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

Khanh B. Trang[1,2], Alessandra Chesi[1,3], Sylvanus Toikumo[4,5], James A. Pippin[1,2], Matthew C. Pahl[1,2], Joan M. O'Brien[7,8], Laufey T. Amundadottir[9], Kevin M. Brown[9], Wenli Yang[10,11], Jaclyn Welles[10,12], Dominic Santoleri[10,12], Paul M. Titchenell[10,12], Patrick Seale[10,11], Babette S. Zemel[6,13], Yadav Wagley[14], Kurt D. Hankenson[14], Klaus H. Kaestner[10,15], Stewart A. Anderson[16,17], Matthew S. Kayser[5,18,19], Andrew D. Wells[1,3,20], Henry R. Kranzler[4,5], Rachel L. Kember[4,5], Struan F.A. Grant[1,2 ,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. Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.
6. Division of Gastroenterology, Hepatology, and Nutrition, Children's Hospital of Philadelphia, PA, USA
7. Scheie Eye Institute, Department of Ophthalmology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, PA, USA
8. Penn Medicine Center for Ophthalmic Genetics in Complex Disease, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, PA, USA
9. Laboratory of Translational Genomics, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA
10. Institute for Diabetes, Obesity and Metabolism, 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

Corresponding author: Struan F.A. Grant

Address:
Children's Hospital of Philadelphia
1102D ARC, 3615 Civic Center Boulevard
Philadelphia, PA 19104
Office telephone: 215-200-0196

Email: grants@chop.edu

ABSTRACT

Recent genome-wide association studies (GWAS) revealed shared genetic components among substance use disorders (SUDs). However, the extent of underlying shared causal variants, effector genes, and cellular contexts, remain unclear. We integrated 3D genomic datasets (high-resolution promoter-focused Capture-C/Hi-C, ATAC-seq, RNA-seq) from 59 diverse human cell types with recent GWAS summary statistics for alcohol (AUD), tobacco (TUD), opioid (OUD), and cannabis use disorder (CanUD). Using stratified LD regression, we determined the proportion of SNP heritability attributable to features in these cell types. We observed significant enrichments ($P<0.05$) in 25 cell types. Heritability reached 4.2-fold enrichment for CanUD in iPSC-derived cortical neurons and 6-fold for AUD/TUD in neural progenitors, confirming their relevance for functional exploration. Additionally, pancreatic cell types, notably insulin-secreted beta cells, showed heritability enrichment for TUD (4.8-fold) and CanUD (5.8-fold), suggesting genomic overlap with metabolic processes. Further investigation revealed significant positive genetic correlations between AUD, TUD, and CanUD with type 2 diabetes (T2D) (FDR<0.05). A multi-omic integrated assessment of T2D-SUD correlated loci revealed strong correlations between gene expression and chromatin accessibility at corresponding cis-regulatory elements and shared functional modules. Our study provides new genomic insights into SUDs and implicates specific cell types for targeted functional follow-up to pinpoint causal variant-gene mechanisms.

INTRODUCTION

Substance use disorders (SUDs) are a group of treatable psychiatric disorders 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 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 show strong familial inheritance patterns [10], with heritability estimates from twin studies for these disorders averaging ~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 for both use and misuse [17, 18]. OUD heritability is around 50%, with 38% of variation attributed to opioid-specific genetic factors [19]. In addition to genetic influences on

specific SUDs [20, 21], broader heritable factors influence SUD susceptibility in general [22].

Multiple large genome-wide association studies (GWAS) of SUDs conducted over the past ten years have yielded 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 the *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 both 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, a comprehensive mechanistic understanding of effector genes and the cellular contexts at GWAS loci remains limited. While some studies [46] have begun to dissect specific effector genes and cell types at individual loci, GWAS generally identifies broad genomic regions associated with traits and lacks cell-type resolution. a.

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 [47]. 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 genetic etiology of SUDs.

METHODS

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

ATAC-seq preprocessing and peak calling: Open chromatin regions (OCRs) were called using the Nextflow pipeline *https://nf-co.re/atacseq* [48]. Reads were aligned to the GRCh38/hg38 assembly genome using Chromap [49], duplicates were removed, alignments from all replicates were pooled, and narrow peaks were called using MACS2. ScATAC-seq data from pancreatic cells was analyzed separately [50], fragments from the same cell ID groups were pooled into a single bam file for each cell type per sample, which were converted to fastq and used as input for the Nextflow pipeline.

RNA-seq preprocessing and expression profiling: the detail configurations, steps, and technical details for each dataset are provided in the original studies. In brief, read fragments from fastq files were mapped to genome assembly GRCh38/hg38 using STAR, independently for each replicate and condition. We used GENCODE annotation files for feature annotation and htseq-count for raw read count calculation at each feature. Read counts were transformed into TPM (transcript per million) and normalized internally between replicates/conditions in each individual study. For comparative measurements, we transformed all the expression values into 0-100 scale.

Promoter Capture-C pre-processing and interaction calling: Paired-end reads were pre-processed using the HICUP pipeline [51] with bowtie2 and GRCh37/hg19. Significant promoter interactions were called using unique read pairs using CHICAGO [52]. We analyzed individual fragments (1frag) and binned four fragments to improve long-distance sensitivity [53]. 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 [50], paired-end reads from each replicate were pre-processed using the HICUP v0.7.4 pipeline [51] and aligned by bowtie2 with GRCh38/hg38. The alignment files were

parsed and processed by pairtools v0.3.0 [54] and indexed and compressed by pairix v0.3.7 [55], 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 [56] and normalized using the ICE method [57]. The matrices from different replicates were merged at each resolution using cooler. Mustache v1.0.1 [58] and Fit-Hi-C2 v2.0.7 [59] 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 v47.

Definition of Specifically Expressed Genes set (SEGs): Normalized transcripts per million (TPM) of all measured genes in 46 of 59 cell types was used to perform differential analysis using *DEseq2* package [60], where cell type and system (immune, metabolic, neural and other) were used as variables for the modeling contrast. we used *apeglm* method for effect size (logarithmic fold change estimates) shrinkage [61] to alleviate this phenomenon during the genes ranking. We followed the original method [62], chose 10% of the highest-ranking genes (3000-5000) for each cell type, added 100kb windows on either side of the transcribed region of each gene in the set of specifically expressed genes to construct a “SEGs” annotation corresponding to that cell type.

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/.

Cell type specific partitioned heritability of each trait*:* Our analysis investigated the cell-type specific enrichment of GWAS signals across different regulatory landscapes to elucidate potential mechanisms of disease association, using LDSC v.1.0.1 with *--h2* flag [47] to estimate the SNP-based heritability of each trait. We categorized the genome based on chromatin accessibility and interactions from each cell type into: 1) Total OCR: For a global assessment of enrichment. 2) Promoter OCR: To test the hypothesis that GWAS variants may impact gene expression by directly affecting promoter activity. 3) cREs: To evaluate the contribution of distal regulatory elements, such as enhancers and suppressors identified by chromatin loops, to GWAS associations. As a negative control to ensure the specificity of enrichment in regulatory regions, we also examined open chromatin regions not classified as promoters or looping cREs ("not-cREs/Prom OCRs"), where we expected to observe minimal signal; and as positive controls, sets of cell-type-specifically expressed genes (SEG) [62] for each cell type. 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 63 categories of the full baseline model (v2.2) as baseline control. Two statistics we used to assess how effectively our annotations capture causal variation are heritability enrichment and standardized effect size (τ^*), as previously defined [63]. When conditioning two correlated annotations in a joint S-LDSC model, they may show similar enrichments, but the τ^* for the annotation with a higher true causal variant membership will be larger and more positive.

Genetic correlation analysis: We used LDSC with --rg flag [64] 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 [65]. 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.

Local genetic correlation analysis*: LAVA (Local Analysis of [co]Variant Association) v0.1.0* [66] was used to estimate genetic correlations within local genomic regions across 10 pair-wise combinations of 4 SUDs and T2D. It accounts for the correlation across LD SNPs by converting marginal SNP effects into their combined effects, based on an external LD reference. The reference genome data were the same as *LDSC* analysis, however we used the GWAS without *--merge-alleles* with the list of HapMap3. For locus definition, we followed the LAVA partitioning algorithm manual (*https://github.com/cadeleeuw/lava-partitioning*) with default parameters to partition human genome build 38 into blocks of approximately equal size (~1 Mb) while minimizing the LD between them, resulting in 7351 LD blocks, excluding the extended MHC region chr6: 28,510,120-33,480,577. To account for the sample overlap, we used the pair-wise intercepts from the bivariate global LDSC to create a symmetric matrix, then converted it to a correlation matrix and provided it to LAVA. For each pair of traits, bivariate analysis was performed only for loci in which both traits exhibited univariate signals at $P < 0.05/7351$. 7,343 were detected in univariate analysis as simultaneously significant in at least one pair of traits and were then subjected to bivariate testing, resulting in 20582 bivariate tests conducted in total. *P*-value for local rg $< 0.05/20582$ was considered significant

sPLS-DA and DIABLO Analysis: To investigate the shared genetic pathways and regulatory mechanisms underlying the observed genetic overlap between T2D and SUDs, particularly focusing on pancreatic beta cells and neural cell types, we employed multivariate dimension reduction and integration techniques using the *mixOmics* [67] (v6.18.0) R package. Initial exploratory analysis of cell-type-specific gene expression profiles was performed using *sparse Partial Least Squares Discriminant Analysis (sPLS-DA)*. This supervised method was applied to RNA-seq data from pancreatic alpha and beta cells and 4 neural cell types. This analysis helped to visualize the relationships between different cell types based on their transcriptomic signatures and shared expression patterns. To integrate the transcriptomic and chromatin accessibility data within the genomic loci exhibiting significant genetic correlation between T2D and SUDs (specifically TUD), we utilized the *Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)* framework. This supervised integration approach

was applied to the subset of genes and OCRs, including promoter OCRs and cREs, located within the identified correlated loci. These tools were built to discriminate between different groups (cell types in our case) based on their molecular profiles. They aim to find latent components that maximize the covariance between the omics data and the categorical outcome (cell types), thereby enhancing the separation of cell types.

***Pathway Enrichment Analysis*:** Following the identification of key genes and regulatory elements through sPLS-DA and DIABLO analyses, pathway enrichment analysis was performed using the *pathfinder* [68] and *clusterProfiler* [69] R package. Genes identified as significantly contributing to the separation or integration patterns in the sPLS-DA and DIABLO models were used as input. These tools leverage active subnetworks to identify enriched pathways and gene sets from various databases (e.g., KEGG, Reactome, GO), providing a functional interpretation of the molecular features highlighted by the multivariate analyses. This step aimed to consolidate the list of individual genes and regulatory elements into broader biological contexts, revealing the specific pathways and mechanisms potentially shared between T2D and SUDs in the studied cell types.

RESULTS

Enrichment patterns across varying open chromatin region definitions

Our primary objective was to investigate the enrichment of GWAS-discovered variants associated with the SUDs across the range of cell types for which we have generated 3D genomic data. We employed Stratified Linkage Disequilibrium Score Regression (S-LDSR) [47] across all 59 human cell types, which encompass immune, metabolic, neuronal, and pluripotent stem cell lineages (**Table S1**), utilizing the recent European-population-GWAS summary statistics for TUD, AUD, CanUD, and OUD (**Table S2**). Leveraging our ATAC-seq and chromosome conformation capture datasets, including promoter Capture-C (PCC) and Hi-C, we generated maps of promoter-connected open chromatin (putative cRE) regions to pinpoint specific functional categories of the genome enriched for trait-associated genetic variation.

We assessed the cell-type specific enrichment of GWAS signals in specific genomic regions: total open chromatin regions (total OCRs), promoter-associated open chromatin regions (promoter OCRs), and putative regulatory elements contacting promoters via chromatin loops (cREs, the putative active and poised enhancers/silencers). We hypothesized that trait-associated variants are significantly enriched in cREs and promoter OCRs, reflecting their potential roles in gene regulation (e.g., altering enhancer or promoter activity). Conversely, we expected minimal enrichment in open chromatin regions not classified as cREs or promoters ("not-cREs/Prom OCRs"), which served as negative controls. As positive controls, we also examined sets of cell-type-specifically expressed genes (SEGs). We measured the heritability enrichment for each genetic annotation for each cell type per each trait (Error! Reference source not found.) and their standardized annotation effect size (τ^*, **Figure 1A**) focusing on events with enrichment *P*-value ≤0.05.

For all cell types, the variants within the total set of OCRs (**"Total OCRs"**) showed a positive risk heritability (enrichment >1) for at least one SUD, including in 52 cell types each for AUD, TUD and CanUD, and 50 for OUD. Of these, in only 9 unique cell types (7 for TUD, 4 for AUD, 5 for CanUD and 1 for OUD) were the enrichments statistically significant (*P* ≤0.05).

When limiting the analysis to just promoter OCRs, the enrichments were significantly lower for immune cell types, markedly greater for metabolic and some other cell types. This OCR category also yielded the greatest variability in enrichment (enrichment standard error) in all cell types and SUDs (**Fig. S2**). While the enrichment in promoter OCRs was positively correlated with SEGs in all cell types with available expression profiles (r^2=0.24), there was a negative correlation in neural cells (r^2=-0.1) and other cell types using τ^* (r^2=-0.44) (**Fig. S3**). This suggests that promoter accessibility, which one expects to be associated with gene expression, behaves inconsistently across cell types.

By further constraining the LD enrichment assessment to include only cREs, we observed greater enrichment for most of the cell types for all disorders, lower *P*-values, and moderate narrower dispersion of enrichment ranges across different cell types compared to the promoter OCRs. This approach also shifted the enrichment more toward neural and pancreatic cell types. The original S-LDSR method analyzed enrichment in the 500-bp flanking regions of their regulatory categories [47]. Our expanded analysis of the ±500 bp window for our cREs incorporated more weighted variants into the enrichment, nearly as many as for the total OCRs in most instances. For GWAS signal distributions with sharp peaks, expanding the region by 500 bp has been reported to dilute the signal and increase *P*-values and enrichment standard errors without increasing heritability enrichment [70]. However, for SUDs, this expansion yielded stronger signals using LDSC statistics, lower *P*-values and enrichment standard errors, and generally 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, we observed generally lower enrichment than for the cREs with their expanded regions. The inconsistencies observed highlight that even when considering DEG and not_cRE/Prom_OCRs as reference points, the other OCR categories demonstrate context-specific and metric-

dependent correlations, reflecting the complex interplay between different regulatory elements and gene expression across cell types.

Collectively, these findings support a model in which genetic risk for SUDs is primarily mediated through distal regulatory elements closely linked to SEGs, rather than promoters or general chromatin accessibility, emphasizing the importance of precise regulatory annotation in functional genomic analyses. Full results are presented in **Table S3**.

Cell Type Prioritization Based on τ^*

Some cell types showed positive enrichment across all OCR categories, albeit at different levels of significance with the differingfrequencies differed for the four SUDs: 25 cell types for AUD, 29 for TUD, 6 for OUD, and 24 for CanUD. Multiple neural lineages demonstrated statistically significant τ^* values for at least one trait. For example, iPSC-derived neurons exhibited positive τ^* values for OUD and TUD, indicating a preferential localization of risk variants in their cREs. Likewise, hESC-derived hypothalamic neurons showed τ^* enrichment for AUD, CanUD, and OUD. Among pancreatic cell types, particularly the alpha and beta cells, modest τ^* enrichment was detected for TUD and CanUD, suggesting involvement of metabolic regulatory pathways. This was corroborated by similar enrichment in adipocytes and liver-derived cell types (e.g., HepG2, hepatocytes), aligning with the known metabolic consequences and systemic impacts of chronic substance use. In immune cells, dendritic cells (e.g., cDC_CD1c), monocytes, and various T-cell subsets (e.g., activated T-helper Th17, T-reg, Tr1) exhibited enriched τ^* for multiple SUDs, especially AUD and CanUD. These findings reinforce the growing evidence of immune-neural cross-talk and neuroinflammation in addiction vulnerability [71].

The iPSC-derived cortical neurons and pancreatic beta cells are two cell types that consistently showed positive enrichment and high adjusted contributions across all OCR and cRE categories for most of the SUDs. Most of the enrichment was significantly

positive (**Fig. 1A**), with a few exceptions that included pancreatic acinar cREs with CanUD, osteoblast cREs with OUD and adipocyte OCRs with OUD.

Diverse enrichment patterns across SUDs

After establishing the regulatory annotation categories and cell types most broadly enriched for heritability, we turned our attention to how these enrichments manifest in a trait-specific manner. All four SUDs shared enrichment in multiple neuronal lineages, indicating a common neurobiological basis to these traits. However, trait-specific differences also emerged, reflecting the relatively low genetic correlations among the four traits (**Fig. 2A**).

The AUD enrichment was more evident in neural cell types, although several of these—e.g., conventional dendritic cells (cDC_CD1c), regulatory T-cells (Treg and Tr1), and monocytes—exhibited significant positive τ^* in cRE-based annotations. In the neural component, iPSC-derived cortical neurons and neural progenitor cells and embryonic stem cell (ESC)-derived hypothalamic NPCs and neurons showed significant positive enrichment in cREs and extended categories. Astrocytes, beta cells and iPSC-derived neural progenitors showed significant enrichment within “not-cREs/Prom OCRs” but with low and non-significant standardized effects.

TUD was also highly significantly enriched in neural cell types, which was greater within the cREs and promoter OCR regions and less within the control OCR categories. Interestingly, the pancreatic alpha and beta cell types revealed significant enrichment consistently across all OCR region categories, while some immune cell types displayed significant enrichment for the “not-cREs/Prom OCRs” regions.

OUD showed a different enrichment pattern for OCR categories, exhibiting the broadest cell-type enrichment spectrum, and the greatest number of negative enrichments. The metabolic cell types showed enhanced enrichment with the “not-cREs/Prom OCRs” regions. Notably, both pancreatic alpha and beta cells showed positive τ^*, along with adipocytes and preadipocytes, likely reflecting a broader systemic involvement in addiction vulnerability, particularly of neuro-metabolic and endocrine

pathways. This finding supports an integrative perspective on OUD as not purely a brain disorder, but one potentially influenced by genetic regulation of peripheral cell types that modulate physiological responses to opioid exposure. OUD also showed the greatest involvement of immune cell types, especially various T-helper cells, suggesting a multifaceted genetic architecture involving neuro-immune-metabolic axes. CanUD also showed strong involvement of immune cells and osteoblast cells but maintained the positive enrichment within the neural cell type.

The low overall heritability of CanUD (h^2 = 0.015) and OUD (h^2 = 0.028) likely explains why cREs appear to contribute disproportionately to per-SNP heritability in these traits. In these cases, cell-type-specific cREs may capture high-impact variants more effectively than the baseline model, resulting in higher enrichment estimates and stronger τ^* effect sizes for these disorders.

Cortical neurons, neural progenitors, and pancreatic cells as key observations

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 hypothalamic NPCs and neurons, and pancreatic alpha and beta cell types were notably consistent across the four disorders (**Fig. 1B**). Across all six cell types, enrichment and τ^* were consistently highest for cREs ±500 bp, followed by cREs alone, with substantially lower values in promoter OCRs and not_cRE/promoter OCRs. This pattern implicates distal regulatory elements, especially those flanked by broader accessible chromatin, as the principal contributors to the heritability signal of SUDs.

Local genetic correlation yielded polygenic loci between SUDs and type 2 diabetes

The involvement of pancreatic cell types in SUDs prompted us to consider the possibility of pleiotropy between SUDs and metabolic traits, in particular T2D. Using LDSC [64] to estimate global SNP-based genetic correlations (global r_g) among SUDs and T2D, we observed that all positive genetic correlations were statistically significant,

albeit moderate in magnitude (max global T2D-SUDs r_g=0.03) (**Fig. 2A**). Individually, coronary heart disease and OUD yielded the highest correlation (global r_g=0.25).

Because a global correlation represents an average of the shared association across the genome, it can be blunted by local correlations in opposing directions. Thus, we carried out local genetic correlation (local r_g) analyses using LAVA [66] to detect genetic regions associated with each SUD and T2D. LAVA revealed that bivariate local r_gs across 1,487 loci achieved significance after multiple correction (FDR < 0.05). Of the significant correlations, 134 had 95% confidence intervals (CIs) for the explained variance that included a value of 1, indicating that the local genetic signals for those traits are entirely shared, and that most were between CanUD and AUD (83 loci) (yellow bar in **Fig. 2B**). Most of the 1,550 significantly correlations reflected pair-wise correlations between CanUD, AUD and TUD (**Figure 2C**), with the results reflecting genome-wide findings with LDSC.

We evaluated the strength and direction of the local r_gs by calculating the mean of the observed correlation coefficients across all tested loci. In **Fig. 2A** we aligned the mean correlations of all the local r_gs (upper triangle) with the global r_gs (lower triangle) of all 10 pairs of traits (full LAVA results of each locus in **Table S4**). Most pairs showed consistently positive significant global correlation r_g with multiple bivariate significant correlated r_gs loci (red numbers). Despite the low global correlation between T2D and SUDs, there were strong local correlations at several loci, with the strongest being with CanUD (mean r_g=0.34).

The LAVA mean local r_gs yielded a correlation of 0.83 with LDSC estimates (**Fig. 2D**), indicating an overall concordance between the two methods. We observed a slight upward trend in the mean LAVA estimates, indicating that the filtering on univariate h^2 may produce bias in the estimated correlations. These findings convincingly demonstrate that individual genetic loci exhibit varying degrees of heritability and correlation, perhaps even more pronounced in diverse cellular contexts, and may not always be identifiable through genome-wide assessments.

Of the 1550 significantly correlated, 114 were between T2D and at least one SUD. The locus that was most highly genetically correlated, chr4:99187511-99411908, yielded

significant correlations between T2D-AUD (r_g=0.8). More than half of these loci (64 loci) were correlated between T2D and TUD (r_g=0.04 – 0.6), 27 between T2D-AUD and 25 between T2D-CanUD with higher correlation (r_g=0.05 – 0.8 and r_g=0.1 – 0.8 respectively) (**Fig. 2F**). Thus, despite moderate genome-wide overlap, substantial genetic architecture is shared between SUDs and T2D.

Common and distinct regulatory mechanisms in neural and pancreatic cells within the genetic overlap between T2D and SUDs

Our observations suggest pancreatic beta cells are crucial for understanding the genetic overlap between T2D and SUDs, especially TUD, likely due to shared pathways or regulatory mechanisms within these cells. This underscores the need to examine cell-type-specific genetic contributions to fully grasp the intricate link between metabolic disorders and SUDs.

Inspecting the top specifically expressed genes of these 6 cell types showed a relatively large number of genes specific to each cell type (**Fig. S4**) that contribute to their unique identity and function. The 4,717 genes shared by various combinations of neural and pancreatic cell types suggest that there could be common underlying genetic programs.

Indeed, sparse Partial Least Squares discriminant analysis (sPLS-DA) from mixOmics using all expressed genes from the two pancreatic cell types and four neural cell types showed a high degree of similarity between pancreatic and neural cells (**Fig. 3A**). The first two principal components clearly differentiated iPSC-derived neurons and neural progenitor cells (NPCs), driven by genes crucial for neuronal development, signaling, and, for NPCs, chromatin regulation and cell cycle control. A fourth component characterized hESC-derived hypothalamic neurons, highlighting genes involved in membrane lipid composition, transporters, and neurotransmitter signaling (**Fig. 3B**, **Fig. S5**). Pancreatic alpha cells were separated only in component 3, driven by *DCAF6*, a gene linked to primary ciliary function, which may influence insulin secretion via ciliary signaling in beta cells, highlighting its potential relevance to diabetes pathophysiology.

Given these overlaps in genes expression profiles between pancreatic and neural cells, we focused on the subset of genes and OCRs that reside within the 114 GWAS loci that were significantly correlated between T2D and SUDs. These involved ~1500 genes, 3648 promoter OCRs and ~2866 cREs. We used the *Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO)* framework to capture the major biological overlap among cell types. The combination of RNA-seq and ATAC-seq at 2 genomic annotations separated the cell types more granularly via an optimized 5-component model, with high correlations between 3 data blocks (**Fig. 3C-D, Fig. S6A-B**). Each component revealed distinct biological processes and key genes: from membrane dynamics in primary lineages (Component 1) to signaling pathways in hESC-derived neural progenitors (Component 2), and dopamine signaling in hypothalamic neurons (Component 3). Further, cytoskeletal function defined iPSC-derived neural progenitors (Component 4), while membrane/lipid metabolism distinguished pancreatic alpha and beta cells (Component 5) (**Fig. 3E**). This highlights that the molecular signatures differentiating these cell types are rooted in core biological processes essential for their unique identities and functions.

Mapping the cRE sets to the shared biology of SUDs and T2D at each of 114 loci revealed substance-specific functional enrichments (**Fig. 3F**). The AUD-T2D gene set shows enrichment in functions such as alcohol dehydrogenase activity, NAD-retinol dehydrogenase activity, and various carboxypeptidase activities, suggesting potential links between alcohol metabolism, other metabolic processes, and protein modification. The TUD-T2D gene set is associated with functions that include microtubule and tubulin binding, outward rectifier potassium channel activity, and symporter and secondary active transmembrane transporter activity, indicating the involvement of cytoskeletal dynamics, ion transport, and broader transmembrane transport mechanisms in the TUD-T2D context. The CanUD-T2D gene set demonstrates enrichment in functions such as ATP-dependent diacylglycerol kinase activity, mRNA binding (5'-UTR and 3'-UTR), and structural constituent of eye lens, pointing towards roles in lipid metabolism, gene regulation at the mRNA level, and potentially ocular complications in individuals with CanUD and T2D. These findings underscore the biological heterogeneity

underlying SUD-T2D comorbidities and emphasize the need for substance-specific approaches to research and treatment efforts.

Our analysis aimed to identify molecular drivers of multi-omic patterns and cell type relationships within disease-associated loci. Despite advanced analytic approaches, clear separation of neural and pancreatic cell types proved challenging, revealing extensive overlap in their genetic and chromatin landscapes. While some analyses highlighted cell-type-specific drivers related to neuronal development, signaling, or metabolic functions, the overall picture points to a profound shared biological basis, especially within the 114 loci linked to both T2D and SUDs. This significant overlap suggests common underlying genetic programs and regulatory mechanisms contribute to the complex interplay between metabolic and neuropsychiatric disorders, highlighting shared pathways active in pancreatic beta cells and neurons.

DISCUSSION

We sought to identify cellular contexts that mediate the genetic etiology of SUDs by examining enrichment patterns of epigenetic features across diverse cell types. While annotating SNPs based on proximity to trait-associated genes seems intuitive, more extensive analysis is required to pinpoint the effector mechanisms. Because cell-type-specific eQTL mapping remains technically challenging and often tags genes with simpler regulation and less functional relevance [72], we used SEG-based LDSC as a more tractable control. Although limited by broad SNP windows that dilute signal specificity, this approach provides a practical baseline for assessing cell-type relevance.

Our approach, which integrated GWAS summary statistics with ATAC-seq and promoter Capture C/Hi-C data, focusing on experimentally defined regulatory regions specific to certain cell types, allows for testing/generating more direct hypotheses about the underlying biological mechanisms. Utilizing LD score regression to quantify polygenic effects and confounding factors, we assessed 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 [70, 73-79]. By partitioning SUD heritability onto cell-type-specific chromatin accessibility maps, we directly tested the hypothesis that genetic variants falling within regions actively regulating gene expression *in specific cell types* contribute disproportionately to SUD risk. This approach simultaneously implicates cell types and highlights the importance of their active regulatory landscapes. Although GWAS SNPs are broadly enriched in regulatory elements, the critical insight gained from our approach is cell-type specificity. While enrichment in "open chromatin" is to be expected, identifying which specific cell types' open chromatin regions are disproportionately enriched for SUD heritability deepens our understanding by pinpointing the cellular contexts where regulatory dysregulation likely contributes to SUD pathology. 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 [80].

Notably, cortical neurons and pancreatic cells were consistently enriched across all OCR categories for all SUDs, suggesting a potentially crucial role for them in the pathophysiology of SUDs. In particular, 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 [81-87].

The implicated involvement of pancreatic 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 [88-90]. 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, an individual's genetic predisposition can influence the susceptibility of their pancreatic tissue to alcohol-induced toxicity, making them more vulnerable to these diseases. Alcohol-induced pancreatic damage can cause beta-cell dysfunction, impair insulin secretion, and dysregulate glucagon secretion, contributing to metabolic abnormalities and the development of diabetes [91-93]. Opioid drugs exert direct effects on pancreatic alpha and beta cells, e.g., by modulating insulin and glucagon secretion, potentially contributing to the glucose dysregulation and metabolic disturbances observed in opioid users [94, 95]. 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, thereby also implicating the endocannabinoid system in the regulation of pancreatic hormone secretion and glucose metabolism [96-98]. A recent study also uncovered a direct link between pancreatic cells and nicotine use, mediated by the transcription factor TCF7L2, the gene for which was implicated in T2D through GWAS [99]. 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.

Consistent with these findings, pancreatic beta cells share features with the brain, including a common set of expressed genes, which suggests an evolutionary link between the two [100]. During pancreatic organogenesis, the depression of polycomb proteins enables beta cells to present with a neuronal gene expression program [101], 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 [102]. Given the shared features of pancreatic beta cells and neurons that extend from transcription to post-transcriptional regulation [103], understanding the interactions between the brain and pancreas could aid in understanding the relationship between addiction and metabolic health.

Recognizing SUDs and T2D as potentially interconnected disorders with shared biological mechanisms opens avenues for more integrated and holistic therapeutic approaches. Instead of treating these conditions in silos, a more effective paradigm might involve screening individuals with one condition for the presence of the other and developing treatment plans that address the comorbidity. The insights gained from our study, particularly regarding the role of pancreatic cells and metabolic pathways, point towards promising new therapeutic targets. Among these, glucagon-like peptide-1 (GLP-1) agonists stand out as a class of drugs with significant potential. Already well-established for their efficacy in managing T2D by improving glycemic control, promoting weight loss, and offering cardiovascular benefits, GLP-1 agonists exert their effects by mimicking the action of the naturally occurring GLP-1 hormone [104]. This hormone plays a crucial role in glucose-dependent insulin secretion from pancreatic beta cells, suppressing glucagon release, slowing gastric emptying, and increasing satiety. Emerging research further suggests a potential role for GLP-1 agonists in the treatment of SUDs [105, 106]. GLP-1 receptors are expressed in brain regions involved in reward processing and addiction, including the ventral tegmental area (VTA) and nucleus

accumbens [107, 108]. Preclinical studies have demonstrated that GLP-1 agonists can modulate dopaminergic signaling in these areas, leading to reduced drug-seeking behavior and decreased intake of substances like alcohol, nicotine, and even opioids [109-111]. While clinical trials investigating the direct use of GLP-1 agonists for SUD treatment are still in early phases, the existing data and biological rationale are compelling. The application of GLP-1 agonists in individuals with comorbid SUDs and T2D or obesity could offer a dual therapeutic benefit, simultaneously addressing metabolic dysregulation and potentially reducing substance use and craving. Their established safety profile and widespread use in T2D and obesity make them an attractive option for repurposing in this context. In a recent placebo-controlled trial, the GLP-1 receptor antagonist, semaglutide—a drug approved for treating obesity and T2D—significantly reduced heavy drinking in individuals with AUD. Studies of other incretin drugs like semaglutide are currently underway [112]. Further research is warranted to fully elucidate the mechanisms by which GLP-1 agonists influence addictive behaviors and to determine their efficacy and optimal use in diverse SUD populations, particularly in those with co-occurring metabolic conditions.

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 [113-117].

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 to SUDs lies. These observations warrant further research aimed at unraveling the underlying causes of these disorders in specific cellular contexts. This could offer new avenues to advance our understanding of the etiology and pathophysiology of SUDs and aid in the development of more effective targeted interventions. The potential for a single class of medication to positively impact both SUDs and T2D underscores the value of understanding the shared biological underpinnings of these complex disorders, as

revealed by integrative genomics approaches like the one employed in our study. This could pave the way for more effective, targeted, and integrated interventions for individuals struggling with both SUDs and metabolic disease.

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 Altimmune and Clearmind Medicine; a consultant to Sobrera Pharmaceuticals, Altimmune, and Lilly; the recipient of research funding and medication supplies for an investigator-initiated study from Alkermes and a company-initiated study from Altimmune; a member of the American Society of Clinical Psychopharmacology's Alcohol Clinical Trials Initiative, which was supported in the last three years by Eli Lilly, Imbrium, and Indivior; and an inventor on U.S. provisional patent

“Multi-ancestry Genome-wide Association Meta-analysis of Buprenorphine Treatment Response.”

REFERENCES

1. WHO, *Global status report on alcohol and health 2018.* 2018.
2. SAMHSA, *National Survey on Drug Use and Health.* 2022: p. Table 2.25—Alcohol use in lifetime: among people aged 12 or older; by age group and demographic characteristics, 2021 and 2022.
3. Esser MB, S.A., Liu Y, Naimi TS, *Deaths from Excessive Alcohol Use — United States, 2016–2021*. MMWR and Morbidity and Mortality Weekly Report. p. 154–161.
4. WHO, *WHO global report on trends in prevalence of tobacco use 2000-2025.* 2021.
5. Ahmad FB, C.J., Rossen LM, Sutton P, *Provisional drug overdose death counts*. 2021, National Vital Statistics System: National Center for Health Statistics.
6. Hasin, D.S., et al., *Prevalence of Marijuana Use Disorders in the United States Between 2001-2002 and 2012-2013.* JAMA Psychiatry, 2015. **72**(12): p. 1235-1242.
7. WHO, *The health and social effects of nonmedical cannabis use.* 2016.
8. Lopez-Quintero, C., et al., *Probability and predictors of transition from first use to dependence on nicotine, alcohol, cannabis, and cocaine: results of the National Epidemiologic Survey on Alcohol and Related Conditions (NESARC).* Drug Alcohol Depend, 2011. **115**(1-2): p. 120-30.
9. Winters, K.C. and C.Y. Lee, *Likelihood of developing an alcohol and cannabis use disorder during youth: association with recent use and age.* Drug Alcohol Depend, 2008. **92**(1-3): p. 239-47.
10. Prom-Wormley, E.C., et al., *The genetic epidemiology of substance use disorder: A review.* Drug Alcohol Depend, 2017. **180**: p. 241-259.
11. Heath, A.C., et al., *Genetic and environmental contributions to alcohol dependence risk in a national twin sample: consistency of findings in women and men.* Psychol Med, 1997. **27**(6): p. 1381-96.
12. Kendler, K.S., *Twin studies of psychiatric illness: an update.* Arch Gen Psychiatry, 2001. **58**(11): p. 1005-14.
13. Koopmans, J.R., et al., *The influence of religion on alcohol use initiation: evidence for genotype X environment interaction.* Behav Genet, 1999. **29**(6): p. 445-53.
14. Viken, R.J., et al., *Longitudinal analyses of the determinants of drinking and of drinking to intoxication in adolescent twins.* Behav Genet, 1999. **29**(6): p. 455-61.
15. Sullivan, P.F. and K.S. Kendler, *The genetic epidemiology of smoking.* Nicotine Tob Res, 1999. **1 Suppl 2**: p. S51-7; discussion S69-70.
16. Agrawal, A., et al., *The genetics of addiction-a translational perspective.* Transl Psychiatry, 2012. **2**(7): p. e140.
17. Agrawal, A. and M.T. Lynskey, *The genetic epidemiology of cannabis use, abuse and dependence.* Addiction, 2006. **101**(6): p. 801-12.
18. Verweij, K.J., et al., *Genetic and environmental influences on cannabis use initiation and problematic use: a meta-analysis of twin studies.* Addiction, 2010. **105**(3): p. 417-30.
19. Deak, J.D. and E.C. Johnson, *Genetics of substance use disorders: a review.* Psychological Medicine, 2021. **51**(13): p. 2189-2200.

20. Tsuang, M.T., et al., *Co-occurrence of abuse of different drugs in men: the role of drug-specific and shared vulnerabilities.* Arch Gen Psychiatry, 1998. **55**(11): p. 967-72.
21. Kendler, K.S., J. Myers, and C.A. Prescott, *Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence.* Arch Gen Psychiatry, 2007. **64**(11): p. 1313-20.
22. Kendler, K.S., et al., *The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women.* Arch Gen Psychiatry, 2003. **60**(9): p. 929-37.
23. Johnson, E.C., Y. Chang, and A. Agrawal, *An Update on the Role of Common Genetic Variation Underlying Substance Use Disorders.* Current Genetic Medicine Reports, 2020. **8**(2): p. 35-46.
24. Hancock, D.B., et al., *Human Genetics of Addiction: New Insights and Future Directions.* Current Psychiatry Reports, 2018. **20**(2): p. 8.
25. Edenberg, H.J. and J.N. McClintick, *Alcohol Dehydrogenases, Aldehyde Dehydrogenases, and Alcohol Use Disorders: A Critical Review.* Alcohol Clin Exp Res, 2018. **42**(12): p. 2281-2297.
26. Gelernter, J., et al., *Genome-wide association study of alcohol dependence:significant findings in African- and European-Americans including novel risk loci.* Molecular Psychiatry, 2014. **19**(1): p. 41-49.
27. Kranzler, H.R., et al., *Genome-wide association study of alcohol consumption and use disorder in 274,424 individuals from multiple populations.* Nature Communications, 2019. **10**(1): p. 1499.
28. Walters, R.K., et al., *Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders.* Nat Neurosci, 2018. **21**(12): p. 1656-1669.
29. Clarke, T.K., et al., *Genome-wide association study of alcohol consumption and genetic overlap with other health-related traits in UK Biobank (N=112□117).* Molecular Psychiatry, 2017. **22**(10): p. 1376-1384.
30. Gelernter, J., et al., *Genome-wide Association Study of Maximum Habitual Alcohol Intake in >140,000 U.S. European and African American Veterans Yields Novel Risk Loci.* Biological Psychiatry, 2019. **86**(5): p. 365-376.
31. Liu, M., et al., *Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use.* Nature Genetics, 2019. **51**(2): p. 237-244.
32. Sanchez-Roige, S., et al., *Genome-wide association study of alcohol use disorder identification test (AUDIT) scores in 20 328 research participants of European ancestry.* Addict Biol, 2019. **24**(1): p. 121-131.
33. Xu, K., et al., *Genomewide Association Study for Maximum Number of Alcoholic Drinks in European Americans and African Americans.* Alcohol Clin Exp Res, 2015. **39**(7): p. 1137-47.
34. Zhou, H., et al., *Association of OPRM1 Functional Coding Variant With Opioid Use Disorder: A Genome-Wide Association Study.* JAMA Psychiatry, 2020. **77**(10): p. 1072-1080.

35. Sanchez-Roige, S., et al., *Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts.* Am J Psychiatry, 2019. **176**(2): p. 107-118.
36. Toikumo, S., et al., *Multi-ancestry meta-analysis of tobacco use disorder identifies 461 potential risk genes and reveals associations with multiple health outcomes.* Nat Hum Behav, 2024.
37. Saunders, G.R.B., et al., *Genetic diversity fuels gene discovery for tobacco and alcohol use.* Nature, 2022. **612**(7941): p. 720-724.
38. Levey, D.F., et al., *Multi-ancestry genome-wide association study of cannabis use disorder yields insight into disease biology and public health implications.* Nat Genet, 2023. **55**(12): p. 2094-2103.
39. Demontis, D., et al., *Genome-wide association study implicates CHRNA2 in cannabis use disorder.* Nat Neurosci, 2019. **22**(7): p. 1066-1074.
40. Johnson, E.C., et al., *A large-scale genome-wide association study meta-analysis of cannabis use disorder.* The Lancet Psychiatry, 2020. **7**(12): p. 1032-1045.
41. Cheng, Z., et al., *Genome-wide Association Study Identifies a Regulatory Variant of RGMA Associated With Opioid Dependence in European Americans.* Biol Psychiatry, 2018. **84**(10): p. 762-770.
42. Gelernter, J., et al., *Genome-Wide Association Study of Opioid Dependence: Multiple Associations Mapped to Calcium and Potassium Pathways.* Biological Psychiatry, 2014. **76**(1): p. 66-74.
43. Polimanti, R., et al., *Leveraging genome-wide data to investigate differences between opioid use vs. opioid dependence in 41,176 individuals from the Psychiatric Genomics Consortium.* Mol Psychiatry, 2020. **25**(8): p. 1673-1687.
44. Deak, J.D., et al., *Genome-wide association study in individuals of European and African ancestry and multi-trait analysis of opioid use disorder identifies 19 independent genome-wide significant risk loci.* Molecular Psychiatry, 2022. **27**(10): p. 3970-3979.
45. Kember, R.L., et al., *Cross-ancestry meta-analysis of opioid use disorder uncovers novel loci with predominant effects in brain regions associated with addiction.* Nature Neuroscience, 2022. **25**(10): p. 1279-1287.
46. Sey, N.Y.A., et al., *Chromatin architecture in addiction circuitry identifies risk genes and potential biological mechanisms underlying cigarette smoking and alcohol use traits.* Mol Psychiatry, 2022. **27**(7): p. 3085-3094.
47. Finucane, H.K., et al., *Partitioning heritability by functional annotation using genome-wide association summary statistics.* Nature Genetics, 2015. **47**(11): p. 1228-1235.
48. Patel, H., et al., *nf-core/atacseq: [2.1.2] - 2022-08-07*. 2023, Zenodo.
49. Zhang, H., et al., *Fast alignment and preprocessing of chromatin profiles with Chromap.* Nature Communications, 2021. **12**(1): p. 6566.
50. Su, C., et al., *3D chromatin maps of the human pancreas reveal lineage-specific regulatory architecture of T2D risk.* Cell Metabolism, 2022. **34**(9).
51. Wingett, S., et al., *HiCUP: pipeline for mapping and processing Hi-C data.* F1000Res, 2015. **4**: p. 1310.

52. Cairns, J., et al., *CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data.* Genome Biology, 2016. **17**(1): p. 1-17.
53. C, S., et al., *Restriction enzyme selection dictates detection range sensitivity in chromatin conformation capture-based variant-to-gene mapping approaches.* Human genetics, 2021. **140**(10).
54. Open2C, et al., *Pairtools: From sequencing data to chromosome contacts.* PLOS Computational Biology, 2024. **20**(5): p. e1012164.
55. S, L. *A tool for indexing and querying on a block-compressed text file containing pairs of genomic coordinates.* 2020; Available from: https://github.com/4dn-dcic/pairix.
56. Abdennur, N., et al., *cooler*.
57. Imakaev, M., et al., *Iterative correction of Hi-C data reveals hallmarks of chromosome organization.* Nature Methods, 2012. **9**(10): p. 999-1003.
58. Roayaei Ardakany, A., et al., *Mustache: multi-scale detection of chromatin loops from Hi-C and Micro-C maps using scale-space representation.* Genome Biology, 2020. **21**(1): p. 1-17.
59. A, K., B. S, and A. F, *Identifying statistically significant chromatin contacts from Hi-C data with FitHiC2.* Nature protocols, 2020. **15**(3).
60. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biology, 2014. **15**(12): p. 550.
61. Zhu, A., J.G. Ibrahim, and M.I. Love, *Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences.* Bioinformatics, 2018. **35**(12): p. 2084-2092.
62. Finucane, H.K., et al., *Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types.* Nature genetics, 2018. **50**(4): p. 621-629.
63. Gazal, S., et al., *Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection.* Nature Genetics, 2017. **49**(10): p. 1421-1427.
64. Bulik-Sullivan, B., et al., *An atlas of genetic correlations across human diseases and traits.* Nature genetics, 2015. **47**(11): p. 1236-1241.
65. Suzuki, K., et al., *Genetic drivers of heterogeneity in type 2 diabetes pathophysiology.* Nature, 2024. **627**(8003): p. 347-357.
66. Werme, J., et al., *An integrated framework for local genetic correlation analysis.* Nature Genetics, 2022. **54**(3): p. 274-282.
67. Rohart, F., et al., *mixOmics: An R package for 'omics feature selection and multiple data integration.* PLOS Computational Biology, 2017. **13**(11): p. e1005752.
68. Ulgen, E., O. Ozisik, and O.U. Sezerman, *pathfindR: An R Package for Comprehensive Identification of Enriched Pathways in Omics Data Through Active Subnetworks.* Frontiers in Genetics, 2019. **10**.
69. Wu, T., et al., *clusterProfiler 4.0: A universal enrichment tool for interpreting omics data.* The Innovation, 2021. **2**(3): p. 100141.
70. Trang, K.B., et al., *3D genomic features across >50 diverse cell types reveal insights into the genomic architecture of childhood obesity.* medRxiv, 2024.

71. Akinyemi, D.E., R. Chevre, and O. Soehnlein, *Neuro-immune crosstalk in hematopoiesis, inflammation, and repair.* Trends in Immunology, 2024. **45**(8): p. 597-608.
72. Mostafavi, H., et al., *Systematic differences in discovery of genetic effects on gene expression and complex traits.* Nat Genet, 2023. **55**(11): p. 1866-1875.
73. Littleton, S.H., et al., *Variant-to-function analysis of the childhood obesity chr12q13 locus implicates rs7132908 as a causal variant within the 3' UTR of FAIM2.* Cell Genom, 2024. **4**(5): p. 100556.
74. Chesi, A., et al., *Genome-scale Capture C promoter interactions implicate effector genes at GWAS loci for bone mineral density.* Nature Communications, 2019. **10**(1): p. 1-11.
75. Su, C., et al., *3D promoter architecture re-organization during iPSC-derived neuronal cell differentiation implicates target genes for neurodevelopmental disorders.* Progress in Neurobiology, 2021. **201**: p. 102000.
76. Lasconi, C., et al., *Variant-to-gene-mapping analyses reveal a role for pancreatic islet cells in conferring genetic susceptibility to sleep-related traits.* Sleep, 2022. **45**(8).
77. Pahl, M.C., et al., *Cis-regulatory architecture of human ESC-derived hypothalamic neuron differentiation aids in variant-to-gene mapping of relevant complex traits.* Nature Communications, 2021. **12**(1): p. 1-12.
78. Su, C., et al., *Mapping effector genes at lupus GWAS loci using promoter Capture-C in follicular helper T cells.* Nature Communications, 2020. **11**(1): p. 3294.
79. Conery, M., et al., *GWAS-informed data integration and non-coding CRISPRi screen illuminate genetic etiology of bone mineral density.* bioRxiv, 2024.
80. de la Torre-Ubieta, L., et al., *The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis.* Cell, 2018. **172**(1): p. 289-304.e18.
81. Koob, G.F. and N.D. Volkow, *Neurocircuitry of Addiction.* Neuropsychopharmacology, 2010. **35**(1): p. 217-238.
82. Koob, G.F. and N.D. Volkow, *Neurobiology of addiction: a neurocircuitry analysis.* Lancet Psychiatry, 2016. **3**(8): p. 760-773.
83. Koob, G.F., *Neurobiology of Opioid Addiction: Opponent Process, Hyperkatifeia, and Negative Reinforcement.* Biol Psychiatry, 2020. **87**(1): p. 44-53.
84. Xu, C., H.H. Loh, and P.Y. Law, *Effects of addictive drugs on adult neural stem/progenitor cells.* Cell Mol Life Sci, 2016. **73**(2): p. 327-48.
85. Robinson, T.E. and B. Kolb, *Structural plasticity associated with exposure to drugs of abuse.* Neuropharmacology, 2004. **47 Suppl 1**: p. 33-46.
86. Montoya-Filardi, A. and M. Mazón, *The addicted brain: imaging neurological complications of recreational drug abuse.* Radiologia, 2017. **59**(1): p. 17-30.
87. Bhattacherjee, A., et al., *Cell type-specific transcriptional programs in mouse prefrontal cortex during adolescence and addiction.* Nature Communications, 2019. **10**(1): p. 4169.
88. Ojo, O., et al., *The Effects of Substance Abuse on Blood Glucose Parameters in Patients with Diabetes: A Systematic Review and Meta-Analysis.* Int J Environ Res Public Health, 2018. **15**(12).

89. Bruggeman, B.S., et al., *Substance Use Affects Type 1 Diabetes Pancreas Pathology: Implications for Future Studies.* Front Endocrinol (Lausanne), 2021. **12**: p. 778912.
90. Reece, A.S. and G.K. Hulse, *Sociodemographically Stratified Exploration of Pancreatic Cancer Incidence in Younger US Patients: Implication of Cannabis Exposure as a Risk Factor.* Gastroenterology Insights, 2023. **14**(2): p. 204-235.
91. Kim, J.Y., et al., *Chronic alcohol consumption potentiates the development of diabetes through pancreatic β-cell dysfunction.* World journal of biological chemistry, 2015. **6**(1): p. 1-15.
92. Steiner, J.L., K.T. Crowell, and C.H. Lang, *Impact of Alcohol on Glycemic Control and Insulin Action.* Biomolecules, 2015. **5**(4): p. 2223-2246.
93. Nelson, N.G., et al., *Joint and separate exposure to alcohol and Δ9-tetrahydrocannabinol produced distinct effects on glucose and insulin homeostasis in male rats.* Scientific Reports, 2019. **9**(1): p. 12025.
94. Moustafa, S.R., *The immune-opioid axis in prediabetes: predicting prediabetes with insulin resistance by plasma interleukin-10 and endomorphin-2 to kappa-opioid receptors ratio.* Diabetology & Metabolic Syndrome, 2021. **13**(1): p. 61.
95. Koekkoek, L.L., et al., *The Clash of Two Epidemics: the Relationship Between Opioids and Glucose Metabolism.* Current Diabetes Reports, 2022. **22**(7): p. 301-310.
96. Song, D., et al., *Delayed presentation of cannabis induced pancreatitis.* Clin Case Rep, 2022. **10**(3): p. e05595.
97. Qi, M., et al., *Chronic marijuana usage by human pancreas donors is associated with impaired islet function.* PLOS ONE, 2021. **16**(10): p. e0258434.
98. Cortes-Justo, E., S.H. Garfias-Ramírez, and A. Vilches-Flores, *The function of the endocannabinoid system in the pancreatic islet and its implications on metabolic syndrome and diabetes.* Islets, 2023. **15**(1): p. 1-11.
99. Duncan, A., et al., *Habenular TCF7L2 links nicotine addiction to diabetes.* Nature, 2019. **574**(7778): p. 372-377.
100. Arntfield, M.E. and D. van der Kooy, *β-Cell evolution: How the pancreas borrowed from the brain: The shared toolbox of genes expressed by neural and pancreatic endocrine cells may reflect their evolutionary relationship.* Bioessays, 2011. **33**(8): p. 582-7.
101. van Arensbergen, J., et al., *Derepression of Polycomb targets during pancreatic organogenesis allows insulin-producing beta-cells to adopt a neural gene activity program.* Genome Res, 2010. **20**(6): p. 722-32.
102. Devaskar, S.U., et al., *Insulin gene expression and insulin synthesis in mammalian neuronal cells.* J Biol Chem, 1994. **269**(11): p. 8445-54.
103. Juan-Mateu, J., et al., *Neuron-enriched RNA-binding Proteins Regulate Pancreatic Beta Cell Function and Survival.* J Biol Chem, 2017. **292**(8): p. 3466-3480.
104. Wang, W., et al., *Associations of semaglutide with first-time diagnosis of Alzheimer's disease in patients with type 2 diabetes: Target trial emulation using nationwide real-world data in the US.* Alzheimer's & Dementia, 2024. **20**(12): p. 8661-8672.

105. Xie, Y., T. Choi, and Z. Al-Aly, *Mapping the effectiveness and risks of GLP-1 receptor agonists.* Nature Medicine, 2025. **31**(3): p. 951-962.
106. Merkel, R., et al., *A novel approach to treating opioid use disorders: Dual agonists of glucagon-like peptide-1 receptors and neuropeptide Y2 receptors.* Neuroscience & Biobehavioral Reviews, 2021. **131**: p. 1169-1179.
107. Cummings, J.L., et al., *evoke and evoke+: design of two large-scale, double-blind, placebo-controlled, phase 3 studies evaluating efficacy, safety, and tolerability of semaglutide in early-stage symptomatic Alzheimer's disease.* Alzheimer's Research & Therapy, 2025. **17**(1): p. 14.
108. Du, H., et al., *The mechanism and efficacy of GLP-1 receptor agonists in the treatment of Alzheimer's disease.* Front Endocrinol (Lausanne), 2022. **13**: p. 1033479.
109. Qeadan, F., A. McCunn, and B. Tingey, *The association between glucose-dependent insulinotropic polypeptide and/or glucagon-like peptide-1 receptor agonist prescriptions and substance-related outcomes in patients with opioid and alcohol use disorders: A real-world data analysis.* Addiction, 2025. **120**(2): p. 236-250.
110. Herman, R.J. and H.D. Schmidt, *Targeting GLP-1 receptors to reduce nicotine use disorder: Preclinical and clinical evidence.* Physiol Behav, 2024. **281**: p. 114565.
111. Jerlhag, E., *GLP-1 Receptor Agonists: Promising Therapeutic Targets for Alcohol Use Disorder.* Endocrinology, 2025. **166**(4).
112. Hendershot, C.S., et al., *Once-Weekly Semaglutide in Adults With Alcohol Use Disorder: A Randomized Clinical Trial.* JAMA Psychiatry, 2025. **82**(4): p. 395-405.
113. Meredith, L.R., et al., *Immune treatments for alcohol use disorder: A translational framework.* Brain, Behavior, and Immunity, 2021. **97**: p. 349-364.
114. Crews, F.T. and R.P. Vetreno, *Mechanisms of neuroimmune gene induction in alcoholism.* Psychopharmacology, 2016. **233**(9): p. 1543-1557.
115. Erickson, E.K., et al., *Neuroimmune signaling in alcohol use disorder.* Pharmacology Biochemistry and Behavior, 2019. **177**: p. 34-60.
116. Osterndorff-Kahanek, E.A., et al., *Chronic Ethanol Exposure Produces Time- and Brain Region-Dependent Changes in Gene Coexpression Networks.* PLOS ONE, 2015. **10**(3): p. e0121522.
117. Fan, X., et al., *Drinking alcohol is associated with variation in the human oral microbiome in a large study of American adults.* Microbiome, 2018. **6**(1): p. 59.

FIGURE LEGENDS

Figure 1. Substance use disorders conditional effect size (τ*) in annotation across diverse cell types. A. Bar plots (left) depict total number of open chromatin regions (OCR x10,000) identified by bulk (blue) or single-cell (red) ATAC-seq and the proportion of those OCR contacting a gene promoter (putative cRE) as measured by HiC (green) or promoter capture-C (orange). The dot plots depict significant enrichment (p-value ≤ 0.05) within OCR annotations for each cell type across 4 SUDs as determined by S-LDSC analysis. The colors of the dots correspond to conditional effect sizes τ* values, with dots featuring an asterisk indicating FDR ≤ 0.05. The size of the dots corresponds to p-values in -log10. **B**. Dot plots depicting heritability enrichment (x-axis) and conditional effect sizes τ* (y-axis) for 4 neural and 2 pancreatic cell types, shaped by disorders, colored by annotations, and sized by percentage of contributed SNPs, white asterisks inside dots indicate significant enrichments p-values ≤0.05 and red circles indicate τ* FDR ≤ 0.05 from LDSC analysis.

Figure 2. Genetic heritability correlations between T2D and SUDs. A. Correlation matrix between the disorders, lower triangle was for genome-wide scale with values displayed for FDR≤0.05, upper triangle is for the average local correlations with numbers of loci displaying significant FDR≤0.05 correlations; the values are red when the correlations are significant on both genome-wide and local level, black if either. **B.** Barplot shows numbers of loci tested (grey), found correlated (blue), and reached a significant FDR≤0.05 (green) for each pair-wise of disorders; yellow bars are the number of loci where the CI-95 of correlation between the pair reached 1. **C.** The chord diagram shows number of significantly correlated loci between 5 disorders. **D.** Scatter points showing distributions of correlations across loci for each pair of disorders, blue boxplots were significant correlated loci, red lines indicate the correlation level at genome-wide scale. **E.** Dotplot shows agreement between genome-wide and average local correlations, colored by P-value at genome-wide scale, sized by number of loci. **F.** Heatmap of 114 loci showed significant correlation between T2D and AUD, TUD and CanUD.

Figure 3. Characterization of regulatory landscapes and transcriptomics in neural and pancreatic cells relevant to T2D-SUD overlaps. A. Cell types projection from sPLS-DA analysis performed using total expression profiles of 6 cell types into 4 optimized components space; ellipse represents 95% confidence intervals around and colored by each sample class; 1st component separated iPSC-derived neurons, 2nd component separated iPSC-derived NPC, 3rd component separated 2 pancreatic cell types and 4th component separated hESC-derived hypothalamus neurons from the rest (more in Fig. S5A) but all combination failed to separate pancreatic cells from neural cell types. **B.** Enriched pathways from a few genes drove the separation of components 2 and 4 from panel A. **C.** Sample plot from multiblock DIABLO analysis performed using 3 blocks of RNA-seq profiles and ATAC-seq profiles from Promoter OCRs and cREs within 114 loci displayed significant correlation between T2D and SUDs; the representative 3 combinations of 5 optimized components showed the most separation power, colored similar to panel A: 1st component effectively separated primary lineages (neural progenitors, differentiated neurons, and pancreatic cells), 2nd component focused on distinguishing hESC-derived neural progenitor cells, 3rd component characterized hESC-derived hypothalamic neurons, 4th component separated iPSC-derived neural progenitor cells, 5th component differentiated pancreatic alpha and beta cells and neural cells; similar to panel A, still failed to distinguish pancreatic cell from naurall cell types. **D.** Circos plot represents the correlations greater than 0.8 between variables of different types, represented on the side quadrants, internal connecting lines show the positive (negative) correlations, the outer lines show the expression levels of each variable in each cell type. **E.** GO-terms enrichments of genes drove the separations with 5 optimized DIABLO components. **F.** Network of enriched GO terms with genes from 114 loci, separated by pair-wise correlation between T2D and SUDs.

SUPPLEMETARY FIGURE LEGENDS

Figure S1. Full S-LDSC enrichment for SUDs across diverse cell types across all annotations**.** Bar plots (left) is the same in Figure 1A. The dot plots depict heritability enrichment for each cell type across 4 SUD traits as determined by LDSC analysis. Whiskers represent enrichment standard errors, with colors matched for HiC vs. capture-C. The colors of the dots correspond to p-values in -log10 of each disorder, with dots featuring a red asterisk indicating a significant FDR ≤ 0.05. The size of the dots corresponds to the proportion of SNP contribution to heritability. Dashed line at 1 indicates no enrichment.

Figure S2. Distributions of different metrics from S-LDSC within each cell type system, colored by annotations, with two-tailed T-test p-values for each pairwise comparison: ****, <0.001, ***, <0.005, **, <0.01 and * <0.05

Figure S3. Correlation between pair-wise annotations for 3 S-LDSC metrics: enrichment, τ*, and p-values, across all cell types and segregated into cell systems. The correlation between SEG and not_cRE/Prom_OCRs in neural cells is strongly negative using τ* (r^2=-0.41) aligning with expectations, yet it is positive in "other cell types" (r^2=0.76), which is counterintuitive when comparing positive and negative control category, suggesting that distinct additional functions of the GWAS variants attributed to the phenotype are present in different cellular system settings.

Figure S4. Venn diagram of top 10% specially expressed genes from 4 neural cell types and 2 pancreatic cell types, numbers are genes within each corresponding intersection.

Figure S5. Sparse Partial Least Squares discriminant analysis (sPLS-DA) on whole expression profiles of 4 neural and 2 pancreatic cell types. A. Tuning the number of components in sPLS-DA on the gene expression data**.** For each component, repeated 50 iterations x3-fold cross-validation (leave-one-out) is used to evaluate the PLS-DA classification performance (overall and balanced error rate BER), for each type of prediction distance (max.dist, centroids.dist and mahalanobis.dist); bars show the

standard deviation across the repeated folds; shows that the error rate reaches a minimum from 4 components. **B.** Tuning keepX for the sPLS-DA performed on the gene expression data, each colored line represents the balanced error rate (y-axis) per component across all tested keepX values (x-axis) with the standard deviation based on the repeated cross-validation folds. The diamond indicates the optimal keepX value on a particular component which achieves the lowest classification error rate as determined with a one-sided T-test, values represented for a given component (e.g. comp 1 to 2) include the optimal keepX value chosen for the previous component (comp 1); optimal keepX parameter according to minimal error rate are 9 (comp1), 30 (comp2), 1 (comp3) and 20 (comp4). **C**. Samples are projected into the space spanned by the first four components, showing the rest of the combinations from Fig. 3A, none of the combinations of all 4 components enables us to discriminate all cell types. **D.** Correlation circle plot representing the genes selected by sPLS-DA, truncated to the first 10 characters, for each combination of pair-wise components; the plots show complete polarized loadings where the nominated genes drive the discrimination force for singular component. **E.** Network of enriched terms and genes from Fig 3B., up/down correspond to direction of fold-change expression of the discriminated cell types versus the rest.

Figure S6. Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) framework using 3 blocks of RNA-seq profiles and ATAC-seq profiles from Promoter OCRs and cREs within 114 loci displayed significant correlation between T2D and SUDs. A. Sample plot from multiblock DIABLO analysis showing the remaining combinations of 5 optimized components, complemented for Fig 3C. **B.** Diagnostic plot where cell types are represented based on the specified component for each data set (mRNA, Promoter OCRs and cREs), bottom triangle numbers indicate the correlation coefficients between the block data sets, cREs correlated best with RNA and Promoter OCRs in the first component, while Promoter OCRs and RNA correlated best in the 2nd and 3rd components. **C.** Clustered Image Map for the variables selected by multiblock DIABLO on each component, Euclidean

distance and Complete linkage methods are used, the CIM represents samples in rows and selected features in columns (indicated by their data type at the top of the plot).

TABLE LEGENDS

Table S1. **Data resources of 59 cell types**

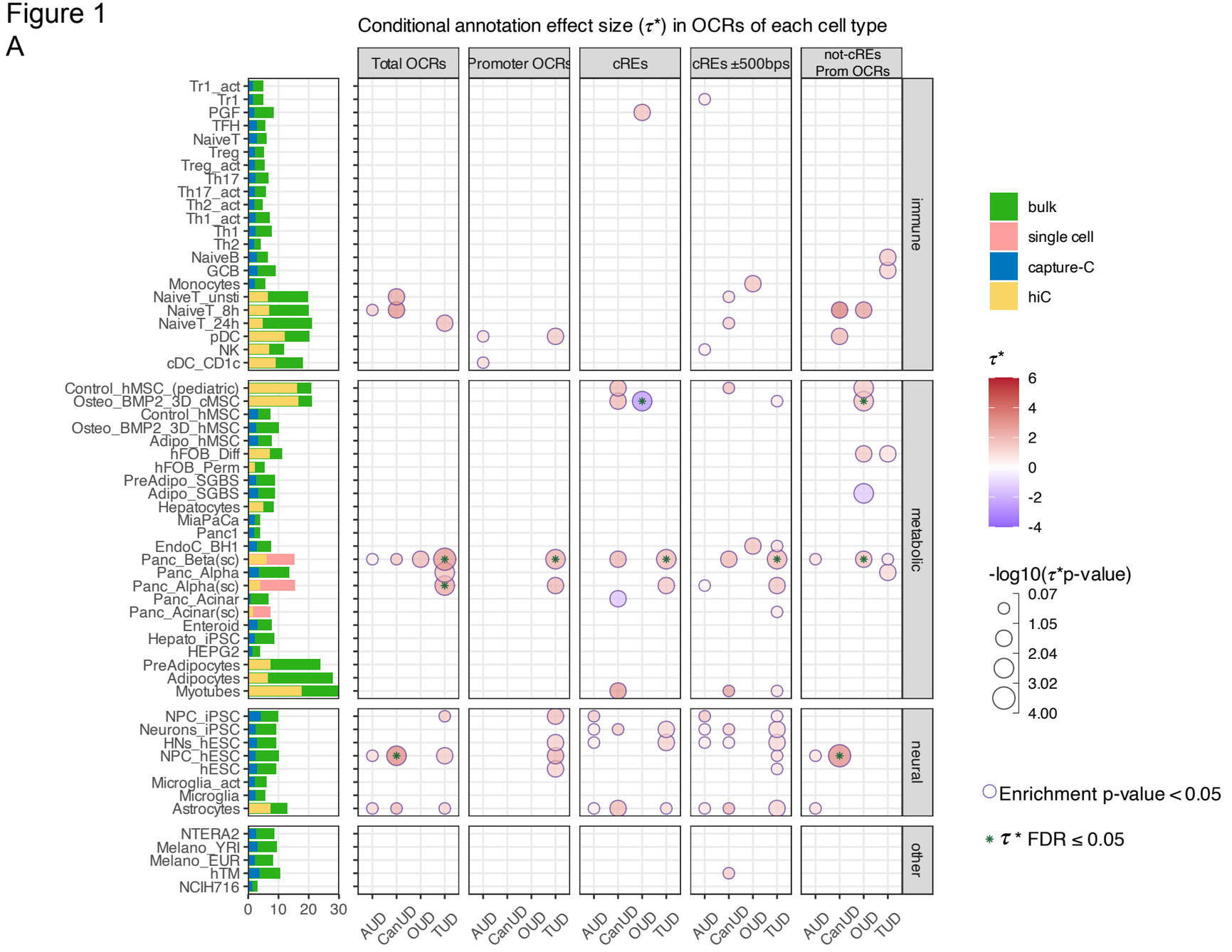
Table S2. **4 substance use disorders GWAS information**

Table S3. **Full S-LDSC results**

Table S4. **Full LAVA results**

Figure 1

A


Conditional annotation effect size (τ*) in OCRs of each cell type
Total OCRs
Promoter OCRs
cREs
cREs ±500bps
not-cREs Prom OCRs
immune
metabolic
neural
other
Tr1_act
Tr1
PGF
TFH
NaiveT
Treg
Treg_act
Th17
Th17_act
Th2_act
Th1_act
Th1
Th2
NaiveB
GCB
Monocytes
NaiveT_unsti
NaiveT_8h
NaiveT_24h
pDC
NK
cDC_CD1c
Control_hMSC_(pediatric)
Osteo_BMP2_3D_cMSC
Control_hMSC
Osteo_BMP2_3D_hMSC
Adipo_hMSC
hFOB_Diff
hFOB_Perm
PreAdipo_SGBS
Adipo_SGBS
Hepatocytes
MiaPaCa
Panc1
EndoC_BH1
Panc_Beta(sc)
Panc_Alpha
Panc_Alpha(sc)
Panc_Acinar
Panc_Acinar(sc)
Enteroid
Hepato_iPSC
HEPG2
PreAdipocytes
Adipocytes
Myotubes
NPC_iPSC
Neurons_iPSC
HNs_hESC
NPC_hESC
hESC
Microglia_act
Microglia
Astrocytes
NTERA2
Melano_YRI
Melano_EUR
hTM
NCIH716
0 10 20 30
AUD CanUD OUD TUD
bulk
single cell
capture-C
hiC
τ*
6 4 2 0 -2 -4
-log10(τ*p-value)
0.07 1.05 2.04 3.02 4.00
Enrichment p-value < 0.05
* τ* FDR ≤ 0.05

B

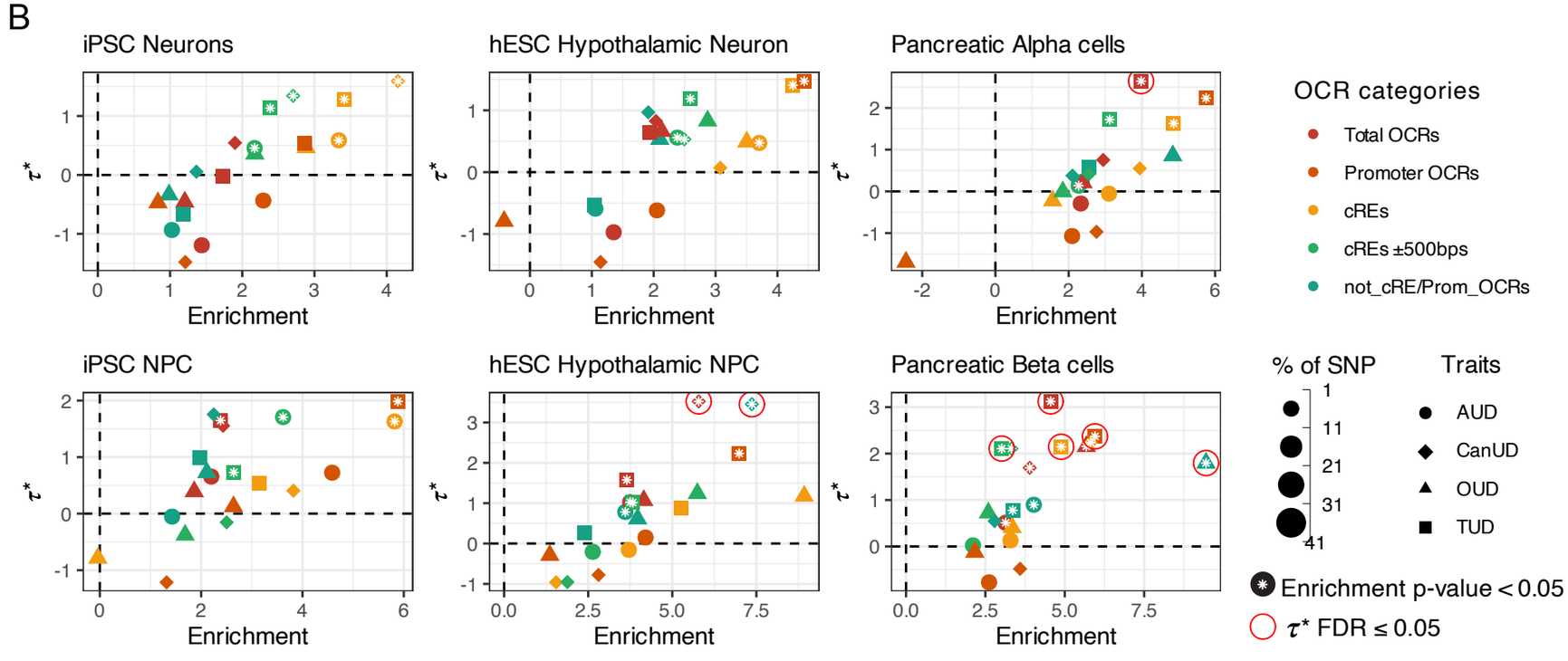

iPSC Neurons
hESC Hypothalamic Neuron
Pancreatic Alpha cells
iPSC NPC
hESC Hypothalamic NPC
Pancreatic Beta cells
τ*
Enrichment
OCR categories
Total OCRs
Promoter OCRs
cREs
cREs ±500bps
not_cRE/Prom_OCRs
% of SNP
1 11 21 31 41
Traits
AUD
CanUD
OUD
TUD
Enrichment p-value < 0.05
τ* FDR ≤ 0.05

Figure 2

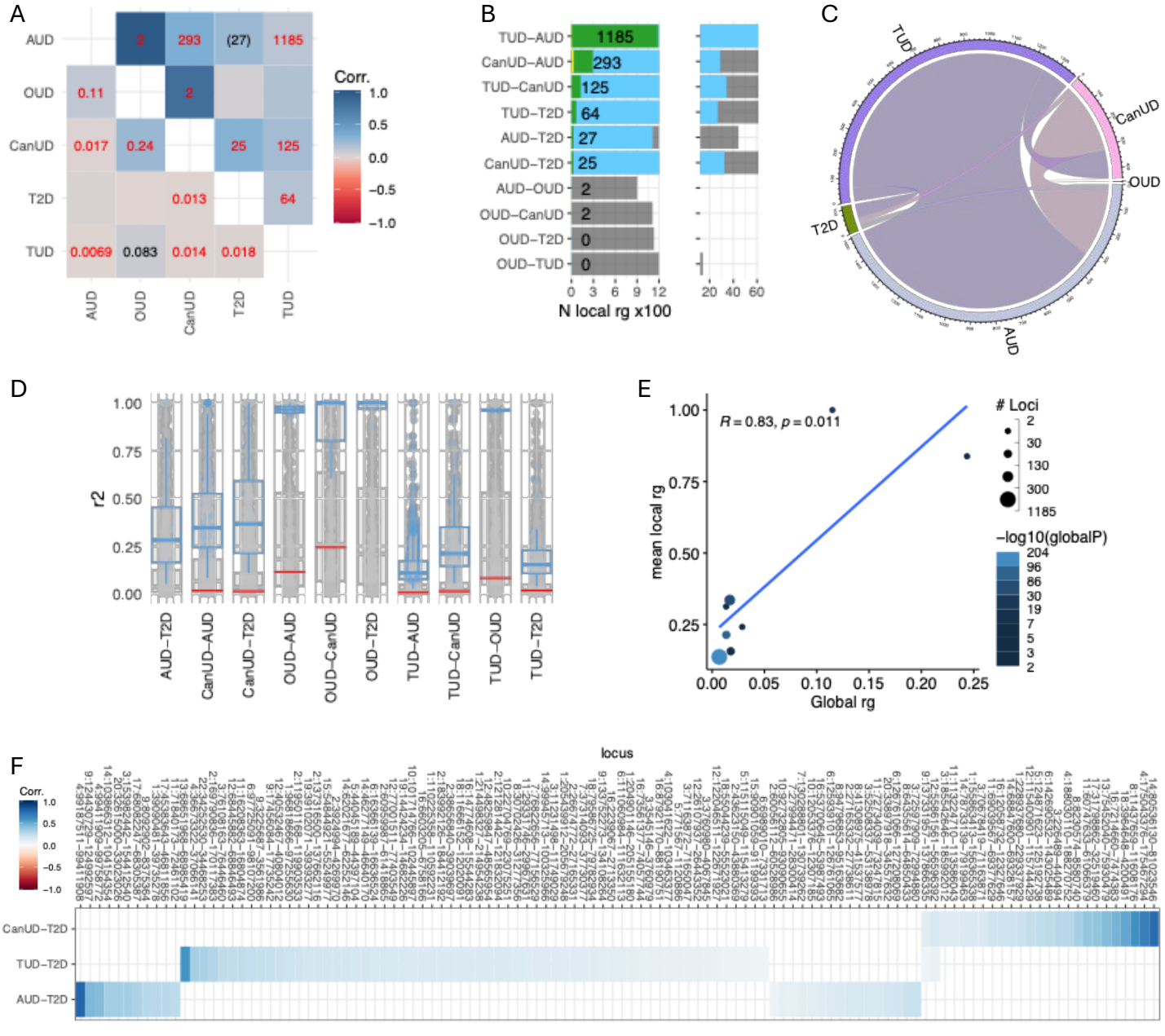
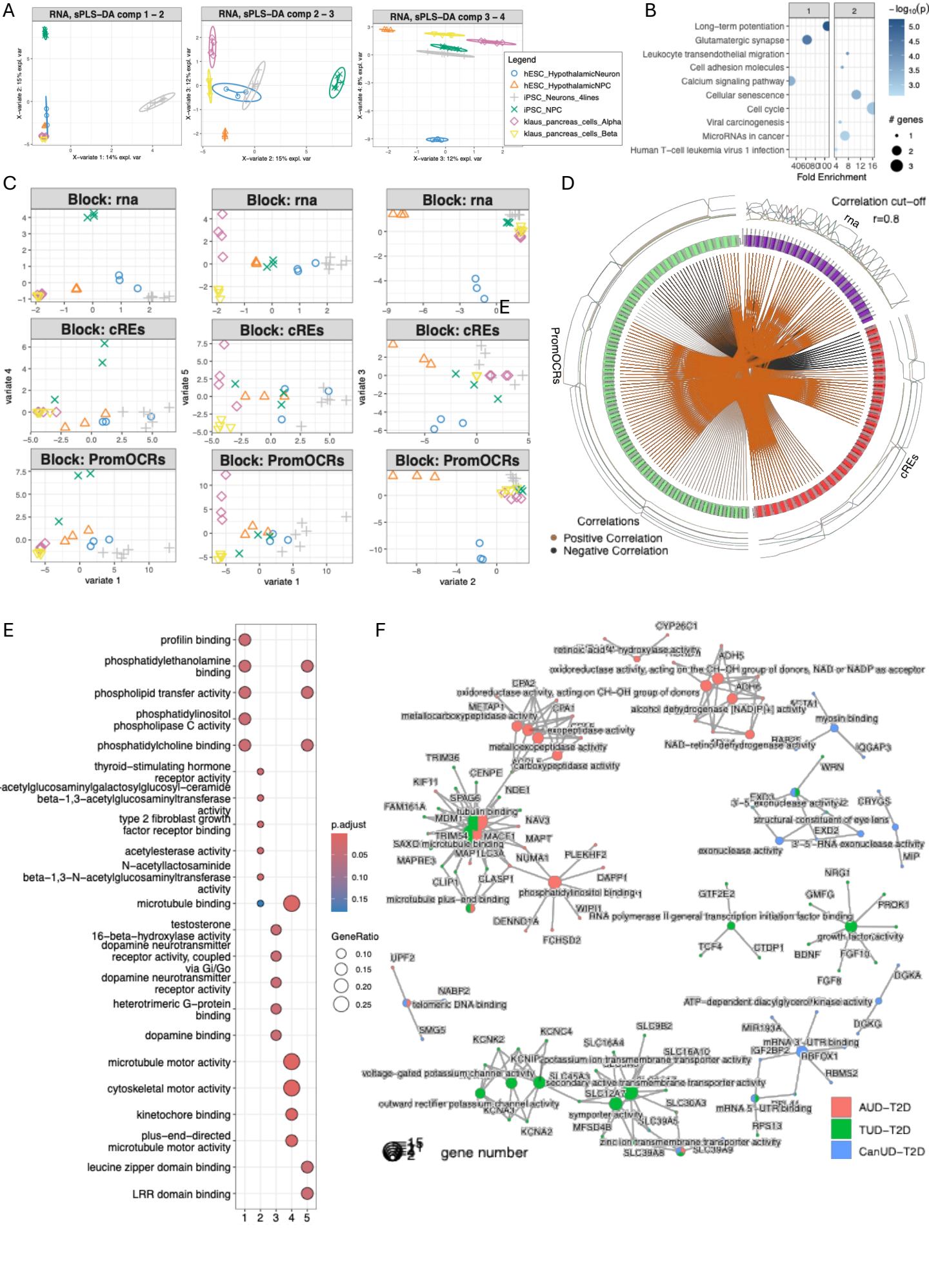

A
AUD
OUD
CanUD
T2D
TUD
2
293
(27)
1185
0.11
2
0.017
0.24
25
125
0.013
64
0.0069
0.083
0.014
0.018
Corr.
1.0
0.5
0.0
−0.5
−1.0
B
TUD−AUD 1185
CanUD−AUD 293
TUD−CanUD 125
TUD−T2D 64
AUD−T2D 27
CanUD−T2D 25
AUD−OUD 2
OUD−CanUD 2
OUD−T2D 0
OUD−TUD 0
0 3 6 9 12
20 40 60
N local rg x100
C
TUD
CanUD
OUD
T2D
AUD
D
r2
1.00
0.75
0.50
0.25
0.00
AUD−T2D
CanUD−AUD
CanUD−T2D
OUD−AUD
OUD−CanUD
OUD−T2D
TUD−AUD
TUD−CanUD
TUD−OUD
TUD−T2D
E
R = 0.83, p = 0.011
mean local rg
Global rg
Loci
2
30
130
300
1185
−log10(globalP)
204
96
86
30
19
7
5
3
2
F
locus
Corr.
CanUD−T2D
TUD−T2D
AUD−T2D

Figure 3


A
RNA, sPLS-DA comp 1 – 2
RNA, sPLS-DA comp 2 – 3
RNA, sPLS-DA comp 3 – 4
X-variate 1: 14% expl. var
X-variate 2: 15% expl. var
X-variate 3: 12% expl. var
Legend
hESC_HypothalamicNeuron
hESC_HypothalamicNPC
iPSC_Neurons_4lines
iPSC_NPC
klaus_pancreas_cells_Alpha
klaus_pancreas_cells_Beta
B
Long-term potentiation
Glutamatergic synapse
Leukocyte transendothelial migration
Cell adhesion molecules
Calcium signaling pathway
Cellular senescence
Cell cycle
Viral carcinogenesis
MicroRNAs in cancer
Human T-cell leukemia virus 1 infection
Fold Enrichment
genes
C
Block: rna
Block: cREs
Block: PromOCRs
variate 1
variate 2
variate 3
variate 4
variate 5
D
Correlation cut-off
r=0.8
rna
PromOCRs
cREs
Correlations
Positive Correlation
Negative Correlation
E
profilin binding
phosphatidylethanolamine binding
phospholipid transfer activity
phosphatidylinositol phospholipase C activity
phosphatidylcholine binding
thyroid-stimulating hormone receptor activity
-acetylglucosaminylgalactosylglucosyl-ceramide beta-1,3-acetylglucosaminyltransferase activity
type 2 fibroblast growth factor receptor binding
acetylesterase activity
N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase activity
microtubule binding
testosterone 16-beta-hydroxylase activity
dopamine neurotransmitter receptor activity, coupled via Gi/Go
dopamine neurotransmitter receptor activity
heterotrimeric G-protein binding
dopamine binding
microtubule motor activity
cytoskeletal motor activity
kinetochore binding
plus-end-directed microtubule motor activity
leucine zipper domain binding
LRR domain binding
F
p.adjust
GeneRatio
gene number
AUD-T2D
TUD-T2D
CanUD-T2D