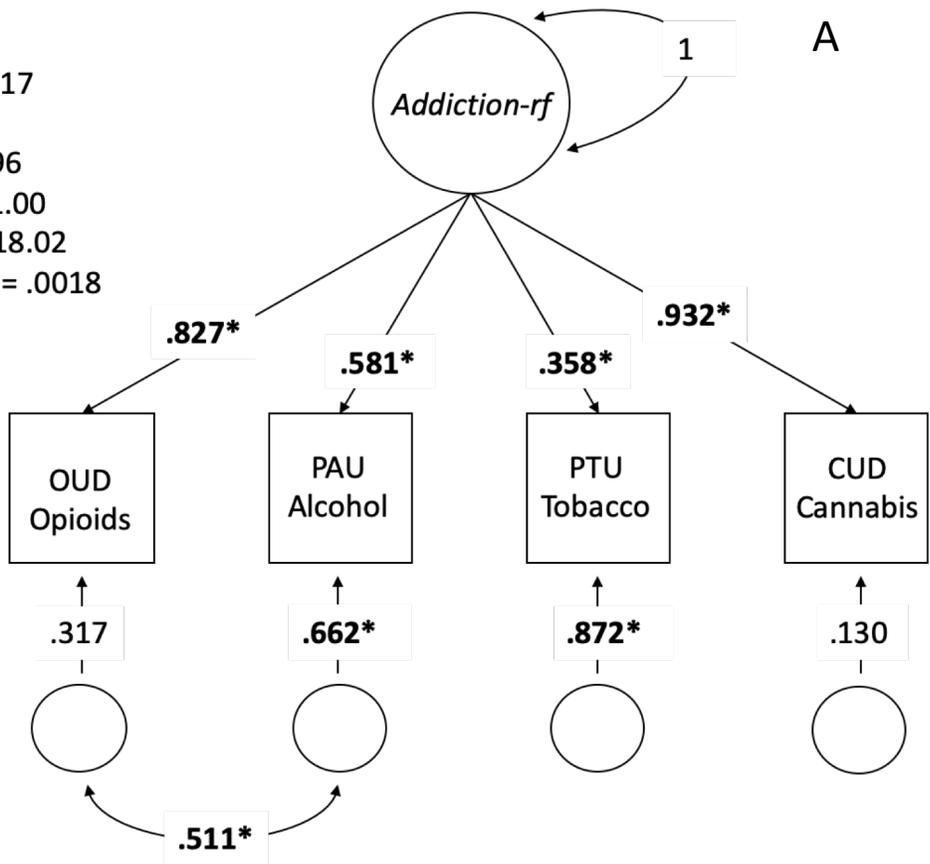


**Supplemental Figure 6. Model Specification for Mediation Models Linking non-substance psychopathology and *Addiction-rf*.** In panel A, we tested whether neurobiological stage-based constructs mediated the association between the *Addiction-rf*-Factor and non-substance psychopathology. In mediation model two, we tested whether neurobiological stage-based constructs had remaining (direct) associations with *Addiction-rf* beyond non-substance psychopathology variables.

## References

1. Walters, R. K. *et al.* Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nat. Neurosci.* **21**, 1656–1669 (2018).
2. Grotzinger, A. D. *et al.* Genetic architecture of 11 major psychiatric disorders at biobehavioral, functional genomic, and molecular genetic levels of analysis. *medRxiv* vol. 18 18–19 (2020).

$\chi^2 = .017$   
 $df = 1$   
 $P = .896$   
 $CFI = 1.00$   
 $AIC = 18.02$   
 $SRMR = .0018$



$\chi^2 = .474$   
 $df = 1$   
 $p = .491$   
 $CFI = 1.00$   
 $AIC = 54.474$   
 $SRMR = 0.0037$

