Shared and unique 3D genomic features of substance use disorders across multiple cell types

Khanh B. Trang1,2, Alessandra Chesi1,3, Sylvanus Toikumo4,5, James A. Pippin1,2, Matthew C. Pahl1,2, Joan M. O’Brien7,8, Laufey T. Amundadottir9, Kevin M. Brown9, Wenli Yang10,11, Jaclyn Welles10,12, Dominic Santoleri10,12, Paul M. Titchenell10,12, Patrick Seale10,11, Babette S. Zemel6,13, Yadav Wagley14, Kurt D. Hankenson14, Klaus H. Kaestner10,15, Stewart A. Anderson16,17, Matthew S. Kayser5,18,19, Andrew D. Wells1,3,20, Henry R. Kranzler4,5, Rachel L. Kember4,5, Struan F.A. Grant1,2,10,15,21

1. Center for Spatial and Functional Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
2. Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA
3. Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
4. Mental Illness Research, Education and Clinical Center, Crescenz Veterans Affairs Medical Center, Philadelphia, PA, USA
5. Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, PA, USA
6. Scheie Eye Institute, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
7. Penn Medicine Center for Ophthalmic Genetics in Complex Disease, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
8. Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
9. Institute for Diabetes, Obesity and Metabolism, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
10. Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
11. Department of Cell and Developmental Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
12. Department of Physiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
13. Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
14. Department of Orthopedic Surgery, University of Michigan Medical School Ann Arbor, MI, USA
15. Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
16. Department of Child and Adolescent Psychiatry, Children's Hospital of Philadelphia, Philadelphia, PA, USA
17. Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA
18. Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
19. Chronobiology Sleep Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
20. Institute for Immunology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
21. Division of Endocrinology and Diabetes, The Children’s Hospital of Philadelphia, Philadelphia, PA, USA

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Recent genome-wide association studies (GWAS) have revealed shared genetic components among alcohol, opioid, tobacco and cannabis use disorders. However, the extent of the underlying shared causal variants and effector genes, along with their cellular context, remain unclear. We leveraged our existing 3D genomic datasets comprising high-resolution promoter-focused Capture-C/Hi-C, ATAC-seq and RNA-seq across >50 diverse human cell types to focus on genomic regions that coincide with GWAS loci. Using stratified LD regression, we determined the proportion of genome-wide SNP heritability attributable to the features assayed across our cell types by integrating recent GWAS summary statistics for the relevant traits: alcohol use disorder (AUD), tobacco use disorder (TUD), opioid use disorder (OUD) and cannabis use disorder (CanUD). Statistically significant enrichments ($P<0.05$) were observed in 14 specific cell types, with heritability reaching 9.2-fold for iPSC-derived cortical neurons and neural progenitors, confirming that they are crucial cell types for further functional exploration. Additionally, several pancreatic cell types, notably pancreatic beta cells, showed enrichment for TUD, with heritability enrichments up to 4.8-fold, suggesting genomic overlap with metabolic processes. Further investigation revealed significant positive genetic correlations between T2D with both TUD and CanUD (FDR<0.05) and a significant negative genetic correlation with AUD. Interestingly, after partitioning the heritability for each cell type’s cis-regulatory elements, the correlation between T2D and TUD for pancreatic beta cells was greater ($r=0.2$) than the global genetic correlation value. Our study provides new genomic insights into substance use disorders and implicates cell types where functional follow-up studies could reveal causal variant-gene mechanisms underpinning these disorders.
INTRODUCTION

Substance use disorders (SUDs) are a group of treatable psychiatric disorders that are associated with a variety of adverse outcomes. SUDs pose a pervasive global health challenge, affecting individuals across diverse demographic groups and populations. SUD severity varies from mild to severe and adverse outcomes generally correlate with the severity of the disorder. Alcohol, socially accepted and widely available, is one of the most commonly misused substances worldwide [1], leading to significant health, social, and economic losses. Excessive alcohol consumption is a major cause of premature death and disability, particularly among individuals aged 20-39, with about 13.5% of deaths in this group linked to alcohol. In the United States, 10.5% of individuals aged 12 and older suffer from alcohol use disorder (AUD), with over 178,000 annual alcohol-related deaths reported by the Centers for Disease Control and Prevention [2,3]. Tobacco use disorder (TUD), affecting 22.3% of the global population, is the most prevalent substance use disorder, causing nearly 8 million preventable deaths annually due to diseases such as cardiovascular and respiratory disorders, as well as various cancers [4]. The prevalence of opioid use disorder (OUD) has surged over the past three decades, driven by prescription opioid misuse and illicit opioids like fentanyl, resulting in approximately 81,806 US opioid overdose deaths in 2022 [5]. Cannabis use disorder (CanUD) has become more prominent with increasing legalization and is associated with morbidity, including certain cancers, cognitive impairments, and schizophrenia, contributing to societal problems that include reduced productivity and accidents [6-9].

SUDs have strong familial inheritance patterns [10], with heritability estimates from twin studies for these disorders of averaging approximately 50%. AUD heritability estimates range from 0.50 to 0.64 [11,12] and are higher for the diagnosis than traits like alcohol use initiation (0.30–0.40) [13] and frequency of consumption (0.37–0.47) [14], suggesting a greater genetic influence on the progression to problematic use. TUD heritability ranges from 0.30 to 0.70, depending on the assessment method used [15,16]. CanUD heritability ranges from 0.51 to 0.59, with environmental and genetic contributors reported across both use and misuse [17,18]. OUD heritability is around 50%, with 38% of variation attributed to opioid-specific genetic factors [19]. Studies also
highlight that in addition to specific genetic influences for use disorders [20,21], there are broader heritable factors influencing SUD susceptibility in general [22].

Over the past ten years, there have been multiple large genome-wide association studies (GWAS) of SUDs, yielding a growing number of genome-wide significant loci [23,24]. Substantial progress has been made in identifying AUD loci, starting with the genes encoding the alcohol metabolic enzymes ADH1B and ALDH2 [25-33]. Recently, larger samples have yielded additional loci, include DRD2, GCKR, KLB and SLC39A8 [27,29,31,32,34,35]. While genetic factors play a significant role in shaping smoking behaviors, recent GWAS have primarily focused on identifying risk variants associated with nicotine consumption and TUD. These studies consistently link nicotine dependence to cholinergic nicotinic receptor genes, especially at CHRNA5-CHRNA3-CHRNB4 locus [24], and a variant at the DNMT3B locus with heavy smoking and lung cancer risk. Recent multi-ancestral meta-analyses revealed 72 independent risk loci for TUD [36], and more than a thousand loci associated with various smoking phenotypes [37]. Due to smaller available samples for GWAS, fewer replicable loci have been uncovered for CanUD. The largest meta-GWAS of CanUD to date [38] identified several significant loci unique to individual ancestries and in the multi-ancestry analysis, with key findings that include loci near SLC36A2, SEMA6D, MCCC2, LRRC3B, PDE4B, LAMB2, FOXP1, GABRB1, MAGI2, SCAI, DRD2 and ZFHX3 and confirming loci near CHRNA2 [39] and FOXP2 [40]. GWAS for OUD associations have revealed CNIH3, KCNG2, APBB2, RGMA, KCNC1, and OPRM1 loci [34,41-43]. Furthermore, a large-scale GWAS integrating data from seven cohorts identified OUD variant associations at OPRM1, FURIN, and 18 other loci using multi-trait methods [44]. A second cross-ancestry meta-analysis uncovered 14 loci for OUD, 12 of which were novel, including RABEPK, FBXW4, NCAM1 and KCNN1 [45].

Despite this progress, we are still far from a mechanistic understanding of the effector genes and the cellular contexts operating at these genetic loci, as GWAS can only identify a genomic region associated with a trait, rather than a specific effector gene, and is cell-type agnostic.

Tools like partitioned Linkage Disequilibrium Score Regression (LDSC) quantify the proportion of genome-wide SNP-heritability for a trait attributable to functional genomics.
categories of one’s choice, using information from all SNPs and explicitly modeling LD
[46]. By combining 3D chromatin maps (Hi-C, Capture-C) with matched chromatin
accessibility data (ATAC-seq), we sought to elucidate the specific cellular settings in
which genetic variation contributes to the risk of AUD, TUD, OUD, and CanUD using
partitioned LDSC. Capitalizing on existing GWAS results and our genomics datasets for
>50 cell types, and leveraging LDSC, we report significant enrichments of the SUD
GWAS signals in regulatory elements of specific cell types, shedding light on the
etiology of susceptibility to SUDs.
METHODS

Data and resource: Table S1 lists the datasets used in prior studies. The original published studies provided configurations and technical details in ATAC-seq, Hi-C, and Capture-C library generation.

ATAC-seq preprocessing and peak calling: Open chromatin regions (OCRs) were called using the ENCODE ATAC-seq pipeline (https://github.com/ENCODE-DCC/atac-seq-pipeline). Reads were aligned to the GRCh37/hg19 or GRCh38/hg38 assembly genome using bowtie2 [47], duplicates were removed, alignments from all replicates were pooled, and narrow peaks were called using MACS2. We lifted all coordinates from GRCh37/hg19 to GRCh38/hg38 to ensure consistency between datasets [48].

Promoter Capture-C pre-processing and interaction calling: Paired-end reads were pre-processed using the HICUP pipeline [49] with bowtie2 and GRCh37/hg19. Significant promoter interactions were called using unique read pairs using CHICAGO [50]. We analyzed individual fragments (1frag) and binned four fragments to improve long-distance sensitivity [51]. Interactions with CHICAGO score > 5 at either 1-fragment or 4-fragment resolution were considered significant. Interactions from both resolutions were merged and lifted from GRCh37/hg19 to GRCh38/hg38.

Hi-C pre-processing and interaction calling. As described in our recent study [52], paired-end reads from each replicate were pre-processed using the HICUP v0.7.4 pipeline [49] and aligned by bowtie2 with GRCh38/hg38. The alignment files were parsed and processed by pairtools v0.3.0 [53] and indexed and compressed by pairix v0.3.7 [54], then converted to Hi-C matrix binary format .cool at multiple resolutions (500 bp; 1, 2, 4, 10, 40, and 500 kbp; and 1 Mbp) by cooler v0.8.11 [55] and normalized using the ICE method [56]. The matrices from different replicates were merged at each resolution using cooler. Mustache v1.0.1 [57] and Fit-Hi-C2 v2.0.7 [58] were used to call significant cis-interaction loops from merged replicate matrices at three resolutions—1 kb, 2 kb, and 4 kb—with the significance threshold at p-value < 0.1 and FDR < 1×10−6, respectively. The identified interaction loops were merged between the two tools, and then merged across resolutions prioritizing higher resolution for overlaps.
**Definition of cis-Regulatory Elements (cREs):** We intersected ATAC-seq OCRs of each cell type with chromatin interaction loops determined by Hi-C/Capture-C of the same cell type, and with promoters (-1,500/+500 bp of TSS) defined by GENCODE v40.

**Reformatting of the GWAS summary statistics for the four substance use disorders:** Table S2 lists the studies from which we drew European-population GWAS summary statistics for each of the four SUDs. We applied `--merge-alleles` with the list of HapMap3 variants to standardize all the GWAS summary statistics files. The baseline model LD scores, plink files, allele frequencies, HapMap3 variants list and regression weight files for the European 1000 genomes project phase 3 in GRCh38 were downloaded from [https://alkesgroup.broadinstitute.org/LDSCORE/GRCh38/](https://alkesgroup.broadinstitute.org/LDSCORE/GRCh38/).

**Cell type specific partitioned heritability of each trait:** We used LDSC v.1.0.1 with `--h2` flag [46] to estimate the SNP-based heritability of each trait within 5 sets of input regions from each cell type: (1) OCRs, (2) OCRs at gene promoters, (3) cREs, (4) cREs with an expanded window of ±500 bp, (5) OCRs that were not cREs and not at a gene promoter. Each set of input regions from each cell type was used to create the annotation, which in turn was used to compute annotation-specific LD scores for each cell type region of interest. These annotation-specific LD scores were used with 53 categories of the full baseline model (v2.2) to compute partitioned heritability.

**Genetic correlation analysis:** We used LDSC with `--rg` flag [59] to compute the genetic correlations between each of the SUDs and type 2 diabetes (T2D) using European-ancestry meta-analysis summary statistics from the most recent T2D GWAS [60]. The genetic correlation between each pair of traits was computed with unconstrained intercepts. We computed global genetic correlations using standardized variants from the GWAS of each trait. We then partitioned the variants of each trait into cREs of each cell type and recomputed the cell type-specific genetic correlation between each SUD and T2D.
RESULTS

Enrichment patterns across varying open chromatin region definitions

To investigate the enrichment of GWAS-discovered genetic variants associated with the SUDs across the different cell types, we employed Stratified Linkage Disequilibrium Score Regression (S-LDSR) [46] across all cis-regulatory elements (cREs) [61,62] identified through ATAC-seq and chromatin capture/Hi-C analysis for each cell type (Figure 1). We integrated our repertoire of 59 cell types (Table S1) with the recent European-population-GWAS summary statistics for TUD, AUD, CanUD and OUD (Table S2).

We assessed cell-type specific enrichment of GWAS signals in three genomic regions: 1) Total OCR: Open chromatin regions defined by ATAC-seq; 2) Promoter OCR: OCRs overlapping a gene promoter; 3) cREs: OCRs not overlapping a promoter but showing a chromatin loop with a promoter (putative enhancers/suppressors). Different GWAS variants can affect a phenotype by regulating gene expression through mechanisms like altering enhancer function (cREs) or transcription factor binding at promoters (Promoter OCRs). As a control, we also assessed open chromatin regions not overlapping a cRE or promoter ("not-cREs/Prom OCRs"), expecting no significant enrichments.

For all cell types, the variants within the total set of OCRs (Fig.1 “Total OCRs” columns) showed a positive risk heritability (Enrichment >1) for at least one disorder, including in 44 cell types for AUD, 51 for TUD, 53 for CanUD, and 31 for OUD. Of these, only 14 unique cell types (12 for TUD, 4 for AUD, 6 for CanUD and 2 for OUD) showed statistically significant enrichments (P<0.05). When limiting the analysis to OCRs at gene promoters (Fig.1 “Promoter OCRs” columns), the enrichments were significantly decreased for immune cell types, while being increased significantly for metabolic and other cell types. This OCR category also yielded the greatest variability of enrichment (Enrichment standard error) in all cell types and all disorders (Fig S1). Further constraining the LD enrichment assessment to include only OCRs that putatively regulate gene expression via chromatin contacts with gene promoters (Figure 1 “cREs” column) significantly increased the enrichment for most of the cell types for...
all disorders, decreased the $P$-values, and moderately narrowed the dispersion of enrichment ranges across different cell types compared to the promoter OCRs. The original S-LDSR method analyzed enrichment in the 500-bp flanking regions of their regulatory categories [46]. Our expanded analysis of the ±500 bp window for our cREs incorporated more weighted variants in the enrichment (Fig S1 – Prop.SNP), nearly as many as for the total OCRs in most instances. For GWAS signal distributions with sharp peaks, expanding the region by 500 bp dilutes the signal and increases $P$-values and enrichment standard errors without increasing heritability enrichment, as previously observed [63]. However, for SUDs, this expansion yielded stronger signals in LDSC statistics, lower $P$-values and enrichment standard errors, and greater enrichment across cell types. This suggests that the greater effect size of genetic signals for these disorders spans broader regions than our cREs.

Conversely, when we analyzed OCRs located outside both cREs and promoters (“not-cREs/Prom OCRs”) as a control assessment (Fig S2), we observed generally lower enrichment than for the cREs with their expanded regions. However, an opposite pattern of enrichment was observed across the cell types ($R = -0.01$ to -0.2 comparing enrichment of “cREs” versus “not-cREs/Prom OCRs”, Fig S3), suggesting distinct functions of the GWAS variants attributed to the phenotype within different cellular system settings. (Detailed results in Table S3).

Some cell types showed positive enrichment across all OCR categories, albeit at different levels of significance, though the numbers differed for the four disorders: 25 cell types for AUD, 29 for TUD, 6 for OUD, and 24 for CanUD. The iPSC-derived cortical neurons and pancreatic beta cells are two cell types that consistently showed positive enrichment across all OCR categories for all four disorders.

**Diverse enrichment patterns of SUDs across cell types and chromatin regions**

The general patterns mentioned above varied across the different SUDs. The AUD enrichment was more evident in the neural cell types (Fig S4A). iPSC-derived cortical neurons and neural progenitor cells and embryonic stem cell (ESC)-derived hypothalamic NPCs and neurons showed significantly positive enrichment across
almost all OCR categories. Only astrocytes showed significant enrichment within “not-cREs/Prom OCRs”. On the other hand, several liver and pancreatic cell types consistently displayed positive enrichment in all OCR categories. Interestingly, hepatocytes and pancreatic beta cells showed significant enrichment for AUD within total OCRs and “not-cREs/Promoter OCRs”, while pancreatic alpha cells showed significant enrichment for the expanded cRE regions.

TUD also revealed highly significant enrichment within neural cell types, with greater strength within the cREs and promoter OCR regions, and less enrichment within the control OCR categories (Fig S4B). Interestingly, the pancreatic alpha and beta cell types displayed significant enrichment consistently across all OCR region categories. Skeletal myotubes, melanocytes and hMSCs also showed enrichment for TUD loci within promoter OCRs and cRE types, while some naïve T-cell types displayed significant enrichment for the “not-cREs/Prom OCRs” regions.

CanUD displayed a contrasting enrichment pattern for OCR categories (Fig S4C). The neural cell types showed enhanced enrichment with the ‘not promoter-OCRs or cREs’ regions. Metabolic cell types, including several pancreatic, liver, myotube, and osteoblast cell types, yielded significant enrichment for cRE regions. This disorder also showed the highest involvement of immune cell types, especially various T-helper cells.

OUD signals had the most sporadic pattern and the greatest number of negative enrichments (Fig S4D). Nonetheless, we observed significant positive enrichments for cortical neurons and pancreatic beta cells.

**Cortical neurons, neural progenitors, and pancreatic cells as key players in SUD pathophysiology and therapeutic targets**

Despite the observed diversity in patterns across distinct SUDs, several cell types consistently exhibited significant levels of enrichment. Notably, iPSC-derived cortical neurons and neural progenitors, ESC-derived hypothalamus NPCs and neurons, and pancreatic alpha and beta cell types were notably consistent across the four disorders (Fig S5). The observed pattern of enrichment within cREs and their expanded regions
suggest a pronounced association with the genetic architecture underlying SUDs. Conversely, there was a significantly lower level of enrichment within the contrasting OCR outside of both cREs and promoters ("not-cREs/Prom OCRs").

The involvement of pancreatic cell types in SUDs prompted us to consider the possibility of pleiotropy between SUDs and metabolic diseases such as type 2 diabetes. The moderate genome-wide positive genetic correlation (R=0.1) between T2D and both TUD and CanUD was significant, whereas it was significantly negative with AUD (Fig S6). Interestingly, after partitioning the heritability to each cell type's cREs, the correlation between T2D and TUD within pancreatic beta cells was significantly higher than the genome-wide correlation value (R=0.2) (Fig S6A). The significance was established with a bootstrap t-test yielding a p-value of 7e-48 (Fig S6B). This implies that pancreatic beta cells are crucial in understanding the genetic overlap between T2D and TUD, possibly due to shared pathways or regulatory mechanisms active in these cells. This finding highlights the importance of examining cell-type-specific genetic contributions to understand more fully the complex relationship between metabolic disorders and SUDs.
DISCUSSION

We sought to identify cellular contexts mediating the genetic etiology of SUDs by examining enrichment patterns of epigenetic features across diverse cell types. Our approach integrated GWAS summary statistics with ATAC-seq, and promoter Capture C/Hi-C data. Utilizing LD score regression, quantifying polygenic effects and confounding factors, we quantified the contribution of SNP-based heritability to SUD pathogenesis. The observed positive heritability enrichment across various open chromatin features spanning multiple cell types revealed the involvement of metabolic and neural systems.

Within our repertoire of cell types, we hypothesized that most of the GWAS signals would reside in cREs or gene promoters, as shown previously [61-68]. The enrichment patterns observed for AUD and TUD generally aligned with this. However, for CanUD and OUD, the GWAS signals exhibited a distinct pattern not fully captured by our cREs. This complexity makes it challenging to identify and characterize all relevant regulatory elements contributing to the susceptibility to these disorders. The presumed temporal nature of open chromatin regions within a given cell type likely also contributes to less observed enrichment within certain cell type cREs [69].

Notably, cortical neurons and pancreatic cells were consistently positively enriched across all OCR categories for all SUDs, suggesting their potentially crucial role in the pathophysiology of SUDs. The enrichment of cortical neurons and neural progenitor cells in SUDs aligns with accumulating evidence that implicates these cell types in the neurobiology of addiction [70-76].

The apparent involvement of pancreatic alpha and beta cells in SUDs represents a novel and intriguing insight into their pathology, underscoring the interplay between metabolic regulation and addictive behaviors. Chronic exposure to substances like alcohol, opioids, and cannabinoids can disrupt pancreatic function, leading to dysregulated glucose metabolism and increased risk of metabolic disorders like diabetes [77-79]. Chronic alcohol consumption increases risk of pancreatic diseases such as pancreatitis and pancreatic cancer, likely due to the toxic effects of alcohol on pancreatic tissue. However, genetic predisposition can influence the susceptibility of pancreatic tissue to alcohol-induced toxicity, making certain individuals more vulnerable.
Alcohol-induced pancreatic damage can impair insulin secretion, cause beta-cell dysfunction, and dysregulate glucagon secretion, contributing to metabolic abnormalities and the development of diabetes [80-82]. Opioid drugs exert direct effects on pancreatic alpha and beta cells. For example, opioids modulate insulin and glucagon secretion by pancreatic islet cells, potentially contributing to the glucose dysregulation and metabolic disturbances observed in opioid users [83,84]. Endocannabinoid receptors, including cannabinoid receptor 1 (CB1), are expressed in pancreatic alpha and beta cells, suggesting a direct influence of cannabinoids on pancreatic function. The activation of CB1 receptors can impair insulin secretion from beta cells and stimulate glucagon release from alpha cells, implicating the endocannabinoid system in the regulation of pancreatic hormone secretion and glucose metabolism [85-87]. Furthermore, a recent study uncovered a direct link between pancreatic cells and nicotine use, mediated by the key diabetes-associated transcription factor TCF7L2 [88]. The study showed that TCF7L2 regulates nicotine intake by modulating nicotinic receptors in the habenula, a brain region involved in stress responses. This regulation influences blood glucose levels through signaling via the autonomic nervous system to the pancreas, with chronic nicotine use disrupting normal glucose regulation.

Interestingly, pancreatic beta cells share features with the brain, including a common set of expressed genes, which suggests an evolutionary link [89]. During pancreatic organogenesis, the depression of Polycomb enables beta cells to present with a neuronal gene expression program [90], with the overall beta cells’ gene expression pattern and chromatin marks being closer to neuronal tissue types than other tissues. Insulin-producing neurons precede beta cells phylogenetically, with rodents, human (in vitro), and most invertebrates having neurons that produce insulin to manage blood glucose levels [91]. Given the shared features of pancreatic beta cells and neurons that extend from transcription to post-transcriptional regulation [92], understanding brain-pancreas interactions could aid in understanding the relationship between addiction and metabolic health.

Several cell types exhibited a degree of negative heritability enrichment, with all traits yielding a negative enrichment in at least one cell type. Significant negative
enrichments were observed in a few cell types, for different disorders, and for different OCR categories. However, these cell types displayed typical levels of enrichment for other traits and those traits demonstrated normal enrichment ranges across other cell types. This suggests that the observed negative enrichments are unlikely to stem from misclassified alleles in trait summaries or biased genetic regions of interest within cell types.

Our findings in immune cell types are consistent with numerous studies that have shown the influence of SUDs on the immune response system and associated cell types. For instance, there are immunomodulatory effects of chronic alcohol consumption, including alterations in cytokine production, heightened CNS inflammation, impaired immune cell function, and increased susceptibility to infection [93-97]. Similarly, tobacco use has been linked to systemic inflammation, immune dysregulation, and heightened risk of autoimmune diseases [98-104]. Additionally, opioid misuse has been shown to disrupt immune homeostasis, leading to immunosuppression [83,105-109]. Finally, cannabis use has been associated with immunomodulatory effects, such as altered cytokine profiles, impaired T-cell function, and dysregulated immune cell signaling [110-114].

In considering these effects, it is important to factor in the effective sample sizes of the GWAS efforts, which can introduce noise and contribute to negative enrichments observed in regression analyses. This limitation, inherent to partial linkage regression methodologies, highlights the need for careful interpretation of disease variant enrichment in specific cellular contexts.

In conclusion, by elucidating relevant regulatory regions across various cell types, we gained insight into the genetic underpinnings of SUDs. Our study has indicated in which specific cellular environments the genetic susceptibility for SUDs appear to lie. These observations warrant further research aimed at unraveling the underlying causes of these disorders in specific cellular contexts. This could offer new avenues for advancing our understanding of SUDs and aiding in the development of more effective targeted interventions.
ACKNOWLEDGEMENTS
S.F.A.G. is the Daniel B. Burke Endowed Chair for Diabetes Research.

AUTHOR CONTRIBUTIONS
Khanh B. Trang: Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization
Alessandra Chesi: Formal analysis, Investigation, Data Curation, Writing- Reviewing and Editing
Sylvanus Toikumo: Resources, Data Curation, Writing- Reviewing and Editing
James A. Pippin: Resources
Matthew C. Pahl: Methodology, Writing- Reviewing and Editing
Joan M. O’Brien: Resources
Laufey T. Amundadottir: Resources
Kevin M. Brown: Resources
Wenli Yang: Resources
Jaclyn Welles: Resources
Dominic Santoleri: Resources
Paul M. Titchenell: Resources
Patrick Seale: Resources
Babette S. Zemel: Resources
Yadav Wagley: Resources
Kurt D. Hankenson: Resources
Klaus H. Kaestner: Resources
Stewart A. Anderson: Resources
Matthew S. Kayser: Funding acquisition, Writing- Reviewing and Editing
Andrew D. Wells: Resources, Writing- Reviewing and Editing
Henry R. Kranzler: Supervision, Funding acquisition, Writing- Reviewing and Editing
Rachel L. Kember: Supervision, Funding acquisition, Writing- Reviewing and Editing
Struan F.A. Grant: Conceptualization, Supervision, Funding acquisition, Writing- Reviewing and Editing

FUNDING
This work was supported by NIAAA grant R01 AA030056 and the Mental Illness Research, Education and Clinical Center of the Crescenz VAMC.
A.C. was supported by NHGRI grant R35 HG011959.
S.F.A.G., W.Y., K.H.K and P.S were supported by NIDDK grant UM1 DK126194.
S.F.A.G. and B.S.Z were supported by NICHD grant R01 HD100406.

COMPETING INTERESTS
Dr. Kranzler is a member of advisory boards for Dicerna Pharmaceuticals, Sophrosyne Pharmaceuticals, Enthion Pharmaceuticals, and Clearmind Medicine; a consultant to Sobrera Pharmaceuticals and Alimmune; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes; a member of the American Society of Clinical Psychopharmacology’s Alcohol Clinical Trials Initiative, which was supported in the last three years by Alkermes, Dicerna, Ethypharm,

The remaining authors have nothing to disclose.
REFERENCES

2. SAMHSA. National Survey on Drug Use and Health. 2022: Table 2.25—Alcohol use in lifetime: among people aged 12 or older; by age group and demographic characteristics, 2021 and 22.
7. WHO. The health and social effects of nonmedical cannabis use. 2016.


FIGURE LEGENDS

Figure 1: Heritability enrichment of substance use disorders across diverse cell types

1st left column: Bar-plot shows total number of open chromatin regions (OCRs) identified by ATAC-seq for each cell type on bulk cells (green), or on single cell (pink); proportion of cREs identified by hi-C (yellow), or Capture-C (blue).

12 dot-plot panels display the heritability enrichment (LDSC analysis) for each cell type across 4 SUDs in 3 categories of OCRs:

- **Total OCRs**: all OCRs identified by ATAC-seq
- **Promoter OCRs**: OCRs overlapped with gene promoters
- **cREs**: of each cell type

Whiskers represent enrichment standard errors. Grey-scaled dots correspond to P-values in -log10, with red asterisks indicating significant $P$-values ($< 0.05$). Dot size corresponds to the proportion of SNP contribution to heritability. Dashed line at 1 indicates no enrichment.