

Word Count: 3386

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26

## Epigenetic and Genetic Profiling of Comorbidity Patterns among Substance Dependence Diagnoses

Gita A. Pathak<sup>1,2</sup>, Robert H. Pietrzak<sup>1,2,3</sup>, AnnMarie Lacobelle<sup>1,2</sup>, Cassie Overstreet<sup>1,2</sup>, Frank R. Wendt<sup>1,2,4</sup>, Joseph D. Deak<sup>1,2</sup>, Eleni Friligkou<sup>1,2</sup>, Yaira Nunez<sup>1,2</sup>, Janitza L. Montalvo-Ortiz<sup>1,2</sup>, Daniel F. Levey<sup>1,2</sup>, Henry R. Kranzler<sup>5</sup>, Joel Gelernter<sup>1,2</sup>, Renato Polimanti<sup>1,2</sup>

<sup>1</sup>Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

<sup>2</sup>U.S Department of Veteran Affairs Connecticut Healthcare System, West Haven, CT, USA

<sup>3</sup>Department of Social and Behavioral Sciences, Yale School of Public Health, New Haven, CT, USA

<sup>4</sup>Regeneron Genetics Center, Regeneron Pharmaceuticals, Tarrytown, NY, USA

<sup>5</sup> Center for Studies of Addiction, University of Pennsylvania Perelman School of Medicine and the Mental Illness Research, Education and Clinical Center, Crescenz VAMC, Philadelphia PA, USA

**Keywords:** Latent class, Methylation, Opioid, Nicotine, Cannabis, Alcohol, Nicotine, Substance Use Disorder, Addiction

**Correspondence:** Renato Polimanti, PhD. Yale University School of Medicine, Department of Psychiatry, 60 Temple, Suite 7A, New Haven, CT 06510, USA. Phone: +1 (203) 737-6391; E-mail: [renato.polimanti@yale.edu](mailto:renato.polimanti@yale.edu)

27 **ABSTRACT**

28 *Objective*—This study investigated the genetic and epigenetic mechanisms underlying  
29 the comorbidity patterns of five substance dependence diagnoses (SDs; alcohol, AD;  
30 cannabis, CaD; cocaine, CoD; opioid, OD; tobacco, TD).

31 *Methods*—A latent class analysis (LCA) was performed on 31,197 individuals (average  
32 age 42±11 years; 49% females) from six cohorts to identify comorbid DSM-IV SD  
33 patterns. In subsets of this sample, we tested SD-latent classes with respect to  
34 polygenic burden of psychiatric and behavioral traits and epigenome-wide changes in  
35 three population groups.

36 *Results*—An LCA identified four latent classes related to SD comorbidities: AD+TD,  
37 CoD+TD, AD+CoD+OD+TD (i.e., polysubstance use, PSU), and TD. In the epigenome-  
38 wide association analysis, *SPATA4* cg02833127 was associated with CoD+TD, AD+TD,  
39 and PSU latent classes. AD+TD latent class was also associated with CpG sites located  
40 on *ARID1B*, *NOTCH1*, *SERTAD4*, and *SIN3B*, while additional epigenome-wide  
41 significant associations with CoD+TD latent class were observed in *ANO6* and *MOV10*  
42 genes. PSU-latent class was also associated with a differentially methylated region in  
43 *LDB1*. We also observed shared polygenic score (PGS) associations for PSU, AD+TD,  
44 and CoD+TD latent classes (i.e., attention-deficit hyperactivity disorder, anxiety,  
45 educational attainment, and schizophrenia PGS). In contrast, TD-latent class was  
46 exclusively associated with posttraumatic stress disorder-PGS. Other specific  
47 associations were observed for PSU-latent class (subjective wellbeing-PGS and  
48 neuroticism-PGS) and AD+TD-latent class (bipolar disorder-PGS).

49 *Conclusions*—We identified shared and unique genetic and epigenetic mechanisms  
50 underlying SD comorbidity patterns. These findings highlight the importance of modeling  
51 the co-occurrence of SD diagnoses when investigating the molecular basis of addiction-  
52 related traits.

53

## 54 INTRODUCTION

55 Substance dependence (SD) is a significant public health concern, affecting more than  
56 40 million individuals in the United States <sup>1</sup>. The most prevalent SDs are related to  
57 alcohol (AD), cannabis (CaD), cocaine (CoD), opioid (OD), and tobacco (TD). Patients  
58 with SDs often exhibit misuse of more than one substance, which complicates treatment  
59 and recovery efforts <sup>2</sup>. The comorbidity of SDs exacerbates a range of negative health  
60 outcomes, making it critical to understand the patterns of SD comorbidities <sup>2,3</sup>. The  
61 heterogeneity of SD comorbidity patterns poses a major challenge in identifying the  
62 mechanistic underpinnings of these disorders. SD patterns can stem from a range of  
63 factors, including shared genetic predisposition, environmental influences, and  
64 sociocultural context. Genome-wide association studies (GWAS) and epigenome-wide  
65 association studies (EWAS) uncovered mechanisms contributing to the predisposition to  
66 SDs <sup>4-10</sup>. To date, these previous efforts largely focused on single SDs, limiting our  
67 understanding of the real world in which patients are generally affected by multiple SDs.

68 This study identified comorbidity patterns among AD, CaD, CoD, OD, and TD using a  
69 latent class analysis (LCA, a method that aims to identify more homogeneous  
70 subgroups in heterogeneous data). Then, we investigated SD-latent classes with  
71 respect to methylation changes and polygenic burdens related to psychiatric and  
72 behavioral traits (**FIGURE 1**). The findings shared and unique epigenetic and genetic  
73 profiles that underlie SD comorbidity patterns.

74

## 75 METHODS

### 76 *Cohorts*

77 We investigated six cohorts with information regarding DSM-IV diagnoses of AD, CaD,  
78 CoUD, OD, and TD (**TABLE 1**). In addition to the Yale-Penn cohort <sup>11-15</sup>, we analyzed five  
79 datasets available from NCBI's Database of Genotypes and Phenotypes (dbGaP):  
80 'Study of Addiction: Genetics and Environment' [phs000092] <sup>16</sup>, 'Genome-Wide  
81 Association Study of Heroin Dependence' [phs000277] <sup>17</sup>, 'Genetic Architecture of  
82 Smoking and Smoking Cessation' [phs000404] <sup>18</sup>, 'CIDR, NCI, NIDA Sequencing of

83 Targeted Genomic Regions Associated with Smoking' [phs000813]<sup>18</sup>, and 'Nicotine  
84 Addiction Genetics and Correlates' [phs001299]<sup>19</sup>. Overall, we analyzed SD data  
85 collected from 31,197 individuals (average age 42±11 years; 49% females).

86

### 87 *Latent Class Analysis*

88 LCA was performed in R version 4.1 using the poLCA R package<sup>20</sup> for the primary  
89 investigation and Mplus software<sup>21</sup> for validation. The number of latent classes was  
90 identified based on the lowest Akaike Information Criterion (AIC), Bayesian Information  
91 Criterion (BIC), Likelihood ratio/deviance statistic ( $G^2$ ), and Pearson Chi-square  
92 goodness of fit statistic ( $\chi^2$ ). The probability threshold to assign participants with SDs in  
93 classes was identified using MANOVA by comparing the probability of SD cases across  
94 the latent classes identified in the best-fitting model.

95

### 96 *Genetic Data*

97 Genotype data from the six studies was cleaned by removing individuals with  
98 mismatched biological sex, low genotyping rate, heterozygosity, and relatedness. Single  
99 nucleotide polymorphisms (SNPs) were removed based on minor allele frequency  
100 (MAF<1%), Hardy-Weinberg equilibrium ( $p<1\times 10^{-6}$ ), and sample missingness (<10%).  
101 Continental genetic ancestry was estimated against a combined reference panel  
102 including 1000 Genomes Project and Human Genome Diversity Project<sup>22</sup>. Genetic  
103 relatedness and within-ancestry principal components were generated using KING<sup>23</sup>.  
104 Genotype data from each study was imputed using TopMed Imputation server<sup>24</sup>.  
105 Because large-scale psychiatric/behavioral GWAS are present only for populations of  
106 European descent, we limited the PGS analysis to this population group. After quality  
107 control (QC), genetic data from 7,659 individuals of European descent were available  
108 for PGS testing.

109

### 110 *Epigenetic Data*

111 DNA methylation data were available for a subset of the Yale-Penn participants (n=886).  
112 Briefly, DNA was extracted from whole blood of Yale-Penn participants collected at the  
113 time of recruitment using Paxgene Blood DNA Kit (Qiagen, MD, USA). Bisulfite  
114 conversion of the extracted DNA was performed using the EZ-96 DNA methylation kit  
115 (Zymo Research, CA, USA) and subsequently processed on the Infinium®  
116 MethylationEPIC BeadChip array (Illumina, San Diego, CA USA) which assesses the  
117 methylation status of over 850,000 CpG sites. Raw signal intensity data (IDAT) files  
118 from the array were uploaded and processed in R Studio. These files provided beta  
119 values, indicating DNA methylation levels at specific CpG sites, with values ranging  
120 from 0 (unmethylated) to 1 (fully methylated). Primary QC steps removed CpG sites with  
121 low detection p-values, missing beads, close proximity to SNPs, multi-hit sites, and non-  
122 autosomal sites. The beta values of the remaining CpG sites were normalized using the  
123 beta mixture quantile (BMIQ) method<sup>25</sup>. Batch effects were identified using singular  
124 vector decomposition. Technical batch effects related to array and slide were corrected  
125 using ComBat function in sva R package<sup>26</sup>. DNA Methylation data were used to  
126 estimate blood cell-type compositions (i.e., CD8+T, CD4+T, natural killer, B, monocytes,  
127 and neutrophils)<sup>27</sup>, and smoking status<sup>28</sup>.

128

### 129 *Epigenome-wide association analysis*

130 After QC, we tested 657,226 CpG sites for differential methylation with respect to SD-  
131 latent classes in three population groups (European-descent, EUR, n=481; African-  
132 descent, AFR n=339; Admixed-Americans, AMR n=66). Association analysis for CpG  
133 sites was performed on M-values (transformed beta values) using empirical Bayes  
134 methods implemented in the limma R package<sup>29</sup>. The analysis was adjusted for age,  
135 sex, genotype-derived principal components 1-10, methylation-based smoking score<sup>28</sup>,  
136 and proportions of blood cell types (i.e., CD8+T, CD4+T, natural killer, B, monocytes,  
137 and neutrophils). Inflation was calculated using the QQperm R package<sup>30</sup>. For the  
138 ancestry-stratified EWAS of each SD-latent class, the minimum sample size was 15.  
139 The meta-analysis across all ancestries was performed using GWAMA<sup>31</sup> to improve  
140 statistical power. For each EWAS cross-ancestry meta-analysis, genomic control

141 correction was applied to epigenome-wide association statistics when lambda was  
142 >1.05. Differentially methylated regions were identified using the dmrff R package<sup>32</sup>.  
143 Regions were defined as consisting of 2 to 30 CpG sites within 500 base pairs. False  
144 discovery rate (FDR  $q < 0.05$ ) was used to adjust for multiple testing. The CpG sites were  
145 further investigated using information available from the EWAS Catalogue<sup>33</sup> (available  
146 at <https://www.ewascatalog.org/>). We also assessed brain-blood concordance of the  
147 CpG sites identified using BECon application<sup>34</sup>.

148

### 149 *Polygenic Score Analysis*

150 PGSs were derived from GWAS of attention deficit/hyperactivity disorder (ADHD;  
151  $N=225,534$ )<sup>35</sup>, anxiety ( $N=1,096,458$ )<sup>36</sup>, autism spectrum disorder ( $N=46,350$ )<sup>37</sup>,  
152 bipolar disorder ( $N=413,466$ )<sup>38</sup>, depression ( $N=1,035,760$ )<sup>39</sup>, educational attainment  
153 ( $N=765,283$ )<sup>40</sup>, neuroticism ( $N=380,000$ )<sup>41</sup>, posttraumatic stress disorder (PTSD;  
154  $N=1,222,882$ )<sup>42</sup>, schizophrenia ( $N=320,404$ )<sup>43</sup>, and subjective well-being ( $N=298,420$ )  
155<sup>44</sup>. The polygenic scores were calculated using PRS-CS<sup>45</sup>, with the 1000 Genomes  
156 reference panel for linkage disequilibrium, scaled to have mean of 0 and a unit of 1  
157 standard deviation. In each cohort, PGSs were tested with respect to SD-latent classes,  
158 including age, sex, and top-ten within-ancestry principal components as covariates. For  
159 each SD-latent class, cohort-specific PGS associations were meta-analyzed using meta  
160 R package<sup>46</sup>. Bonferroni-corrected threshold of  $p < 1.25 \times 10^{-3}$  was used to adjust for the  
161 number of PGS association tests performed ( $n=40$ ).

162

## 163 **RESULTS**

### 164 *Latent class analysis of SDs*

165 With respect to comorbidity patterns among DSM-IV based AD, CaD, CoD, OD and TD,  
166 the model with five latent classes was identified as the best fitting because of the lowest  
167 value combination of AIC, BIC,  $G^2$ , and  $\chi^2$  metrics (**SUPPLEMENTARY TABLE 1**).

168 Considering a posterior probability  $\geq 70\%$  (**FIGURE 2; SUPPLEMENTARY FIGURE 1**), we  
169 stratified the sample of 31,197 subjects across the five SD-latent classes. Because of

170 the high case-posterior probability for AD and TD (97% and 75%, respectively), 6,487  
171 participants were included in the AD+TD-latent class. A total of 1,170 participants were  
172 assigned to the CoD+TD-latent class, because of the high case-posterior probability of  
173 these SDs (85% and 70%, respectively). We also observed 2,090 participants related to  
174 a polysubstance use disorder (PSU)-latent class, which had high case-posterior  
175 probabilities for AD (95%), CoD (99%), OD (100%), and TD (96%). An additional latent  
176 class including 1,162 individuals showed high posterior probability only for TD (100%).  
177 Finally, there were 11,759 participants assigned to a latent class with high control-  
178 posterior probability for all SDs (i.e., 89% AD, 80% CaD, 94% CoD, 72%TD, 99% OD).  
179 CaD did not show high case-posterior probability in any of the SD-latent classes  
180 **(SUPPLEMENTARY TABLE 2)**.

181

### 182 *Epigenome-wide association study of SD latent classes*

183 Relative to the control-latent class group, we performed cross-ancestry EWAS meta-  
184 analyses of AD+TD, CoD+TD, PSU-latent classes in 886 individuals, identifying seven  
185 CpG associations (**FIGURE 3; SUPPLEMENTARY FIGURE 2; SUPPLEMENTARY TABLES 3-6**).  
186 TD-latent class was excluded from this analysis because of the small sample size. After  
187 FDR multiple testing correction (FDR  $q < 0.05$ ), *SPATA4* cg02833127 (location: 1<sup>st</sup> exon,  
188 5'UTR) was associated with the three SD-latent classes investigated (i.e., AD+TD  $\beta =$   
189  $0.49$ ,  $P = 9.9 \times 10^{-50}$ ; CoD-TD  $\beta = -0.3$ ,  $P = 2.93 \times 10^{-11}$ ; PSU  $\beta = -0.24$ ,  $P = 3.19 \times 10^{-11}$ ). Four  
190 additional CpG sites were associated with the AD+TD-latent class: *SIN3B* cg06815056  
191 (location: TSS200;  $\beta = 0.23$ ,  $P = 2.91 \times 10^{-7}$ ), *ARID1B* cg19436567 (location: 1<sup>st</sup> exon;  
192  $\beta = 0.22$ ,  $P = 1.47 \times 10^{-7}$ ), *SERTAD4* cg20270863 (location: TSS200, 5'UTR;  $\beta = 0.16$ ,  
193  $P = 3.42 \times 10^{-7}$ ) and *NOTCH1* cg13725899 (location: gene body;  $\beta = 0.16$ ,  $P = 8.69 \times 10^{-8}$ ).  
194 Two additional FDR-significant CpG associations were detected with respect to  
195 CoD+TD latent class: *MOV10* cg08355353 (Location gene body;  $\beta = 0.43$ ,  $P = 1.08 \times 10^{-7}$ )  
196 and *ANO6* cg09909775 (Location: 1stExon, 5'UTR;  $\beta = 0.50$ ;  $P = 2.03 \times 10^{-7}$ ). Considering  
197 information available from the EWAS catalog (**SUPPLEMENTARY TABLE 7**), these CpG  
198 sites were previously identified in the context of aging (cg08355353, cg06815056,  
199 cg13725899, cg19436567, and cg02833127), rheumatoid arthritis (cg06815056),

200 Alzheimer's disease (cg13725899), renal carcinoma (cg20270863), and molecular  
201 regulation (cg06815056, cg13725899, and cg19436567).

202 We observed cross-ancestry heterogeneity in the association between *SPATA4*  
203 cg02833127 and AD+TD-latent class ( $I^2=88\%$ , heterogeneity  $p < 2 \times 10^{-4}$ ) where the  
204 effect was more negative in EUR ( $\beta = -0.60$ , 95%CI=-0.67 – -0.52) than in AFR ( $\beta = -$   
205 0.33, 95%CI=-0.43 – -0.23). Conversely, no cross-ancestry heterogeneity was observed  
206 in *SPATA4* cg02833127 association with CoD+TD and PSU-latent classes and also for  
207 the other FDR-significant CpG associations (heterogeneity- $p > 0.05$ ; **FIGURE 3**).

208 In addition to CpG association, we also observed one FDR-significant differentially  
209 methylated region (chr10:102120371-102120478;  $Z=5.466$ ;  $P= 4.60 \times 10^{-8}$ ) related to  
210 PSU-latent class, spanning 2 CpG sites (cg15320455 – TSS200; cg17106419 –  
211 TSS200) within *LDB1* gene (**SUPPLEMENTARY TABLE 8**).

212 Comparing brain-blood methylation patterns of the identified CpG sites  
213 (**SUPPLEMENTARY FIGURE 2**), we observed correlation estimates in the top-90<sup>th</sup>  
214 percentile for *MOV10* cg08355353 ( $\rho=0.36$  in the Brodmann area (BA) 7), *SERTAD4*  
215 cg20270863 ( $\rho=0.38$  in BA20), *SIN3B* cg06815056 ( $\rho=-0.45$  in BA20), *SPATA4*  
216 cg02833127 ( $\rho=-0.44$  in BA7), and *ARID1B* cg19436567 ( $\rho=0.49$  in BA10;  $\rho=-$   
217 0.35 in BA20)

218

### 219 *Polygenic burden associations with SD-latent classes*

220 Assessing the pleiotropy across SD-latent classes, we observed different patterns of  
221 psychiatric-behavioral PGS associations (**FIGURE 4**; **SUPPLEMENTARY TABLE 9**). After  
222 Bonferroni multiple testing correction ( $p < 1.25 \times 10^{-3}$ ), PTSD was the only PGS  
223 associated across the four SD-latent classes (**SUPPLEMENTARY TABLE 9**). TD-latent  
224 class was only associated with PTSD PGS (Odds ratio, OR=1.32,  $p=7.24 \times 10^{-4}$ ). In  
225 contrast, PSU, AD+TD, and CoD+TD latent classes showed shared PGS associations  
226 with ADHD (positive relationship), anxiety (positive relationship), educational attainment  
227 (inverse relationship), and schizophrenia (positive effect) with comparable effect sizes  
228 (**SUPPLEMENTARY TABLE 9**). In addition to shared pleiotropic mechanisms, we also



229 observed specific PGS associations. PSU-latent class was uniquely associated with  
230 PGS of neuroticism ( $OR=1.21$ ;  $P=2.62 \times 10^{-6}$ ) and subjective wellbeing ( $OR=0.86$ ;  
231  $P=2.12 \times 10^{-4}$ ) with opposite effect directions. AD+TD-latent class was uniquely  
232 associated with bipolar disorder-PGS ( $OR=1.14$ ;  $P=1.39 \times 10^{-5}$ ).

233 Because PGS associations were estimated from the meta-analysis of cohorts with  
234 different characteristics (**TABLE 1**), we also estimated cross-cohort heterogeneity within  
235 the PGS effects. After Bonferroni correction, there was significant heterogeneity  
236 (heterogeneity  $Q-P < 1.25 \times 10^{-3}$ ) in the association between PTSD PGS and AD+TD-  
237 latent class ( $I^2=0.89$  ;  $\tau^2=0.05$ ;  $Q=28.4$ ,  $Q-df=3$ ,  $Q-p=3 \times 10^{-6}$ ) and between  
238 educational attainment PGS and TD-latent class ( $I^2=0.84$ ;  $\tau^2=0.12$ ;  $Q=18.9$ ,  $Q-df=2$ ,  
239  $Q-P=2.81 \times 10^{-4}$ ). In the first, the heterogeneity was driven by two-studies, phs000404  
240 and phs001299 (**SUPPLEMENTARY TABLES 9 AND 10**). The meta-analyzed effect  
241 association between educational attainment PGS and TD-latent class was only  
242 nominally significant ( $OR=0.82$ ,  $P=0.01$ ) and its Bonferroni-significant heterogeneity was  
243 driven by two studies, Yale-Penn and phs000092 (**SUPPLEMENTARY TABLES 9 AND 10**).

244

## 245 **DISCUSSION**

246 SDs have a pervasive negative impact on individuals, families, and society at large.  
247 More severe consequences are experienced by individuals with co-occurring SDs,  
248 which represent the largest portion of affected individuals<sup>47</sup>. To date, most genomic  
249 studies modeled singular SD cases without considering co-occurring SDs<sup>9, 10</sup>. Recently,  
250 large-scale GWAS datasets have been used to investigate the shared pathogenesis  
251 across SDs and related phenotypes<sup>48, 49</sup>. While the findings of these studies increased  
252 our understanding of the molecular mechanisms responsible for polysubstance  
253 comorbidities, more efforts are needed to dissect the dynamics underlying different  
254 patterns of SD co-occurrence. Applying LCA, previous observational studies of  
255 nationally representative samples, treatment-seeking populations, and cohorts enriched  
256 for SD cases identified latent classes reflecting different polysubstance patterns<sup>50-54</sup>.  
257 Building on this evidence, we identified four SD-latent classes in a sample of >30,000

258 individuals and then investigated genetic and epigenetic mechanisms associated with  
259 the different SD comorbidity patterns.

260 Investigating epigenome-wide differences, our study revealed several differentially  
261 methylated sites with shared and distinct associations across SD-latent classes.  
262 *SPATA4* cg02833127 was consistently hypomethylated among cases in the AD+TD,  
263 CoD+TD, and PSU latent classes. This CpG site has been associated with aging  
264 trajectories from birth to late adolescence with methylation decreasing later in life <sup>55</sup>.  
265 *SPATA4* locus has been also identified in a large-scale GWAS of educational  
266 attainment<sup>40</sup>. While the function of this locus is still unknown, a study observed aging-  
267 related changes in mice overexpressing this gene <sup>56</sup>. In this context, *SPATA4*  
268 cg02833127 association across SD-latent classes points to the possible impact of SD  
269 comorbidities on aging-related regulatory mechanisms. We observed cross-ancestry  
270 heterogeneity in the association of this CpG site with AD+TD latent class with greater  
271 effect in EUR than AFR. This supports that SD comorbidities may affect biological  
272 pathways differently among population groups. Alternatively, cross-ancestry  
273 heterogeneity of the association between *SPATA4* cg02833127 and AD+TD latent class  
274 may reflect the contribution of environmental factors acting differently across population  
275 groups.

276 The AD+TD-latent class was uniquely associated with four additional CpG sites. Among  
277 them, *NOTCH1* cg13725899 was previously identified by multiple EWAS<sup>57-60</sup>. Two of  
278 which were conducted using brain tissues and linked this CpG site with fetal brain  
279 development <sup>58</sup> and Alzheimer's disease <sup>60</sup>. Other *NOTCH1* CpG sites have been  
280 reported in EWAS of AD <sup>7</sup> and tobacco smoking <sup>61</sup>. The *NOTCH1* locus was also  
281 identified in GWAS of brain-related outcomes, including cortical thickness and  
282 Alzheimer's disease<sup>62, 63</sup>. The protein product of this gene plays an important role in a  
283 signaling pathway involved in neurodevelopmental processes<sup>64</sup>. *NOTCH1* may play a  
284 role in SD pathogenesis through its negative regulation of GABA receptors <sup>65</sup>. Also, the  
285 other genes mapping to AD+TD-associated CpG sites (i.e., *ARID1B* and *SIN3B*) appear  
286 to play important regulatory roles in the central nervous system<sup>66, 67</sup>. Additionally, there

287 were DNA methylation changes in *ARID1B* and *SIN3B* have been previously linked to  
288 cigarette smoking<sup>68-71</sup> and alcohol use<sup>72-74</sup>.

289 The CoD+TD latent class showed specific epigenetic associations in *ANO6* and  
290 *MOV10*. The latter locus has been linked to regulatory processes related to neuronal  
291 development and function<sup>75</sup>. Multiple GWAS identified *MOV10* in relation to cortical  
292 surface area, cortical thickness, and brain connectivity<sup>76-78</sup>. In contrast, *ANO6* gene  
293 does not appear to play a specific role in brain function, but is involved in phospholipid  
294 regulation occurring in various biological systems<sup>79</sup>. Previous EWAS identified other  
295 *ANO6* and *MOV10* CpG sites as associated with tobacco smoking<sup>80, 81</sup>.

296 In addition to identifying individual CpG sites, we also observed a differentially  
297 methylated region mapping to *LDB1* that was associated with the PSU-latent class. This  
298 locus appears to be required for the early development of the dorsal telencephalon and  
299 the thalamus<sup>82</sup>. Additionally, *LDB1* regulates gene expression in olfactory sensory  
300 neurons<sup>83</sup>. The brain gene expression of this locus has been identified as negatively  
301 correlated with cocaine-seeking response in rats<sup>84</sup>.

302 Overall, our EWAS findings support the association of SD-latent classes with genes  
303 involved in brain developmental processes. The applicability of our blood-based findings  
304 into brain mechanisms is supported by the high blood-brain correlation observed for  
305 several of the CpG sites identified (i.e., *MOV10* cg08355353, *SERTAD4* cg20270863,  
306 *SIN3B* cg06815056, *SPATA4* cg02833127, and *ARID1B* cg19436567). In some cases,  
307 genes we identified here (but with different CpG sites) were reported as associated by  
308 previous EWAS of tobacco smoking. Because we controlled for the effect of tobacco  
309 smoking on DNA methylation, the epigenetic associations identified in our study likely  
310 reflect SD comorbidities rather than the co-occurrence of tobacco smoking. Indeed, we  
311 did not identify any association between SD-latent classes and *AHRR* locus, which is  
312 the most established epigenetic biomarker for tobacco smoking<sup>85</sup>.

313 We also identified genetic associations linking SD comorbidity patterns to other  
314 psychiatric disorders. PTSD was the only PGS showing Bonferroni significant  
315 association across the four SD-latent classes. This may support the role of mechanisms  
316 related to trauma response across SD comorbidity patterns. This is also in line with the

317 well-known impact of stress response on biological mechanisms involved in SD  
318 pathogenesis<sup>86</sup>. TD-latent class was only associated with PTSD PGS. This may be due  
319 to the small sample size of the TD-latent class. Alternatively, it may also reflect the  
320 limited pleiotropy of this SD pattern with other mental illnesses. With respect to the other  
321 PGS associations shared among the remaining SD-latent classes, the relationships  
322 observed with respect to neurodevelopmental disorders (i.e., ADHD and schizophrenia)  
323 appear to converge with the EWAS findings, highlighting the potential role of altered  
324 neurodevelopmental processes in the shared pathogenesis among SD comorbidity  
325 patterns. Consistent with this hypothesis, previous studies highlighted how early-life  
326 adversity affects the development of brain reward and stress circuits increasing  
327 addiction vulnerability<sup>87</sup>.

328 In addition to shared pleiotropic mechanisms, we also observed PGS associations  
329 specific to certain SD-latent classes. Bipolar-disorder PGS was uniquely associated with  
330 the AD+TD latent class. Multiple studies pointed to neural reward circuit dysfunction as  
331 the risk factor shared between bipolar disorder and SDs<sup>88</sup>. The PSU-latent class  
332 showed two specific PGS associations: negative association with subjective well-being,  
333 and positive association with neuroticism. These may be related to poorer outcomes  
334 that are seen in individuals affected by the comorbidities of four SDs than those  
335 observed in subjects with only two comorbid SDs<sup>89, 90</sup>. In a previous study, a genetically  
336 inferred addiction factor shared across AD, TD, CaD, and TD was highly genetically  
337 correlated with neuroticism<sup>90</sup>.

338 Our study has multiple limitations. The cohorts investigated were specifically recruited to  
339 study SD genetics, overrepresenting the samples for cases. This may reduce the  
340 generalizability of the SD-latent classes identified. The PGS analysis was limited to  
341 individuals of European descent, because of the lack of large-scale GWAS to derive  
342 statistically powerful PGS for other population groups. In the epigenome-wide analysis,  
343 we conducted a cross-ancestry meta-analysis assessing heterogeneity among the  
344 population groups. However, we had limited power to investigate ancestry-specific  
345 epigenetic effects. Similarly, we focused only on five main SDs because of the limited

346 information on other substances in the cohorts investigated (e.g., hallucinogens,  
347 inhalants, sedatives, and amphetamines).

348 Notwithstanding these limitations, the present study provides new insights into genetic  
349 and epigenetic mechanisms contributing to in SD comorbidity patterns. Our findings  
350 support the potential role of brain developmental processes on SD pathogenesis and  
351 suggest possible mechanisms that differentiate the co-occurrence of different SD  
352 combinations. Building on this evidence, further studies are needed to extend these  
353 findings elucidating the dynamics underlying polysubstance use disorders.

354

### 355 **ACKNOWLEDGEMENTS**

356 This study is supported by the National Institute on Drug Abuse, R33 DA047527. GAP  
357 acknowledges support from the Yale Biological Sciences Training Program (T32  
358 MH014276), Alzheimer's Association (AARF-22-967171), NIH National Institute of Aging  
359 (K99AG078503), Yale Franke Fellowship in Science & Humanities, and Yale Women's  
360 Faculty Forum Award. RP acknowledges grants from the National Institute of Mental  
361 Health (RF1 MH132337) and One Mind Rising Star Award. JDD acknowledges support  
362 by the National Institute on Drug Abuse K01 DA058807. DFL is funded by a Career  
363 Development Award from the US Department of Veterans Affairs Office of Research and  
364 Development (1IK2BX005058). HK acknowledges support from the Department of  
365 Veterans Affairs (VISN 4 MIRECC and I01 BX004820). JLMO acknowledges support  
366 from U.S. Department of Veterans Affairs via 1IK2CX002095 and NIDA R21DA050160.  
367 JG reports support from the Department of Veterans Affairs (5IO1CX001849-04 and the  
368 VISN 1 New England MIRECC) and NIH/NIDA (2R01DA037974, 1R01DA058862-01).

369

### 370 **CONFLICT OF INTEREST**

371 RP received a research grant from Alkermes outside the scope of the present study. RP  
372 and JG are paid for their editorial work on the journal Complex Psychiatry. JG and HRK  
373 are holders of U.S. patent 10,900,082 titled: "Genotype-guided dosing of opioid  
374 agonists," issued 26 January 2021. HRK is a member of advisory boards for Dicerna

375    Pharmaceuticals, Sophrosyne Pharmaceuticals, Enthion Pharmaceuticals, and  
376    Clearmind Medicine; a consultant to Sobrera Pharmaceuticals and Altimune; the  
377    recipient of research funding and medication supplies for an investigator-initiated study  
378    from Alkermes; a member of the American Society of Clinical Psychopharmacology's  
379    Alcohol Clinical Trials Initiative, which was supported in the last three years by  
380    Alkermes, Dicerna, Ethypharm, Lundbeck, Mitsubishi, Otsuka, and Pear Therapeutics.  
381    FRW is an employee of Regeneron Pharmaceuticals with no conflict of interest related  
382    to any intellectual property of the company. The other authors have no competing  
383    interests to report.

384

**Table 1: Characteristics of the cohorts investigated.**

	<b>N</b>	<b>Age, mean±SD</b>	<b>Female %</b>	<b>AD, N</b>	<b>CaD, N</b>	<b>CoD, N</b>	<b>OD, N</b>	<b>TD, N</b>
<b>phs000092</b>	4,121	39±9	54	1,946	753	1,130	304	1,856
<b>phs000277</b>	3,515	46±10	42	55	4	0	0	205
<b>phs000404</b>	1,527	37±6	59	204	222	149	21	1,239
<b>phs000813</b>	2,969	52±7	60	516	354	210	57	1,844
<b>phs001299</b>	3,508	50±14	51	493	134	6	32	1,367
<b>Yale-Penn</b>	15,557	40±12	46	7,481	3,897	8,662	4,379	8,219
<b>Total</b>	31,197	42±11	49	10,695	5,364	10,157	4,793	14,730

385

386

## 387 **FIGURE LEGENDS**

388 **Figure 1. Study Design.** An overview of the study design showing five substance  
389 dependence (SD) diagnoses assessed in six different cohorts and used to identify latent  
390 classes related to their comorbidities. Each of the SD latent classes was compared  
391 against the control group in the epigenome-wide association study and estimating  
392 genetic burden of ten psychiatric and behavioral traits.

393 **Figure 2. Distribution of case- and control-posterior probability across SDs in the**  
394 **five-latent class model.** The x-axis (top) shows each of the substance dependence  
395 diagnoses, and the y-axis shows the probability value of cases and controls for each of  
396 the five SDs across each latent class shown on the left. Details of the statistical  
397 comparisons are shown in Supplementary Figure 1.

398 **Fig 3. Epigenome-wide association study of SD latent classes.** Forest Plots show  
399 meta-analyzed effect sizes of the significant CpG sites across ancestral groups; EUR –  
400 European descent; AFR – African descent; AMR – Admixed Americans.

401 **Fig 4. Association of polygenic scores of psychiatric and behavioral traits with SD**  
402 **latent classes.** Forest plot showing meta-analyzed associations of ten traits with SUD  
403 latent classes. The x-axis shows odds ratio as points and 95% confidence intervals as  
404 lines. The Bonferroni-significant associations are marked with green, while others are  
405 indicated in grey. Details regarding PGS associations are shown in Supplementary  
406 Table 9.

407



408 **REFERENCES**

- 409 1. Substance Use & Substance Use Disorders.  
410 [https://wwwnc.cdc.gov/travel/yellowbook/2024/additional-](https://wwwnc.cdc.gov/travel/yellowbook/2024/additional-considerations/substance-use)  
411 [considerations/substance-use](https://wwwnc.cdc.gov/travel/yellowbook/2024/additional-considerations/substance-use), 2024, Accessed Date Accessed 2024 Accessed.
- 412 2. Abuse NloD. Common Comorbidities with Substance Use Disorders Research  
413 Report. NCBI: Bethesda (MD), 2020.
- 414 3. Owens PL, Heslin KC, Fingar KR, Weiss AJ. Co-occurrence of Physical Health  
415 Conditions and Mental Health and Substance Use Conditions Among Adult  
416 Inpatient Stays, 2010 Versus 2014. *Healthcare Cost and Utilization Project*  
417 *(HCUP) Statistical Briefs*: Rockville (MD), 2006.
- 418 4. Stephenson M, Bollepalli S, Cazaly E, Salvatore JE, Barr P, Rose RJ *et al.*  
419 Associations of Alcohol Consumption With Epigenome-Wide DNA Methylation  
420 and Epigenetic Age Acceleration: Individual-Level and Co-twin Comparison  
421 Analyses. *Alcohol Clin Exp Res* 2021; **45**(2): 318-328.
- 422 5. Montalvo-Ortiz JL, Cheng Z, Kranzler HR, Zhang H, Gelernter J. Genomewide  
423 Study of Epigenetic Biomarkers of Opioid Dependence in European- American  
424 Women. *Sci Rep* 2019; **9**(1): 4660.
- 425 6. Poisel E, Zillich L, Streit F, Frank J, Friske MM, Foo JC *et al.* DNA methylation in  
426 cocaine use disorder-An epigenome-wide approach in the human prefrontal  
427 cortex. *Front Psychiatry* 2023; **14**: 1075250.
- 428 7. Lohoff FW, Roy A, Jung J, Longley M, Rosoff DB, Luo A *et al.* Epigenome-wide  
429 association study and multi-tissue replication of individuals with alcohol use  
430 disorder: evidence for abnormal glucocorticoid signaling pathway gene  
431 regulation. *Mol Psychiatry* 2021; **26**(6): 2224-2237.
- 432 8. Kember RL, Hartwell EE, Xu H, Rotenberg J, Almasy L, Zhou H *et al.* Phenome-  
433 wide Association Analysis of Substance Use Disorders in a Deeply Phenotyped  
434 Sample. *Biol Psychiatry* 2023; **93**(6): 536-545.
- 435 9. Deak JD, Johnson EC. Genetics of substance use disorders: a review. *Psychol*  
436 *Med* 2021; **51**(13): 2189-2200.
- 437 10. Gelernter J, Polimanti R. Genetics of substance use disorders in the era of big  
438 data. *Nat Rev Genet* 2021; **22**(11): 712-729.
- 439 11. Gelernter J, Kranzler HR, Sherva R, Almasy L, Koesterer R, Smith AH *et al.*  
440 Genome-wide association study of alcohol dependence: significant findings in  
441 African- and European-Americans including novel risk loci. *Mol Psychiatry* 2014;  
442 **19**(1): 41-49.
- 443 12. Gelernter J, Kranzler HR, Sherva R, Koesterer R, Almasy L, Zhao H *et al.*  
444 Genome-wide association study of opioid dependence: multiple associations  
445 mapped to calcium and potassium pathways. *Biol Psychiatry* 2014; **76**(1): 66-74.

- 446 13. Gelernter J, Kranzler HR, Sherva R, Almasy L, Herman AI, Koesterer R *et al.*  
447 Genome-wide association study of nicotine dependence in American populations:  
448 identification of novel risk loci in both African-Americans and European-  
449 Americans. *Biol Psychiatry* 2015; **77**(5): 493-503.
- 450 14. Gelernter J, Sherva R, Koesterer R, Almasy L, Zhao H, Kranzler HR *et al.*  
451 Genome-wide association study of cocaine dependence and related traits:  
452 FAM53B identified as a risk gene. *Mol Psychiatry* 2014; **19**(6): 717-723.
- 453 15. Sherva R, Wang Q, Kranzler H, Zhao H, Koesterer R, Herman A *et al.* Genome-  
454 wide Association Study of Cannabis Dependence Severity, Novel Risk Variants,  
455 and Shared Genetic Risks. *JAMA Psychiatry* 2016; **73**(5): 472-480.
- 456 16. Rice JP, Hartz SM, Agrawal A, Almasy L, Bennett S, Breslau N *et al.* CHRN3 is  
457 more strongly associated with Fagerstrom test for cigarette dependence-based  
458 nicotine dependence than cigarettes per day: phenotype definition changes  
459 genome-wide association studies results. *Addiction* 2012; **107**(11): 2019-2028.
- 460 17. Nelson EC, Lynskey MT, Heath AC, Wray N, Agrawal A, Shand FL *et al.* ANKK1,  
461 TTC12, and NCAM1 polymorphisms and heroin dependence: importance of  
462 considering drug exposure. *JAMA Psychiatry* 2013; **70**(3): 325-333.
- 463 18. Bierut LJ, Madden PA, Breslau N, Johnson EO, Hatsukami D, Pomerleau OF *et al.*  
464 Novel genes identified in a high-density genome wide association study for  
465 nicotine dependence. *Hum Mol Genet* 2007; **16**(1): 24-35.
- 466 19. Pergadia ML, Agrawal A, Heath AC, Martin NG, Bucholz KK, Madden PA.  
467 Nicotine withdrawal symptoms in adolescent and adult twins. *Twin Res Hum*  
468 *Genet* 2010; **13**(4): 359-369.
- 469 20. Linzer DA, Lewis JB. poLCA: An R Package for Polytomous Variable Latent  
470 Class Analysis. *Journal of Statistical Software* 2011; **42**(10): 1 - 29.
- 471 21. Finch WH, Bronk KC. Conducting Confirmatory Latent Class Analysis Using  
472 Mplus. *Structural Equation Modeling: A Multidisciplinary Journal* 2011; **18**(1):  
473 132-151.
- 474 22. Wendt FR, Pathak GA, Vahey J, Qin X, Koller D, Cabrera-Mendoza B *et al.*  
475 Modeling the longitudinal changes of ancestry diversity in the Million Veteran  
476 Program. *Hum Genomics* 2023; **17**(1): 46.
- 477 23. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust  
478 relationship inference in genome-wide association studies. *Bioinformatics* 2010;  
479 **26**(22): 2867-2873.
- 480 24. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A *et al.* Next-  
481 generation genotype imputation service and methods. *Nat Genet* 2016; **48**(10):  
482 1284-1287.
- 483 25. Teschendorff AE, Marabita F, Lechner M, Bartlett T, Tegner J, Gomez-Cabrero D  
484 *et al.* A beta-mixture quantile normalization method for correcting probe design

- 485 bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013; **29**(2):  
486 189-196.
- 487 26. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for  
488 removing batch effects and other unwanted variation in high-throughput  
489 experiments. *Bioinformatics* 2012; **28**(6): 882-883.
- 490 27. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson  
491 HH *et al.* DNA methylation arrays as surrogate measures of cell mixture  
492 distribution. *BMC Bioinformatics* 2012; **13**: 86.
- 493 28. Bollepalli S, Korhonen T, Kaprio J, Anders S, Ollikainen M. EpiSmokEr: a robust  
494 classifier to determine smoking status from DNA methylation data. *Epigenomics*  
495 2019; **11**(13): 1469-1486.
- 496 29. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W *et al.* limma powers  
497 differential expression analyses for RNA-sequencing and microarray studies.  
498 *Nucleic Acids Res* 2015; **43**(7): e47.
- 499 30. Turner SD. qqman: an R package for visualizing GWAS results using Q-Q and  
500 manhattan plots. *Journal of Open Source Software* 2018; **3**.
- 501 31. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-  
502 analysis. *BMC Bioinformatics* 2010; **11**(1): 288.
- 503 32. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K. dmrff:  
504 identifying differentially methylated regions efficiently with power and control.  
505 *bioRxiv* 2018: 508556.
- 506 33. Battram T, Yousefi P, Crawford G, Prince C, Sheikhal Babaei M, Sharp G *et al.*  
507 The EWAS Catalog: a database of epigenome-wide association studies.  
508 *Wellcome Open Res* 2022; **7**: 41.
- 509 34. Edgar RD, Jones MJ, Meaney MJ, Turecki G, Kobor MS. BECon: a tool for  
510 interpreting DNA methylation findings from blood in the context of brain.  
511 *Translational psychiatry* 2017; **7**(8): e1187.
- 512 35. Demontis D, Walters GB, Athanasiadis G, Walters R, Therrien K, Nielsen TT *et al.*  
513 Genome-wide analyses of ADHD identify 27 risk loci, refine the genetic  
514 architecture and implicate several cognitive domains. *Nat Genet* 2023; **55**(2):  
515 198-208.
- 516 36. Friligkou E, Løkhammer S, Cabrera-Mendoza B, Shen J, He J, Deiana G *et al.*  
517 Gene Discovery and Biological Insights into Anxiety Disorders from a Multi-  
518 Ancestry Genome-wide Association Study of >1.2 Million Participants.  
519 *medRxiv* 2024: 2024.2002.2014.24302836.
- 520 37. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H *et al.* Identification  
521 of common genetic risk variants for autism spectrum disorder. *Nat Genet* 2019;  
522 **51**(3): 431-444.

- 523 38. Mullins N, Forstner AJ, O'Connell KS, Coombes B, Coleman JRI, Qiao Z *et al.*  
524 Genome-wide association study of more than 40,000 bipolar disorder cases  
525 provides new insights into the underlying biology. *Nat Genet* 2021; **53**(6): 817-  
526 829.
- 527 39. Als TD, Kurki MI, Grove J, Voloudakis G, Therrien K, Tasanko E *et al.* Depression  
528 pathophysiology, risk prediction of recurrence and comorbid psychiatric disorders  
529 using genome-wide analyses. *Nat Med* 2023; **29**(7): 1832-1844.
- 530 40. Okbay A, Wu Y, Wang N, Jayashankar H, Bennett M, Nehzati SM *et al.* Polygenic  
531 prediction of educational attainment within and between families from genome-  
532 wide association analyses in 3 million individuals. *Nat Genet* 2022; **54**(4): 437-  
533 449.
- 534 41. Kim Y, Saunders GRB, Giannelis A, Willoughby EA, DeYoung CG, Lee JJ.  
535 Genetic and neural bases of the neuroticism general factor. *Biol Psychol* 2023;  
536 **184**: 108692.
- 537 42. Nievergelt CM, Maihofer AX, Atkinson EG, Chen CY, Choi KW, Coleman JRI *et al.*  
538 Genome-wide association analyses identify 95 risk loci and provide insights  
539 into the neurobiology of post-traumatic stress disorder. *Nat Genet* 2024; **56**(5):  
540 792-808.
- 541 43. Trubetskoy V, Pardinas AF, Qi T, Panagiotaropoulou G, Awasthi S, Bigdeli TB *et al.*  
542 Mapping genomic loci implicates genes and synaptic biology in schizophrenia.  
543 *Nature* 2022; **604**(7906): 502-508.
- 544 44. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG, Fontana MA *et al.*  
545 Genetic variants associated with subjective well-being, depressive symptoms,  
546 and neuroticism identified through genome-wide analyses. *Nat Genet* 2016;  
547 **48**(6): 624-633.
- 548 45. Ge T, Chen C-Y, Ni Y, Feng Y-CA, Smoller JW. Polygenic prediction via Bayesian  
549 regression and continuous shrinkage priors. *Nature Communications* 2019;  
550 **10**(1): 1776.
- 551 46. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a  
552 practical tutorial. *Evid Based Ment Health* 2019; **22**(4): 153-160.
- 553 47. Bailey AJ, McHugh RK. Why do we focus on the exception and not the rule?  
554 Examining the prevalence of mono- versus polysubstance use in the general  
555 population. *Addiction* 2023; **118**(10): 2026-2029.
- 556 48. Hatoum AS, Johnson EC, Colbert SMC, Polimanti R, Zhou H, Walters RK *et al.*  
557 The addiction risk factor: A unitary genetic vulnerability characterizes substance  
558 use disorders and their associations with common correlates.  
559 *Neuropsychopharmacology* 2022; **47**(10): 1739-1745.
- 560 49. Hatoum AS, Colbert SMC, Johnson EC, Huggett SB, Deak JD, Pathak G *et al.*  
561 Multivariate genome-wide association meta-analysis of over 1 million subjects

- 562 identifies loci underlying multiple substance use disorders. *Nat Ment Health*  
563 2023; **1**(3): 210-223.
- 564 50. Patrick ME, Berglund PA, Joshi S, Bray BC. A latent class analysis of heavy  
565 substance use in Young adulthood and impacts on physical, cognitive, and  
566 mental health outcomes in middle age. *Drug Alcohol Depend* 2020; **212**: 108018.
- 567 51. Schepis TS, McCabe SE. The latent class structure of substance use in US  
568 adults 50 years and older. *Int J Geriatr Psychiatry* 2021; **36**(12): 1867-1877.
- 569 52. Agrawal A, Lynskey MT, Madden PA, Bucholz KK, Heath AC. A latent class  
570 analysis of illicit drug abuse/dependence: results from the National  
571 Epidemiological Survey on Alcohol and Related Conditions. *Addiction* 2007;  
572 **102**(1): 94-104.
- 573 53. Chan G, Connor J, Hall W, Leung J. The changing patterns and correlates of  
574 population-level polysubstance use in Australian youth: a multi-group latent class  
575 analysis of nationally representative samples spanning 12 years. *Addiction* 2020;  
576 **115**(1): 145-155.
- 577 54. Rodriguez AS, Robinson LD, Kelly PJ, Hudson S. Polysubstance use classes  
578 and health outcomes among women attending specialist substance use  
579 treatment services. *Drug Alcohol Rev* 2022; **41**(2): 488-500.
- 580 55. Mulder RH, Neumann A, Cecil CAM, Walton E, Houtepen LC, Simpkin AJ *et al.*  
581 Epigenome-wide change and variation in DNA methylation in childhood:  
582 trajectories from birth to late adolescence. *Hum Mol Genet* 2021; **30**(1): 119-134.
- 583 56. Li Z, Xu K, Zhao S, Guo Y, Chen H, Ni J *et al.* SPATA4 improves aging-induced  
584 metabolic dysfunction through promotion of preadipocyte differentiation and  
585 adipose tissue expansion. *Aging Cell* 2021; **20**(1): e13282.
- 586 57. Islam SA, Goodman SJ, Maclsaac JL, Obradovic J, Barr RG, Boyce WT *et al.*  
587 Integration of DNA methylation patterns and genetic variation in human pediatric  
588 tissues help inform EWAS design and interpretation. *Epigenetics Chromatin*  
589 2019; **12**(1): 1.
- 590 58. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CC, O'Donovan MC *et al.*  
591 Methylomic trajectories across human fetal brain development. *Genome Res*  
592 2015; **25**(3): 338-352.
- 593 59. Inkster AM, Yuan V, Konwar C, Matthews AM, Brown CJ, Robinson WP. A cross-  
594 cohort analysis of autosomal DNA methylation sex differences in the term  
595 placenta. *Biol Sex Differ* 2021; **12**(1): 38.
- 596 60. Zhang L, Silva TC, Young JI, Gomez L, Schmidt MA, Hamilton-Nelson KL *et al.*  
597 Epigenome-wide meta-analysis of DNA methylation differences in prefrontal  
598 cortex implicates the immune processes in Alzheimer's disease. *Nat Commun*  
599 2020; **11**(1): 6114.
- 600 61. Tsai PC, Glastonbury CA, Eliot MN, Bollepalli S, Yet I, Castillo-Fernandez JE *et*  
601 *al.* Smoking induces coordinated DNA methylation and gene expression changes

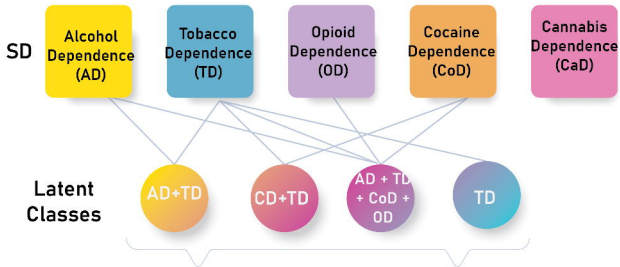
- 602 in adipose tissue with consequences for metabolic health. *Clin Epigenetics* 2018;  
603 **10**(1): 126.
- 604 62. Harper JD, Fan KH, Aslam MM, Snitz BE, DeKosky ST, Lopez OL *et al.* Genome-  
605 Wide Association Study of Incident Dementia in a Community-Based Sample of  
606 Older Subjects. *J Alzheimers Dis* 2022; **88**(2): 787-798.
- 607 63. He L, Loika Y, Park Y, Genotype Tissue Expression c, Bennett DA, Kellis M *et al.*  
608 Exome-wide age-of-onset analysis reveals exonic variants in ERN1 and SPPL2C  
609 associated with Alzheimer's disease. *Transl Psychiatry* 2021; **11**(1): 146.
- 610 64. Lampada A, Taylor V. Notch signaling as a master regulator of adult  
611 neurogenesis. *Front Neurosci* 2023; **17**: 1179011.
- 612 65. Ni T, Zhu L, Wang S, Zhu W, Xue Y, Zhu Y *et al.* Medial prefrontal cortex Notch1  
613 signalling mediates methamphetamine-induced psychosis via Hes1-dependent  
614 suppression of GABA(B1) receptor expression. *Mol Psychiatry* 2022; **27**(10):  
615 4009-4022.
- 616 66. Latypova X, Vincent M, Molle A, Adebambo OA, Fourgeux C, Khan TN *et al.*  
617 Haploinsufficiency of the Sin3/HDAC corepressor complex member SIN3B  
618 causes a syndromic intellectual disability/autism spectrum disorder. *Am J Hum*  
619 *Genet* 2021; **108**(5): 929-941.
- 620 67. Moffat JJ, Jung EM, Ka M, Jeon BT, Lee H, Kim WY. Differential roles of ARID1B  
621 in excitatory and inhibitory neural progenitors in the developing cortex. *Sci Rep*  
622 2021; **11**(1): 3856.
- 623 68. Zhu G, Liu J, Li Y, Huang H, Chen C, Wu D *et al.* ARID1B Deficiency Leads to  
624 Impaired DNA Damage Response and Activated cGAS-STING Pathway in Non-  
625 Small Cell Lung Cancer. *J Cancer* 2024; **15**(9): 2601-2612.
- 626 69. Carnes MU, Quach BC, Zhou L, Han S, Tao R, Mandal M *et al.* Smoking-  
627 informed methylation and expression QTLs in human brain and colocalization  
628 with smoking-associated genetic loci. *medRxiv* 2023.
- 629 70. Cardenas A, Ecker S, Fadadu RP, Huen K, Orozco A, McEwen LM *et al.*  
630 Epigenome-wide association study and epigenetic age acceleration associated  
631 with cigarette smoking among Costa Rican adults. *Sci Rep* 2022; **12**(1): 4277.
- 632 71. Mishra PP, Mishra BH, Raitoharju E, Mononen N, Viikari J, Juonala M *et al.* Gene  
633 Set Based Integrated Methylome and Transcriptome Analysis Reveals Potential  
634 Molecular Mechanisms Linking Cigarette Smoking and Related Diseases.  
635 *OMICS* 2023; **27**(5): 193-204.
- 636 72. Ponomarev I. Epigenetic control of gene expression in the alcoholic brain.  
637 *Alcohol Res* 2013; **35**(1): 69-76.
- 638 73. Dugue PA, Wilson R, Lehne B, Jayasekara H, Wang X, Jung CH *et al.* Alcohol  
639 consumption is associated with widespread changes in blood DNA methylation:  
640 Analysis of cross-sectional and longitudinal data. *Addict Biol* 2021; **26**(1):  
641 e12855.

- 642 74. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai PC, Reynolds LM *et al.* A DNA  
643 methylation biomarker of alcohol consumption. *Mol Psychiatry* 2018; **23**(2): 422-  
644 433.
- 645 75. Skariah G, Seimetz J, Norsworthy M, Lannom MC, Kenny PJ, Elrakhawy M *et al.*  
646 Mov10 suppresses retroelements and regulates neuronal development and  
647 function in the developing brain. *BMC Biol* 2017; **15**(1): 54.
- 648 76. Hofer E, Roshchupkin GV, Adams HHH, Knol MJ, Lin H, Li S *et al.* Genetic  
649 correlations and genome-wide associations of cortical structure in general  
650 population samples of 22,824 adults. *Nat Commun* 2020; **11**(1): 4796.
- 651 77. Grasby KL, Jahanshad N, Painter JN, Colodro-Conde L, Bralten J, Hibar DP *et*  
652 *al.* The genetic architecture of the human cerebral cortex. *Science* 2020;  
653 **367**(6484).
- 654 78. Sha Z, Schijven D, Fisher SE, Francks C. Genetic architecture of the white  
655 matter connectome of the human brain. *Sci Adv* 2023; **9**(7): eadd2870.
- 656 79. Schreiber R, Ousingsawat J, Wanitchakool P, Sirianant L, Benedetto R, Reiss K  
657 *et al.* Regulation of TMEM16A/ANO1 and TMEM16F/ANO6 ion currents and  
658 phospholipid scrambling by Ca(2+) and plasma membrane lipid. *J Physiol* 2018;  
659 **596**(2): 217-229.
- 660 80. Sikdar S, Joehanes R, Joubert BR, Xu CJ, Vives-Usano M, Rezwan FI *et al.*  
661 Comparison of smoking-related DNA methylation between newborns from  
662 prenatal exposure and adults from personal smoking. *Epigenomics* 2019; **11**(13):  
663 1487-1500.
- 664 81. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR *et al.*  
665 Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* 2016; **9**(5):  
666 436-447.
- 667 82. Kinare V, Pal S, Tole S. LDB1 Is Required for the Early Development of the  
668 Dorsal Telencephalon and the Thalamus. *eNeuro* 2019; **6**(1).
- 669 83. Monahan K, Horta A, Lomvardas S. LHX2- and LDB1-mediated trans interactions  
670 regulate olfactory receptor choice. *Nature* 2019; **565**(7740): 448-453.
- 671 84. Towers EB, Shapiro DA, Abel JM, Bakhti-Suroosh A, Kupkova K, Auble DT *et al.*  
672 Transcriptional Profile of Exercise-Induced Protection Against Relapse to  
673 Cocaine Seeking in a Rat Model. *Biol Psychiatry Glob Open Sci* 2023; **3**(4): 734-  
674 745.
- 675 85. Philibert R, Dogan M, Beach SRH, Mills JA, Long JD. AHRR methylation predicts  
676 smoking status and smoking intensity in both saliva and blood DNA. *Am J Med*  
677 *Genet B Neuropsychiatr Genet* 2020; **183**(1): 51-60.
- 678 86. Milivojevic V, Sinha R. Central and Peripheral Biomarkers of Stress Response for  
679 Addiction Risk and Relapse Vulnerability. *Trends Mol Med* 2018; **24**(2): 173-186.

- 680 87. Levis SC, Baram TZ, Mahler SV. Neurodevelopmental origins of substance use  
681 disorders: Evidence from animal models of early-life adversity and addiction. *Eur*  
682 *J Neurosci* 2022; **55**(9-10): 2170-2195.
- 683 88. Bart CP, Titone MK, Ng TH, Nusslock R, Alloy LB. Neural reward circuit  
684 dysfunction as a risk factor for bipolar spectrum disorders and substance use  
685 disorders: A review and integration. *Clin Psychol Rev* 2021; **87**: 102035.
- 686 89. Zeng J, You L, Yang F, Luo Y, Yu S, Yan J *et al.* A meta-analysis of the neural  
687 substrates of monetary reward anticipation and outcome in alcohol use disorder.  
688 *Hum Brain Mapp* 2023; **44**(7): 2841-2861.
- 689 90. Peechatka AL, Whitton AE, Farmer SL, Pizzagalli DA, Janes AC. Cigarette  
690 craving is associated with blunted reward processing in nicotine-dependent  
691 smokers. *Drug Alcohol Depend* 2015; **155**: 202-207.
- 692
- 693

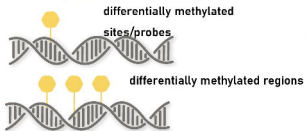


N = 31,197 (average age 42 yrs  $\pm$  11; Females = 49%)



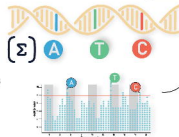
### Epigenome wide association study

N=886



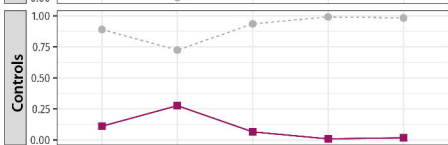
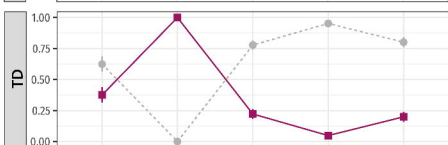
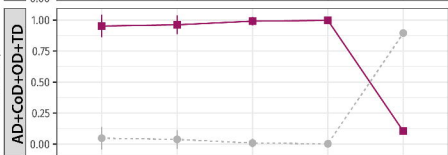
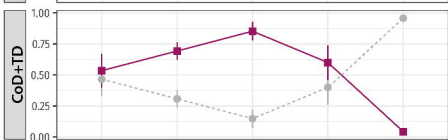
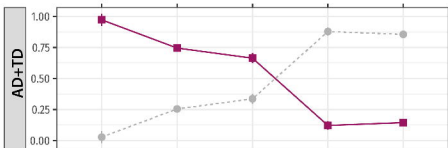
### Polygenic burden of psychiatric and behavioral traits

N=7,659



Attention deficit/Hyperactivity Disorder  
Anxiety  
Autism Spectrum Disorder  
Bipolar Disorder  
Depression  
Educational Attainment  
Neuroticism  
Post-traumatic stress disorder  
Schizophrenia  
Subjective well-being

Alcohol Dependence (AD)  
 Tobacco Dependence (TD)  
 Cocaine Dependence (CoD)  
 Opioid Dependence (OD)  
 Cannabis Dependence (CaD)



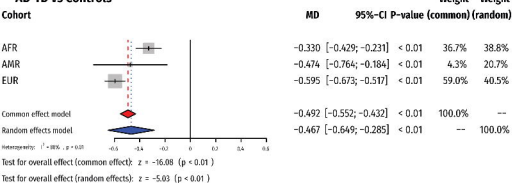
Dependency

Case

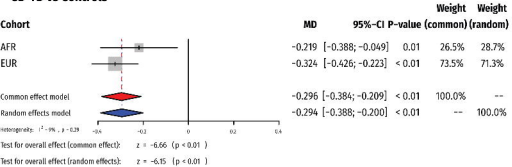
Control

## cg02833127 (SPATA4)

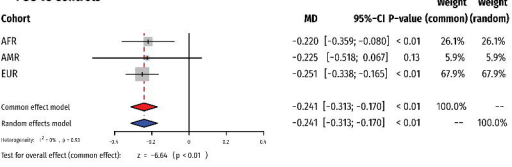
### AD-TD vs Controls



### CD-TD vs Controls

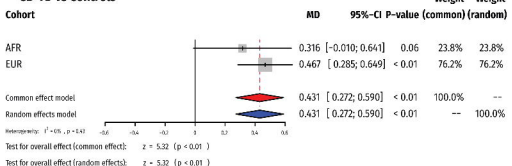


### PSU vs Controls



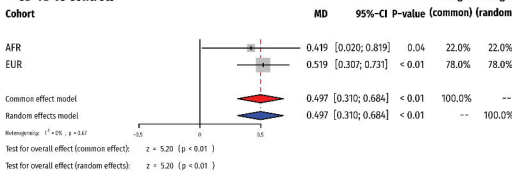
## cg08355353 (MOV10)

### CD-TD vs Controls



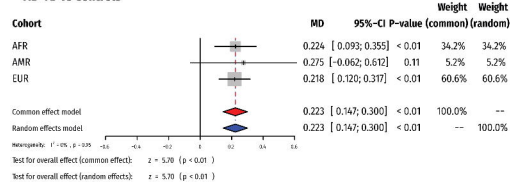
## cg09099775 (ANO6)

### CD-TD vs Controls



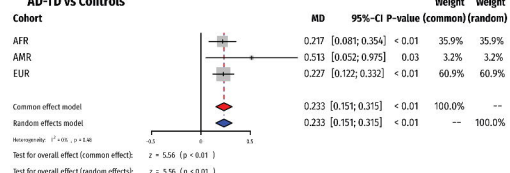
## cg19436567 (ARID1B)

### AD-TD vs Controls



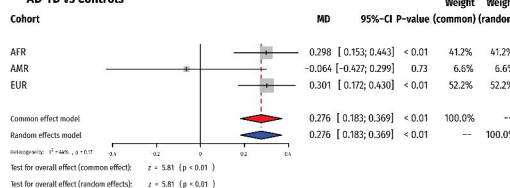
## cg06815056 (SIN3B)

### AD-TD vs Controls



## cg13725899 (NOTCH1)

### AD-TD vs Controls



## cg20270863 (C1orf133/SERTAD4)

### AD-TD vs Controls

