Widespread genetic effects and sex differences play a crucial role in addiction.

Authors:

Michael C. Saul[1,2], Jared R. Bagley[2,3], Lauren S. Bailey[2,3], Udita Datta[1,2], Price E. Dickson[1,2], Rainy Dodd[1,2], Leona H. Gagnon[1,2], Violet M. Kimble[4], Michael Leonardo[1,2], Sam-Moon Kim[2,5], Ashley Olson[1,2], Tyler Roy[1,2], Sarah A. Schoenrock[2,6], Troy Wilcox[1,2], J. David Jentsch[2,3], Ryan W. Logan[2,5], Colleen A. McClung[2,5], Vivek M. Philip[1,2], Laura G. Reinholdt[1,2], Stacey J. Sukoff Rizzo[1,2,7], Lisa M. Tarantino[2,6,8], and Elissa J. Chesler[1,2,*].

Affiliations: [1]The Jackson Laboratory of Mammalian Genetics, Bar Harbor, ME USA; [2]Center for Systems Neurogenetics of Addiction at The Jackson Laboratory, Bar Harbor, ME USA; [3]Binghamton University Department of Psychology, Binghamton, NY USA; [4]Drew University Department of Neuroscience, Madison, NJ USA; [5]University of Pittsburgh School of Medicine Department of Psychiatry, Pittsburgh, PA USA; [6]University of North Carolina-Chapel Hill Department of Genetics, Chapel Hill, NC USA; [7]University of Pittsburgh School of Medicine Department of Neurobiology, Pittsburgh, PA USA; [8]University of North Carolina-Chapel Hill Eshelman School of Pharmacy Division of Pharmacotherapy and Experimental Therapeutics, Chapel Hill, NC USA.

[*]To whom correspondence should be addressed: elissa.chesler@jax.org

Keywords: addiction, sex differences, genetics, cocaine, genomics.

Abstract

Though risk for cocaine use disorder, like most complex disease, is subject to considerable inter-individual variation, the sources of that variation – including genetics and sex – are frequently ignored in non-human animal studies. Here, we studied both males and females of eight different inbred mouse strains whose reproducible genomes capture 90% of the genetic diversity mice. In this population, individual differences explain a substantial proportion of variance in important cocaine-related behavioral, physiological, and striatum transcriptional responses traits. Individual differences thus represent a crucial source of biological information about addiction mechanisms missing in typical studies.

Introduction

Addictions are a highly prevalent complex disease, leading to a public health crisis associated with substantial morbidity and mortality. Illicit substance use disorders afflict 1 in 14 young adults in the United States[1] and drug overdoses are now the leading cause of accidental death among American adults under 55[2]. Genetic variation and sex differences are both known to influence addiction vulnerability; cocaine use disorder is highly heritable ($H^2 = 0.71$)[3] and substance use behaviors show sex differences in both humans and other animals[4,5]. Consequently, the neurobiology underlying addiction cannot be understood completely without consideration of genetic background and sex.

However, like many complex diseases, behavioral and genomic studies of addiction-related phenotypes often utilize only males of a single inbred mouse strain[6] or outbred rat populations confounded by vendor[7]. While these experimental methodologies certainly produce valuable biological insights, ignoring sex differences and genetic variation limits their generalizability and does not reduce experimental noise[8,9]. Inclusion of genetic variability and both males and females in rodent studies has the power to identify addiction-relevant targets, delineate variants' effects on co-regulation and co-expression networks, and define druggable network nodes. Rodent genetic variation does not capture precise human variants, but it can be exploited to determine underlying mechanisms in addiction-relevant processes. As a further benefit, genotypic and phenotypic precision allows genetics studies in rodents to be performed at orders of magnitude lower cost than human GWAS.

To assess genetics, sex, and their interaction on cocaine-related phenotypes in mice, we undertook a large-scale comprehensive evaluation of behavioral, physiological, and brain transcriptomic measures in both male and female mice from the eight inbred founder strains of the Diversity Outbred (DO) mouse heterogenous stock[10]. The genomes represented by these strains capture approximately 90% of the genetic diversity in *Mus musculus*[11].

Results

In the eight founder strains for the DO, we surveyed behavioral and physiological correlates of future cocaine use – including multiple novelty response behaviors, circadian rhythm

38 phenotypes observed in cells, and reversal learning as a measure of reward learning and
39 impulsivity. We directly measured cocaine-related behaviors such as initial locomotor
40 sensitivity to cocaine and intravenous self-administration (IVSA) of cocaine. Finally, we
41 measured the striatum transcriptome response to cocaine in all eight founder strains with
42 RNAseq. Our work represents the first time that many of these behavioral, physiological, and
43 molecular traits have been studied using methods powered to detect sex and genotype effects
44 and the extent to which these interact by moderating one another.

45 Response to novelty predicts psychostimulant addiction-related phenotypes in both humans[12]
46 and mice[13]. We first assessed differences in behavioral traits related to exploration and

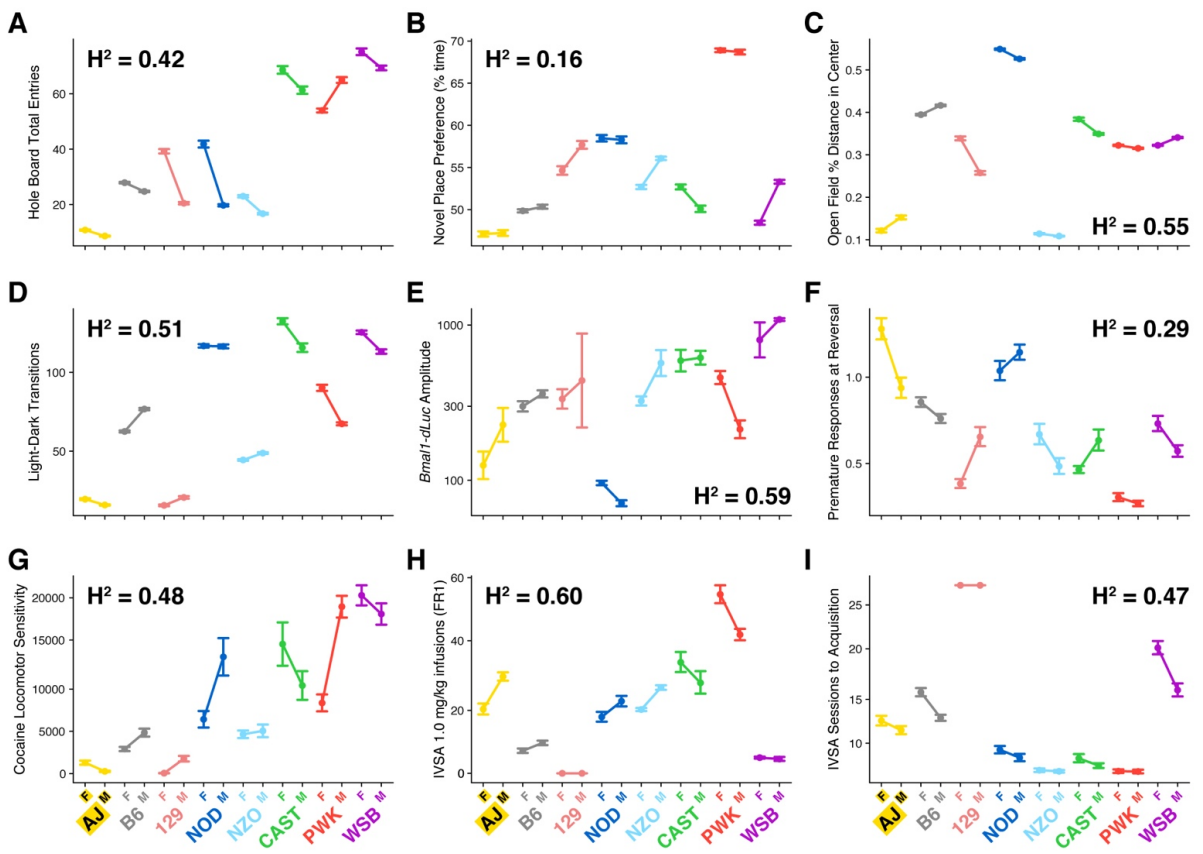

Figure 1: Heritable differences in cocaine-related behavioral and physiological traits. All plots show the mean ± the standard error in both sexes with females on the left. **A)** Hole board total entries, **B)** Novel place preference (% time in novel zone), **C)** Open field (% time in center of open field), **D)** Transitions between light and dark in the light-dark box, **E)** Circadian rhythm fibroblast *Bmal1-dLuc* luminescence amplitude, **F)** Reversal learning premature responses to the hole rewarded during acquisition, **G)** Initial locomotor sensitivity to cocaine (cm moved on Day 3 – Day 2), **H)** Cocaine intravenous self-administration number of infusions at 1.0 mg/kg FR1, and **I)** Sessions to acquisition of cocaine intravenous self-administration.

47 response to novelty. These behavioral risk measures were moderately to strongly heritable
48 among these strains (H^2 = 0.16-0.55, **Figure 1A-D**, **Supplementary Table S1**). Further, for the
49 phenotype of total nose pokes in the hole board, there was a significant sex difference ($F_{1,597}$ =
50 10.69, p = 0.0014). We detected strain-by-sex interactions in the total number of hole board
51 nose pokes ($F_{7,597}$ = 2.74, p = 0.0083) and in the proportion of distance traveled in the center of
52 the open field ($F_{7,684}$ = 2.09, p = 0.043), indicating that sex differences in these novelty response
53 traits are moderated by genetic background. Total nose pokes in the hole board (**Figure 1A**) and
54 novel place preference (**Figure 1B**) showed the largest differences in wild-derived strains,
55 demonstrating the importance of their expanded phenotypic range[14]. Transitions between the
56 sides of the light-dark box also showed extreme phenotypes in wild-derived mouse strains
57 (**Figure 1D**).

58 Circadian rhythm and reward-related behavioral phenotypes are co-inherited and rhythm
59 disruptions are linked to development and progression of substance use disorders[15]. The
60 molecular clock system directly influences expression of dopamine receptors in the striatum[16].
61 A cell-based assay on fibroblasts derived from each of the founder strains in which a *Bmal1-*
62 *dLuc* reporter was utilized for circadian measurement of luciferase bioluminescence to assess
63 differences in circadian rhythmicity[17,18]. We found very high heritability of the amplitude of
64 these rhythmic patterns (H^2 = 0.59, **Figure 1E**), but no significant sex differences or strain x sex
65 interactions. These results suggest that genetic differences in the molecular clock are one
66 potential mechanism for individual differences in addiction-related phenotypes.

67 Reversal learning tasks evaluate impulsive and compulsive behaviors that predict addiction
68 liability[19]. Within reversal learning paradigms, one measure of impulsivity is the number of
69 premature responses produced during the reversal phase of the task. Premature reversal
70 responses in these showed moderate heritability across this panel of mice (H^2 = 0.29, **Figure**
71 **1F**), but no significant sex differences or strain x sex interactions. We observed the lowest rate
72 of premature responding in the wild-derived PWK/PhJ strain.

73 We next sought to assess individual differences in initial locomotor response to cocaine, a
74 behavioral phenotype that predicts subsequent drug use in humans[20]. In these diverse mouse
75 strains, initial cocaine sensitivity showed strong heritability (H^2 = 0.48 for initial sensitivity,

76 **Figure 1G**). Wild-derived strains, particularly WSB/EiJ and PWK/PhJ, exhibited the highest initial
77 sensitivity to cocaine.

78 Operant drug self-administration procedures directly quantify reinforced responding for drug[13].
79 Heritability of cocaine IVSA traits in the founders of the DO was very strong (H^2 = 0.47 for
80 sessions to acquisition of IVSA; H^2 = 0.60 for total infusions at FR-1 1.0 mg/kg, **Figure 1H-I**). The
81 high heritability of infusions earned is quite similar to the observed heritability of human
82 cocaine use disorder[3]. Of note, 129S1/SvImJ mice do not acquire IVSA and do not take any
83 cocaine during the acquisition phase. The largest phenotypic range manifests in wild-derived
84 mice; PWK/PhJ mice acquire IVSA behavior very quickly and self-administer the most cocaine –
85 about four times as much as C57BL/6J – while WSB/EiJ mice self-administer the least amount of
86 cocaine relative to the other strains that do acquire IVSA behavior. The broad phenotypic range
87 of volitional cocaine-taking behaviors in these genetically diverse mouse strains represents a
88 clear opportunity to study mechanisms underlying variation in the physiology of addiction.

89 Because addiction-related phenotypes are highly heritable and sometimes exhibit sex
90 differences in a manner dependent on genetics, we next sought to assess heritable differences
91 in the molecular response to cocaine. Male and female mice of the eight DO founder strains
92 were given multiple injections of cocaine or saline over 19 days. Samples for bulk RNAseq were
93 collected from striatum at least 24 hours after the final injection of repeated administration of
94 either cocaine or vehicle (sham). First, we documented that in the absence of cocaine, most
95 transcripts show moderate to high baseline heritability across the founder strains (median H^2 =
96 0.29) and that subspecies of origin explains the greatest amount of variation in expression
97 (**Figure 2A**). Using a linear modeling approach[21,22], we found few expression differences
98 attributable solely to the effect of cocaine treatment. Instead, significant effects of cocaine
99 arose in interaction with individual differences such as genetic background. The greatest
100 quantity of differentially expressed transcripts was observed in the strain-by-sex-by-drug
101 treatment three-way interaction (1,282 genes at $q < 0.01$, **Figure 2B, Supplementary Table S2**).
102 Some strains responded more strongly than others; 129S1/SvImJ and PWK/PhJ had many
103 expressed genes influenced by cocaine and sex-by-cocaine interactions (PWK/PhJ: 89 genes at q
104 < 0.01, **Figure 2C**) while the commonly used C57BL/6J strain had very few (4 genes at q < 0.01,

Figure 2C). Because statistical power was approximately equal for all strains involved in this study, these differences likely reflect real individual differences in the brain's sex-specific responses to cocaine.

Many significant effects of cocaine on gene expression would have been missed in had only single strain been analyzed. For instance, we identified differential expression of the serotonin receptor 1D gene *Htr1d* in the strain-by-sex-by-cocaine interaction factor. This gene is upregulated in males of two strains but downregulated in males of another (**Figure 2D**).

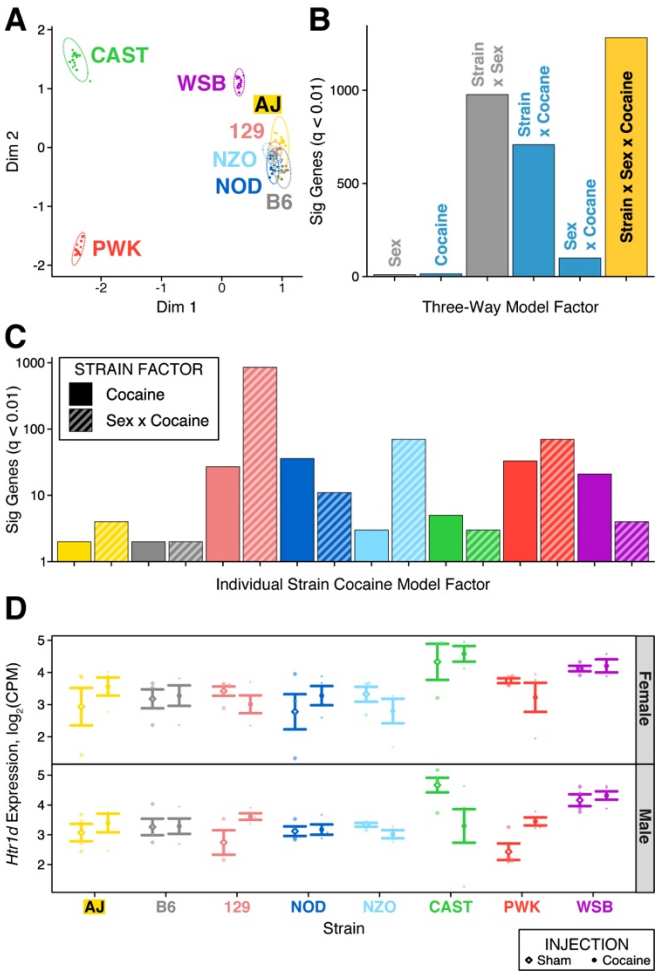

Figure 2: Individual differences in striatum cocaine response transcriptomes. **A)** High heritability of baseline transcriptome partitions the subspecies in a multidimensional scaling plot, **B)** In a three-way analysis of strain, sex, and cocaine, many genes showed significant expression changes in response to cocaine in interaction factors; the strain x sex x cocaine interaction factor identified 1,282 DEGs. In contrast, there were very few significant DEGs in the main effect of cocaine. **C)** Two-way models show the magnitude of transcriptional responses to cocaine differs by strain. B6 has few differentially expressed genes while 129 and PWK have many. The y-axis is $\log_{10}$. **D)** The serotonin 1D receptor gene *Htr1d* is differentially expressed in males of 129S1/SvImJ, PWK/PhJ, and CAST/EiJ strains. CAST/EiJ males show the opposite directionality of differential expression to PWK/PhJ and 129S1/SvImJ males.

Discussion

Some of these heritable transcriptome differences we observed recapitulate previous work showing heritable differences in pharmacokinetics of cocaine[23] – for example, mice closely related to the high cocaine taking PWK/PhJ strain are known to have some of the highest brain concentrations of cocaine shortly after injection. These pharmacokinetic differences do not account for the strong differences between strains such as 129S1/SvImJ and A/J, whose brain cocaine pharmacokinetic profiles are very

133 similar [23]. Further research on highly diverse mice may resolve the mechanisms driving
134 individual differences between strains with closely matched pharmacokinetic profiles.

135 Genetic variation is a valuable resource for the discovery of biological mechanisms of
136 addiction[24,25]. Similar to humans, individual differences among mice greatly influence
137 behavioral, physiological, and transcriptomic cocaine-related traits. For many of these traits,
138 individual differences explain a substantial proportion of the variation. Individual variation in
139 addiction related traits is a largely untapped resource that would have otherwise remained
140 unknown had we only examined a single inbred strain. Because the vast majority of the
141 biomedical literature is built on limited genetic diversity, such individual differences represent a
142 crucial opportunity to find new, clinically relevant genes and mechanisms.

143 Methods

144 Standard Operating Procedures

145 All methodologies used in this work are documented in depth in the Center for Systems
146 Neurogenetics of Addiction's Standard Operating Procedures (SOPs)
147 (https://www.jax.org/research-and-faculty/research-centers/systems-neurogenetics/data-
148 resources). The brief methods appearing below summarize these SOPs and reference specific
149 SOPs.

150 Animals

151 Mice from the following strains were surveyed in these experiments: A/J (JAX stock #000646),
152 C57BL/6J (JAX stock #000664), 129S1/SvImJ (JAX stock #002448), NOD/ShiLtJ (JAX stock
153 #001976), NZO/HlLtJ (JAX stock #002105), CAST/EiJ (JAX stock #000928), PWK/PhJ (JAX stock
154 #003715), and WSB/EiJ (JAX stock #001145). These strains are the founders of the DO
155 heterogeneous stock and Collaborative Cross recombinant inbred strains. Surveys of these eight
156 strains can demonstrate statistical heritability patterns that justify further dissection using the
157 derived resources[10]. The mice used in the Research Animal Facility at The Jackson Laboratory
158 came from breeding colonies maintained in the Research Animal Facility. These colonies were
159 derived from production colonies at The Jackson Laboratory and breeders were replaced with
160 animals from The Jackson Laboratory's production colony at least every five generations. The

161 mice used in the Jentsch Lab at Binghamton University were shipped to the Jentsch Lab from
162 either the Chesler Lab colonies or from production colonies in The Jackson Laboratory. The
163 studies described utilized a total of 1,085 mice (**Supplementary Table S3**).

164 ## Mouse Husbandry and Housing

165 All procedures were approved by the Jackson Laboratory of Mammalian Genetics institutional
166 animal care and use committee. Mice were housed according to the CSNA animal housing SOP
167 (https://www.jax.org/-/media/jaxweb/files/research-and-faculty/tools-and-resources/system-
168 neurogenetics/csna-animal-housing.pdf).

169 ## Cocaine

170 Cocaine hydrochloride was provided by the National Institute on Drug Abuse Drug Supply
171 Program Division of Therapeutics and Medical Consequences (catalog number: 9041-001).
172 Cocaine was stored in powder form until it was formulated into 0.9% Saline (100 mg/mL clear
173 solution) in various concentrations specific to each experiment according to their individual
174 SOPs.

175 ## Novelty Response Behavioral Phenotypes

176 Open field, light-dark box, hole board, and novel place preference behavioral paradigms were
177 conducted in order with one test per day on consecutive days during the light phase of the
178 light:dark cycle. Open field data were collected for 60 minutes according to the SOP
179 (https://www.jax.org/-/media/jaxweb/files/research-and-faculty/tools-and-resources/system-
180 neurogenetics/open-field-
181 assay.pdf?la=en&hash=32DDAFF2B17B2D4961C136C5616C4982AC23EC3B). Light-dark data
182 were collected for 20 minutes with the mouse starting in the light side of the chamber facing
183 the dark side according to the SOP (https://www.jax.org/-/media/jaxweb/files/research-and-
184 faculty/tools-and-resources/system-neurogenetics/light-dark-
185 assay.pdf?la=en&hash=A63CF8D22EB7936CF6C69A3178373981F4016675). Hole board data
186 were collected for 20 minutes according to the SOP (https://www.jax.org/-
187 /media/jaxweb/files/research-and-faculty/tools-and-resources/system-neurogenetics/hole-
188 board-assay.pdf?la=en&hash=EC343A797D37209CF64D34E6031608A511D8E15D). Novel place

189 preference included a five minute acclimation period to a center chamber, a 10 minute
190 exposure period to a randomized exposure side, a five minute acclimation period, then testing
191 for 20 minutes for preference for novel or exposure side of the novel place preference
192 apparatus according to the SOP (https://www.jax.org/-/media/jaxweb/files/research-and-
193 faculty/tools-and-resources/system-neurogenetics/novelty-place-preference-
194 assay.pdf?la=en&hash=B5D2D0FC9028B408E84729C0C8832C580AB8E039). All mice from the
195 Center for Systems Neurogenetics of Addiction were tested through this novelty pipeline prior
196 to any other test and were then randomized and assigned into either reversal learning, cocaine
197 locomotor sensitization, or cocaine intravenous self-administration. The novelty study
198 produced observations from a total of 783 mice.

199 ### *Bmal1-dLuc* Circadian Rhythm Data

200 Data for circadian rhythm were measured in primary fibroblast cultures generated from skin
201 biopsies in the founders. To isolate fibroblasts, ear biopsies (1 mm in diameter) were digested
202 in Dulbecco's Modified Eagle's Medium (DMEM, HyClone) containing 2.5 mg/ml collagenase D
203 (Gibco) and 1.25 mg/ml pronase (Millipore) for 90 mins and then plated in DMEM growth
204 media containing 10% Fetal Bovine Serum (FBS, HyClone), 292 μg/ml L-glutamine (HyClone),
205 100 units/ml penicillin (Hyclone) and 100 μg/ml streptomycine (HyClone). *Bmal1-dLuc* reporter
206 was delivered to fibroblasts by lentiviral-mediated gene delivery (VectorBuilder). Following
207 synchronization of rhythms by 15 μM forskolin (Sigma) for 2 hrs, the temporal patterns of
208 *Bmal1-dLuc* bioluminescence was recorded for ~70 secs at intervals of 10 mins over 6-7 days
209 from fibroblast cultures in DMEM recording media containing 15 μM forskolin, 25 mM HEPES
210 (Gibco), 292 μg/ml L-glutamine, 100 units/ml penicillin, 100 μg/ml streptomycine, and 10μM
211 luciferin (Promega) by an automated 32-channel luminometer (Lumicycle, ActiMetrics) in a
212 standard tissue culture incubator at 32°C. The amplitude of bioluminescence rhythms was
213 determined from baseline-subtracted data using the damped sine fit and Levenberg-Marquardt
214 algorithm[26]. The circadian study produced observations from a total of 56 mice.

215 ## Reversal Learning

216 Data for reversal learning were collected using the SOP (https://www.jax.org/-
217 /media/jaxweb/files/research-and-faculty/tools-and-resources/system-neurogenetics/reversal-
218 learning-assay.pdf?la=en&hash=8484E47B170462960E11C1FAEEE6FF3CE6FDFC08). The
219 reversal learning data produced observations from a total of 202 mice.

220 ## Initial Locomotor Sensitivity to Cocaine

221 Data for initial locomotor sensitivity were collected as described previously[27] using data from
222 days 1-3 in the SOP for locomotor behavioral sensitization to cocaine (https://www.jax.org/-
223 /media/jaxweb/files/research-and-faculty/tools-and-resources/system-neurogenetics/cocaine-
224 locomotor-sensitization-
225 assay.pdf?la=en&hash=9E5D4C248C3BCCAD947C164AE81663C13A77EB0D). Briefly, mice were
226 placed into the open field arena for 30 minutes, removed, and injected i.p. with either saline
227 (days 1-2) or 10 mg/kg cocaine (day 3) and returned to the open field arena for 60 minutes.
228 Distance moved after injection on day 3 minus day 2 was uses as a measure of initial locomotor
229 sensitivity to cocaine. The sensitization study produced observations from a total of 230 mice.

230 ## Cocaine Intravenous Self-Administration

231 Prior to cocaine intravenous self-administration, mice were implanted with a jugular catheter
232 and allowed 10 days for post-operative recovery. In an operant conditioning paradigm, mice
233 were allowed to acquire cocaine self-administration at 1.0 mg/kg, then evaluated for dose-
234 response effects at eight different doses. After a final stabilizing dose at 1.8 mg/kg, responses
235 during seven days of withdrawal were recorded. Finally, cued reinstatement was recorded for
236 two days. Self-administration in these eight mouse strains was performed according to v1.0 of
237 the CSNA's SOP (https://www.jax.org/-/media/jaxweb/files/research-and-faculty/tools-and-
238 resources/system-neurogenetics/intravenous-self-administration-ivsa-
239 paradigm.pdf?la=en&hash=FA64135F219C7DF65937A1CF9270301B0E771836). The
240 intravenous self-administration study produced observations from a total of 217 mice.

Data Deposit

Data for each phenotype will be deposited in the Mouse Phenome Database (MPD)[28] upon publication.

Heritability Calculations

For each trait, heritability was calculated from linear models using the isogenic strain as the independent categorical variable using the following equation:

$$h^2 = \frac{MS_{strain}}{MS_{strain} + (n_{mean} - 1) * MS_{resid}}$$

where MS_{strain} is the mean square of the strain effect, n_{mean} is the mean number of samples within each strain, and MS_{resid} is the mean square of the residuals. For the reversal learning data, an additional additive covariate of site was included in the model to account for inter-lab variation. This term was not utilized in the heritability calculation.

For some traits such as number of infusions at FR-1 1.0 mg/kg cocaine self-administration, a single strain such as 129S1/SvImJ showed little to no variation, which may upwardly bias heritability calculations. For these traits, heritability was calculated both with and without the low variance strain. The data reported in the paper rely upon the inclusive calculation, but results of both methods of calculation are reported for completeness (see **Supplementary Table S1**).

RNAseq

Striatum tissue was collected during the light stage of the light:dark cycle between 24 and 48 hours after the final injection in the cocaine behavioral sensitization protocol according to the SOP (https://www.jax.org/-/media/jaxweb/files/research-and-faculty/tools-and-resources/system-neurogenetics/post-sensitization-tissue-collection.pdf?la=en&hash=9E6CD8DEB39606B791A5D25F6CD0611EF14D96A7). Tissue was collected for both sexes of each founder strain exposed to either sham (saline) or 10 mg/kg IP cocaine.

266 RNA was isolated from striatum tissue using the MagMAX mirVana Total RNA Isolation Kit
267 (ThermoFisher) and the KingFisher Flex purification system (ThermoFisher). Tissues were lysed
268 and homogenized in TRIzol Reagent (ThermoFisher). After the addition of chloroform, the RNA-
269 containing aqueous layer was removed for RNA isolation according to the manufacturer's
270 protocol, beginning with the RNA bead binding step.

271 RNA concentration and quality were assessed using the Nanodrop 2000 spectrophotometer
272 (Thermo Scientific) and the RNA Total RNA Nano assay (Agilent Technologies). 2µl of diluted
273 1:1000 diluted ERCC Spike-in Control Mix 1 (Ambion by Life Technologies) was added to 100ng
274 of each RNA sample prior to library construction. Libraries were prepared by the Genome
275 Technologies core service at The Jackson Laboratory using the KAPA RNA Hyper Prep Kit with
276 RiboErase (HMR) (KAPA Biosystems), according to the manufacturer's instructions. Briefly, the
277 protocol entails depletion of ribosomal RNA (rRNA), RNA fragmentation, first and second strand
278 cDNA synthesis, ligation of Illumina-specific adapters containing a unique barcode sequence for
279 each library, magnetic bead size selection, and PCR amplification. Libraries were checked for
280 quality and concentration using the D5000 ScreenTape assay (Agilent Technologies) and
281 quantitative PCR (KAPA Biosystems), according to the manufacturers' instructions.

282 RNAseq libraries were pooled and sequenced by Novogene in 150 bp paired-end format on an
283 Illumina NovaSeq 6000 sequencer targeting 90 million read pairs per sample. Sequencing
284 achieved a median read depth of 132 million reads. The resultant reads were determined to be
285 of consistently high quality using fastqc v0.11.3 and MultiQC v1.2.

286 Reads were generated from raw data and demultiplexed using BCL2Fastq v2.18.0.12,
287 concatenated by sample, and aligned with the STAR aligner v2.6.1[29] to the GRCm38 mouse
288 reference genome with v94 of the Ensembl transcriptome. Transcript-level quantification was
289 estimated using RSEM v1.3.0[30] on a transcriptome BAM file produced as an output of this
290 alignment. The data were imported into R v3.5.1 and summarized to the gene level using
291 tximport v1.10.1[31], TMM-normalized using edgeR v3.24.3[21], and imported into limma v3.38.3[32]
292 using the $\log_2$-transformation function voom. We compared multivariate approaches modeling
293 with interaction factors between edgeR and voom+limma approaches and found that
294 voom+limma performs better than edgeR for controlling false negatives. Upon initial

295 examination of the findings, we identified intermittent contamination with choroid plexus,
296 which potentially derives from the ventricular aspect of the dorsal striatum. Correcting for this
297 contamination necessited an additive covariate for choroid plexus consisting of log-mean CPM
298 values of *Kl* and *Ttr* expression, unambiguous markers for choroid plexus[33]. These values were
299 $\log_2$ transformed for work in limma. For An overall model for all strains included this choroid
300 plexus factor as a nuisance variable plus the main effects of strain, sex, and cocaine injection
301 and all of their interactions. Individual models included the choroid plexus nuisance variable
302 plus sex, cocaine injection, and sex:cocaine injection interaction. Correction for local false
303 discovery rates utilized the qvalue package in R v2.14.1[34]. Because brain transcriptional changes
304 are subtle[35], all results reported are at $q < 0.01$ with no fold-change cutoff (**Supplementary**
305 **Table S2**).

306 Raw data and transcript-level expression estimates will be deposited in the Gene Expression
307 Omnibus [36] upon publication (accession number: GSEXXXXX).

308 Supplemental Material

309 **Supplementary Table S1**: Statistical test results for heritability and genotype-by-sex
310 interactions.

311 **Supplementary Table S2**: Differential expression results for all tests discussed.

312 **Supplementary Table S3**: Sample sizes for all strain and sex combinations for studies reported.

313 Raw behavioral data will be deposited in the Mouse Phenome Database upon publication.

314 Gene expression data will be deposited in the Gene Expression Omnibus upon publication.

315 All scripts, code, and metadata used for analysis are deposited in GitHub (repository:
316 github.com/msaul/csna_founders_survey_2020)

317 Author Contributions

318 EJC, PED, LMT, SAS, JDJ, RWL, CAM, LGR, VMP, SJSR, and CSNA conceived the studies. SJSR, PED,
319 JRB, LSB, SAS, RD, ML, AO, TR, TW, and LHG designed and implemented the behavioral
320 experiments. SMK designed and implemented the circadian transcriptional experiment. JRB,

321 LSB, UD, PED, ML, SMK, AO, TR, SAS, TW, LHG, VMP, and MCS analyzed the behavioral data.
322 SMK, MCS, and VMP analyzed the circadian transcriptional data. MCS, VMK, and VMP analyzed
323 the RNAseq data. MCS, UD, PED, JRB, LSB, SAS, TW, LHG, VMP, and EJC interpreted the
324 behavioral results. SMK, MCS, and VMP interpreted the results of the circadian transcriptional
325 experiment. MCS, VMK, VMP, and EJC interpreted the RNAseq results. MCS and EJC wrote the
326 manuscript.

327 Acknowledgements

328 We gratefully acknowledge Technologies service at The Jackson Laboratory Genome for their
329 expert assistance on the RNAseq experiment, Surgical Services at The Jackson Laboratory for
330 their work on the intravenous self-administration paradigm, and the NIDA Drug Supply for
331 providing cocaine. We would further like to thank Stephen Krasinski for critical reading and
332 comments during manuscript preparation, Rob Williams for helpful discussion on the analysis,
333 and C. Herbert Pratt for scientific program management. This work was funded by NIH P50
334 DA039841 (Center for Systems Neurogenetics of Addiction) to EJC, LMT, JDJ, RWL CAM, VMP,
335 LGR, and SJSR as well as NIH R01 DA037927 to EJC and K99 DA043573 to PED.

336 Literature Cited

337 1. SAMHSA. *The 2017 National Survey on Drug Use and Health (NSDUH)*.
338 https://www.samhsa.gov/data/nsduh/reports-detailed-tables-2017-NSDUH (2017).

339 2. Kochanek, K. D., Murphy, S. L., Xu, J. & Arias, E. Deaths: Final Data for 2017. *Natl. Vital*
340 *Stat. Rep.* **68**, 1–77.

341 3. Goldman, D., Oroszi, G. & Ducci, F. The genetics of addictions: uncovering the genes. *Nat.*
342 *Rev. Genet.* **6**, 521–532 (2005).

343 4. Becker, J. B. & Chartoff, E. Sex differences in neural mechanisms mediating reward and
344 addiction. *Neuropsychopharmacology* **44**, 166–183 (2019).

345 5. Becker, J. B., Perry, A. N. & Westenbroek, C. Sex differences in the neural mechanisms
346 mediating addiction: a new synthesis and hypothesis. *Biol. Sex Differ.* **3**, 14 (2012).

347 6. Walker, D. M. *et al.* Cocaine self-administration alters transcriptome-wide responses in the
348 brain's reward circuitry. *Biol. Psychiatry* (2018).

349 7. Fitzpatrick, C. J. *et al.* Variation in the Form of Pavlovian Conditioned Approach Behavior
350 among Outbred Male Sprague-Dawley Rats from Different Vendors and Colonies: Sign-
351 Tracking vs. Goal-Tracking. *PLoS ONE* **8**, e75042 (2013).

352 8. Prendergast, B. J., Onishi, K. G. & Zucker, I. Female mice liberated for inclusion in
353 neuroscience and biomedical research. *Neurosci. Biobehav. Rev.* **40**, 1–5 (2014).

354 9. Tuttle, A. H., Philip, V. M., Chesler, E. J. & Mogil, J. S. Comparing phenotypic variation
355 between inbred and outbred mice. *Nat. Methods* **15**, 994–996 (2018).

356 10. Saul, M. C., Philip, V. M., Reinholdt, L. G. & Chesler, E. J. High-Diversity Mouse
357 Populations for Complex Traits. *Trends Genet.* **0**, (2019).

358 11. Roberts, A., De Villena, F. P.-M., Wang, W., McMillan, L. & Threadgill, D. W. The
359 polymorphism architecture of mouse genetic resources elucidated using genome-wide
360 resequencing data: implications for QTL discovery and systems genetics. *Mamm. Genome*
361 **18**, 473–481 (2007).

362 12. Ersche, K. D., Turton, A. J., Pradhan, S., Bullmore, E. T. & Robbins, T. W. Drug addiction
363 endophenotypes: impulsive versus sensation-seeking personality traits. *Biol. Psychiatry* **68**,
364 770–773 (2010).

365 13. Dickson, P. E. *et al.* Association of novelty-related behaviors and intravenous cocaine self-
366 administration in Diversity Outbred mice. *Psychopharmacology (Berl.)* **232**, 1011–1024
367 (2015).

368 14. Wahlsten, D., Metten, P. & Crabbe, J. A rating scale for wildness and ease of handling
369 laboratory mice: results for 21 inbred strains tested in two laboratories. *Genes Brain Behav.*
370 **2**, 71–79 (2003).

371 15. Logan, R. W., Williams III, W. P. & McClung, C. A. Circadian rhythms and addiction:
372 Mechanistic insights and future directions. *Behav. Neurosci.* **128**, 387–412 (2014).

373 16. Ozburn, A. R. *et al.* Direct Regulation of Diurnal Drd3 Expression and Cocaine Reward by
374 NPAS2. *Biol. Psychiatry* **77**, 425–433 (2015).

375 17. Ramanathan, C. *et al.* Cell Type-Specific Functions of Period Genes Revealed by Novel
376 Adipocyte and Hepatocyte Circadian Clock Models. *PLoS Genet.* **10**, e1004244 (2014).

377 18. Kim, S.-M., Neuendorff, N., Chapkin, R. S. & Earnest, D. J. Role of Inflammatory Signaling
378 in the Differential Effects of Saturated and Poly-unsaturated Fatty Acids on Peripheral
379 Circadian Clocks. *EBioMedicine* **7**, 100–111 (2016).

380 19. Izquierdo, A. & Jentsch, J. D. Reversal learning as a measure of impulsive and compulsive
381 behavior in addictions. *Psychopharmacology (Berl.)* **219**, 607–620 (2012).

382 20. de Wit, H. & Phillips, T. J. Do initial responses to drugs predict future use or abuse?
383 *Neurosci. Biobehav. Rev.* **36**, 1565–1576 (2012).

384 21. Chen, Y., Lun, A. T. L. & Smyth, G. K. Differential Expression Analysis of Complex RNA-
385 seq Experiments Using edgeR. in *Statistical Analysis of Next Generation Sequencing Data*

386 (eds. Datta, S. & Nettleton, D.) 51–74 (Springer International Publishing, 2014).
387 doi:10.1007/978-3-319-07212-8_3.

388 22. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. Robust
389 Hyperparameter Estimation Protects Against Hypervariable Genes and Improves Power to
390 Detect Differential Expression. *Ann. Appl. Stat.* **10**, 946–963 (2016).

391 23. Wiltshire, T. *et al.* Initial locomotor sensitivity to cocaine varies widely among inbred mouse
392 strains: Initial locomotor sensitivity to cocaine in inbred mice. *Genes Brain Behav.* **14**, 271–
393 280 (2015).

394 24. Kumar, V. *et al.* C57BL/6N mutation in cytoplasmic FMRP interacting protein 2 regulates
395 cocaine response. *Science* **342**, 1508–1512 (2013).

396 25. Ruan, Q. T. *et al.* A Mutation in Hnrnph1 That Decreases Methamphetamine-Induced
397 Reinforcement, Reward, and Dopamine Release and Increases Synaptosomal hnRNP H and
398 Mitochondrial Proteins. *J. Neurosci.* **40**, 107–130 (2020).

399 26. Izumo, M., Johnson, C. H. & Yamazaki, S. Circadian gene expression in mammalian
400 fibroblasts revealed by real-time luminescence reporting: Temperature compensation and
401 damping. *Proc. Natl. Acad. Sci.* **100**, 16089–16094 (2003).

402 27. Schoenrock, S. A. *et al.* Characterization of genetically complex Collaborative Cross mouse
403 strains that model divergent locomotor activating and reinforcing properties of cocaine.
404 *Psychopharmacology (Berl.)* (2020) doi:10.1007/s00213-019-05429-3.

405 28. Bogue, M. A. *et al.* Mouse Phenome Database: a data repository and analysis suite for
406 curated primary mouse phenotype data. *Nucleic Acids Res.* gkz1032 (2019)
407 doi:10.1093/nar/gkz1032.

408 29. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
409 (2013).

410 30. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or
411 without a reference genome. 16 (2011).

412 31. Soneson, C., Love, M. & Robinson, M. Differential analyses for RNA-seq: transcript-level
413 estimates improve gene-level inferences. *F1000Research* **4**, (2016).

414 32. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and
415 microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

416 33. Sathyanesan, M. *et al.* A molecular characterization of the choroid plexus and stress-induced
417 gene regulation. *Transl. Psychiatry* **2**, e139 (2012).

418 34. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl.*
419 *Acad. Sci.* **100**, 9440–9445 (2003).

420 35. Hitzemann, R. *et al.* Introduction to Sequencing the Brain Transcriptome. in *International*
421 *Review of Neurobiology* vol. 116 1–19 (Elsevier, 2014).

422 36. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic*
423 *Acids Res.* **41**, D991–D995 (2012).

424