Genome-wide association study of cocaine self-administration behavior in Heterogeneous Stock rats

Montana Kay Lara[1], Lieselot L. G. Carrette[1], Thiago Missfeldt Sanches[1], Oksana Polesskaya[1], Alicia Avelar[1], Angela Beeson[5], Hassiba Beldjoud[1], Brent Boomhower[1], Molly Brennan[1], Denghui Chen[1], Lindsey China[1], Apurva S. Chitre[1], Dana Elizabeth Conlisk[2], McKenzie Fannon[1], Benjamin B. Johnson[1], Elaine Keung[1], Adam Kimbrough[3], Jenni Kononoff[2], Angelica Renee Martinez[1], Lisa Maturin[1], Khai-Minh Nguyen[1], Alex Morgan[1], Joseph Mosquera[1], Dyar Othman[1], Sonja L. Plasil[1], Jarryd Ramborger[1], Paul Schweitzer[2], Sharona Sedighim[1], Osborne Seshie[5], Kokila Shankar[1], Benjamin Sichel[1], Sierra Simpson[1], Lauren Cassandra Smith[1], Elizabeth A. Sneddon[1], Lani Tieu[1], Nathan Velarde[1], Selene Zahedi[1,4], Leah C. Solberg Woods[5], Marsida Kallupi[1], Giordano de Guglielmo[1], Abraham A. Palmer[1,6*], Olivier George[1*]

[1] Department of Psychiatry, University of California San Diego, La Jolla, CA, USA
[2] The Scripps Research Institute, La Jolla, CA, USA
[3] Purdue University, West Lafayette, IN, USA
[4] Institut de Neurosciences de la Timone, Aix-Marseille Université, Marseille, 13005, France
[5] Department of Internal Medicine, Section on Molecular Medicine, Wake Forest University School of Medicine, Winston-Salem, NC, USA
[6] Institute for Genomic Medicine, University of California San Diego, La Jolla, CA, USA

* Co-corresponding authors
Abraham A. Palmer email: aapalmer@health.ucsd.edu
Olivier George email: olgeorge@health.ucsd.edu

Keywords: GWAS, Heterogeneous Stock rats, cocaine, cocaine use disorder, self-administration, *ces1*

Abstract

Background
Cocaine use disorder (CUD) is a major public health crisis with detrimental individual and societal effects. The specific genes mediating CUD remain largely unknown.

Methods
We conducted a genome-wide association study (GWAS) using outbred N/NIH Heterogeneous Stock (HS; n = 836, female = 415, male = 421) rats. We examined multiple CUD-related phenotypes that captured acquisition of self-administration, escalation of intake, and compulsive-like responding.

Results
Consistent with the existing literature, these traits were phenotypically and genetically correlated and exhibited modest heritability (h^2 = 0.07 – 0.16). We identified six genome-wide significant associations. One locus on chromosome 19 was associated with the variable time between cocaine infusions (post infusion interval) and contains several carboxylesterase genes that are orthologous to the human *CES1* gene; notably, carboxylesterases metabolize cocaine. Three non-synonymous coding variants in the genes *Ces1c* and *Ces1d* were in perfect linkage disequilibrium with this locus, suggesting that one or more of them might be the causal SNP. The other 5 loci also contained promising coding and expression variants, including *Trak2,* a gene previously associated with CUD in human GWAS and *Slc10a7*, *Plcl1*, and *Satb2* which have been associated with alcohol and tobacco use disorder.

Conclusions
This is the largest genetic study of cocaine self-administration ever conducted in any species. Our results replicate previous loci associated with CUD in humans and provide several novel biological insights including the potential of pharmacological strategies targeting carboxylesterases for the treatment of CUD.

Introduction

Cocaine use disorder (CUD) is a pressing public health issue. The 2022 National Survey on Drug Use and Health found that approximately 5 million individuals in the US reported using cocaine in the past year [1]. Notably, 1.4 million people meet the criteria for CUD. The complexity of CUD is further amplified by its frequent comorbidity with other neuropsychiatric disorders, underscoring the challenges in developing effective therapeutic strategies.

CUD is a highly heritable and genetically complex trait. Twin studies estimate the heritability of cocaine dependence to be as high as 70% [2-4], a finding supported by recent comprehensive reviews [5,6]. Complementing these findings, genome-wide association studies (GWAS) have also revealed a significant heritable component, with single nucleotide polymorphism (SNP)-based heritability estimated at 27-30% [7,8]. Despite the strong evidence of the genetic etiology of CUD, few genome-wide significant associations have been reported: *FAM53B, NCOR2* [9], *LINC01411*, *TRAK2*, *LPHN2* [10], and *FAM78B* [11]. The paucity of significant and replicated associations for CUD limits our understanding of this disorder, hampering our ability to identify novel pharmacological targets.

Progress in human studies has been constrained by small sample sizes, complex diagnostic criteria, and the need for cocaine-exposed controls. Rodent models provide a complementary resource for exploring the genetic basis of CUD-like behaviors under tightly controlled experimental conditions. Recent studies using the hybrid mouse diversity panel (HMDP) and BXD recombinant inbred mouse population have identified candidate genes associated with cocaine self-administration [12,13], including *Fam53b,* an ortholog to the human gene (*FAM53B*) implicated in CUD [14].

In addition to inbred strains and recombinant inbred panels, outbred rodent models offer a valuable tool for dissecting the genetic basis of CUD-like behaviors. Outbred populations such as the Heterogeneous Stock (HS) rats provide a high degree of genetic diversity and low levels of linkage disequilibrium, enabling greater mapping precision. Prior work has established the phenotypic diversity of HS rats across a broad range of addiction-relevant behaviors, including cocaine self-administration [15-22]. Importantly, GWAS conducted in HS rats have identified numerous genetic variants associated with clinically relevant phenotypes related to a broad spectrum of human disorders [23-30].

Building on this framework, we expanded the sample used for a phenotypic analysis in de Guglielmo et al. (2024) to include nearly 900 HS rats that had undergone extended access to cocaine intravenous self-administration. This procedure captures key stages of the addiction process, such as initial acquisition of cocaine self-administration, escalation of cocaine intake, motivation to seek cocaine, and compulsive-like cocaine use despite adverse consequences [31-33]. By dissecting both the macrostructure, such as patterns of intake over multiple sessions, and the microstructure of cocaine self-administration, such as the differences in time between drug infusions, we can examine both correlated and

orthogonal aspects of cocaine self-administration. This level of detail is unobtainable in human studies, highlighting a major advantage of performing this work in rats. We then used SNP genotypes at millions of markers to perform a GWAS analysis. In addition to the GWAS results, we used various secondary analysis strategies, such as integration with coding variants and eQTLs to uncover novel genetic drivers of cocaine self-administration behaviors.

Methods and Materials

To measure behaviors related to CUD, HS rats underwent experimental testing for extended access to cocaine self-administration (Figure 1A). More detailed experimental protocols can be found at https://www.protocols.io/workspaces/george-lab and are also described in prior publications [20,21]. All phenotypic and genetic data can be found in the supplementals or UC Library (https://doi.org/10.6075/J0XD1222) [34].

Animals

This study contains data from 20 cohorts of HS rats (approximately 1,125 total experimental rats), spanning over 8 years. The first 17 cohorts were bred at Wake Forest University School of Medicine by Dr. Leah Solberg Woods (NMcwiWFsm:HS #13673907, RRID:RGD_13673907). The last 3 cohorts were bred at the University of California San Diego by Dr. Abraham A. Palmer (McwiWfsmAap:HS #155269102, RRID:RGD_155269102). Rats were shipped or transferred to Dr. Olivier George's lab in batches of 30-60 rats at 3 to 4 weeks of age. RFID chips were inserted into all rats at weaning. The first 6 cohorts were tested at the Scripps Research Institute, and the subsequent 14 cohorts were tested at the University of California San Diego. Each cohort consisted of a similar number of female and male rats. Upon arrival at the George lab, all rats from Wake Forest were kept in quarantine for two weeks. Following quarantine or transfer, rats of the same sex were pair-housed in the vivarium with reverse 12-hour light/dark cycle, temperature and humidity control, and ad libitum access to food and water. All experimental testing was approved by the Institutional Animal Care and Use Committees at the Scripps Research Institute or the University of California San Diego. All rats were treated in compliance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals.

Intravenous catheterization

The detailed protocol for intravenous jugular catheterization can be found at https://www.protocols.io/view/intravenous-jugular-catheterization-for-rats-36wgq5z45gk5/v1. Rats (age 7-9 weeks) were anesthetized with isoflurane vapor (induced at 5% and maintained at 2-3%). The jugular vein was exposed via an incision, and a catheter was inserted through an incision in the back and threaded under the skin to the neck. The catheter was inserted into the vein, secured with sutures, flushed, and checked for blood flow before the surgical site was closed. Post-operative care included administration of antibiotics (Cefazolin 330 mg/kg intramuscular) and analgesics (Flunixin 2.5 mg/kg subcutaneous) administered immediately following surgery and as needed over the next 5 days. Rats were monitored before returning to home cages. After catheterization surgery, rats were allowed to recover for at least one week before testing. Catheters were flushed daily with heparinized saline (10 U/ml of heparin sodium; American Pharmaceutical Partners, Schaumberg, IL, USA) in 0.9% bacteriostatic sodium chloride (Hospira, Lake Forest, IL, USA), which contained 52.4 mg/0.2 ml of Cefazolin).

Cocaine self-administration (escalation)

The detailed protocol for cocaine self-administration can be found at https://www.protocols.io/view/cocaine-iv-self-administration-dm6gpwx3dlzp/v1 [20,21]. All behavior testing was performed during the dark cycle, with the start of the session aligned to the lights going off in the housing room. The operant chambers for self-administration experiments were housed in sound-attenuating cubicles (Med Associates, St. Albans, VT, USA). The chambers were equipped with two retractable levers (active and inactive), and sessions began when these were extended. Infusion pumps were used to deliver cocaine HCl dissolved in 0.9% sterile saline at a dose of 0.5 mg/kg per infusion. Rats were weighed each week and given the appropriate corresponding drug concentration for their weight. Responses on the right (active) lever activated the infusion pump, which delivered 0.1 mL of drug over 6 seconds followed by a 20 second timeout. The timeout period was signaled by illumination of a light above the active lever, and rats would not receive additional drug for active lever presses during this time. Responses on the left (inactive) lever were recorded but had no consequence.

After one week of recovery after surgery, rats (age 8-10 weeks) underwent 10 sessions of short access (ShA) to cocaine self-administration for 2 hours per day, 5 days a week, under a fixed ratio 1 schedule (FR1). After ShA, rats underwent 14 sessions of long access (LgA) to cocaine self-administration for 6 hours per day, 5 days a week (Figure 1A).

Brevital test

After LgA, in order to test catheter patency, rats were administered Brevital (methohexital sodium; 2 mg/kg), which is a barbiturate with short-acting anesthetic effects. Rats with patent intravenous catheters show an immediate loss of muscle tone after Brevital administration. Any rat that failed to react to the Brevital infusion was recorded as a "catheter fail". If a rat had a catheter fail, behavioral data and health and experimental notes of the rat were reviewed by the team to determine the last good experimental session day where the catheter was functional. If this was unclear, the rat was excluded from subsequent trait analysis. These decisions were made by highly trained individuals who were blind to the genotype of the rat in question.

Progressive ratio (motivation)

At the end of both ShA and LgA, rats underwent one and two sessions, respectively, of self-administration on a progressive ratio (PR) schedule of reinforcement (Figure 1A). During PR, response requirements for receiving a single dose of cocaine increased in the following manner: 1, 2, 4, 6, 9, 12, 15, 20, 25, 32, 40, 50, 62, 77, 95, 118, 145, 178...[35]. The PR measure of total presses was defined as the number of active lever presses a rat made (not the number of infusions they received). The session ended once a rat did not complete the required number of presses within a 60 minute period.

Foot shock (compulsive-like responding)

Between the two sessions of PR after LgA, rats underwent a 1-hour foot shock punishment test under the same FR schedule as ShA and LgA self-administration (Figure 1A). Rats received a foot shock (0.3 mA, 0.5 s) for 30% of cocaine infusions over this hour. The

"shock compulsivity" measure was calculated as the total number of rewards in this session after the initial shock. Rats who never received a shock were not included in this trait for downstream genetic analysis.

Phenotype data processing

Behavioral data, including time stamps for active and inactive lever presses and infusions, was recorded by MED-PC IV software. Experimental notes, including any catheter disconnections, sickness, or death, were also recorded. Of the 1,125 rats that underwent behavioral testing, 158 (13%) were excluded due to sickness, death, catheter failure, or other experimental complications, and 131 (11%) were excluded due to missing or unusable genotype data. Thus, data for part or all of the study was available from 836 rats.

Data from MED-PC files were extracted using custom Python scripts. Data were visually inspected by at least 3 people by plotting all summary data for each day for each individual rat. Missing data from catheter disconnections or computer issues was imputed using Soft-Impute [36]. Specifically, we only imputed values for the total number of infusions for short or long access days if the data was missing. We did not impute timestamp data (list of time signatures for infusions or lever presses), or any other behavioral experiment. Imputation used the number of infusions for behavioral data, but did not consider any genetic data. If a rat died during the experiment, data for the days after death were considered missing and were not imputed or used in subsequent behavioral or genetic analysis. In cases of death or sickness, the team visually inspected all behavioral data to determine the likely onset of sickness or cause of death. Only sessions in which the animal appeared healthy were included in the trait analysis. To avoid including data potentially influenced by early signs of illness or experimental complications leading to death, any sessions during the four days immediately preceding death were always excluded, as well as any subsequent sessions after the onset of sickness. All data that passed quality checks, including imputed data, was used in the genetic analysis.

If only a portion of data was kept for a rat, the missing data was denoted with NAs. While we were able to impute total number of infusions on a given day of short or long access, missing timestamp data could not be imputed, thus some calculated traits that rely on timestamp data could not be determined. After applying these quality control measures data from 836 rats were available for the subsequent genetic analysis, however due to missing data that could not be imputed, for some traits, fewer than 836 rats were available for analysis.

Addiction-like trait calculations

All traits included in this study are detailed in Supplemental Table 1. For both ShA and LgA, early and late denote if a trait was derived from the first 3 or the last 3 or 4 sessions, respectively. Average infusions is the mean number of infusions over the designated

sessions. Escalation refers to the increase in the total number of infusions over LgA, and specifically LgA late average infusions is the measure that is most often associated with this trait. Post infusion interval is the median time in between infusions over the designated sessions. Loading is the total number of infusions received in the first 10 minutes over the designated sessions. Titration is the total number of infusions received in the last hour over the designated sessions. Timeout is the average difference between the active lever presses and number of infusions over the designated sessions. PR total presses refers to the total number of active lever presses made in the designated session. For calculated traits that take the sum of infusions across multiple sessions, rats with missing data for any of the days within the time period specified were excluded (e.g., if a rat had NA for session 14 because it was sick, LgA late traits was not calculated). For any trait that was calculated using timestamps (e.g., traits that require total or average numbers of lever presses or infusions), rats without timestamp data were excluded. For shock measures, rats that never received a foot shock (even if they received an infusion) were excluded. The total number of infusions were calculated for both ShA and LgA, and these were summed to produce a grand total infusion measure. The total numbers of rats that had values for each trait can be found in Supplemental Table 1.

Behavioral correlations and PCA

To examine the correlations between all trait measures, phenotypic correlations were determined using Spearman's correlation. These correlations were visualized using Seaborn Clustermap, and hierarchical clustering was used to group traits according to their patterns of similarity. We performed principal component analysis (PCA) on all calculated traits from LgA (11 measures in total), including early and late post infusion interval, average infusions, loading, titration, and timeout, as well as LgA total infusions. These traits spanned multiple behavioral clusters. The first principal component (PC1) explained 54% of the variance of these traits; we termed this PC "LgA PC1 addiction-like behavior" and included it in downstream genetic analysis.

Trait processing

To ensure reproducibility, all traits used for GWAS were processed in the same way. First, we used linear regression to remove the effects of any covariate that explained more than 2% of the variance, including age, cohort, sex, coat color, or weight. Then all traits were quantile normalized as part of our pre-processing GWAS pipeline (https://github.com/sanchestm/GWAS-pipeline). This process addresses any deviations from normality in the traits and allows for the use of the same GWAS significance thresholds for all traits.

Genotyping

To genotype HS rats, we used low-coverage whole genome sequencing. DNA was obtained from spleens, which were collected postmortem, as described in detail by Chen et al. [37]. Detailed protocols for spleen collection, DNA isolation, sample processing, preparing sequencing libraries, and performing the sequencing can be found on protocols.io here: https://www.protocols.io/workspaces/cgord/publications. DNA was extracted using the

DNA Advance kit and sequencing libraries were prepared using the Twist 96-Plex Library Prep kit (TWIST Bioscience). Reads were aligned to the rat reference genome mRatBN7.2 from the Rat Genome Sequencing Consortium (GCA_015227675.2 GCF_015227675.2) using BWA-mem v0.7.17. Mapped sequences were used to construct haplotypes and impute biallelic single nucleotide polymorphism (SNP) genotypes using STITCH v1.6.6 [38]. This process resulted in a high-quality set of 7,358,643 SNPs with imputation quality scores (INFO) $\geq$ 0.9. We additionally removed all SNPs with high missing rates (missing rate > 0.1; 23,389 removed), low minor allele frequencies (MAF 0 - 0.005; 1,887,082 removed, most of which were monomorphic), and extreme deviations from Hardy-Weinberg Equilibrium (HWE p < 1e-10; 2,604 removed). This filtered set of 5,446,333 SNPs in 836 HS rats was used for all downstream analyses[34].

Genetic analysis

We performed a genome-wide association study on the traits described in the prior section using the mixed linear model analysis (MLMA) function from the Genome-wide Complex Trait Analysis (GCTA) software package [39] as described previously [24,27,29]. Briefly, the GWAS pipeline (https://github.com/sanchestm/GWAS-pipeline) uses a genetic relatedness matrix (GRM), which estimates the genetic relationship between individuals based on their genotypes on all autosomes, to control for population structure and accurately determine the genetic contribution to the measured traits. We used the GCTA leave one chromosome out (LOCO) method to avoid proximal contamination [40]. SNP heritability estimates were calculated with GCTA using the GRM and a restricted maximum likelihood (REML) approach [39]. GWAS significance thresholds were calculated using 1,000 permutation tests [41], which we were able to use across all traits because all traits were quantile normalized, making the genotype and phenotype distributions identical. SNPs reached genome-wide significance according to the following threshold: $-\log_{10}(p) = 5.58$ ($\alpha = 0.05$).

Results

HS rats exhibit wide variation in cocaine self-administration

Traits were derived from the following experimental assays: short access (ShA) and long access (LgA) to cocaine self-administration, progressive ratio (PR), and shock (Figure1A). Rats exhibited substantial individual differences in cocaine intake both during ShA and LgA (Figure 1B). While some rats maintained low levels of intake (sorted at the bottom of both ShA and LgA intake graphs in Figure 1B), others exhibited dramatic escalation of intake, some exceeding over 150 infusions per session by the end of LgA. This pattern shows the development of pronounced differences in cocaine intake under both short and long access conditions, which broadly parallels the variability in cocaine use among humans.

To characterize distinct behavioral domains underlying the range of addiction-like traits measured, we performed correlation and cluster analysis (Figure 1C). Almost every pairwise comparison in the matrix was significant. This pervasive covariance replicates our previous results [20] and suggests that the animals' intake patterns, pauses, effort, and shock tolerance reflects a common behavioral construct of addiction-like behavior and are likely at least partially controlled by a common set of heritable and environmental factors. Despite this pervasive covariance, the dendrogram still resolved into several coherent sub-modules. Emergent behavioral clusters with highly correlated traits included early and late acquisition (with mostly ShA-derived measures) and LgA escalation (with mostly LgA-derived measures). The post infusion interval cluster, which reflects the time between infusions, included both ShA and LgA post infusion interval measures, suggesting that it may be a stable trait with a high genetic component. This cluster was moderately anti-correlated with other behavioral clusters. This relationship is expected because a rat with a high intake will spend less time between infusions. Finally, timeout (TO), PR, and shock were clustered together, suggesting that they may represent similar measures of motivation to seek cocaine despite adverse consequences, including lack of access to cocaine (TO), increased effort (PR), and punishment (foot shock).

Principal component analysis of cocaine self-administration

To capture a composite measure of addiction-like behavior, we performed principal component analysis (PCA) on all traits derived from LgA sessions because these traits exhibited the greatest variability. The analyzed traits spanned multiple behavioral domains from the cluster analysis, including early and late measures of average intake, escalation of intake, and post infusion intervals. The first principal component (PC1) explained 55% of the variance among these traits (Figure 1D), providing a robust aggregate measure of addiction-like behavior. The biplot of PC1 and PC2 (Figure 1E) shows the direction of each trait along these axes, and the continuum of z-scored total infusions along PC1. Additionally, the heatmaps of all LgA traits (Figure 1F) highlight the relationship between individual LgA traits and LgA PC1 addiction-like behavior. Furthermore, traits from different

behavioral domains were well-represented (indicated by the behavioral cluster coloring), demonstrating that PC1 captures diverse aspects of LgA cocaine intake.

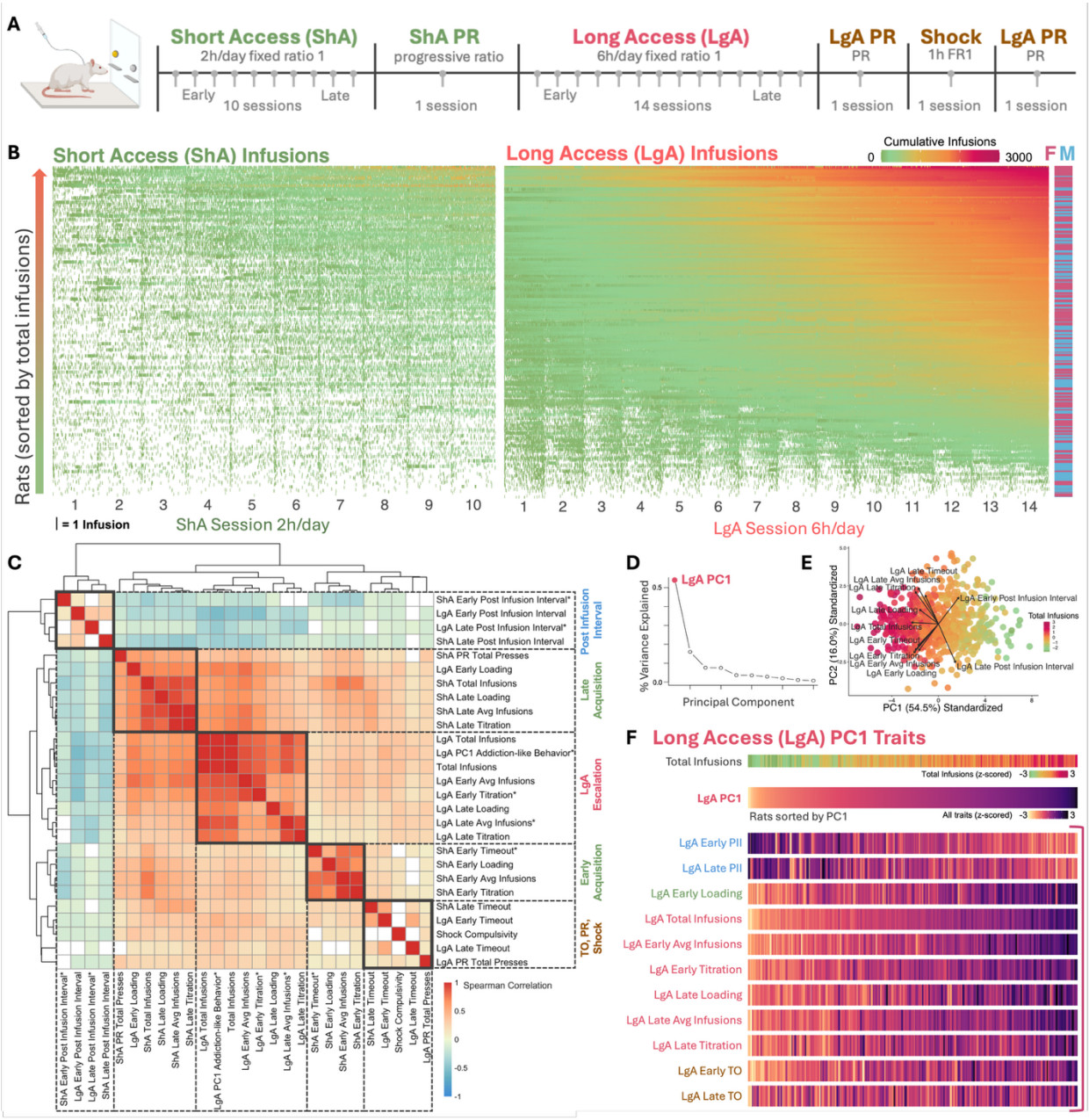

Figure 1. **HS rats show individual differences in addiction-like behaviors.** A) Timeline of extended access to cocaine self-administration, starting with short access (ShA) for 10 sessions 2 hours per day, then progressive ratio (PR) for 1 session, long access (LgA) for 14 sessions 6 hours per day, PR for 1 hour session, footshock punishment for 1 hour session, and a final PR 1 hour session. ShA and LgA traits were categorized as "early" or "late", corresponding to the beginning or end sessions of ShA or LgA. B) Individual infusions are shown over time for each rat. The y-axis of all rats is sorted by total infusions. The x-axis

shows the sessions for both short and long access. One line indicates one infusion. The color of the infusion line reflects the cumulative infusions the rat has received at that timepoint. C) Behavioral correlations were calculated for all traits using Spearman's correlation and traits were clustered using average linkage hierarchical clustering. Clusters are indicated with boxes and labeled based on the traits they contain. Cluster label colors are consistent with color labels for the behavioral experiment they came from, with the post infusion interval being notably separated. Non-significant values are indicated in white. Generally, LgA traits clustered together, and both early and late post infusion intervals clustered separately. D) To aggregate traits for a composite behavioral phenotype of addiction-like behavior, we used PC1 of all LgA traits, which accounted for 55% of the variance. E) Biplot of PC1 and PC2 showing where each LgA trait falls along the PC axes. Each dot represents a rat, and the color reflects the total number of infusions. F) All traits included in LgA PC1 addiction-like behavior are represented, and for each heatmap, the x axis is all rats sorted by their LgA PC1 addiction-like behavior value. The total infusions trait is included for reference, and generally, rats with higher total infusions are towards the right of the x axis. Labels for traits included in LgA PC1 addiction-like behavior are also colored by the behavioral cluster in which they belong.

GWAS

To identify genetic loci associated with cocaine self-administration behaviors, we performed GWAS on all measures from ShA, LgA, PR, and shock (Supplemental Report). Of these 27 traits, there were six genome-wide significant associations ($-\log_{10}(p) = 5.58$, $\alpha = 0.05$) between traits and regions across the HS rat genome. The porcupine plot in Figure 2A summarizes the combined p-values for all 6 traits with significant associations, including ShA early post infusion interval, LgA late average infusions, ShA early timeout, LgA late post infusion interval, LgA PC1 addiction-like behavior, and LgA early titration. Traits from multiple behavioral clusters contributed to significant associations, indicating that genetic influences span multiple aspects of cocaine self-administration.

Heritability

To quantify the proportion of phenotypic variance attributable to SNPs, we calculated SNP-based heritability for each trait. SNP heritability estimates reflect the extent to which variation in the trait is explained by additive effects of all the SNPs in the HS rat genome. Heritabilities for all traits ranged from low to moderate (Figure 2B, Supplemental Table 2). The highest significant heritability ($h^2 = 0.16$, $p = 0.005$) was observed for LgA PC1 addiction-like behavior. Additionally, multiple traits across behavioral domains for ShA and LgA showed significant heritability ranging from $h^2 = 0.07$ to 0.16 ($p < 0.05$). For example, the total intake of cocaine infusions during LgA was heritable (LgA total infusions, $h^2 = 0.13$, $p = 0.009$), which is a measure that is often used across drug self-administration studies to determine resistance or vulnerability to the development of addiction-like phenotypes.

Locus for LgA late post infusion interval

GWAS identified a significant locus on chromosome 19 associated with LgA late post infusion interval, a measure reflecting the time elapsed between successive cocaine infusions. The top SNP within this locus was located at position 19:13,896,276 ($-\log_{10}(p) = 5.61$), and the associated linkage disequilibrium (LD) block spanned from ~13,800,000 bp to ~14,050,000 bp, a region that is approximately 0.25 mb (Figure 2C). Three non-synonymous coding variants were identified within this locus that were in almost perfect LD ($r^2 > 0.99$) with the top SNP. Two of these were missense variants located in the coding region of the *Ces1d* gene: 19:13,883,660 (c.557A>G, p.Q186R) and 19:13,896,831 (c.1269G>A, p.M423I). The third missense variant, 19:13,941,339 (c.1226C>T, p.P409L), was in the coding region of the *Ces1c* gene. The genes *Ces1c* and *Ces1d* encode carboxylesterase 1 enzymes and are orthologous to human *CES1*. The enzyme Carboxylesterase 1 is one of the main enzymes responsible for the metabolism of cocaine [42]. Based on the available data, it is not clear whether these coding variants would be expected to increase or decrease metabolism of cocaine.

In addition to coding variants, several expression quantitative trait loci (eQTL) were identified within this LD block.[18] These eQTLs affect expression of *Rasd2* in the nucleus accumbens core, prelimbic cortex, and whole brain tissue (all with $r^2 = 1.0$). *Rasd2* has been implicated in several relevant phenotypes, including sensitivity to the stimulant effects of cocaine [43]. Finally, the coding region for the gene *Slc6a2,* which encodes the norepinephrine transporter, which is inhibited by cocaine, is also located within this locus; however, we did not identify any coding variants or eQTLs for this gene, suggesting that it is not the cause of this association.

The top SNP in this locus had a minor allele frequency of 48%. Rats with the CC genotype at this SNP exhibited lower LgA post infusion intervals (Figure 2D). Lower post infusion interval values, or those that are in the bottom quartile, correlated with high levels of cocaine intake and greater addiction-like traits.

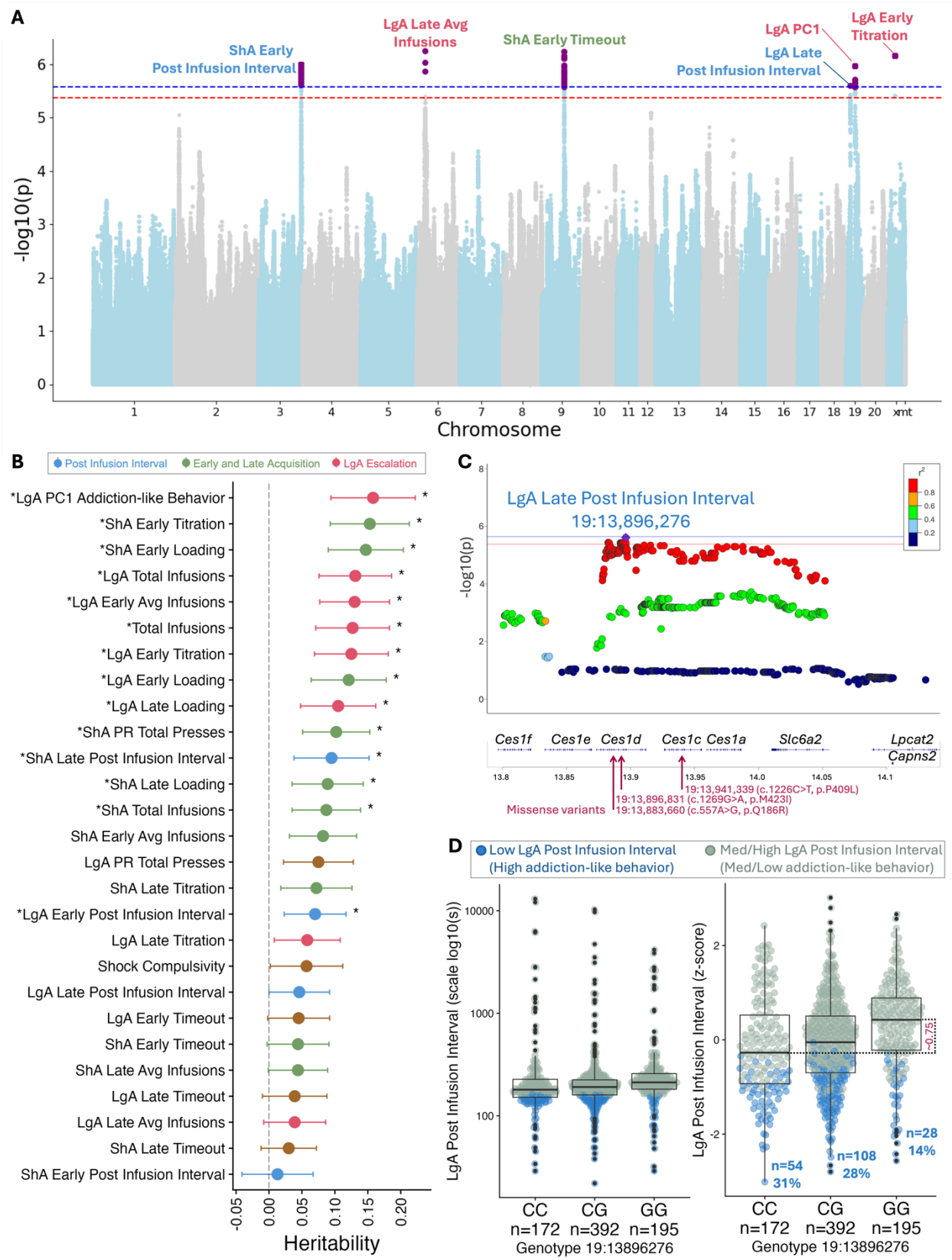
A
ShA Early Post Infusion Interval
LgA Late Avg Infusions
ShA Early Timeout
LgA PC1
LgA Early Titration
LgA Late Post Infusion Interval
-log10(p)
Chromosome
B
Post Infusion Interval
Early and Late Acquisition
LgA Escalation
*LgA PC1 Addiction-like Behavior
*ShA Early Titration
*ShA Early Loading
*LgA Total Infusions
*LgA Early Avg Infusions
*Total Infusions
*LgA Early Titration
*LgA Early Loading
*LgA Late Loading
*ShA PR Total Presses
*ShA Late Post Infusion Interval
*ShA Late Loading
*ShA Total Infusions
ShA Early Avg Infusions
LgA PR Total Presses
ShA Late Titration
*LgA Early Post Infusion Interval
LgA Late Titration
Shock Compulsivity
LgA Late Post Infusion Interval
LgA Early Timeout
ShA Early Timeout
ShA Late Avg Infusions
LgA Late Timeout
LgA Late Avg Infusions
ShA Late Timeout
ShA Early Post Infusion Interval
Heritability
C
LgA Late Post Infusion Interval
19:13,896,276
-log10(p)
Ces1f
Ces1e
Ces1d
Ces1c
Ces1a
Slc6a2
Lpcat2
Capns2
Missense variants
19:13,941,339 (c.1226C>T, p.P409L)
19:13,896,831 (c.1269G>A, p.M423I)
19:13,883,660 (c.557A>G, p.Q186R)
D
Low LgA Post Infusion Interval (High addiction-like behavior)
Med/High LgA Post Infusion Interval (Med/Low addiction-like behavior)
LgA Post Infusion Interval (scale log10(s))
LgA Post Infusion Interval (z-score)
CC n=172
CG n=392
GG n=195
Genotype 19:13896276
n=54 31%
n=108 28%
n=28 14%
-0.75

Figure 2. Genome-wide association study for cocaine self-administration in HS rats. A) The porcupine plot displays all *p*-values combined for traits with a significant association distributed across the entire genome of HS rats. The orange and yellow lines show the significance thresholds derived from the permutation tests: $-\log_{10}(p) = 5.58$ ($\alpha = 0.05$) and $-\log_{10}(p) = 5.38$ ($\alpha = 0.10$), respectively. B) Heritability of all cocaine self-administration traits. Error bars represent standard error; the color reflects the behavioral cluster from which the trait belongs (Figure1B). Stars indicate significant heritabilities ($p < 0.05$). C) Regional association plot zooming into the locus on chromosome 19 that was significantly associated with post infusion interval. The individual dots represent SNPs, and the color of the dots indicate linkage disequilibrium between the top SNP (in purple) and other nearby SNPs. The blue and red lines show the significance thresholds: $-\log_{10}(p) = 5.58$ ($\alpha = 0.05$) and $-\log_{10}(p) = 5.38$ ($\alpha = 0.10$), respectively. Below the x-axis shows the genes mapped to that chromosomal position. There are three missense variants located in this locus. D) Effect plot for Post Infusion Interval depicting the genetic effect of the top SNP, with genotype on the x-axis, including homozygous reference alleles, heterozygous, and homozygous alternate alleles. Both plots show LgA post infusion interval on the y-axis, with the first being $\log_{10}$ scaled, and the second z-scored for better visualization. The z-scored difference (~0.75) of Post Infusion Interval between CC and GG genotypes is indicated between the dashed lines. The blue dots represent rats who were in the bottom quartile for LgA Post Infusion Interval.

Loci for additional cocaine self-administration traits

GWAS identified five additional loci exceeding the genome-wide significance threshold that were associated with CUD-like traits (Figure 2A, Supplemental Report). Other tables include: all significant loci (Supplemental Table 3), the corresponding genes in range (Supplemental Table 4), eQTLs (Supplemental Table 5), and splice QTLs (sQTLS, Supplemental Table 6).

LgA PC1 addiction-like behavior, derived from LgA traits spanning multiple behavioral domain clusters, was significantly associated with a locus on chromosome 19 at 29,567,757 bp ($-\log_{10}(p) = 5.98$). Two eQTLs ($r^2 = 0.9$) were identified in whole brain tissue and the nucleus accumbens core for the gene *Lsm6*, which is involved in RNA binding and splicing. Additionally, four sQTLs for *Lsm6* were identified in the basolateral amygdala, nucleus accumbens core, prelimbic cortex, and whole brain tissue ($r^2 = 0.65$, 0.94, 0.66, and 0.68, respectively). Finally, the gene *Slc10a7*, which has been implicated in smoking [10,44,45] and nicotine dependence [46] was located in this locus.

ShA early post infusion interval, which represents the median time elapsed between infusions during the first three sessions of ShA, was associated with a locus on chromosome 3 (Supplemental Report). The top SNP within this locus was at position 3:163,216,140 ($-\log_{10}(p) = 5.997$) (Supplemental Report). Seven missense variants in perfect LD ($r^2 = 1.0$) with this SNP were identified within the gene *Zfp831,* which encodes a

zinc finger protein. This region also included an eQTL for *Zfp831* in the basolateral amygdala. Additionally, three eQTLs for the gene *Gnas,* which encodes the stimulatory alpha subunit of a G protein, were detected at SNPs in moderate to high LD (r^2 = 0.6 – 1.0) in the lateral habenula, nucleus accumbens core, and the prelimbic cortex. This locus also contained a coding variant for the gene *Ctsz* (r^2 = 1.0), as well as eQTLs for *Ctsz* in the nucleus accumbens core and whole brain tissue (both with r^2 = 1.0).

LgA late average infusions, a measure of escalation of intake, was associated with a locus on chromosome 6, with the peak SNP at position 6:34,284,564 ($-\log_{10}(p) = 6.25$) (Supplemental Report). Six significant coding variants in perfect LD (r^2 = 1.0) were identified in this locus, all in the gene *Rad51ap2*, which plays a role in *Rad51* activity as part of DNA repair pathways and meiotic recombination [47]. Two additional SNPs in high LD with the top SNP (r^2 = 0.73 and 0.75) were sQTLs for the gene *Vsnl1* in brain and prelimbic cortex tissues. *Vsnl1* encodes a protein that is part of the visinin/recoverin subfamily of neuronal calcium sensor proteins [48].

Lastly, ShA early timeout, calculated as the difference between the number of active lever presses and the number of infusions received, captures burst pressing and possibly perseverative behavior. A locus on chromosome 9 was significantly associated with ShA early timeout, with the peak SNP at position 9:57520529, and the implicated region spanned approximately 4 mb (Supplemental Report). While there were no coding variants or eQTLs in this region that were in high LD ($r^2 > 0.7$) with the top SNP, several SNPs in this locus were in moderate LD ($0.5 < r^2 < 0.7$) with the top SNP and are associated with genes that have been implicated in human GWAS for alcohol drinking, nicotine, and cocaine use: *Plcl1, Satb2,* and *Trak2* [10,44,45].

Discussion

To investigate the underlying genetic factors influencing resilience or susceptibility to CUD-related traits, we performed a GWAS in HS rats using an extended access self-administration model. The phenotypes assessed in this study capture key dimensions of addiction-like behaviors. The extended access model allowed us to characterize escalating intake, increased motivation to take the drug, and compulsive-like behavior despite negative consequences [20,32]. We identified multiple significant loci associated with CUD-like behaviors, notably including a region containing coding variants for *Ces1c* and *Ces1d* that encode enzymes involved in cocaine metabolism that is associated with the post infusion interval.

Twin studies of cocaine use, abuse, and dependence have high narrow-sense heritability estimates ranging from ~0.3 – 0.7 [2-4,49,50]. More recent GWAS studies estimate SNP-based heritability for cocaine dependence to be around ~0.3 [7,8]. SNP heritability represents the proportion of phenotypic variance explained by the additive effects of SNPs included in the study. SNP heritability is expected to be lower than heritability estimates from twin studies in part because it only captures variance attributable to common variants. We found heritabilities for CUD-like traits in HS rats ranged from 0.07 to 0.16. Interestingly, the highest heritability was attributed to the principal component derived from multiple CUD-like behavioral domains: LgA PC1 addiction-like behavior. This is notable because PC1 integrates multiple phenotypic dimensions into a robust metric, and the relative high heritability of 16% underscores the utility in mapping using principal components. These findings suggest that genetic influences on addiction-like behavior may become more pronounced as drug exposure escalates, aligning with the progressive nature of substance use disorders in humans.

Human GWAS of CUD remain in their infancy, with sample sizes still numbering in the thousands. These studies are limited by the difficulty of obtaining cases as well as uncertainty about whether controls should have previous exposure to cocaine. Genome-wide significant associations have been reported for a few genes including *FAM53B, NCOR2* [9]; *FAM78B* [11]; and *LINC01411*, *TRAK2*, and *LPHN2* [10]. One of the genes we identified (*Trak2*) is the rat homolog of *TRAK2*, which is one of the few genes that has also been associated with CUD in humans [10]. The fact that some of the genes are not observed across species likely reflects low power in both human and rat GWAS, differences between the CUD phenotype in humans and the phenotypes measured in HS rats, and differences in common, polymorphic alleles present in the human populations that have been studied and HS rats. Given these limitations, the fact that even a single gene has been found in both species is remarkable.

In addition to the overlap of *Trak2* between rats and human CUD GWAS, several other genes we identified have also been identified in human GWAS of other SUDs. For example, *SLC10A7* has also been implicated in smoking initiation [45] and nicotine dependence [46];

PLCL1 has been implicated in the maximum number of alcoholic drinks (Kapoor et al., 2013) and smoking initiation [44,45]; and *SATB2* has been implicated smoking initiation [45].

We identified a genome-wide significant locus on chromosome 19 for LgA post infusion interval, which contained several *Ces1* genes. Three missense variants were present in *Ces1c* and *Ces1d*, which encode enzymes that are involved in the metabolism of various substances, including cocaine [42,51]. As a drug-metabolizing enzyme, CES1 hydrolyzes cocaine into inactive metabolites, influencing both the duration and intensity of cocaine's effects on the brain. At present, it is not clear whether or how these coding variants might alter enzymatic activity. Differences in *Ces1* activity in HS rats may impact CUD-like traits, reflecting the role of this gene in regulating the response to repeated drug exposure. This locus was significantly associated with the elapsed time between infusions, a key behavioral measure linked to the compulsive nature of cocaine self-administration, as hypothesized by the compulsion zone theory [52], which proposes that cocaine self-administration is controlled by pharmacokinetic constraints, where rats maintain cocaine levels above the priming/remission threshold but below the satiety threshold to sustain drug-taking behavior. *CES1* has already been discussed as a potential target for therapeutic interventions [53], and our findings provide further support for studies aimed at modulating addiction-like behaviors by modulating CES1 activity.

In addition to drug metabolism, we identified genes of various biological function that were associated with addiction related traits, as well as expression and splice QTLs that may explain how these loci affect the behavioral differences seen across rats [18]. Central to the process of addiction are neurocircuits involved in reward, motivation, learning, and inhibitory control [54-57]. Within the 6 significant loci we identified in this study, several genes had eQTLs and sQTLs in brain regions heavily involved in these processes [18]. Genes in the implicated loci with significant eQTLs and sQTLs were located in the nucleus accumbens (*Rasd2, Gnas, Ctsz, Lsm6*), which is involved in the dopamine reward circuit associated with binging and intoxication, as well as the prelimbic cortex (*Rasd2, Gnas, Vsnl1, Lsm6*) and amygdala (*Zfp831, Lsm6)*, which play a role in negative emotional states and stress responses in the withdrawal stage [54,55,58-63]. Critically, these circuits involve dopamine signaling, which can be altered as a result of chronic drug use [64-70]. Among the genes identified in this study, several may be involved in this process. For example, *Rasd2,* which encodes a Ras protein enriched in the striatum, has been shown to modulate dopamine neurotransmission, and have effects on motor response after cocaine exposure [43,71,72]. *Gnas*, which encodes the stimulatory alpha subunit of G protein coupled receptors, including dopamine receptors, has been shown to regulate dopamine receptor signaling and have downstream effects after cocaine exposure [73-76]. Lastly, *Trak2,* which encodes the trafficking kinesin protein 2, interacts and transports GABA receptors in the brain [77,78]. The GABAergic system is crucial for inhibitory neuronal activity that regulates dopamine reward and reinforcement pathways in addiction [79].

There is growing evidence that there are shared genetic drivers underlying multiple substance use disorders (SUDs) or addiction-related behaviors [80,81]. Much of this genetic signal appears to represent non-substance specific factors that may reflect behavioral under-control [82], which may partially act through their ability to influence initiation of drug taking. In contrast, the procedure used in our study assured that all rats have an opportunity to try cocaine. Therefore, our GWAS is intended to capture substance-specific genetic factors. The identification of *Ces1c* and *Ces1d*, which influence cocaine metabolism, reflect this bias towards cocaine-specific factors.

Many prior animal studies have examined responses to drug exposure. For examine, Zhou, et al. [83], examined gene expression at the single cell level using some of the same HS rats that are included in this GWAS. That study highlighted differences in gene expression that were observed after 4 weeks of cocaine abstinence in rats that self-administered either large or small amounts of cocaine. In contrast, the present study sought to identify genes that predispose rats to voluntarily self-administer large or small amounts of cocaine. These two approaches highlight the difference between temporally variable states (cocaine exposure) and stable traits (genotype).

To our knowledge, this is the largest study of extended access to cocaine self-administration in outbred rats. Our GWAS identified multiple loci significantly associated with CUD-like traits in HS rats, including escalation of intake of cocaine, time elapsed between cocaine infusions, and an aggregate measure of addiction-like behavior. This study highlights the utility of rich behavioral phenotyping and using genetically diverse rats to identify novel genetic loci associated with the multifaceted process of drug use and the transition to addiction-like behaviors. Samples from some of the rats used in these studies are available thought the Cocaine biobank (https://www.cocainebiobank.org/) [21]. Genes associated with CUD in humans remain limited, however our GWAS identified one gene (*Trak2*) that has also been identified by human GWAS of CUD, and the novel identification of *Ces1* offers a fresh avenue for future studies.

Acknowledgements
Funding was provided by the following NIH grants: P50DA037844, P30DA060810, U01DA051234, and U01DA043799.

Disclosures
The authors of this paper declare that they have no relevant disclosures.

References

1 Substance Abuse and Mental Health Services Administration. Key substance use and mental health indicators in the United States: Results from the 2022 National Survey on Drug Use and Health (HHS Publication No. PEP23-07-01-006, NSDUH Series H-58). (Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration, 2023).

2 Kendler, K. S. & Prescott, C. A. Cocaine use, abuse and dependence in a population-based sample of female twins. *Br J Psychiatry* **173**, 345-350 (1998). https://doi.org/10.1192/bjp.173.4.345

3 Kendler, K. S., Karkowski, L. M., Neale, M. C. & Prescott, C. A. Illicit Psychoactive Substance Use, Heavy Use, Abuse, and Dependence in a US Population-Based Sample of Male Twins. *Archives of General Psychiatry* **57**, 261-269 (2000). https://doi.org/10.1001/archpsyc.57.3.261

4 van den Bree, M. B., Johnson, E. O., Neale, M. C. & Pickens, R. W. Genetic and environmental influences on drug use and abuse/dependence in male and female twins. *Drug Alcohol Depend* **52**, 231-241 (1998). https://doi.org/10.1016/s0376-8716(98)00101-x

5 Fernàndez-Castillo, N., Cabana-Domínguez, J., Corominas, R. & Cormand, B. Molecular genetics of cocaine use disorders in humans. *Molecular Psychiatry* **27**, 624-639 (2022). https://doi.org/10.1038/s41380-021-01256-1

6 Pierce, R. C. *et al.* Environmental, genetic and epigenetic contributions to cocaine addiction. *Neuropsychopharmacology* **43**, 1471-1480 (2018). https://doi.org/10.1038/s41386-018-0008-x

7 Huggett, S. B. & Stallings, M. C. Cocaine'omics: Genome-wide and transcriptome-wide analyses provide biological insight into cocaine use and dependence. *Addict Biol* **25**, e12719 (2020). https://doi.org/10.1111/adb.12719

8 Cabana-Domínguez, J., Shivalikanjli, A., Fernàndez-Castillo, N. & Cormand, B. Genome-wide association meta-analysis of cocaine dependence: Shared genetics with comorbid conditions. *Prog Neuropsychopharmacol Biol Psychiatry* **94**, 109667 (2019). https://doi.org/10.1016/j.pnpbp.2019.109667

9 Gelernter, J. *et al.* Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Mol Psychiatry* **19**, 717-723 (2014). https://doi.org/10.1038/mp.2013.99

10 Sun, J., Kranzler, H. R., Gelernter, J. & Bi, J. A genome-wide association study of cocaine use disorder accounting for phenotypic heterogeneity and gene–environment interaction. *J Psychiatry Neurosci* **45**, 34-44 (2020). https://doi.org/10.1503/jpn.180098

11 Sherva, R. *et al.* Genome-wide association study of phenotypes measuring progression from first cocaine or opioid use to dependence reveals novel risk genes. *Explor Med* **2**, 60-73 (2021). https://doi.org/10.37349/emed.2021.00032

12 Bagley, J. R., Khan, A. H., Smith, D. J. & Jentsch, J. D. Extreme phenotypic diversity in operant response to intravenous cocaine or saline infusion in the hybrid mouse diversity panel. *Addict Biol* **27**, e13162 (2023). https://doi.org/10.1111/adb.13162

13 Khan, A. H. *et al.* Genetic pathways regulating the longitudinal acquisition of cocaine self-administration in a panel of inbred and recombinant inbred mice. *Cell Rep* **42**, 112856 (2023). https://doi.org/10.1016/j.celrep.2023.112856
14 Dickson, P. E. *et al.* Systems genetics of intravenous cocaine self-administration in the BXD recombinant inbred mouse panel. *Psychopharmacology (Berl)* **233**, 701-714 (2016). https://doi.org/10.1007/s00213-015-4147-z
15 Hansen, C. & Spuhler, K. Development of the National Institutes of Health genetically heterogeneous rat stock. *Alcohol Clin Exp Res* **8**, 477-479 (1984). https://doi.org/10.1111/j.1530-0277.1984.tb05706.x
16 Solberg Woods, L. C. & Palmer, A. A. Using Heterogeneous Stocks for Fine-Mapping Genetically Complex Traits. *Methods Mol Biol* **2018**, 233-247 (2019). https://doi.org/10.1007/978-1-4939-9581-3_11
17 Ishiwari, K. *et al.* Environmental enrichment promotes adaptive responding during tests of behavioral regulation in male heterogeneous stock rats. *Sci Rep* **14**, 4182 (2024). https://doi.org/10.1038/s41598-024-53943-y
18 Munro, D. *et al.* The regulatory landscape of multiple brain regions in outbred heterogeneous stock rats. *Nucleic Acids Res* **50**, 10882-10895 (2022). https://doi.org/10.1093/nar/gkac912
19 Gancarz, A. M. *et al.* Reward maximization assessed using a sequential patch depletion task in a large sample of heterogeneous stock rats. *Sci Rep* **13**, 7027 (2023). https://doi.org/10.1038/s41598-023-34179-8
20 de Guglielmo, G. *et al.* Large-scale characterization of cocaine addiction-like behaviors reveals that escalation of intake, aversion-resistant responding, and breaking-points are highly correlated measures of the same construct. *Elife* **12** (2024). https://doi.org/10.7554/eLife.90422
21 Carrette, L. L. G. *et al.* The Cocaine and Oxycodone Biobanks, Two Repositories from Genetically Diverse and Behaviorally Characterized Rats for the Study of Addiction. *eNeuro* **8** (2021). https://doi.org/10.1523/eneuro.0033-21.2021
22 Kallupi, M. *et al.* Nociceptin attenuates the escalation of oxycodone self-administration by normalizing CeA-GABA transmission in highly addicted rats. *Proc Natl Acad Sci U S A* **117**, 2140-2148 (2020). https://doi.org/10.1073/pnas.1915143117
23 Chitre, A. S. *et al.* Genome-wide association study in a rat model of temperament identifies multiple loci for exploratory locomotion and anxiety-like traits. *Front Genet* **13**, 1003074 (2022). https://doi.org/10.3389/fgene.2022.1003074
24 Chitre, A. S. *et al.* Genome-Wide Association Study in 3,173 Outbred Rats Identifies Multiple Loci for Body Weight, Adiposity, and Fasting Glucose. *Obesity (Silver Spring)* **28**, 1964-1973 (2020). https://doi.org/10.1002/oby.22927
25 Fowler, S. *et al.* Genome-wide association study finds multiple loci associated with intraocular pressure in HS rats. *Front Genet* **13**, 1029058 (2022). https://doi.org/10.3389/fgene.2022.1029058
26 Gunturkun, M. H. *et al.* Genome-Wide Association Study on Three Behaviors Tested in an Open Field in Heterogeneous Stock Rats Identifies Multiple Loci Implicated in

Psychiatric Disorders. *Front Psychiatry* **13**, 790566 (2022). https://doi.org/10.3389/fpsyt.2022.790566

27 Kuhn, B. N. *et al.* Genome-wide association study reveals multiple loci for nociception and opioid consumption behaviors associated with heroin vulnerability in outbred rats. *bioRxiv* (2024). https://doi.org/10.1101/2024.02.27.582340

28 King, C. P. *et al.* Genomic Loci Influencing Cue-Reactivity in Heterogeneous Stock Rats. *bioRxiv* (2024). https://doi.org/10.1101/2024.03.13.584852

29 Lara, M. K. *et al.* Genome-wide association study of delay discounting in Heterogeneous Stock rats. *Genes Brain Behav* **23**, e12909 (2024). https://doi.org/10.1111/gbb.12909

30 Hong-Le, T. *et al.* Genetic Mapping of Multiple Traits Identifies Novel Genes for Adiposity, Lipids, and Insulin Secretory Capacity in Outbred Rats. *Diabetes* **72**, 135-148 (2023). https://doi.org/10.2337/db22-0252

31 George, O., Mandyam, C. D., Wee, S. & Koob, G. F. Extended access to cocaine self-administration produces long-lasting prefrontal cortex-dependent working memory impairments. *Neuropsychopharmacology* **33**, 2474-2482 (2008). https://doi.org/10.1038/sj.npp.1301626

32 Edwards, S. & Koob, G. F. Escalation of drug self-administration as a hallmark of persistent addiction liability. *Behav Pharmacol* **24**, 356-362 (2013). https://doi.org/10.1097/FBP.0b013e3283644d15

33 Koob, G. F. *et al.* Neurobiological mechanisms in the transition from drug use to drug dependence. *Neurosci Biobehav Rev* **27**, 739-749 (2004). https://doi.org/10.1016/j.neubiorev.2003.11.007

34 Palmer, A. A. Heterogeneous Stock (HS) Rat Genotypes, Version 7. (2025). https://doi.org/https://doi.org/10.6075/J0XD1222

35 Richardson, N. R. & Roberts, D. C. S. Progressive ratio schedules in drug self-administration studies in rats: a method to evaluate reinforcing efficacy. *Journal of Neuroscience Methods* **66**, 1-11 (1996). https://doi.org/https://doi.org/10.1016/0165-0270(95)00153-0

36 Mazumder, R., Hastie, T. & Tibshirani, R. Spectral Regularization Algorithms for Learning Large Incomplete Matrices. *J Mach Learn Res* **11**, 2287-2322 (2010).

37 Chen, D. *et al.* A Cost-effective, High-throughput, Highly Accurate Genotyping Method for Outbred Populations. *bioRxiv* (2024). https://doi.org/10.1101/2024.07.17.603984

38 Davies, R. W., Flint, J., Myers, S. & Mott, R. Rapid genotype imputation from sequence without reference panels. *Nature Genetics* **48**, 965-969 (2016). https://doi.org/10.1038/ng.3594

39 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011). https://doi.org/10.1016/j.ajhg.2010.11.011

40 Cheng, R., Parker, C. C., Abney, M. & Palmer, A. A. Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3 (Bethesda)* **3**, 1861-1867 (2013). https://doi.org/10.1534/g3.113.007948

41 Cheng, R. & Palmer, A. A. A simulation study of permutation, bootstrap, and gene dropping for assessing statistical significance in the case of unequal relatedness. *Genetics* **193**, 1015-1018 (2013). https://doi.org/10.1534/genetics.112.146332
42 Redinbo, M. R., Bencharit, S. & Potter, P. M. Human carboxylesterase 1: from drug metabolism to drug discovery. *Biochemical Society Transactions* **31**, 620-624 (2003). https://doi.org/10.1042/bst0310620
43 Napolitano, F. *et al*. The striatal-enriched protein Rhes is a critical modulator of cocaine-induced molecular and behavioral responses. *Sci Rep* **9**, 15294 (2019). https://doi.org/10.1038/s41598-019-51839-w
44 Kapoor, M. *et al*. A meta-analysis of two genome-wide association studies to identify novel loci for maximum number of alcoholic drinks. *Hum Genet* **132**, 1141-1151 (2013). https://doi.org/10.1007/s00439-013-1318-z
45 Saunders, G. R. B. *et al*. Genetic diversity fuels gene discovery for tobacco and alcohol use. *Nature* **612**, 720-724 (2022). https://doi.org/10.1038/s41586-022-05477-4
46 Drgon, T. *et al*. Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers. *Mol Med* **15**, 21-27 (2009). https://doi.org/10.2119/molmed.2008.00096
47 Kovalenko, O. V., Wiese, C. & Schild, D. RAD51AP2, a novel vertebrate- and meiotic-specific protein, shares a conserved RAD51-interacting C-terminal domain with RAD51AP1/PIR51. *Nucleic Acids Research* **34**, 5081-5092 (2006). https://doi.org/10.1093/nar/gkl665
48 Liebl, M. P. *et al*. Dimerization of visinin-like protein 1 is regulated by oxidative stress and calcium and is a pathological hallmark of amyotrophic lateral sclerosis. *Free Radical Biology and Medicine* **72**, 41-54 (2014). https://doi.org/https://doi.org/10.1016/j.freeradbiomed.2014.04.008
49 Kendler, K. S., Aggen, S. H., Tambs, K. & Reichborn-Kjennerud, T. Illicit psychoactive substance use, abuse and dependence in a population-based sample of Norwegian twins. *Psychol Med* **36**, 955-962 (2006). https://doi.org/10.1017/s0033291706007720
50 Kendler, K. S., Myers, J. & Prescott, C. A. Specificity of genetic and environmental risk factors for symptoms of cannabis, cocaine, alcohol, caffeine, and nicotine dependence. *Arch Gen Psychiatry* **64**, 1313-1320 (2007). https://doi.org/10.1001/archpsyc.64.11.1313
51 Wang, D. *et al*. Human carboxylesterases: a comprehensive review. *Acta Pharm Sin B* **8**, 699-712 (2018). https://doi.org/10.1016/j.apsb.2018.05.005
52 Norman, A. B. & Tsibulsky, V. L. The compulsion zone: a pharmacological theory of acquired cocaine self-administration. *Brain Res* **1116**, 143-152 (2006). https://doi.org/10.1016/j.brainres.2006.07.092
53 Her, L. & Zhu, H. J. Carboxylesterase 1 and Precision Pharmacotherapy: Pharmacogenetics and Nongenetic Regulators. *Drug Metab Dispos* **48**, 230-244 (2020). https://doi.org/10.1124/dmd.119.089680
54 Koob, G. F. & Volkow, N. D. Neurocircuitry of addiction. *Neuropsychopharmacology* **35**, 217-238 (2010). https://doi.org/10.1038/npp.2009.110

55 Fang, Y. *et al.* Neurobiological mechanisms and related clinical treatment of addiction: a review. *Psychoradiology* **2**, 180-189 (2022). https://doi.org/10.1093/psyrad/kkac021
56 Volkow, N. D., Wang, G. J., Fowler, J. S. & Tomasi, D. Addiction circuitry in the human brain. *Annu Rev Pharmacol Toxicol* **52**, 321-336 (2012). https://doi.org/10.1146/annurev-pharmtox-010611-134625
57 Nestler, E. J. The neurobiology of cocaine addiction. *Sci Pract Perspect* **3**, 4-10 (2005). https://doi.org/10.1151/spp05314
58 Kalivas, P. W. & O'Brien, C. Drug addiction as a pathology of staged neuroplasticity. *Neuropsychopharmacology* **33**, 166-180 (2008). https://doi.org/10.1038/sj.npp.1301564
59 McGregor, A. & Roberts, D. C. Dopaminergic antagonism within the nucleus accumbens or the amygdala produces differential effects on intravenous cocaine self-administration under fixed and progressive ratio schedules of reinforcement. *Brain Res* **624**, 245-252 (1993). https://doi.org/10.1016/0006-8993(93)90084-z
60 Salamone, J. D., Correa, M., Farrar, A. & Mingote, S. M. Effort-related functions of nucleus accumbens dopamine and associated forebrain circuits. *Psychopharmacology (Berl)* **191**, 461-482 (2007). https://doi.org/10.1007/s00213-006-0668-9
61 Scofield, M. D. *et al.* The Nucleus Accumbens: Mechanisms of Addiction across Drug Classes Reflect the Importance of Glutamate Homeostasis. *Pharmacological reviews* **68**, 816-871 (2016). https://doi.org/10.1124/pr.116.012484
62 Ahmed, S. H. & Kenny, P. J. Cracking the molecular code of cocaine addiction. *Ilar j* **52**, 309-320 (2011). https://doi.org/10.1093/ilar.52.3.309
63 Chow, J. J. *et al.* Neurobiology of the incubation of drug craving: An update. *Pharmacological reviews* **77**, 100022 (2025). https://doi.org/10.1016/j.pharmr.2024.100022
64 Wise, R. A. & Robble, M. A. Dopamine and Addiction. *Annu Rev Psychol* **71**, 79-106 (2020). https://doi.org/10.1146/annurev-psych-010418-103337
65 Anderson, S. M. & Pierce, R. C. Cocaine-induced alterations in dopamine receptor signaling: Implications for reinforcement and reinstatement. *Pharmacology & Therapeutics* **106**, 389-403 (2005). https://doi.org/https://doi.org/10.1016/j.pharmthera.2004.12.004
66 Browne, C. J. *et al.* Cocaine and morphine induce shared and divergent transcriptional regulation in nucleus accumbens D1 and D2 medium spiny neurons. *Mol Psychiatry* (2025). https://doi.org/10.1038/s41380-025-03004-1
67 Martínez-Rivera, F. J. *et al.* Transcriptional characterization of cocaine withdrawal versus extinction within nucleus accumbens in male rats. *Nat Commun* **16**, 2886 (2025). https://doi.org/10.1038/s41467-025-58151-4
68 Thibeault, K. C., Kutlu, M. G., Sanders, C. & Calipari, E. S. Cell-type and projection-specific dopaminergic encoding of aversive stimuli in addiction. *Brain Res* **1713**, 1-15 (2019). https://doi.org/10.1016/j.brainres.2018.12.024

69 Siciliano, C. A., Calipari, E. S., Ferris, M. J. & Jones, S. R. Adaptations of presynaptic dopamine terminals induced by psychostimulant self-administration. *ACS Chem Neurosci* **6**, 27-36 (2015). https://doi.org/10.1021/cn5002705
70 Siciliano, C. A. & Jones, S. R. Cocaine Potency at the Dopamine Transporter Tracks Discrete Motivational States During Cocaine Self-Administration. *Neuropsychopharmacology* **42**, 1893-1904 (2017). https://doi.org/10.1038/npp.2017.24
71 Serra, M. *et al.* Involvement of the Protein Ras Homolog Enriched in the Striatum, Rhes, in Dopaminergic Neurons' Degeneration: Link to Parkinson's Disease. *Int J Mol Sci* **22** (2021). https://doi.org/10.3390/ijms22105326
72 Vitucci, D. *et al.* Rasd2 Modulates Prefronto-Striatal Phenotypes in Humans and 'Schizophrenia-Like Behaviors' in Mice. *Neuropsychopharmacology* **41**, 916-927 (2016). https://doi.org/10.1038/npp.2015.228
73 Beaulieu, J.-M. & Gainetdinov, R. R. The Physiology, Signaling, and Pharmacology of Dopamine Receptors. *Pharmacological reviews* **63**, 182-217 (2011). https://doi.org/10.1124/pr.110.002642
74 Self, D. W. *et al.* Involvement of cAMP-Dependent Protein Kinase in the Nucleus Accumbens in Cocaine Self-Administration and Relapse of Cocaine-Seeking Behavior. *The Journal of Neuroscience* **18**, 1848-1859 (1998). https://doi.org/10.1523/jneurosci.18-05-01848.1998
75 Zhao, N., Wang, H. Y. & Dow-Edwards, D. Cocaine exposure during the early postnatal period diminishes medial frontal cortex Gs coupling to dopamine D1-like receptors in adult rat. *Neurosci Lett* **438**, 159-162 (2008). https://doi.org/10.1016/j.neulet.2008.04.014
76 Yano, H. *et al.* Gs- versus Golf-dependent functional selectivity mediated by the dopamine D1 receptor. *Nature Communications* **9**, 486 (2018). https://doi.org/10.1038/s41467-017-02606-w
77 Beck, M. *et al.* Identification, Molecular Cloning, and Characterization of a Novel GABAA Receptor-associated Protein, GRIF-1*. *Journal of Biological Chemistry* **277**, 30079-30090 (2002). https://doi.org/https://doi.org/10.1074/jbc.M200438200
78 Brickley, K., Smith, M. J., Beck, M. & Stephenson, F. A. GRIF-1 and OIP106, members of a novel gene family of coiled-coil domain proteins: association in vivo and in vitro with kinesin. *J Biol Chem* **280**, 14723-14732 (2005). https://doi.org/10.1074/jbc.M409095200
79 Nutt, P. D. J., Nestor, D. L. J., Nutt, D. & Nestor, L. in *Addiction* 0 (Oxford University Press, 2013).
80 Deak, J. D. & Johnson, E. C. Genetics of substance use disorders: a review. *Psychol Med* **51**, 2189-2200 (2021). https://doi.org/10.1017/s0033291721000969
81 Hatoum, A. S. *et al.* Multivariate genome-wide association meta-analysis of over 1 million subjects identifies loci underlying multiple substance use disorders. *Nat Ment Health* **1**, 210-223 (2023). https://doi.org/10.1038/s44220-023-00034-y
82 Poore, H. E. *et al.* A multivariate approach to understanding the genetic overlap between externalizing phenotypes and substance use disorders. *Addict Biol* **28**, e13319 (2023). https://doi.org/10.1111/adb.13319

83 Zhou, J. L. *et al*. Single-nucleus genomics in outbred rats with divergent cocaine addiction-like behaviors reveals changes in amygdala GABAergic inhibition. *Nat Neurosci* **26**, 1868-1879 (2023). https://doi.org/10.1038/s41593-023-01452-y