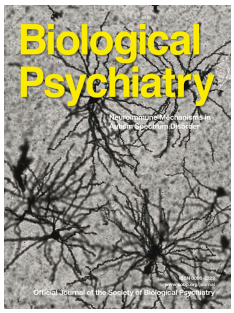
Accepted Manuscript



Cocaine self-administration alters transcriptome-wide responses in the brain's reward circuitry

Deena M. Walker, Hannah M. Cates, Yong-Hwee E. Loh, Immanuel Purushothaman, Aarthi Ramakrishnan, Kelly M. Cahill, Casey K. Lardner, Arthur Godino, Hope G. Kronman, Jacqui Rabkin, Zachary S. Lorsch, Philipp Mews, Marie A. Doyle, Jian Feng, Benoit Labonté, Ja Wook Koo, Rosemary C. Bagot, Ryan W. Logan, Marianne L. Seney, Erin S. Calipari, Li Shen, Eric J. Nestler

PII: S0006-3223(18)31447-1

DOI: 10.1016/j.biopsych.2018.04.009

Reference: BPS 13520

To appear in: *Biological Psychiatry*

Received Date: 16 October 2017

Revised Date: 15 March 2018

Accepted Date: 17 April 2018

Please cite this article as: Walker D.M., Cates H.M., Loh Y.-H.E., Purushothaman I., Ramakrishnan A., Cahill K.M., Lardner C.K., Godino A., Kronman H.G., Rabkin J., Lorsch Z.S., Mews P., Doyle M.A., Feng J., Labonté B., Koo J.W., Bagot R.C., Logan R.W., Seney M.L., Calipari E.S., Shen L. & Nestler E.J., Cocaine self-administration alters transcriptome-wide responses in the brain's reward circuitry, *Biological Psychiatry* (2018), doi: 10.1016/j.biopsych.2018.04.009.

Cocaine self-administration alters transcriptome-wide responses in the brain's reward circuitry

Short Title: Cocaine Self-Administration Primes the Transcriptome

Deena M. Walker[1,*], Hannah M. Cates[1,*], Yong-Hwee E. Loh[1,*], Immanuel Purushothaman[1], Aarthi Ramakrishnan[1], Kelly M. Cahill[5,] Casey K. Lardner[1], Arthur Godino[1], Hope G. Kronman[1], Jacqui Rabkin[1], Zachary S. Lorsch[1], Philipp Mews[1], Marie A. Doyle[1], Jian Feng[1,‡], Benoit Labonté[1,‡], Ja Wook Koo[1,‡], Rosemary C. Bagot[1,‡], Ryan W. Logan[2,3,4], Marianne L. Seney[2,3], Erin S. Calipari[6,†¶], Li Shen[1,†] and Eric J. Nestler[1,†,¶]

[1]Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA
[2]Department of Psychiatry, University of Pittsburgh Medical School, Pittsburgh, PA 15213, USA
[3]Translational Neuroscience Program, University of Pittsburgh Medical School, Pittsburgh, PA 15213, USA
[4]Center for Systems Neurogenetics of Addiction, The Jackson Laboratory, Bar Harbor, ME 14609, USA
[5]Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA
[6] Department of Pharmacology, Vanderbilt Center for Addiction Research, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

[¶]To whom Correspondence should be addressed:
Eric J. Nestler, Fishberg Department of Neuroscience and Friedman Brain Institute, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, Box 1065, New York, New York 10029, USA. Phone: (212) 659-5996; email: eric.nestler@mssm.edu

Erin S. Calipari, Department of Pharmacology, Vanderbilt Center for Addiction Research, Vanderbilt Brain Institute, Vanderbilt University School of Medicine, 865F Light Hall, 2215 Garland Avenue, Nashville, TN 37232, USA. Phone: 615-343-5792. email: erin.calipari@vanderbilt.edu

[‡]Present Addresses: JF, Department of Biological Science, Program in Neuroscience, Florida State University, Tallahassee, FL; JWK, Department of Neural Development and Disease, Korea Brain Research Institute, Daegu, Republic of Korea; BL, Department of Psychiatry and Neurosciences, Faculty of Medicine, Laval University, Québec City, Québec, G1J2G3 Canada; RCB, Department of Psychology, McGill University, Montréal, Québec, H3A 1B1 Canada.

[†,*]Authors contributed equally to this project.

Abstract: 249 words (250 limit)

Article Body: 3998 (4,000 limit, excluding references)

Figures: 7; Tables: 0

Supplemental Information: Text, 8 Figures, 8 Tables

48 Keywords: RNA-sequencing, gene expression, nucleus accumbens, prefrontal cortex,

49 basolateral amygdala, dorsal striatum, ventral hippocampus

ABSTRACT

BACKGROUND: Global changes in gene expression underlying circuit and behavioral dysregulation associated with cocaine addiction remain incompletely understood. Here, we show how a history of cocaine self-administration (SA) “re-programs” transcriptome-wide responses throughout the brain’s reward circuitry at baseline and in response to context and/or cocaine re-exposure after prolonged withdrawal (WD).

METHODS: We assigned male mice to one of six groups: saline/cocaine SA + 24 hr WD; or saline/cocaine SA + 30 d WD + an acute saline/cocaine challenge within the previous drug-paired context. RNA-sequencing was conducted on six interconnected brain reward regions. Using pattern analysis of gene expression and factor analysis of behavior, we identified genes that are strongly associated with addiction-related behaviors and uniquely altered by a history of cocaine SA. We then identified potential upstream regulators of these genes.

RESULTS: We focused on three Patterns of gene expression that reflect responses to: a) acute cocaine, b) context re-exposure, and c) drug + context re-exposure. These Patterns revealed region-specific regulation of gene expression. Further analysis revealed that each of these gene expression Patterns correlated with an “Addiction Index”—a composite score of several addiction-like behaviors during cocaine SA—in a region-specific manner. CREB and nuclear receptor families were identified as key upstream regulators of genes associated with such behaviors.

CONCLUSIONS: This comprehensive picture of transcriptome-wide regulation in the brain’s reward circuitry by cocaine SA and prolonged WD provides new insight into the molecular basis of cocaine addiction, which will guide future studies of the key molecular pathways involved.

INTRODUCTION

Addiction arises from genetic and environmental factors, which determine individual responses to initial and repeated drug exposure at the molecular, cellular, and circuit levels (1). A key feature of addiction is the ability for drug or drug-associated cues to trigger relapse, even after periods of prolonged abstinence (2). It is hypothesized that susceptibility to relapse depends on long-term neuroadaptations within the brain's reward circuitry (3-5).

Behavioral responses to cocaine self-administration (SA) after withdrawal (WD) and re-exposure to drug or contextual cues are well characterized in rodent models. However, the underlying molecular mechanisms remain elusive. Most studies investigating transcriptional changes associated with long-term WD followed by cocaine/context re-exposure have focused on candidate genes within one or two brain regions. These studies have found that long-term WD from cocaine SA is associated with changes in growth factors and their signaling cascades (6-9), neurotransmitter and neuropeptide systems (10, 11), and immediate early genes (10, 12).

The few studies investigating transcriptome-wide changes after short-term WD from cocaine SA (13, 14), or long-term WD but without re-exposure (15), focused primarily on nucleus accumbens (NAc), ventral tegmental area (VTA) (14), or prefrontal cortex (PFC) (13, 15). No study has characterized transcriptome-wide changes across multiple interconnected brain reward regions. Furthermore, no transcriptomic study has compared multiple stages of WD plus drug/context re-exposure, while leveraging individual variability to identify genes transcriptome-wide associated with addiction-related behaviors.

Here, we performed RNA-sequencing on six reward-related brain regions in mice with a history of saline or cocaine SA. We profiled the transcriptome in these regions after short- and long-term WD with drug/context re-exposure. We hypothesized that a history of cocaine SA "re-programs" the transcriptome, resulting in "priming" or "desensitization" of molecular targets upon re-exposure to drug-related context ± cocaine.

METHODS AND MATERIALS

See supplemental information for detailed methods.

Experimental animals. Male C57BL/6J mice were obtained from The Jackson Laboratory. All experiments were conducted in accordance with guidelines of the Institutional Animal Care and Use Committee at Mount Sinai.

RNA-Sequencing. Brain regions were dissected rapidly and frozen on dry ice. RNA extraction, library preparation, and RNA-seq were conducted as described (16-18). Multiple targets were validated by qPCR using TaqMan assays (Figure S1; ThermoFisher, Foster City, CA).

Statistics and Bioinformatics.

Behavior: Behaviors were analyzed using ANOVA or Kruskall-Wallis tests depending on homozygosity of variance. All analyses were conducted using SPSS Statistical Software, V24 (IBM, Armonk, NY).

Transcriptomic Analysis: Pairwise differential expression comparisons were performed as reported (16, 17) using Voom-Limma (19); a significance threshold of fold change>1.15 and nominal $p<0.05$ were applied.

Factor Analysis and Linear Modeling: Factor analysis was used to reduce the dimensions of interdependent behavioral variables. The transformed behavioral data were then used as continuous covariates to predict gene expression in linear models. A composite “Addiction Index” (AI) of 3 factors (Figure 4; Supplemental Figures S3,S4) most closely associated with SA behaviors was calculated (Supplemental Methods). Regression analysis was conducted using Voom-Limma to determine AI associations with gene expression (19).

124

125 All other bioinformatic analyses were conducted as reported (16-18, 20, 21).

126

RESULTS

Cocaine Self-Administration Behavior:

Figure 1 provides an outline of experimental procedures, which are explained in detail in Supplemental Methods. To determine how a history of cocaine SA influences circuit-wide transcriptomes, RNA-seq was performed on PFC, dorsal striatum (DStr), NAc, basolateral amygdala (BLA), ventral hippocampus (vHIP), and VTA, obtained from the following six groups of male mice (Figure 1A): saline SA + 24 hr WD (S24, n=5-8); cocaine SA + 24 hr WD (C24, n=5-8); saline SA + 30 d WD + saline re-exposure (SS, n=5-8); saline SA + 30 d WD + cocaine exposure (SC, n=5-8); cocaine SA + 30 d WD + saline exposure (CS, n=3-7); and cocaine SA + 30 d WD + cocaine re-exposure (CC, n=5-7). Supplemental Methods provides a complete breakdown of sample size by brain region.

Gene Up- and Downregulation as a Function of History of Cocaine SA and Drug Re-Exposure:

Previous work demonstrates that repeated, non-contingent cocaine injections cause gene “priming” or “desensitization” in NAc upon cocaine re-exposure after prolonged WD (22, 23). We therefore used RNA-seq to investigate this phenomenon genome-wide and analyze transcriptomic changes throughout the reward circuitry in response to drug re-exposure after cocaine SA. Baseline transcriptional effects of cocaine SA were established by differential gene expression profiling in each brain region. Figure 1B shows pairwise comparisons of each cocaine treatment group with their saline controls (C24 vs. S24; SC, CS and CC vs. SS) and numbers of differentially expressed genes (DEGs; $p<0.05$ and fold change>15%) in each brain region (Supplemental Table S1).

To focus on genes that were *uniquely* altered following context/drug re-exposure after WD, we compared all groups to the same baseline (S24); Figures 1C, 2A; detailed description of pattern identification in Supplemental Methods). Figures 2B-D show heatmaps of DEG

patterns within each brain region for all comparisons (C24, SS, SC, CS, and CC vs. S24). This approach revealed two key findings: 1) most DEGs change in the same direction across all re-exposure paradigms (SS – CC); and 2) the *magnitude* of change for these transcripts was significantly different depending on the animals' history of cocaine SA and re-exposure (Figure 2B-D).

We focused on three patterns associated with drug use: first-ever exposure to cocaine (SC; Pattern A; Figure 2B), re-exposure to cocaine-paired context (CS, Pattern B, Figure 2C), and re-exposure to cocaine-paired context + cocaine (CC, Pattern C, Figure 2D). Each Pattern includes genes that were both differentially expressed from S24 ($p<0.05$; fold change>15%) and distinct from all other groups. Supplemental Table S2 provides complete gene lists for each pattern. Figure 3A-C shows the number of up- and downregulated DEGs in each Pattern, with a cell type analysis of DEGs shown in Supplemental Table S7.

One challenge in devising treatments for addiction is that many genes show different, sometimes opposite, regulation across brain regions. It was therefore of interest to identify specific transcripts that show similar directional changes across brain regions. Fisher's exact tests (FETs) to compare overlap of DEGs associated with Patterns A–C (Figure 3E-F; Supplemental Table S3) revealed significant overlap of upregulated genes across brain regions in Patterns A–C and identified 2 transcripts that are upregulated across a majority of brain regions in Pattern A (*Atp5j2* and *Sox18*). In Pattern C, overlap of 7 downregulated genes occurred in DStr, NAc and BLA, 2 of which were also downregulated in VTA (*Lmtk3* and *Map4k2*). All genes with fold-change >15% from each Pattern were validated by qPCR in 3 brain regions (Figure 3G-I; Supplemental Figure S1). Therefore, we used fold-change cutoff of 15% for all comparisons.

Predicted Upstream Regulators Have Unique Gene Targets Based on Cocaine SA History and Re-Exposure Across Brain Regions:

We hypothesized that these Pattern-associated genes might have common upstream regulators across brain regions, which could serve as potential targets for therapeutic intervention. Exploration of upstream regulators was conducted using Ingenuity Pathway Analysis (IPA; Qiagen Fredrick, MD) for each brain region and each Pattern. Comparison analysis was conducted to identify upstream regulators shared across brain regions (Figure 3J-L). Only those upstream regulators with an activation z-score>2 and p-value<0.01 in at least one brain region were included.

Seven molecules (CREB1, EGF, TGFB1, CREM, VEGF, HNF4A, and TCF7L2) were predicted as upstream regulators in Pattern C and at least 1 other Pattern. Notably, CREB1 was a predicted upstream regulator across all 3 Patterns (highlighted in red, Figure 3J-L). CREB1 was the top upstream regulator in Patterns A and C and a predicted upstream regulator of genes in PFC, NAc, and BLA for Patterns A, B, and C (Figure 3J-L). CREB1 is activated by initial cocaine exposure and is critical for synaptic plasticity involved in cocaine reward (24, 25). Therefore, the prediction that CREB1 is an upstream regulator of genes responding to an acute dose of cocaine in all brain regions (Pattern A) validates our pattern identification methodology (Figure 3J). It should be noted that each gene list is unique for a Pattern within a brain region. Therefore, the finding that CREB1 is a predicted upstream regulator in all 3 Patterns in PFC, NAc, and BLA suggests that a history of cocaine SA with drug/context re-exposure results in different targets for CREB1 in these regions. TGFB1, CREM, EGF, and VEGF were predicted upstream regulators of patterns associated with an acute dose of cocaine, with or without a history of cocaine SA (Patterns A & C; highlighted in orange, Figure 3J & L). Finally, HNF4A, a nuclear receptor, and TCF7L2 were predicted upstream regulators in Patterns associated with cocaine SA + WD (Patterns B & C; highlighted in purple, Figure 3K & L). Molecular pathway analysis also identified biological processes associated with the three Patterns (Supplemental Figure S2).

Association of Gene Expression Regulation with Behavioral Features of Cocaine SA

We next studied whether individual differences in cocaine SA behavior contributed to the regulation of gene expression observed across brain regions and gene expression Patterns. We used exploratory factor analysis to reduce multidimensional behavioral data to factors associated with interrelated variables (Figures 1E, 4A; Supplemental Figure S3). We identified 3 factors that are associated with SA behaviors and reflect important components of addiction: Factor 1 – cocaine intake and infusion; Factor 3 – discrimination between active and inactive levers; and Factor 4 – consummatory regulation (altered intake between FR1 and FR2; Figures 1E and 4A).

To simplify these measures of addiction-related behaviors, we calculated a composite score, or "addiction index" (AI), for each animal (Figure 4B; Supplemental Methods). Individual data are presented for each factor (behavior: Figure 4D, G & J; factor values: Figure 4E, H & K). If an animal scored high on all 3 factors (e.g., ▲ in the cocaine SA group), it has a high AI. However, if an animal scored low on one factor (e.g., × does not discriminate between active and inactive levers and ■ does not increase lever pressing when moved to FR2) their AI is lower. Factor 2 was not included in the AI because it represents differences in total lever pressing (Supplemental Figure S4), a behavior more reflective of locomotor activity and not SA *per se*. Use of this factor analysis and calculated AI scores illustrates their utility in identifying key components of complex behavioral datasets and in discriminating between baseline individual differences in behavior and those driven specifically by cocaine SA.

We used linear modeling to identify genes associated with AI scores (Figures 1F and 5A; Supplemental Table S4) to test the hypothesis that individual differences in SA behavior are associated with transcriptional regulation. We noted that the direction of expression changes in genes associated with AI scores were similar across all four 30 d WD groups (Supplemental Figure S5). Because we observed changes in *magnitude* but not direction in genes categorized

as Patterns, we hypothesized a similar effect would be observed in genes associated with AI scores. We calculated magnitude change by subtracting the log fold-change in expression of SS vs. S24 from all other comparisons (SC, CS and CC vs. S24; Figure 5B). This allowed us to adjust for gene expression differences observed between the two saline control groups. For example, if a gene is further downregulated after cocaine re-exposure, it has a negative value (blue). However, if the downregulation is blunted in comparison to that of the SS controls, it has a positive value (yellow). Heatmaps of genes significantly associated with AI scores are displayed ($p<0.05$, $|slope|>0.2$) ranked by -log(p-value) and sign of slope (red=positive association; gray=negative association).

The heatmaps reveal that a history of cocaine SA (Patterns B & C) augments the transcriptional response observed in the SS groups of those genes positively and negatively associated with AI in all 6 brain regions (Figure 5C-H). The same is not true after an animal's first dose of cocaine. Notably, in NAc, the transcriptional response of genes associated with AI is attenuated when compared to the SS group (Figure 5E). These data suggest that one dose of cocaine has little impact on genes associated with addiction-related behaviors.

We next used FETs to identify specific transcripts positively or negatively associated with AI across brain regions (Figure 5H). More transcripts overlapped across brain regions in our pair-wise comparisons than in the Patterns. Notably, genes encoding AP-1 transcription factors, including *Fos*, *Fosb,* and *Fosl2* were associated with AI in the BLA, vHIP, and NAc. This is consistent with prior work implicating AP-1 as an important transcriptional mediator of drug action (25). Genes associated with AI were enriched for neuronal-specific transcripts in all regions (Supplemental Table S7). Six transcripts (*Hspb1, Dnajc3, Mpdz, Tmem252, Lcn2,* and *Hspa1b*) were positively associated across 5 brain regions. Notably, Lipocalin 2 (*Lcn2*) was associated with AI all regions except the VTA, where there was a trend (slope=1.84; p-value=0.07), suggesting that *Lcn2* may be a potential novel therapeutic target for addiction.

Upstream regulator analysis identified 192 molecules predicted to regulate genes associated with AI (Figure 5J; Supplemental Table S5). RICTOR was the top-predicted regulator in PFC, DStr, vHIP, and VTA, and CREB1 (highlighted in red) was a predicted upstream regulator of genes in PFC, NAc, BLA, and vHIP. Finally, HNF4A was a predicted upstream regulator in 4 out of 6 brain regions. Notably, CREB1 and HNF4A were both predicted in cocaine SA + WD Patterns (Patterns B and C).

Transcriptome-Wide Expression Profiles Dependent on a History of Cocaine SA and Re-Exposure Reflect Region-Specific Roles in Addiction-Related Behaviors:

To determine if genes associated with AI overlap with genes changed in the condition defining each Pattern of gene expression, we used rank rank hypergeometric overlap (RRHO) analysis, which compares large datasets in a threshold-free manner (16, 21, 26, 27) (Figure 1F & 6). In each brain region, there was significant overlap of genes up- and downregulated in Patterns B and C —Patterns related to cocaine SA—and genes positively and negatively associated with AI, respectively. This finding is supported by FETs on filtered lists (left) showing significant overlap of up- and downregulated genes in Patterns B in all brain regions except NAc. In contrast, overlap between Pattern A—associated with initial, acute cocaine exposure—and AI was absent or far weaker. This is similar to SS vs. S24 comparisons (Supplemental Figure S7) in all brain regions except vHIP, where AI overlaps strongly with Pattern A (Figure 6E). Additionally, each region showed some Pattern-specific associations with the AI (Figure 6). Notably, NAc displayed strong associations with Pattern C (Figure 6C) only and BLA showed the strongest associations with Pattern B (Figure 6D).

Motif Analysis Reveals Nuclear Receptors as Important Regulators of Transcription After a History of Cocaine SA

We conducted HOMER motif analysis on genes associated with AI and categorized as either Pattern B or C for each brain region (Figures 1F and 7A; Supplemental Table S6). We found enrichment of several putative transcription factor binding sites implicated previously in reward-associated behaviors (SMAD, E2F, CREB, EGR, and AP1 families) across multiple brain regions (7, 8, 28-33). Interestingly, the nuclear receptor (NR) family was predicted in every brain region. HNF4A (NR2A1) was a predicted regulator in Patterns associated with a history of cocaine SA (Figure 3J-L; Patterns B and C) and genes associated with AI (Figure 5I). NRs have recently been identified as critical for CREB-regulated learning and memory in hippocampus (34) and important for aspects of cocaine SA in NAc (35). This, in combination with the prediction of CREB as an upstream regulator across all 3 Patterns and AI, raised the hypothesis that NRs may influence CREB transcriptional regulation in a context-dependent manner throughout the brain.

Because NR family members are associated with AI across all brain regions and show region-specific alterations in expression (Figure 7B), we considered the possibility that the region-specific association of NRs with AI, coupled with known regulation of CREB activity and binding, could influence the magnitude of expression of addiction-related genes after a history of cocaine SA. We used *in silico* analysis to test the hypothesis that CREB and NRs could potentially interact to influence expression in a context-specific manner. We identified proximally located CREB and NR binding sites (MatInspector, Genomatix, Germany) in a representative gene, *Lcn2*, that was positively associated with AI across multiple brain regions (Figure 7C;). Hypothetical transcription factor binding states in each brain region are presented based on region-specific NR expression, association with AI, and known binding data from the MatInspector database (Figure 7C & D). This illustrates the concept that different NRs could influence CREB-induced transcriptional regulation in a region-specific manner. For example, there are two regions in the promoter of *Lcn2* where CREB and NR binding motifs occur within 50 bp of each other. In NAc and VTA, different NRs are expressed and/or associated with AI.

306 Thus, two putative binding states are represented: 1) In NAc, NR2B1 binds near CREB in the
307 more distal binding zone, while both NR3C4 and NRC3C bind near CREB in the more proximal
308 binding zone; 2) in VTA, because NR2B1 is negatively associated with AI, it is not available to
309 bind, while NR4A2 is positively associated and available (Figure 7C). This Figure serves to
310 illustrate just one hypothetical mechanism by which the same upstream regulator (e.g., CREB)
311 can have different downstream effects across brain regions and behavioral histories.
312 Furthermore, this analysis serves as an example of how our extensive datasets can be used
313 moving forward.
314

DISCUSSION

These data provide the first unbiased assessment of gene regulation across various time-points of cocaine SA—short- and long-term WD— and two different re-exposure paradigms in six interconnected brain reward regions. While prior studies have investigated transcriptional responses to cocaine re-exposure after SA (6-12), these have not done so transcriptome-wide across a range of brain regions. Furthermore, this study is particularly powerful as we used individual variability to identify transcripts associated with aspects of cocaine SA behavior. We leveraged two statistical approaches (pattern identification and factor analysis) to characterize novel gene expression patterns throughout the reward circuitry that are sensitive to drug re-exposure after prolonged WD from cocaine SA.

Traditional methods of analyzing RNA-seq data have focused on pair-wise comparisons to identify DEGs when compared to a single control group. Our dataset contained two control groups, so pair-wise comparison using each condition's control (S24 and SS) could not uncover all transcriptional differences. Therefore, we utilized a novel approach to identify patterns of expression that reflect differences from both baselines and identified transcripts that were *uniquely* altered by either context re-exposure alone or context + drug re-exposure. This revealed that many genes associated with long-term WD and re-exposure were altered in *magnitude* but not direction. Pattern identification therefore allowed us to detect genes that were uniquely altered by acute cocaine (Pattern A), cocaine-paired context (Pattern B), or context + cocaine re-exposure (Pattern C) independent of baseline changes. Furthermore, each gene was only characterized as one pattern per brain region, thus revealing those genes associated uniquely with context- and/or drug-induced relapse.

This pattern identification analysis revealed individual transcripts that are regulated across multiple brain regions and may serve as therapeutic targets for addiction. For example, in Pattern C, two protein kinases (*Lmtk3* and *Map4k2*) are downregulated in DStr, NAc, BLA, and VTA. Knockout of *Lmtk3* increases locomotor activity and dopamine turnover in striatum

(36). Both are involved in actin cytoskeletal remodeling (37, 38) and *Map4k2* has been linked to inflammatory responses (39), two key processes in synaptic plasticity (6, 40). Similarly, transcripts were identified that were associated with AI across multiple brain regions (Figure 5H). Notably, *Lcn2* was positively associated with AI across all 6 brain regions (VTA = trend). LCN2 forms a complex with matrix metalloproteinase 9 (MMP9) and protects it from degradation, thus prolonging its activity (41). MMP9 activity has been shown to be critical for cue- and cocaine-induced reinstatement (42). These transcripts provide valuable information regarding biological processes important for cocaine addiction, and serve as potential brain-wide therapeutic targets.

One key finding of the pattern analysis came from upstream regulator analysis, which showed that many predicted transcriptional regulators were consistent across Patterns and brain regions (Figure 3J-L). This is significant because each gene list is unique for a Pattern within a brain region, suggesting that the targets of these predicted regulators change depending on cocaine history and re-exposure paradigm. This provides a potential mechanism for our hypothesis that a history of cocaine SA "primes" the reward circuitry at the transcriptional level to respond to context/drug re-exposure.

We identified CREB1 as a predicted upstream regulator in Patterns A, B, and C in PFC, BLA, and NAc – brain regions implicated in cue-induced reinstatement (43-45). CREB1 has long been implicated in addiction-related phenomena (24, 25, 46) and is critical for synaptic plasticity and reward learning. Prediction of CREB1 as a regulator of expression in all brain regions upon initial exposure to cocaine validates our pattern identification methodology.

Individual differences in SA behavioral responses correlate with gene expression changes following WD. To date, those correlations have been restricted to drug-taking animals without including saline controls, and none have been performed transcriptome-wide (47, 48). Two limitations of previous analyses are: 1) false positives/negatives due to constraints in statistical analysis of small sample sizes typical of RNA-seq experiments, and 2) the inability to

use all available SA behavioral data in correlation analysis (e.g., saline animals cannot be correlated with intake). To understand how individual differences in cocaine SA behavior might influence the transcriptional landscape after long-term WD and re-exposure, we used factor analysis to generate a composite AI that incorporates variability in SA behaviors associated with addiction-like outcomes and discriminates between saline and cocaine animals (Figure 4). This allowed us to use the saline controls in our linear model to account for baseline differences in behavior and substantially increased our sample size, reducing the likelihood of false discovery.

The greater transcriptional response in Patterns B and C drive association with the AI in a region-specific manner (Figure 5B-G). This is further reflected in the RRHO analyses (Figure 6). Thus, context is exceptionally important for the transcriptional component of relapse, and the response appears to be region-specific. RRHOs highlight which Pattern of gene expression contributes to AI in each brain region, thus showing which Pattern most reflects addiction-related behaviors. Together, our data suggest that transcriptional reprogramming occurs during long-term WD and is associated with the degree of the addictive phenotype.

The high degree of overlap of transcripts associated with AI across brain regions (Figure 5I) suggests once again that there is a suite of transcripts throughout the reward circuitry being targeted by similar upstream regulators. As in the Patterns, CREB1 was a predicted upstream regulator in PFC, NAc, BLA, and vHIP of genes associated with AI (Figure 5J). HNF4A was also a predicted upstream regulator of genes associated with AI and was one of two upstream regulators (TCF7L2) predicted for both Patterns B and C. HNF4A is implicated in epigenetic mechanisms (49-52) and dendritic spine morphology (51). While expression of *Hnf4a* was not detected in our sequencing data, other NRs were. Additionally, many NRs share a consensus sequence and compete for DNA binding (53).

Based on this knowledge, we used HOMER *de novo* motif analysis to identify putative transcription factor binding sites across genes in Patterns B or C that were also associated with AI. Strikingly, NRs were present in every brain region in a similar Pattern-specific manner as

seen by RRHO. Furthermore, CREB1 and other CREB family members were predicted in all brain regions. Thus, we posit that NRs might influence transcriptional regulation by CREB proteins in response to drug/context re-exposure in a region-specific manner.

Using novel analytic approaches followed by upstream-regulator, motif and other *in silico* analyses, we present here candidate genes and transcriptional regulators that might serve as therapeutics for addiction-related disorders. While CREB and NRs are highlighted for follow-up, this serves as just one example for how this vast dataset can be mined in future studies. To conclude, our datasets provide a highly unique resource of transcriptional regulation throughout the brain's reward circuitry and across cocaine SA, WD, and re-exposure. The transcriptional reprogramming that occurs offers valuable information regarding gene expression correlating with high performance on a highly ethologically relevant model of addiction. Thus, this work provides an increasingly complete understanding of the molecular basis of cocaine addiction and allows us to work toward individualized therapeutics.

ACKNOWLEDGEMENTS AND FINANCIAL DISCLOSURES

This work was funded by grants from the National Institute on Drug Abuse (NIDA) P01DA008227 (EJN), R01DA007359 (EJN), K99DA04211 (ESC), K01MH103473 (MLS), K01DA038654 (RWL). The authors reported no biomedical financial interests or potential conflicts of interest. We would like to thank Drs. David Self and Erin Larson for their thoughtful comments regarding the design and execution of mouse self-administration experiments.

REFERENCES:

1. Hyman SE, Malenka RC, Nestler EJ (2006): Neural mechanisms of addiction: the role of reward-related learning and memory. *Annu Rev Neurosci*. 29:565-598.
2. Koob GF, Volkow ND (2016): Neurobiology of addiction: a neurocircuitry analysis. *Lancet Psychiatry*. 3:760-773.
3. Koob GF, Caine SB, Parsons L, Markou A, Weiss F (1997): Opponent process model and psychostimulant addiction. *Pharmacol Biochem Behav*. 57:513-521.
4. Wang JQ, McGinty JF (1999): Glutamate-dopamine interactions mediate the effects of psychostimulant drugs. *Addict Biol*. 4:141-150.
5. White FJ, Kalivas PW (1998): Neuroadaptations involved in amphetamine and cocaine addiction. *Drug Alcohol Depend*. 51:141-153.
6. Cahill ME, Bagot RC, Gancarz AM, Walker DM, Sun H, Wang ZJ, et al. (2016): Bidirectional Synaptic Structural Plasticity after Chronic Cocaine Administration Occurs through Rap1 Small GTPase Signaling. *Neuron*. 89:566-582.
7. Chandra R, Francis TC, Konkalmatt P, Amgalan A, Gancarz AM, Dietz DM, et al. (2015): Opposing role for Egr3 in nucleus accumbens cell subtypes in cocaine action. *J Neurosci*. 35:7927-7937.
8. Gancarz-Kausch AM, Schroeder GL, Panganiban C, Adank D, Humby MS, Kausch MA, et al. (2013): Transforming growth factor beta receptor 1 is increased following abstinence from cocaine self-administration, but not cocaine sensitization. *PLoS One*. 8:e83834.
9. Graham DL, Edwards S, Bachtell RK, DiLeone RJ, Rios M, Self DW (2007): Dynamic BDNF activity in nucleus accumbens with cocaine use increases self-administration and relapse. *Nat Neurosci*. 10:1029-1037.
10. Freeman WM, Patel KM, Brucklacher RM, Lull ME, Erwin M, Morgan D, et al. (2008): Persistent alterations in mesolimbic gene expression with abstinence from cocaine self-administration. *Neuropsychopharmacology*. 33:1807-1817.
11. Lu L, Grimm JW, Shaham Y, Hope BT (2003): Molecular neuroadaptations in the accumbens and ventral tegmental area during the first 90 days of forced abstinence from cocaine self-administration in rats. *J Neurochem*. 85:1604-1613.
12. Larson EB, Akkentli F, Edwards S, Graham DL, Simmons DL, Alibhai IN, et al. (2010): Striatal regulation of DeltaFosB, FosB, and cFos during cocaine self-administration and withdrawal. *J Neurochem*. 115:112-122.
13. Sadakierska-Chudy A, Frankowska M, Jastrzebska J, Wydra K, Miszkiel J, Sanak M, et al. (2017): Cocaine Administration and Its Withdrawal Enhance the Expression of Genes Encoding Histone-Modifying Enzymes and Histone Acetylation in the Rat Prefrontal Cortex. *Neurotox Res*. 32:141-150.
14. Vallender EJ, Goswami DB, Shinday NM, Westmoreland SV, Yao WD, Rowlett JK (2017): Transcriptomic profiling of the ventral tegmental area and nucleus accumbens in rhesus macaques following long-term cocaine self-administration. *Drug Alcohol Depend*. 175:9-23.
15. Freeman WM, Lull ME, Patel KM, Brucklacher RM, Morgan D, Roberts DC, et al. (2010): Gene expression changes in the medial prefrontal cortex and nucleus accumbens following abstinence from cocaine self-administration. *BMC Neurosci*. 11:29.

16. Bagot RC, Cates HM, Purushothaman I, Lorsch ZS, Walker DM, Wang J, et al. (2016): Circuit-wide Transcriptional Profiling Reveals Brain Region-Specific Gene Networks Regulating Depression Susceptibility. *Neuron*. 90:969-983.

17. Bagot RC, Cates HM, Purushothaman I, Vialou V, Heller EA, Yieh L, et al. (2017): Ketamine and Imipramine Reverse Transcriptional Signatures of Susceptibility and Induce Resilience-Specific Gene Expression Profiles. *Biol Psychiatry*. 81:285-295.

18. Labonte B, Engmann O, Purushothaman I, Menard C, Wang J, Tan C, et al. (2017): Sex-specific transcriptional signatures in human depression. *Nat Med*. 23:1102-1111.

19. Law CW, Chen Y, Shi W, Smyth GK (2014): voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 15:R29.

20. Pena CJ, Kronman HG, Walker DM, Cates HM, Bagot RC, Purushothaman I, et al. (2017): Early life stress confers lifelong stress susceptibility in mice via ventral tegmental area OTX2. *Science*. 356:1185-1188.

21. Seney ML, Huo Z, Cahill K, French L, Puralewski R, Zhang J, et al. Opposite molecular signatures of depression in men and women. *Biological Psychiatry*.

22. Damez-Werno D, LaPlant Q, Sun H, Scobie KN, Dietz DM, Walker IM, et al. (2012): Drug experience epigenetically primes Fosb gene inducibility in rat nucleus accumbens. *J Neurosci*. 32:10267-10272.

23. Maze I, Covington HE, 3rd, Dietz DM, LaPlant Q, Renthal W, Russo SJ, et al. (2010): Essential role of the histone methyltransferase G9a in cocaine-induced plasticity. *Science*. 327:213-216.

24. Nestler EJ (2016): Reflections on: "A general role for adaptations in G-Proteins and the cyclic AMP system in mediating the chronic actions of morphine and cocaine on neuronal function". *Brain Res*. 1645:71-74.

25. Robison AJ, Nestler EJ (2011): Transcriptional and epigenetic mechanisms of addiction. *Nat Rev Neurosci*. 12:623-637.

26. Plaisier S.B. TR, Wong J.A., Graeber T.G. (2010): Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res*. 38:169.

27. Stein JL, de la Torre-Ubieta L, Tian Y, Parikshak NN, Hernandez IA, Marchetto MC, et al. (2014): A quantitative framework to evaluate modeling of cortical development by neural stem cells. *Neuron*. 83:69-86.

28. Cates HM, Heller EA, Lardner CK, Purushothaman I, Pena CJ, Walker DM, et al. (2017): Transcription Factor E2F3a in Nucleus Accumbens Affects Cocaine Action via Transcription and Alternative Splicing. *Biol Psychiatry*.

29. Cates HM, Thibault M, Pfau M, Heller E, Eagle A, Gajewski P, et al. (2014): Threonine 149 phosphorylation enhances DeltaFosB transcriptional activity to control psychomotor responses to cocaine. *J Neurosci*. 34:11461-11469.

30. Feng J, Wilkinson M, Liu X, Purushothaman I, Ferguson D, Vialou V, et al. (2014): Chronic cocaine-regulated epigenomic changes in mouse nucleus accumbens. *Genome Biol*. 15:R65.

31. McClung CA, Nestler EJ (2003): Regulation of gene expression and cocaine reward by CREB and DeltaFosB. *Nat Neurosci*. 6:1208-1215.

32. Renthal W, Kumar A, Xiao G, Wilkinson M, Covington HE, 3rd, Maze I, et al. (2009): Genome-wide analysis of chromatin regulation by cocaine reveals a role for sirtuins. *Neuron*. 62:335-348.

33. Vialou V, Maze I, Renthal W, LaPlant QC, Watts EL, Mouzon E, et al. (2010): Serum response factor promotes resilience to chronic social stress through the induction of DeltaFosB. *J Neurosci*. 30:14585-14592.

34. Bridi MS, Hawk JD, Chatterjee S, Safe S, Abel T (2017): Pharmacological Activators of the NR4A Nuclear Receptors Enhance LTP in a CREB/CBP-Dependent Manner. *Neuropsychopharmacology*. 42:1243-1253.

35. Zhang Y, Kong F, Crofton EJ, Dragosljvich SN, Sinha M, Li D, et al. (2016): Transcriptomics of Environmental Enrichment Reveals a Role for Retinoic Acid Signaling in Addiction. *Front Mol Neurosci*. 9:119.

36. Inoue T, Hoshina N, Nakazawa T, Kiyama Y, Kobayashi S, Abe T, et al. (2014): LMTK3 deficiency causes pronounced locomotor hyperactivity and impairs endocytic trafficking. *J Neurosci*. 34:5927-5937.

37. Kyriakis JM (1999): Signaling by the germinal center kinase family of protein kinases. *J Biol Chem*. 274:5259-5262.

38. Xu Y, Zhang H, Lit LC, Grothey A, Athanasiadou M, Kiritsi M, et al. (2014): The kinase LMTK3 promotes invasion in breast cancer through GRB2-mediated induction of integrin beta(1). *Sci Signal*. 7:ra58.

39. Chuang HC, Wang X, Tan TH (2016): MAP4K Family Kinases in Immunity and Inflammation. *Adv Immunol*. 129:277-314.

40. Yirmiya R, Goshen I (2011): Immune modulation of learning, memory, neural plasticity and neurogenesis. *Brain Behav Immun*. 25:181-213.

41. Yan L, Borregaard N, Kjeldsen L, Moses MA (2001): The high molecular weight urinary matrix metalloproteinase (MMP) activity is a complex of gelatinase B/MMP-9 and neutrophil gelatinase-associated lipocalin (NGAL). Modulation of MMP-9 activity by NGAL. *J Biol Chem*. 276:37258-37265.

42. Smith AC, Kupchik YM, Scofield MD, Gipson CD, Wiggins A, Thomas CA, et al. (2014): Synaptic plasticity mediating cocaine relapse requires matrix metalloproteinases. *Nat Neurosci*. 17:1655-1657.

43. Kalivas PW (2009): The glutamate homeostasis hypothesis of addiction. *Nat Rev Neurosci*. 10:561-572.

44. Pickens CL, Airavaara M, Theberge F, Fanous S, Hope BT, Shaham Y (2011): Neurobiology of the incubation of drug craving. *Trends Neurosci*. 34:411-420.

45. See RE (2005): Neural substrates of cocaine-cue associations that trigger relapse. *Eur J Pharmacol*. 526:140-146.

46. Carlezon WA, Jr., Duman RS, Nestler EJ (2005): The many faces of CREB. *Trends Neurosci*. 28:436-445.

47. Tacelosky DM, Alexander DN, Morse M, Hajnal A, Berg A, Levenson R, et al. (2015): Low expression of D2R and Wntless correlates with high motivation for heroin. *Behav Neurosci*. 129:744-755.

541 48. Zhou Y, Leri F, Cummins E, Kreek MJ (2015): Individual differences in gene expression of
542 vasopressin, D2 receptor, POMC and orexin: vulnerability to relapse to heroin-seeking in
543 rats. *Physiol Behav*. 139:127-135.

544 49. Cicchini C, de Nonno V, Battistelli C, Cozzolino AM, De Santis Puzzonia M, Ciafre SA, et al.
545 (2015): Epigenetic control of EMT/MET dynamics: HNF4alpha impacts DNMT3s through
546 miRs-29. *Biochim Biophys Acta*. 1849:919-929.

547 50. Fernandez-Santiago R, Carballo-Carbajal I, Castellano G, Torrent R, Richaud Y, Sanchez-
548 Danes A, et al. (2015): Aberrant epigenome in iPSC-derived dopaminergic neurons from
549 Parkinson's disease patients. *EMBO Mol Med*. 7:1529-1546.

550 51. Liu D, Tang H, Li XY, Deng MF, Wei N, Wang X, et al. (2017): Targeting the HDAC2/HNF-
551 4A/miR-101b/AMPK Pathway Rescues Tauopathy and Dendritic Abnormalities in
552 Alzheimer's Disease. *Mol Ther*. 25:752-764.

553 52. Yamanishi K, Doe N, Sumida M, Watanabe Y, Yoshida M, Yamamoto H, et al. (2015):
554 Hepatocyte nuclear factor 4 alpha is a key factor related to depression and physiological
555 homeostasis in the mouse brain. *PLoS One*. 10:e0119021.

556 53. Olivares AM, Moreno-Ramos OA, Haider NB (2015): Role of Nuclear Receptors in Central
557 Nervous System Development and Associated Diseases. *J Exp Neurosci*. 9:93-121.

558

559

FIGURE LEGENDS

Figure 1: Outline of Experimental Approach and Bioinformatic Analyses: (A) Experimental design and summary of groups. Mice were food trained followed by 5-10 d of FR1 scheduling and 4-5 d of FR2. One group was euthanized 24 h after their last SA session while another cohort of animals were group housed in their home cage for 30 d. After WD, animals were given an injection of saline or cocaine and re-exposed to their original SA chamber for 1 h and euthanized immediately. (B) Data collection and RNA-seq data analysis. RNA-seq was performed on micro-dissections of 6 reward-associated brain regions. Differential expression analysis was performed to identify DEGs compared to their control group (S24 or SS). Number of DEGs per brain region are indicated (Red = greatest; Gray = least). (C) In an effort to identify genes that were uniquely altered by cocaine re-exposure we used pattern analysis and compared all groups to the same baseline (S24). Three patterns were investigated: Pattern A: genes uniquely altered by an initial dose of cocaine (SC; 1 h post-injection); Pattern B: genes uniquely altered by re-exposure to cocaine-paired context (CS); and Pattern C: genes uniquely altered by cocaine re-exposure (CC; 1 h post-injection). (D) Data collection and analysis of cocaine SA behavioral data. Because all animals (saline included), underwent varying numbers of SA trials at FR1, behavioral data was aligned to the day each animal transitioned onto an FR2 schedule (i.e., the last day on FR1). Therefore, data for days 5 - 10 of FR1, but not 1 -4, includes a majority of the animals in the study. In self-administering animals, cocaine (red) acted as a reinforcer as shown by increased active lever (solid line) vs. inactive lever (dotted line) responding on day 3 of FR1 (indicated by *). This did not occur for saline animals (black). Cocaine SA animals began pressing the active lever significantly more than saline (indicated by *) beginning on day 6 of FR1, which continued throughout FR2. Cocaine SA animals (red) received more infusions than their saline counterparts (black) and maintained the same number of infusions after switching to an FR2 schedule, indicating that cocaine was reinforcing lever pressing in these mice. (E) We generated an “addiction index” using exploratory factor analysis

586 to reduce the multi-dimensional behavioral data to “factors” associated with components of
587 cocaine SA behavior. We then combined the 3 factors most strongly associated with an
588 addicted-like phenotype to differentiate between individual animals with high performance
589 across multiple behavioral endpoints. (F) Integration of genes and behaviors to identify
590 transcripts important for the addicted-like phenotype. Enrichment testing reveals transcripts
591 regulated across multiple brain regions. *In silico* analysis of potential upstream regulators of the
592 enriched genes. Rank-rank hypergeometric overlap used to determine if gene expression
593 Patterns are associated with the addiction index within a brain region. Behavioral data were
594 analyzed using Kruskal Wallis followed by Mann-Whitney Nonparametric Test; *p<0.05;
595 **p<0.01; data are presented as mean ± SEM.
596
597 **Figure 2: Gene expression Patterns associated with cocaine exposure.** (A) To reduce the
598 dimensions of our RNA-seq data and identify genes that were uniquely changed by a specific
599 exposure paradigm, we used pattern analysis to categorize genes into Patterns of expression
600 when compared to the same S24 baseline. Categorization of genes affected uniquely by: (B) an
601 initial dose of cocaine (Pattern A); (C) re-exposure to the cocaine-paired context after 30 d WD
602 from cocaine SA (Pattern B); (D) re-exposure to cocaine in the cocaine-paired context after 30 d
603 WD from cocaine SA (Pattern C). Heatmaps show that, for all brain regions, expression of
604 genes categorized in each Pattern is, by definition, most pronounced in the comparison that
605 represents that Pattern (e.g., Pattern A most pronounced in SC vs S24 when compared to other
606 groups).
607
608 **Figure 3: Gene expression patterns associated with cocaine exposure reveal circuit-wide**
609 **transcriptional changes and upstream regulators.** (A-C) Number and percentage of genes
610 up- and downregulated (yellow=>60% up; blue=>60% down) in each brain region for each of the
611 three Patterns defined in Figure 2**.** (D-F) Overlap across brain regions of upregulated (top) and

downregulated (bottom) genes, color-coded for significance. Total number of regulated genes in each region is shown in parentheses. Examples of transcripts up- or downregulated across more than two brain regions are listed in the insets. (G-I) Patterns were validated using qPCR on technical replicates. Patterns were validated for 8 transcripts across 3 brain regions. Representative transcripts from each pattern are presented. Fold-changes of at least 15% in the RNA-seq data were validated using qPCR across all patterns analyzed, supporting use of this fold-change in all analyses. (J-L) Upstream regulator analysis was conducted across brain regions for each Pattern. Five upstream regulators were consistently predicted to regulate genes across brain regions: CREB1 (highlighted in red) is a predicted upstream regulator of all Patterns. Regulators overlapping between Patterns A and C are highlighted in orange and are likely indicative of those important for regulating the response to acute cocaine exposure independent of a history of cocaine SA. Regulators overlapping between Patterns B and C are highlighted in purple and are likely indicative of those important for regulating the response to a cocaine-paired context after a history of cocaine SA. Activation Z-Scores in heatmaps: positive (yellow) = overrepresentation of targets activated by regulator; negative (blue) = overrepresentation of targets repressed by regulator; no direction (black) = no significant enrichment of activated versus repressed targets; white = not a predicted upstream regulator. *p<0.05; **p<0.01; * * = transcripts overlap across multiple brain regions.

Figure 4: Generation of an "addiction index" for individual animals. (A-B) Exploratory factor analysis on multiple behavioral endpoints reduced multi-dimensional behavioral data to 8 "factors." A composite score, or "addiction index (AI)," of those factors most strongly associated with behaviors reflective of an addicted-like phenotype was generated using the individual transformed data for Factors 1, 3, & 4. (C-K) Data for individual animals for each behavior and each factor are presented. Each animal is represented by the same unique shape and color. (C, F, I) Factor loading, or associations, of Factors 1, 3, & 4 with SA behaviors (yellow = positive;

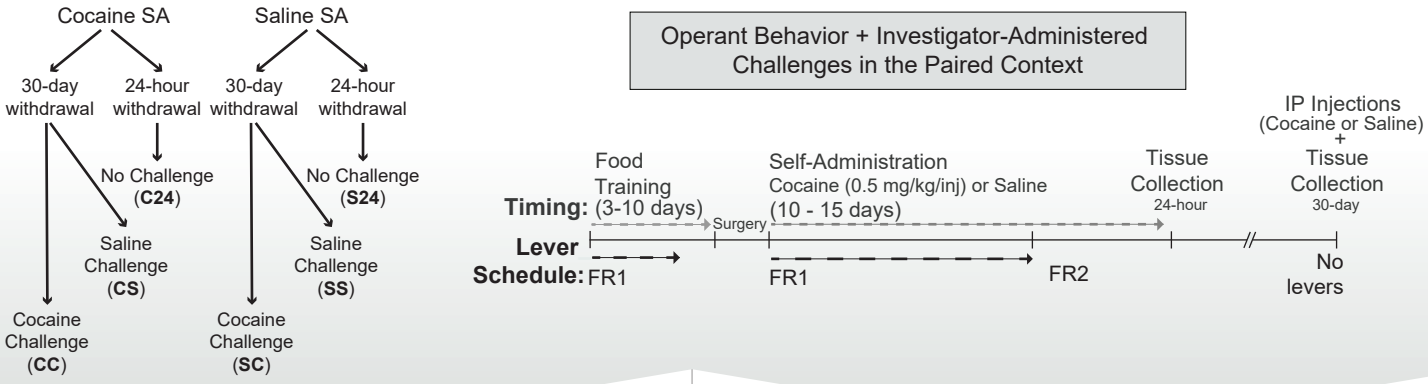
blue = negative) are presented. (D, G, J) Individual data presented for the behaviors associated with each factor. (D) Factor 1 associated with intake and infusions; (G) Factor 3 is positively associated with active lever and negatively associated with inactive lever under an FR2 schedule; (J) Factor 4 is positively associated with FR2 lever presses and negatively associated with lever pressing on an FR1 schedule. (E, H, K) Individual transformed data for Factors 1 (E), 3 (H) and 4 (K). The product of these values was calculated to generate an AI for each individual. An animal must display high performance on all three factors (▲) to have a high AI. By contrast, if an animal performs poorly on one of the behaviors (× or ■) their AI is lower.

Figure 5: Genes associated with the AI are reprogrammed by cocaine SA to be responsive to drug or cocaine-paired context. (A) Linear modeling was used to identify genes associated with the AI within each brain region. Only genes with a slope of at least ±15% and a nominal $p<0.05$ were investigated. Similar to the gene expression Patterns (Figure 2), we observed that directional changes in expression were similar across all re-exposure comparisons (SS, SC, CS & CC vs. S24). Genes that were negatively associated with AI (gray bar) were downregulated and genes positively associated with AI (red bar) were upregulated (Supplemental Table S4). (B-G) Heatmaps were transformed to indicate change in expression from SS controls. Blue = fold change in the negative direction from SS vs. S24 and yellow = fold change in the positive direction from SS vs. S24. Cocaine SA programs those transcripts associated with AI to be hyper-response to context either with or without drug. (H) Overlap of genes positively (left) or negatively (right) associated with AI across brain regions, color-coded for significance. Total number of genes in each brain region listed in parentheses and total number of genes overlapping between regions indicated in corresponding boxes. There is significant overlap of genes associated with the AI across most brain regions. (I) Upstream regulator analysis reveals similar putative transcriptional regulators in genes associated with AI as those associated with specific gene expression Patterns. Colors correspond to regulators

664 overlapping in multiple Patterns (see Figure 3). Activation Z-Scores: positive (yellow) =
665 overrepresentation of targets activated by regulator; negative (blue) = overrepresentation of
666 targets repressed by regulator; no direction (black) = no significant enrichment of activated or
667 repressed targets; white = not a predicted upstream regulator.
668
669 **Figure 6: Overlap of transcriptional profiles related to the AI and gene expression**
670 **Patterns reveals which Pattern contributes most to AI.** A-F) Overlap of genes positively or
671 negatively associated with AI and also up- or downregulated within each gene expression
672 Pattern within the gene lists filtered for significance (Fisher's exact test; left) or transcriptome-
673 wide expression profiles (RRHO plots; right). Overlap of genes associated with AI are specific to
674 brain regions. For example, significant overlap of up- and downregulated genes across Patterns
675 B & C with AI are observed in PFC and VTA. vHIP. BLA and DStr are enriched in genes in
676 Pattern B and NAc only shows enrichment of genes in Pattern C. RRHO plots to the right of
677 each panel reveal significance of overlap between region-specific transcriptional profiles
678 associated with AI for Patterns A-C. A key for these plots is shown to the right.
679
680 **Figure 7: Motif analysis reveals putative role for NRs in controlling region-specific**
681 **cocaine-induced gene expression.** (A) HOMER motif analysis was conducted on genes
682 defined as either Pattern B or C and significantly associated with the AI (lists from Figure 6
683 enrichment tests). Table of putative transcription factor families whose motifs were enriched in
684 at least 4 of 6 brain regions. Members of the NR family were predicted upstream regulators in all
685 brain regions and were Pattern-specific. (B) NR family members are positively (red) and
686 negatively (gray) associated with the AI in a region-specific manner. Black indicates no
687 association and white indicates no detectable expression. Only NRs with a significant
688 association in at least one brain region are displayed. (C) Hypothetical model of transcriptional
689 co-regulation by CREB and NRs in a gene positively associated with AI across all brain regions

690 (VTA = trend). *In silico* analysis of transcription factor binding sites, identified using
691 MatInspector, indicate motifs in close proximity to each other (less than 50 bp), and binding data
692 from the MatInspector database indicate binding of specific NRs within the *Lcn2* promoter.
693 Based on our AI data, we extrapolated possible region-specific binding states that could be
694 regulating the transcriptional response to drug or context re-exposure. Color indicates sub-
695 family of NRs: orange = NR2 subfamily; pink = NR3 subfamily; green = NR4 subfamily. X =
696 negative association with AI.

A. Experimental Design



B. Data Collection & Analysis: RNAseq

RNA isolated 6 brain regions from individual animals → RNA-seq and differential analysis were performed

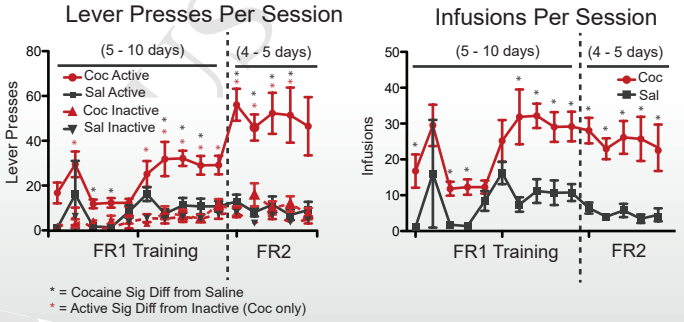
Number of Differentially Expressed Genes in Each Brain Region

PFC		DStr		NAc		BLA		vHIP		VTA		
↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	↑	↓	
559	53	383	37	166	18	37	47	200	11	227	35	C24 v. S24
129	36	79	13	61	56	280	163	457	725	245	97	SC v. SS
114	54	82	49	120	158	202	70	727	1018	113	211	CS v. SS
72	14	185	162	78	220	236	305	279	262	52	76	CC v. SS

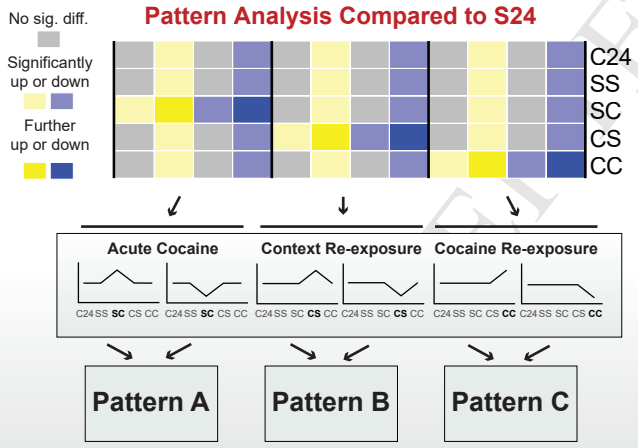
Number of Genes: 0 – 500

D. Data Collection & Analysis: Behavior

Animals placed in operant chamber each day for food or drug SA → Lever presses and cocaine infusions recorded daily for each animal

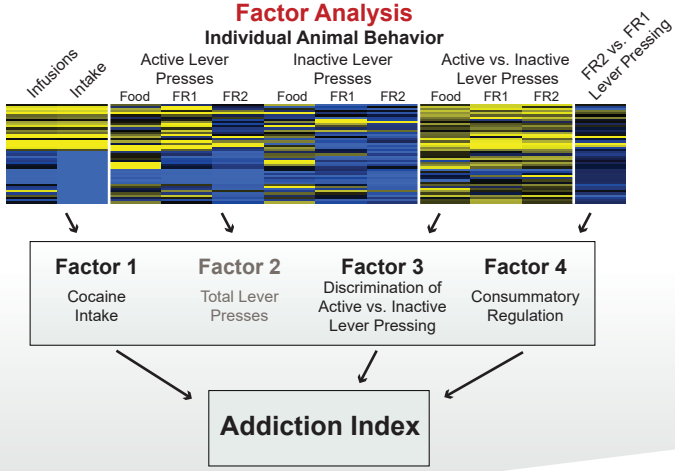


C.



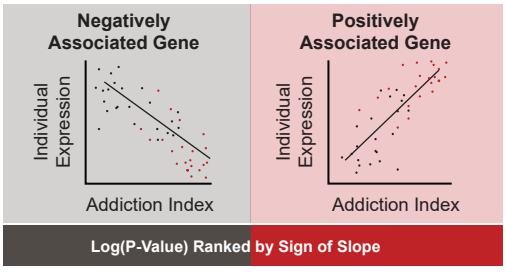
Reducing Dimensions of the Data

E.

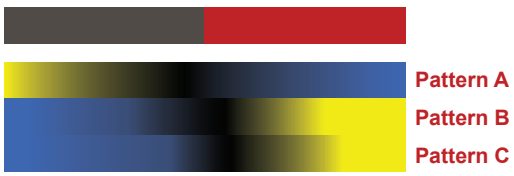


F. Integrating Genes and Behavior

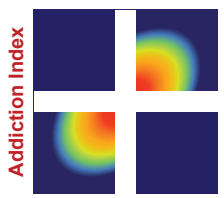
Individual variability used to determine relationship between expression and addiction index



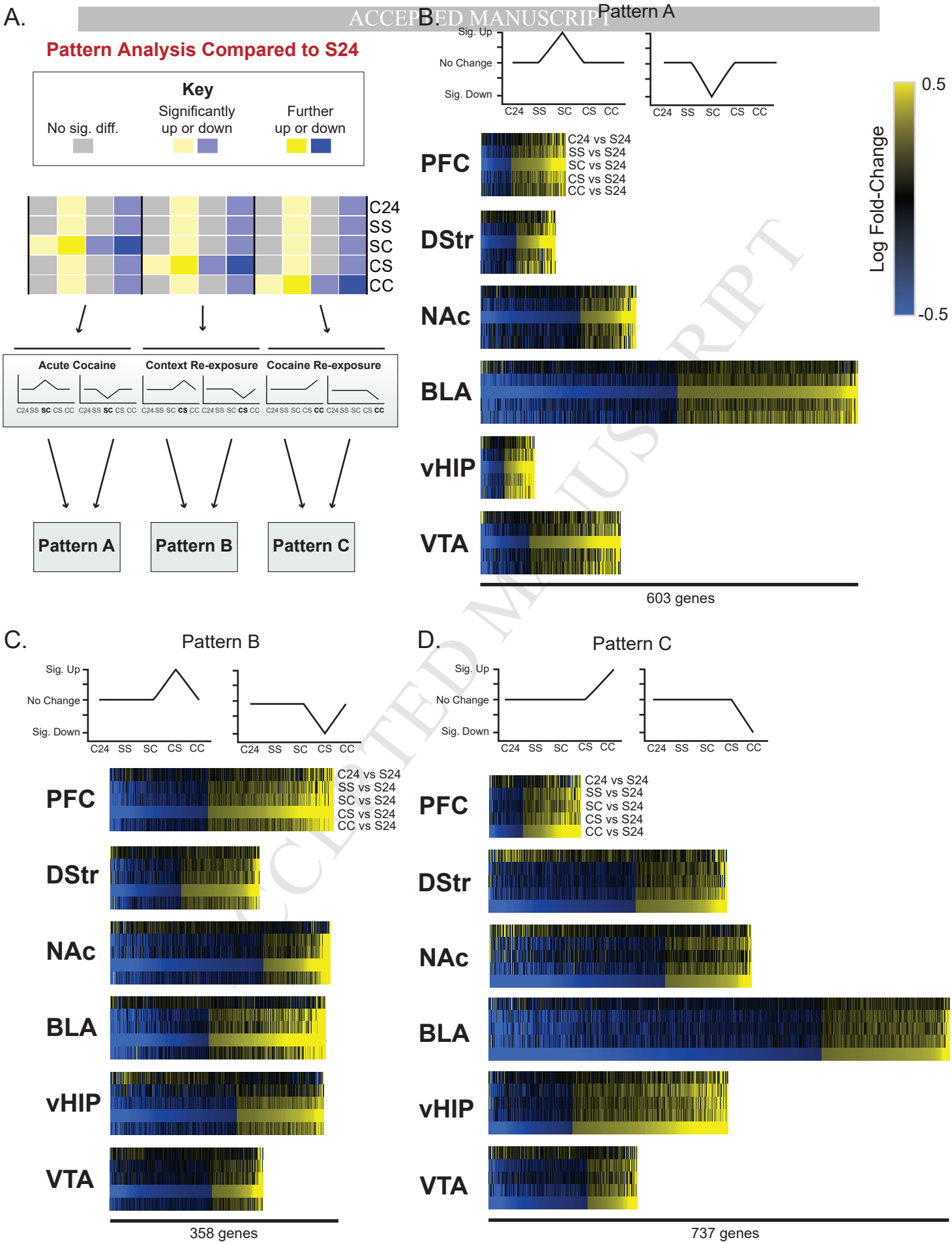
Context-Dependent Expression of Genes Associated with Addiction Index

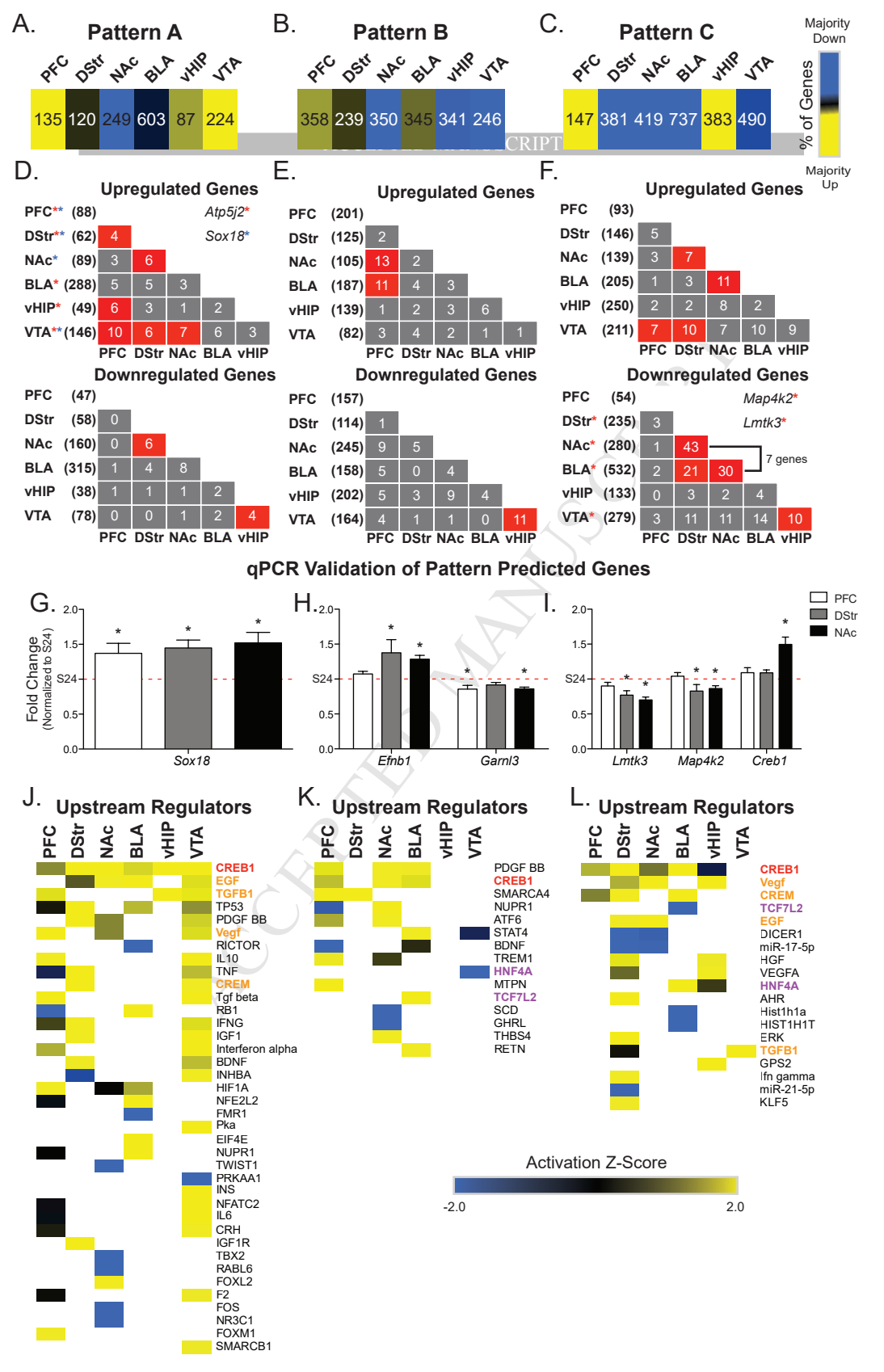


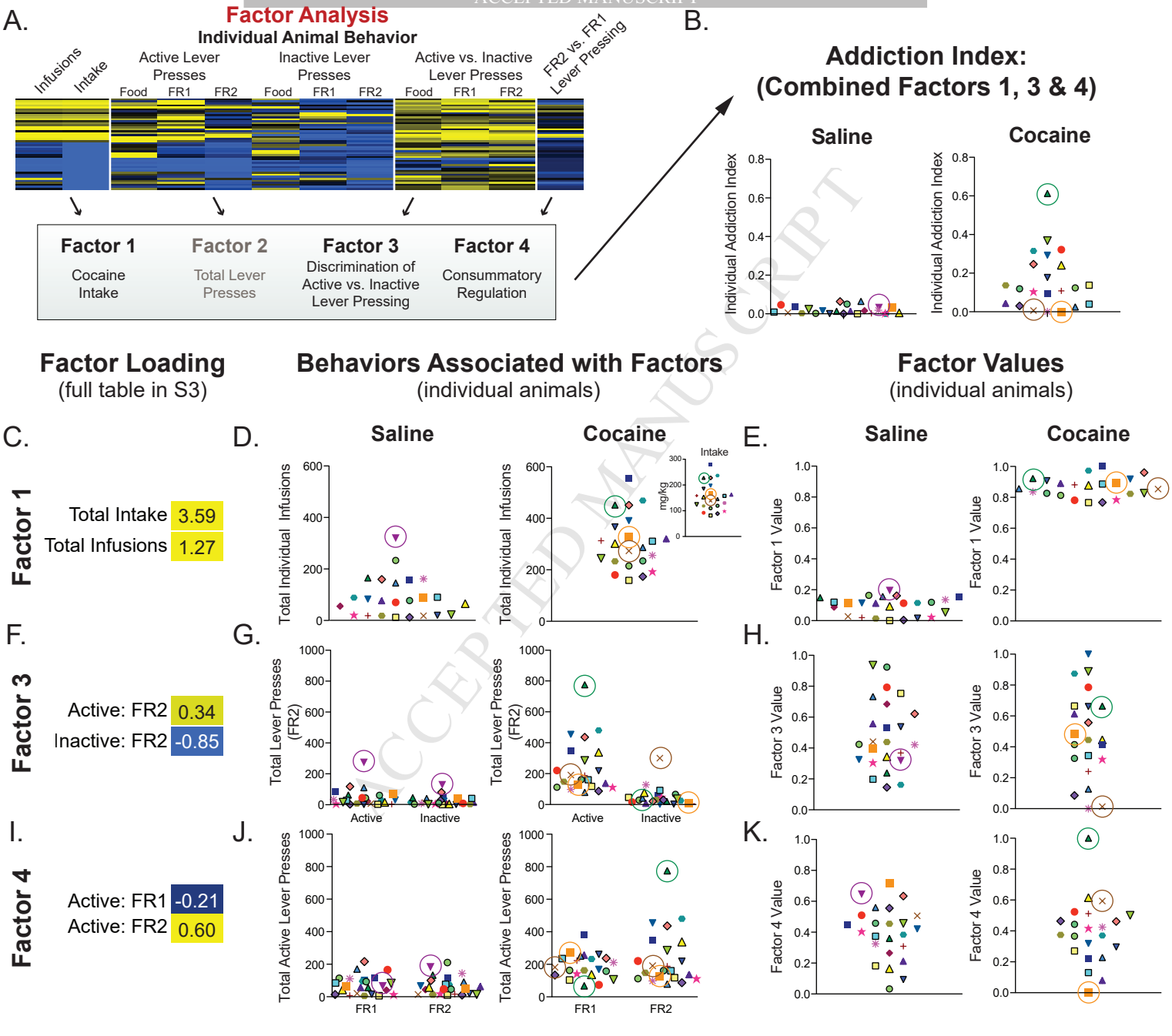
Threshold-Free Overlap of Patterns and Addiction Index

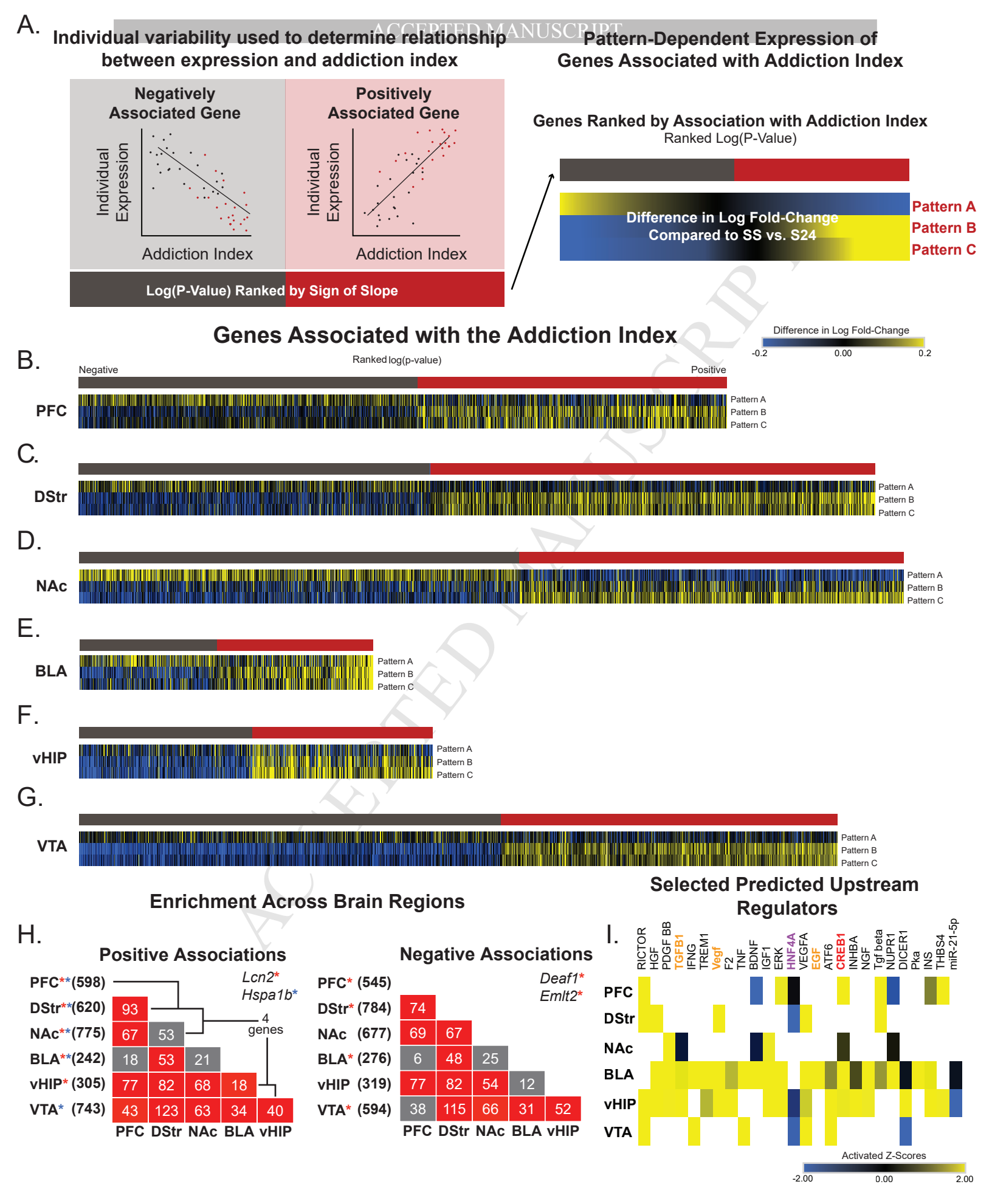


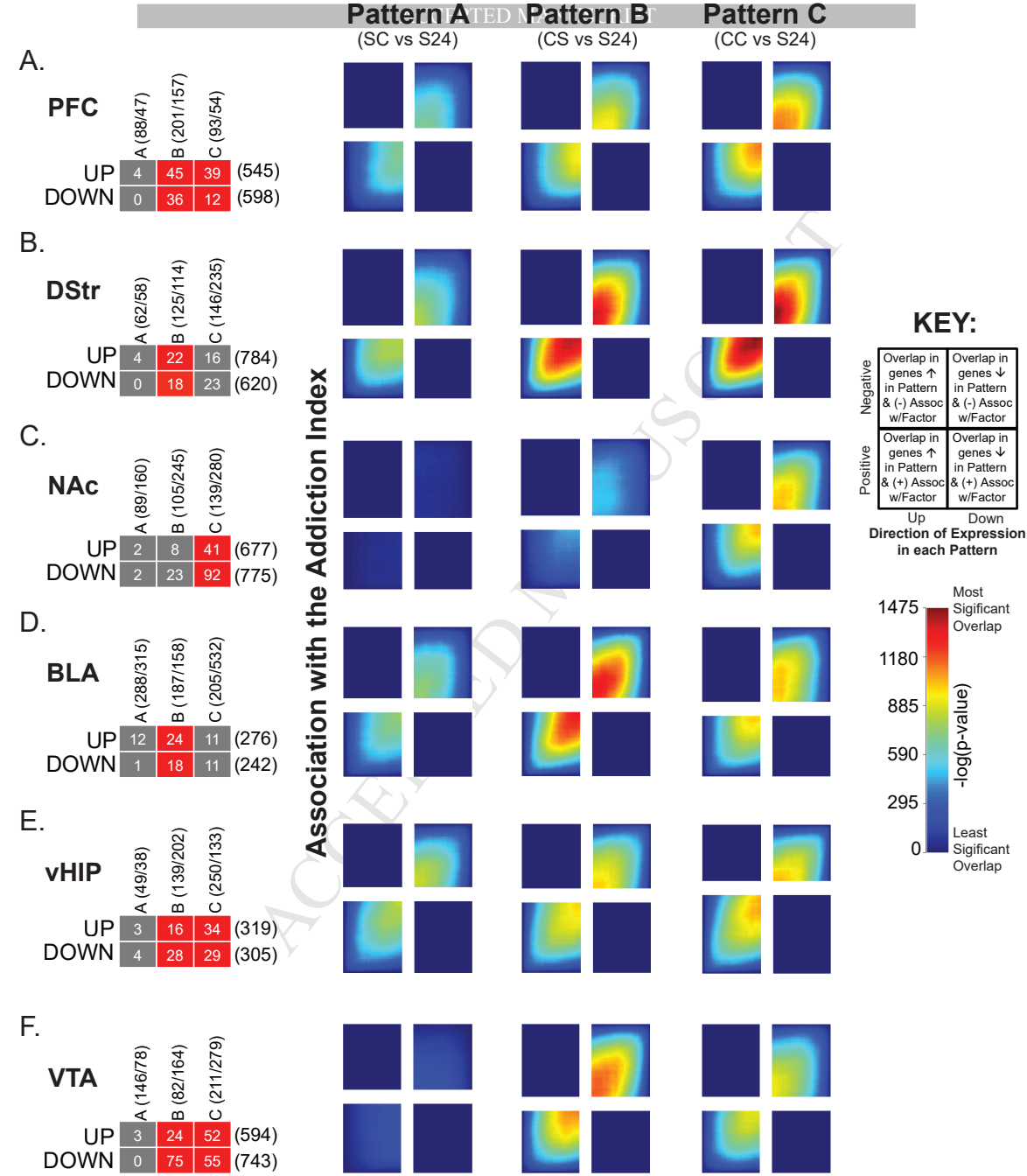
Motif Analysis on Genes Associated with Addiction Index











A.

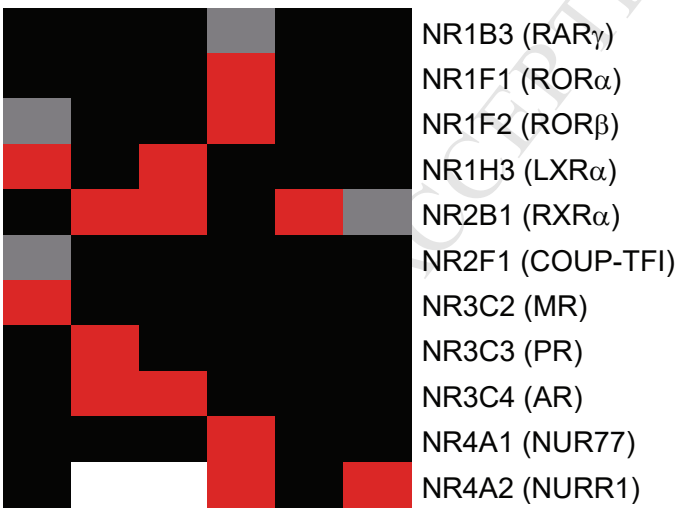
Transcription Factor Motif Enrichment in Genes Associated with the Addiction Index

	PFC	DStr	NAc	BLA	vHIP	VTA
Pattern B	E2F3 EGR1 EGR2 **NR1A2** **NR2E3** SMAD3 SMAD4	ATF1 CREB1 CREB3L2 E2F3 EGR1 **NR2F2** **NR3B3**	EGR1 EGR2 **NR2A1** **NR4A2** **NR5A2** SMAD4 TBOX:SMAD	EGR1 EGR2 **NR2E1** **NR2E3** **NR2F2** **NR3B1** SMAD3	ATF1 CREM E2F4 FOS::JUN **NR2E3** SMAD3 SMAD4	CREB3L2 E2F FOS::JUN **NR2F2** **NR4A1**
Pattern C	ATF1 E2F3 **NR2F2** **NR2F6** **NR3B1** TBOX::SMAD	ATF1 CREB1 E2F1 E2F3 FOS::JUN **NR2E3** **NR3B1** **NR4A1** **NR5A1**	ATF4 E2F4 EGR1 **NR2A1** **NR2E3** **NR3B1**	CREB1 E2F1 E2F2 EGR1 FOS::JUN	**NR3C1** **NR3C2** **NR2A1** **NR2F2** **NR4A1** **NR Class**	CREB1 E2F EGR1 EGR2 FOS

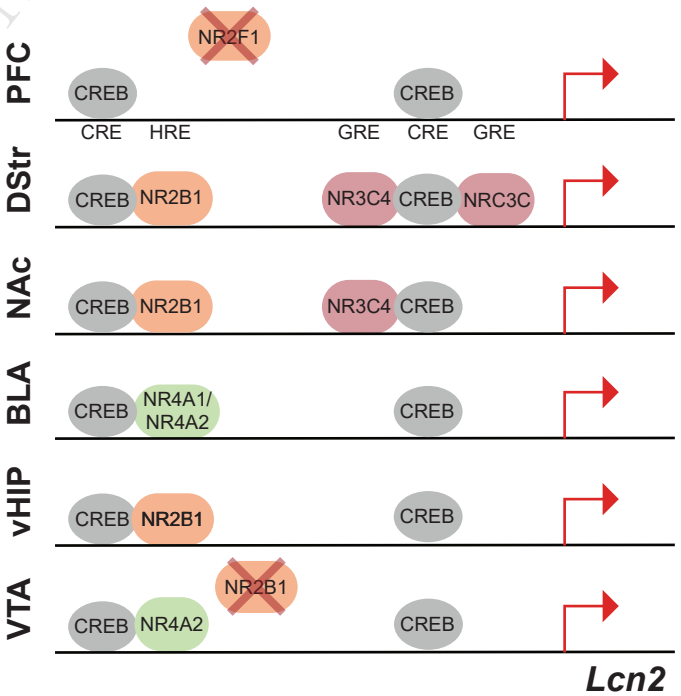
(Nuclear Receptors)

B.

Nuclear Receptors Associated with the Addiction Index



C.



Cocaine Self-Administration Alters Transcriptome-wide Responses in the Brain's Reward Circuitry

Supplemental Information

SUPPLEMENTAL TABLE LEGENDS

All supplemental tables are provided in Excel; See supplemental .zip file to download

Table S1: Differentially expressed genes (DEGs) calculated from pair-wise comparisons. A complete list of DEGs (nominal p-value < 0.05; fold-change ± 15%) in relation to saline controls (either S24 or SS) presented in Figure 1B. Each comparison is presented on a separate tab.

Table S2: Genes categorized by Patterns of expression. A complete list of genes categorized as Pattern A, B, or C in each brain region. Log fold-change for all conditions when compared to the same baseline (S24) are included. Genes were categorized by their expression patterns and a fold-change cut off of ± 15% was applied to each list to identify genes uniquely altered under each re-exposure condition. Each Pattern is presented on a separate tab.

Table S3: Overlap of genes categorized as Pattern A, B, or C across brain regions. A complete list of genes categorized as Patterns A, B, or C that overlap across multiple brain regions. Fisher's exact tests revealed significant enrichment across lists. Comparisons reaching significance after multiple comparison correction (FDR) are bolded. Each pattern and direction of regulation is presented on a separate tab of the table.

Table S4: Overlap of genes associated with the addiction index (AI) across brain regions. A complete list of genes associated with AI that overlap across multiple brain region. Fisher's exact tests revealed significant enrichment across lists. Comparisons reaching significance after multiple comparison correction (FDR) are bolded. Positive and negative associations are presented on separate tabs.

Table S5: Table of predicted upstream regulators of genes associated with AI. A full list of predicted regulators of genes associated with AI and their activation z-scores. Activation z-Scores: positive = overrepresentation of targets activated by regulator; negative =

overrepresentation of targets repressed by regulator; no direction = no significant enrichment of activated or repressed targets; white = not a predicted upstream regulator.

Table S6: Overlap of genes categorized as Pattern A, B or C and associated with the AI within a brain region. A complete list of genes categorized as Pattern A, B, or C that overlap with those associated with AI within each brain region. Fisher's exact tests revealed significant enrichment across lists. Comparisons reaching significance after multiple comparison correction (FDR) are bolded. Comparison of Pattern/AI for each brain region are presented on a separate tab.

Table S7: Cell-type specific enrichment of genes categorized as Pattern A, B, or C or those associated with AI. Fisher's exact tests revealed significant enrichment of cell-type specific genes in those lists of genes categorized as Pattern A, B, or C or genes associated with AI within each brain region. Only comparisons reaching significance after multiple comparison correction (FDR) are presented.

Table S8: Transcriptome-wide associations with Factors 1 – 8. A complete list of the associations and p-values for each gene and Factor across all brain regions is presented. Each brain region is provided on a separate tab.

SUPPLEMENTAL FIGURES

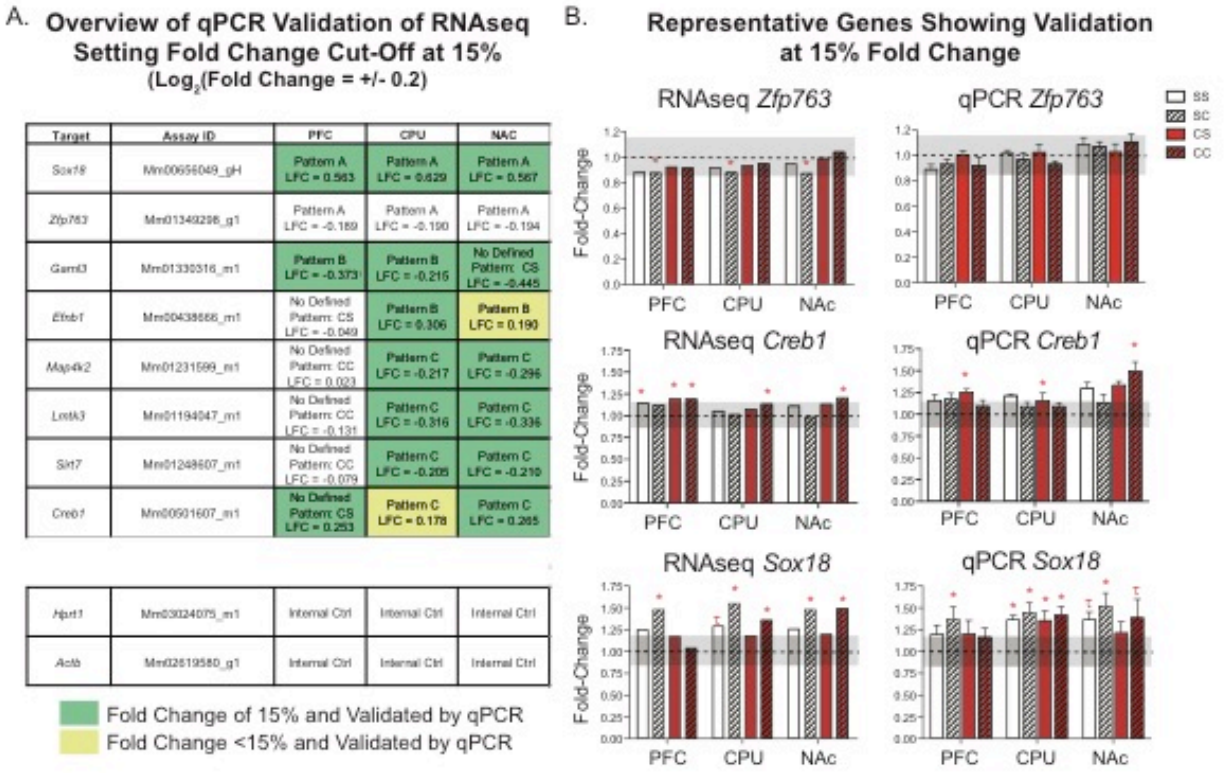


Figure S1: qPCR validation of Patterns in three brain regions reveals that fold changes of at least 15% are replicable. (A) List of 8 genes categorized as Patterns A, B, or C were validated using qPCR on technical replicates of the samples used in the RNA-seq experiment. Only those genes with a fold change of at least 15% were validated. (B-D) Expression of representative transcripts measured by RNA-seq and qPCR. Changes in expression of at least 15% in the RNA-seq data were validated by qPCR. This is exemplified by those changes in *Zfp763* (categorized as Pattern A but with <15% change in expression); *Sox18* and *Creb1* (Categorized as Pattern A or C, respectively with >15% change in expression). Gray shaded area on graphs indicates 15% change from S24. * = $p<0.05$

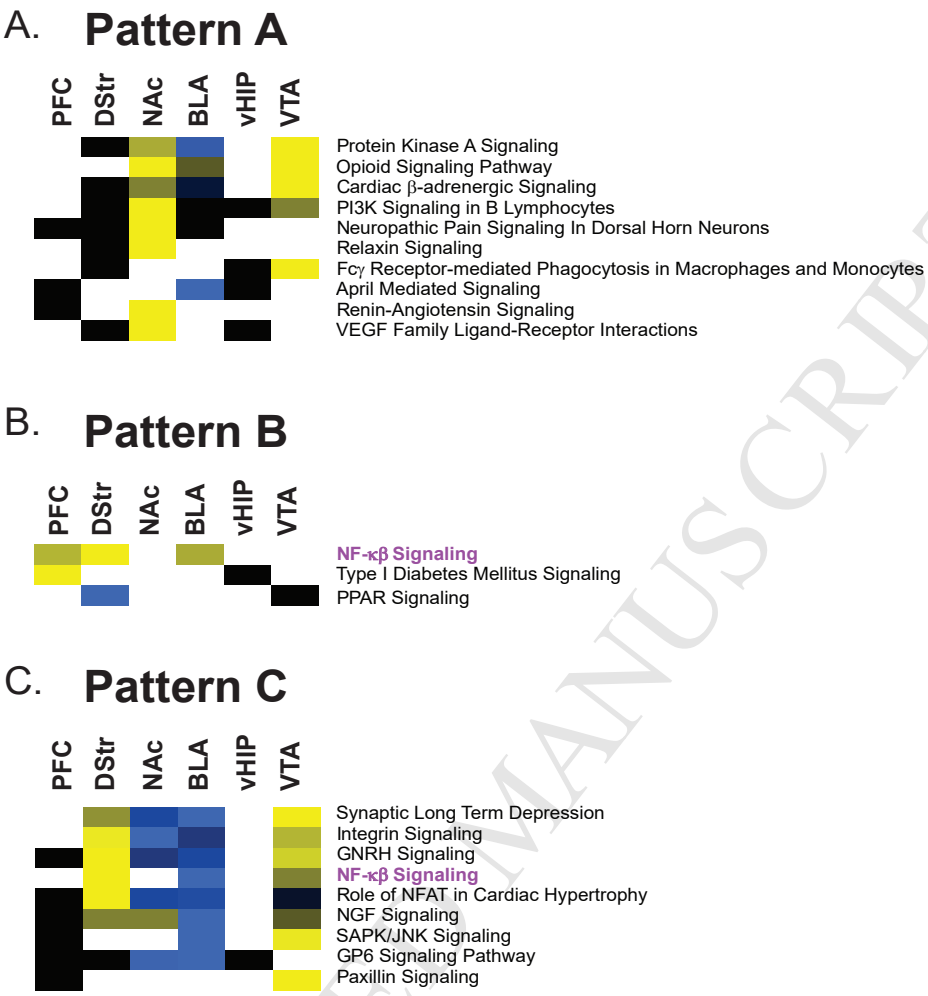
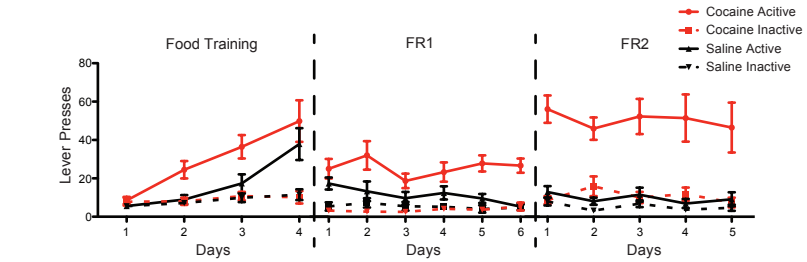


Figure S2: Similar pathways are associated with Patterns of gene expression across brain regions. (A) Pattern A was associated with protein kinase A signaling, while (B) Pattern B was dominated by NFκB and PPAR, a nuclear receptor, signaling. (C) Pathways associated with Pattern C included synaptic long-term depression and NFκB signaling. Pathways associated with both Patterns B and C are highlighted in purple. Only those pathways that met the following criteria were included: at least 1 brain region with an activation z-score>2 and p-value<0.01. Activation z-Scores: positive (yellow) = overrepresentation of targets activated in pathway; negative (blue) = overrepresentation of targets repressed in pathway; no direction (black) = no significant enrichment of activated or repressed targets; white = not a predicted pathway.



B Factor Loading for Behavioral Endpoints

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
	Total Intake	3.59	-0.40	-0.02	-0.03	-0.03	0.00	-0.01	-0.01
	Avg Intake/day	1.87	-0.16	0.07	-0.06	0.03	0.07	-0.01	-0.09
	Intake (Y/N)	0.49	-0.06	-0.01	-0.00	-0.01	-0.01	-0.01	0.01
	Total Infusions	1.27	0.87	-0.01	0.09	-0.09	-0.03	0.15	0.05
	Total Days SA	0.03	-0.14	-0.06	0.03	-0.02	-0.04	0.06	0.10
Total Lever Presses	Active: Food	0.65	0.92	0.19	0.19	-0.57	0.13	-0.06	0.02
	Active: FR1	1.06	0.85	-0.21	-0.21	0.05	-0.01	0.08	0.17
	Active: FR2	1.67	0.97	0.34	0.60	0.22	0.01	0.06	-0.06
	Inactive: Food	0.20	0.45	-0.20	0.37	-0.46	0.65	-0.02	0.06
	Inactive: FR1	0.09	0.38	-1.23	0.33	-0.25	-0.27	0.30	-0.12
	Inactive: FR2	0.61	0.62	-0.85	0.73	0.19	-0.25	-0.88	0.09
% Active Lever Presses	Food	0.06	0.06	0.05	-0.03	-0.02	-0.07	0.00	0.00
	FR1	0.09	0.06	0.10	-0.05	0.02	-0.01	-0.02	0.04
	FR2	0.09	0.05	0.11	-0.02	-0.01	0.02	0.07	-0.01
Lever Presses/Day	Active: Food	0.61	0.71	-0.05	0.21	-0.56	0.11	0.01	0.19
	Active: FR1	0.89	0.97	-0.07	-0.26	0.10	0.07	-0.04	-0.06
	Active: FR2	1.49	0.72	0.27	0.53	0.22	-0.02	0.10	0.01
	Inactive: FR1	0.02	0.49	-0.83	0.22	-0.13	-0.16	0.16	-0.26
	Inactive: FR2	0.48	0.40	-0.70	0.55	0.20	-0.23	-0.65	0.13
	Grand Mean FR1	0.77	1.03	-0.15	-0.40	0.10	0.18	-0.08	0.03
	Grand Mean FR2	1.50	0.75	0.27	0.54	0.24	-0.03	0.09	0.01
	Consummatory Regulation	0.35	-0.33	0.24	0.72	0.07	-0.14	-0.05	0.04

Figure S3: Factor loading for behavioral endpoints used in factor analysis. (A) Behavioral data represented in the factor analysis. All lever pressing data (food training, FR1, FR2; active vs. inactive) were included as variables in the factor analysis. Here we present a subset of the data aligned to the first day of each phase of self-administration. Because all animals had differing numbers of days in each phase, only those days in which the majority (>70%) of the animals in the study are presented. An image of the complete data set is presented in Figure 1D. (B) Factor analysis was used to reduce multidimensional behavioral endpoints to factors. The association of each factor with each behavioral endpoint included in the analysis is displayed. Factors were positively (yellow), negatively (blue), or not associated (black) with each endpoint. These particular associations allowed for the interpretation of the how each factor related to various SA behaviors.

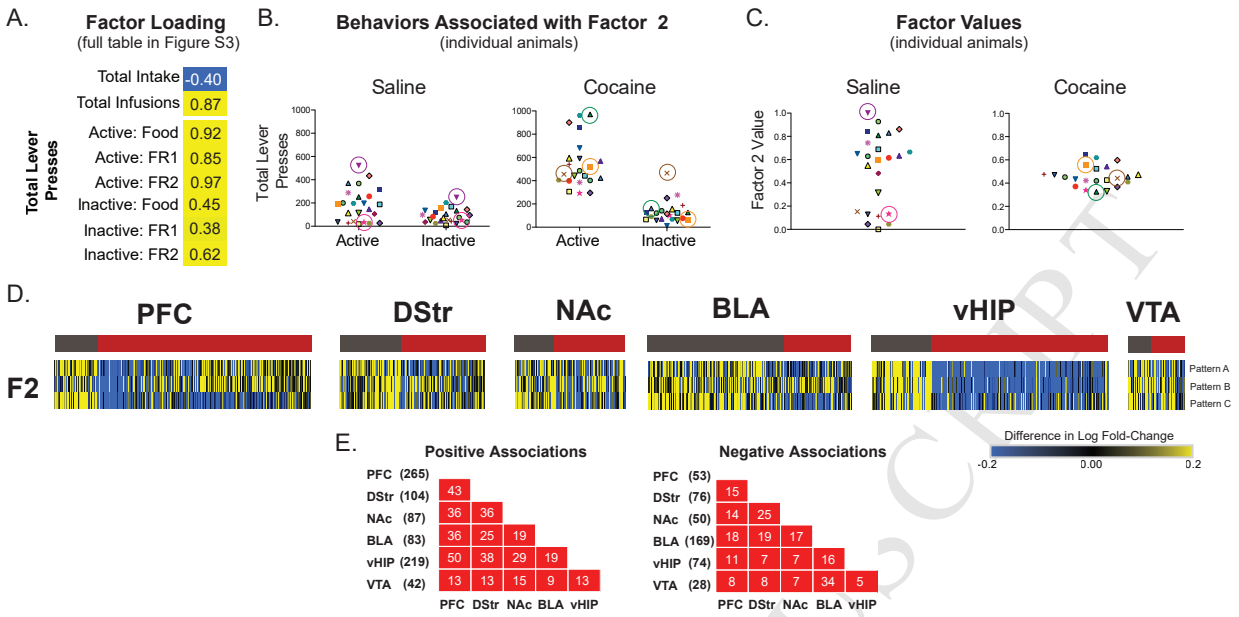


Figure S4: Factor 2 discriminates between baseline differences in saline animals. (A) Factor 2, in the factor analysis, was positively associated with both active and inactive lever pressing and negatively associated with intake. (B) Individual data for total number of lever presses in saline (left) and cocaine (right) for the entire SA experiment, including food training. (C) Individual factor values for Factor 2 for saline (left) or cocaine (right) animals. Animals with the greatest number of lever presses, but no intake, had highest factor value (▼in saline group). Animals with increased lever pressing coupled with high intake (▲in cocaine group) had lower factor values. Finally, those animals with few lever presses and no intake (★ in saline group) had the lowest factor values. (D) Linear modeling was used to identify genes associated with Factor 2 within each brain region. Only genes with a |slope|>0.2 and a nominal p-value of <0.05 were investigated. (D) Genes were ranked by -log p-value signed by the slope of the association with Factor 2. Negative associations with Factor 2 are presented in gray and genes positively associated with Factor 2 are presented in red. (D) Heatmaps presented are transformed to indicate change in expression from SS controls. Blue = fold change in the negative direction from SS vs S24 and yellow = fold change in the positive direction from SS vs S24. These data indicate that changes in expression in transcripts associated with Factor 2 are most robust in the SS vs S24. This highlights the power of factor analysis to extract important information related to baseline behaviors and indicates that those differences are reflected in our transcriptomic data as well. (E) Overlap of genes positively (left) or negatively (right) associated with Factor 2 across brain regions, color-coded for significance. Total number of genes in each brain region listed in

parentheses and total number of genes overlapping between regions indicated in corresponding boxes. There is a high degree of overlap of transcripts associated with Factor 2 in all brain regions.

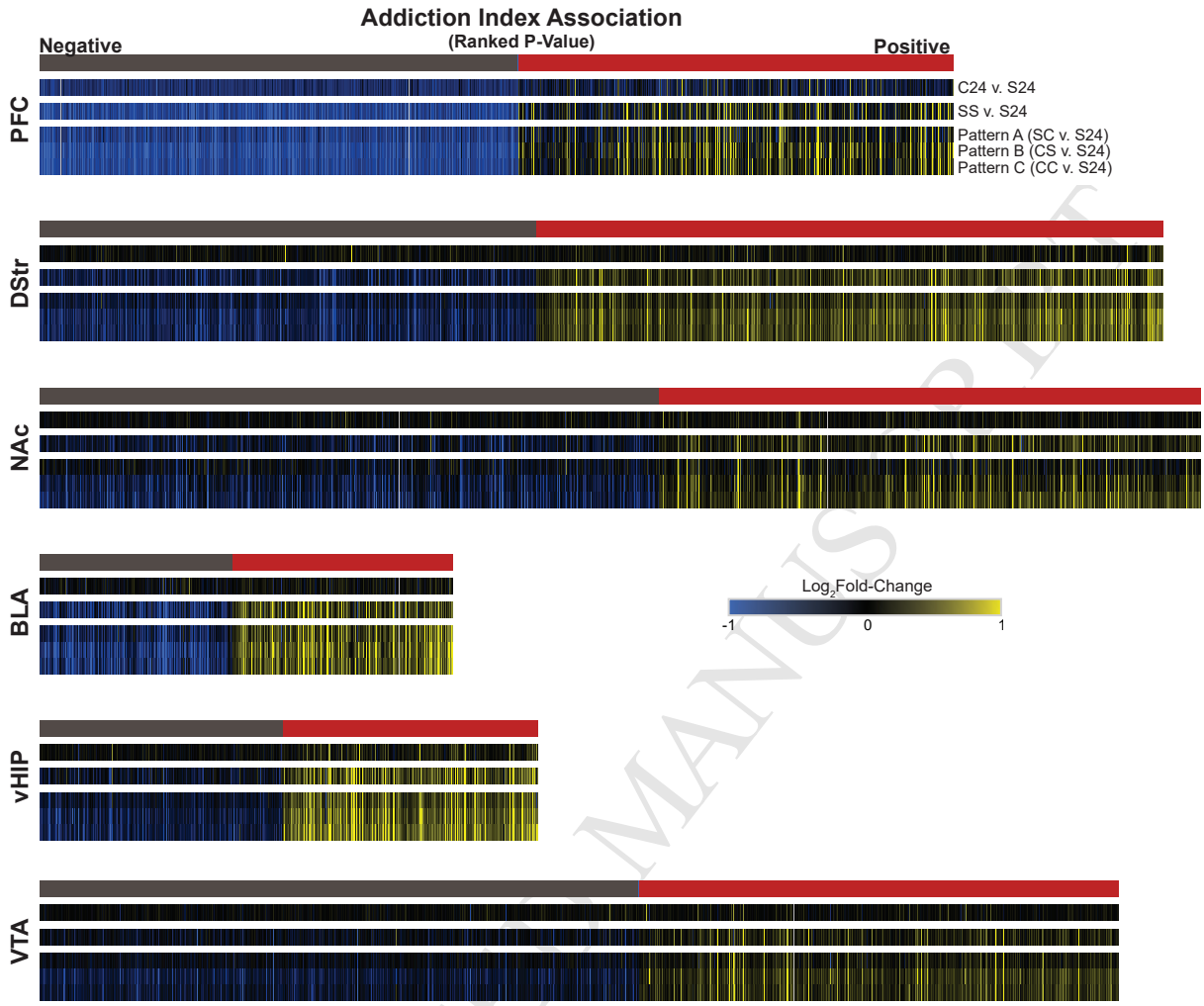


Figure S5: Raw heatmap of addiction index associated genes. (A-F) Raw expression of genes associated with AI in all brain regions for all groups when compared to the same baseline (S24). Log fold-change in expression of genes associated with AI and ranked by the sign of the association and -log(p-value) (gray = negative associations; red = positive associations). In all groups but C24, genes that were negatively associated with AI (gray bar) were downregulated and genes positively associated with AI (red bar) were upregulated. In all brain regions, the strongest response was in comparisons representing either Pattern B or C, suggesting that that transcriptional response to re-exposure to context/cocaine is influenced by addiction-related behaviors during cocaine SA.

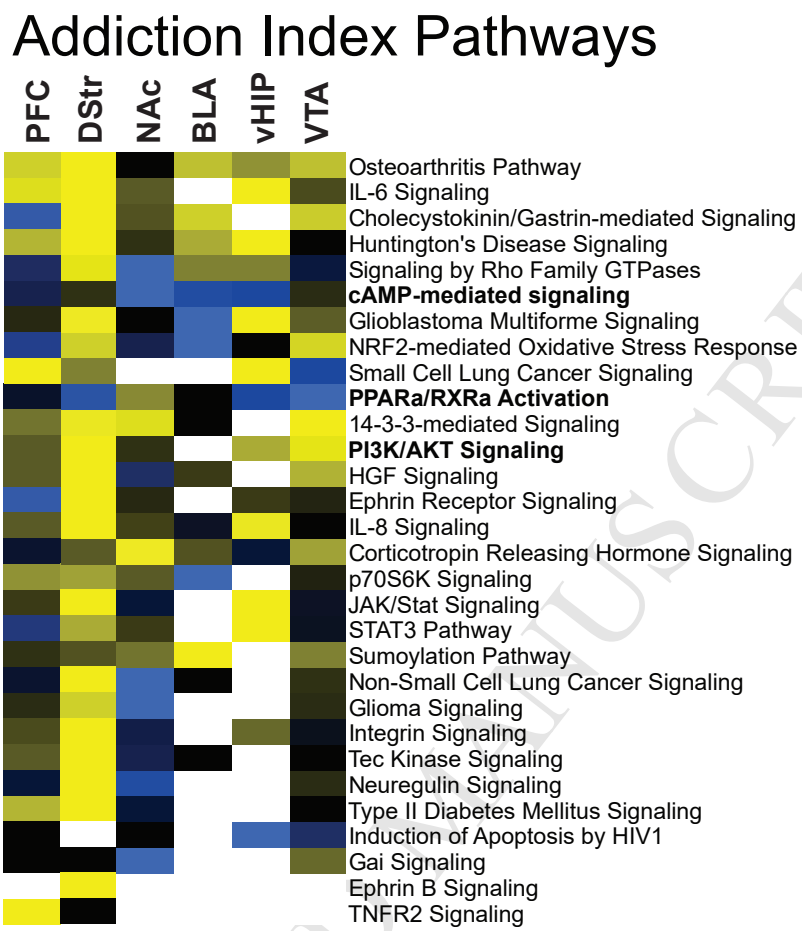


Figure S6: Pathways associated with the addiction index (AI). Ingenuity Pathway Analysis revealed genes associated with the AI, which were enriched for cAMP-mediated signaling, PPARα/RXRα activation, and PI3K/AKT signaling among others. Activation z-Scores: positive (yellow) = overrepresentation of targets activated by regulator; negative (blue) = overrepresentation of targets repressed by regulator; no direction (black) = no significant enrichment of activated or repressed targets; white = not a predicted upstream regulator. Behavioral data analyzed using Kruskal-Wallis followed by Mann-Whitney Nonparametric Test; *p<0.05; **p<0.001; data presented as mean ± SEM.

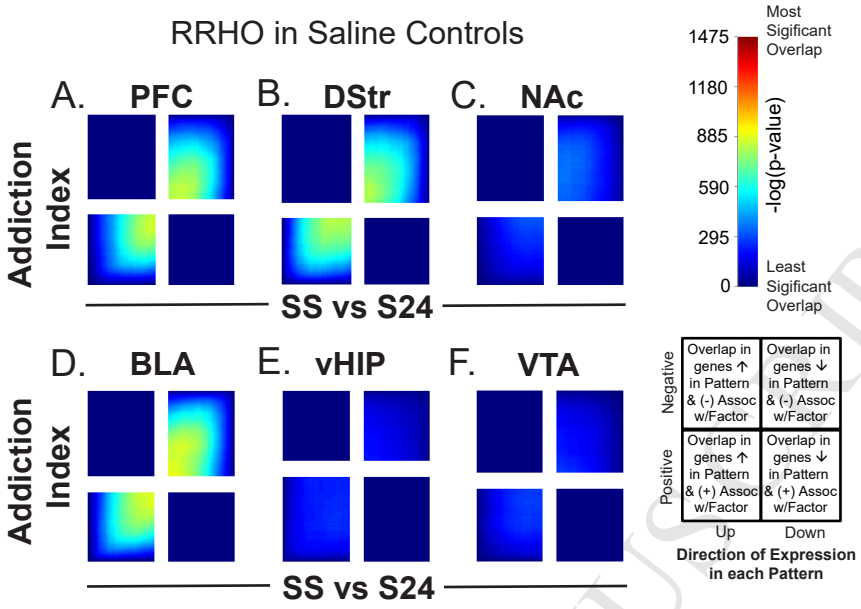


Figure S7: Overlap of transcriptional profiles related to the AI and saline controls. (A-F) RRHO plots reveal little overlap of genes positively or negatively associated with AI and up- or downregulated in saline control animals (SS vs S24). As predicted, little to no overlap of expression profiles was observed in NAc, vHIP and VTA. Overlap of expression was weak in PFC, DStr and BLA and similar to that observed in the comparisons with Pattern A. A key for these plots is provided.

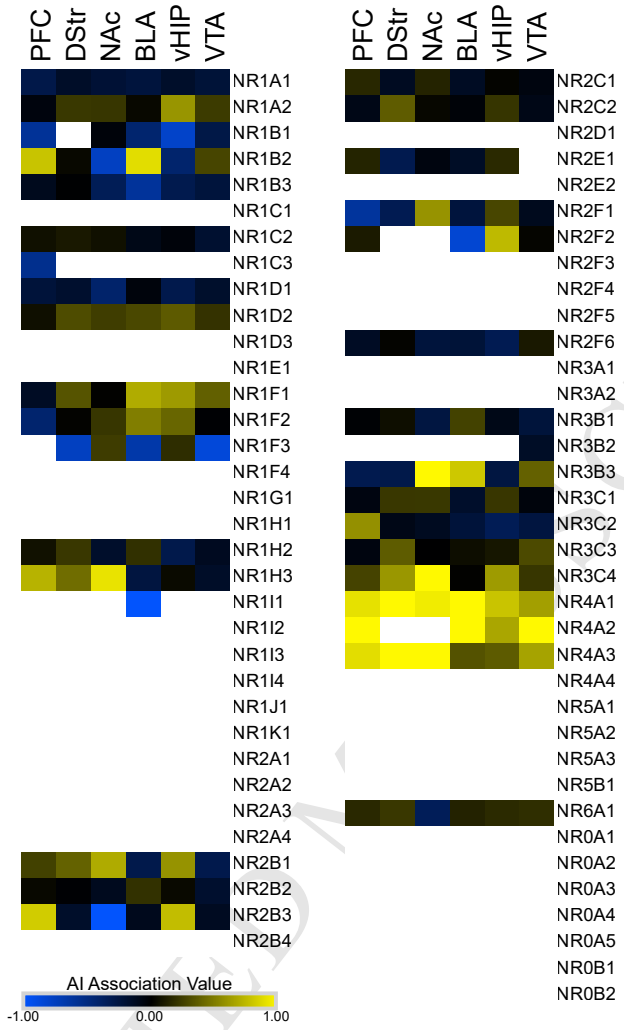


Figure S8: Full list of NR family members associated with the AI. Heatmap of association of all known nuclear receptors with the AI. Members of NR1 – 4 subfamilies are expressed throughout the reward circuitry. Strongest associations are found within NR2B and NR4A subfamilies. Yellow = positive association; Blue = negative association; Black = no significant association; white = expression not detected in our dataset.

SUPPLEMENTAL METHODS

Animals

In order to identify gene expression changes arising from environmental exposures, independent of genome sequence variation, genetically identical male C57BL/6J mice (6-8 wk-old) were used. Due to limitations of the study (e.g., number of operant boxes, number of animals, number of groups), we focused on males to limit the number of cohorts required. Male mice weighing 20-24 g were maintained on a 12 hr reverse light-dark cycle (lights on at 19:00) at 22-25°C with *ad libitum* access to food and water, except during training and testing when access to food was restricted. During self-administration testing mice were food restricted to 95% of their free-feeding weight. Mice were housed 5 per cage prior to jugular vein catheterization surgeries, at which point mice were housed individually. Following SA, those animals included in the withdrawal (WD) groups were rehoused with their original cage mates for the remainder of the experiment with *ad libitum* access to food and water.

Training, Surgery and Self-Administration

Food Training: Following 7-10 d of acclimation in the animal facility, mice were trained initially (3-10 d) for food reinforcement in standard operant chambers (Med Associates, St Albans, USA) equipped with 2 retracting levers (active and inactive), a cue light, and a house light. Animals were placed in operant chambers and illumination of the house light and extension of the levers signaled the beginning of the self-administration session. Active lever presses resulted in food reinforcer delivery followed by a 20 sec time-out period during which a cue light was illuminated and levers were retracted. Responding on the inactive lever was recorded, but resulted in no programmed consequence. Responding on the active lever was reinforced on a fixed-ratio one (FR1) schedule. Animals were considered to have acquired when they exhibited stable responding on the active lever (60% active/total lever presses) and >10 lever presses per 1 hr

session on an FR1 schedule of reinforcement. Once the animals met acquisition criteria, most were moved onto an FR5 schedule to further confirm acquisition of the task.

Cocaine Self-Administration: Following food training, mice were implanted with a jugular catheter (0.3 mm inner and 0.6mm outer diameter) under ketamine (100 mg/kg IP)-xylazine (10 mg/kg IP) anesthesia. Mice were administered MediGel® CPF containing carprofen (5 mg/kg) 1 d pre-op as an analgesic and intravenous ampicillin (0.5mg/kg) for infection prevention for 3 d post-op. In addition to standard chow, DietGel® Recovery (Westbrook ME) was provided to each mouse for 3 d post-op to aid in recovery. Mice were allowed to recover for 3-5 d before testing. Catheters were flushed daily with heparinized saline (10U/ml in 0.9% sterile saline) to ensure catheter patency. After recovery, mice began cocaine SA. For mice self-administering cocaine, active responses (FR1) resulted in a single (0.03ml) infusion of cocaine (0.5 mg/kg/infusion over 3.25 sec; cocaine HCL from the NIDA drug supply) and a discrete light cue was illuminated during the 20-s time-out period. Mice underwent 2 hr daily session for 10-15 d: 5-10 d on an FR1 schedule followed by 4-5 d of FR2 schedule. When animals self-administer drug on low effort schedules of reinforcement they defend a specific blood level of drug. Thus, in the case of changes in dose or FR requirement, animals will adjust responding to continue getting the same relative amount of drug (1) – referred to herein as “consummatory regulation.” In order to confirm that animals were in fact being reinforced by cocaine the FR requirement was increased. As predicted, animals assigned to cocaine SA (n=22), but not saline (n=24), pressed the active over inactive lever throughout the FR1 and FR2 phases (Figure 1D; corrected $p<0.05$). Behavior is aligned to all animals’ first FR2 day in graphs, thus saline extinction is not easily observed.

The experiment was phased such that all six groups of mice were the same age at the time of euthanasia. Thus, animals were run in 2 cohorts. The first cohort was rehoused with their original cage mates and exposed to WD/forced abstinence for 30 d following their final trial. After 30 d of WD/forced abstinence, mice were given an IP injection of either cocaine (10 mg/kg) or

saline, placed back in their original operant chamber with house light illuminated; however, the levers were not extended. Animals were euthanized via cervical dislocation 1 hr after injection. All mice in cohort 1 were given saline injections (IP) for seven days prior to euthanasia to reduce stress in response to handling and injection. The second cohort was euthanized 24 hr after the final SA trial to assess the transcriptional alterations that occur following short-term WD (24 hr). Because all animals were socially isolated during food training and self-administration, cohort 2 was euthanized after prolonged social isolation. Small but significant differences in behavior were observed between the 2 cohorts, which most likely reflect slight differences in training paradigms (Figure 1A & D).

RNA Isolation, Library Preparation, and Sequencing

For all groups, brains were removed and sectioned on ice in a brain block (1 mm thick) and micropunches of six brain regions (PFC, NAc, DStr, vHIP, BLA, and VTA) were snap frozen on dry ice and stored at -80°C until use.

RNA was isolated as previously described (2) using RNAeasy Mini Kit (Qiagen, Fredrick, MD) using a modified protocol from the manufacturer allowing for the separation and purification of small RNAs from total RNA. Briefly, after cell lysis and extraction with QIAzol (Qiagen, Fredrick, MD), small RNAs were collected in the flow-through and purified using the RNeasy MinElute spin columns and total RNA was purified using RNeasy Mini spin columns. Samples were treated with DNAse to rid samples of genomic DNA and run on nanodrop and an Agilent Bioanalyzer 2100 to confirm RNA purity, integrity, and concentration. All samples' RIN>8.

Libraries were prepared using the TruSeq Stranded mRNA HT Sample Prep Kit protocol (Illumina, San Diego, CA). Briefly, poly A selection and fragmentation of 300 ng of RNA was converted to cDNA with random hexamers. Adapters were ligated and samples were size-selected with AMPur XP beads (Beckman Coulter, Brea, CA). Barcode bases (6 bp) were introduced at one end of the adaptors during PCR amplification steps. Library size and

concentration was assessed using Tape Station (Life Technologies, Grand Island, NY) before sequencing. Libraries were pooled for multiplexing (4 pools of ~60 samples with each group and brain region equally represented across each pool) and sequenced on a HighSeq2500 System using V4 chemistry with 50 base pair single-end reads at GeneWiz LLC (South Plainfield, NJ). Each pool was sequenced 8 times with the goal of obtaining ~25 million reads per sample. Initial quality control assessments revealed 43 samples, which did not meet standards for read depth and were excluded from analysis. Therefore, the final number of samples included in the analysis were between 5 – 8 per group apart from the CS group in VTA (N = 3).

qPCR Validation

Technical replicates were used to validate Patterns of expression across three brain regions. RNA (500 ng) from PFC, DStr, and NAc used for RNA-seq was converted to cDNA using High Capacity Reverse Transcriptase Kits (Catalog #: 4368814; ThermoFisher, Foster City, CA) according to manufacturer's protocol. qPCR was run for 8 genes of interest and 2 internal controls (Supplemental Figure S1) using Taqman® gene expression assays (Supplemental Figure1A) and Taqman® Fast Universal Master Mix (Catalog #: 4444964; ThermoFisher, Foster City, CA) on an ABI Quant Studio Flex 7 according to the manufacturer's protocol. Six plates were run for each brain region using the following run parameters: 1 cycle (2 min @ 50°C followed by 2 min @ 95°C); 45 cycles (1 sec @ 95°C followed by 20 sec @ 60°C). Expression within each brain region was analyzed using the comparative Ct method (3). Each sample was normalized to its own internal controls (geometric mean of the Ct values for *Hprt1* and *Actb)* and calibrated to the average ΔCt for the S24 groups. In order to replicate the pair-wise differential expression analysis used for RNA-seq data, a Student's t-test was used to identify genes significantly different from S24.

Statistical and Bioinformatic Analyses

Behavior: Lever-pressing behavior and infusions were analyzed using a Kruskal-Wallis non-parametric test followed by Mann-Whitney Test to identify differences at individual time-points, treatments, or levers (cocaine vs saline; active vs inactive). Other behaviors were analyzed using ANOVA or Kruskal-Wallis tests depending on homozygosity of variance. All analyses were conducted using SPSS Statistical Software, V24 (IBM Analytics, Armonk, NY). To account for malfunctions in the operant chambers during SA sessions (e.g., broken tubing, stuck levers, etc.) we calculated the moving average of lever presses for the first 5 d of FR1 and the last 5 d of FR2 (averaged 3 d together each time). We then subtracted the grand mean of the FR1 moving average from the FR2 moving average as an indicator of consummatory regulation, which was included as a variable in the factor analysis.

Differential Expression Analysis: Sequencing short reads were aligned to the mouse mm10 genome using Tophat2 (4). QC analysis revealed a range of 18-60 million reads per sample with an average mapping rate of 90.2%. Read counts were generated using HtSeq-count against the Encode vM4 annotation. Stochastic outlier selection (5) was utilized to identify outliers prior to differential expression analysis. Samples with an outlier probability of >90% were excluded from analysis (4 samples out of 235 or 1.7%). Three of these belonged to one animal in which 4 of the 6 the brain regions investigated were predicted outliers; therefore, the entire animal was excluded from analysis. Data were filtered for low abundance transcripts by keeping only genes with more than 1 RPKM in at least 80% of samples per group. After filtering, pair-wise differential expression comparisons using Voom Limma were performed (6) and a nominal significance threshold of fold change>1.3 and $p<0.05$ was applied.

Pattern Analysis: Each Pattern included genes that were differentially expressed from S24 ($p<0.05$; fold change>15%) and also different from all other groups. For example, a gene that is

significantly increased in all groups compared to S24 and is further up-regulated by cocaine re-exposure is categorized as Pattern C. Importantly, even when genes are responsive to other stimuli, they are only categorized within Patterns A-C if the magnitude of change is greatest in that Pattern when compared to all other groups. Thus, we identified genes that are uniquely regulated by each stimulus in each brain region. Figure 2 highlights the fact that re-exposure to context alters expression of many genes in the same direction, but suggests that the magnitude of this changes is dependent on both a history of cocaine SA and re-exposure to context/cocaine.

Factor Analysis and Linear Modeling: Factor analysis was used to reduce the dimensions of the interdependent behavioral variables and help account for variability in the data due to differences in training, cohorts, and malfunctions in the operant chambers. All animals were included in the analysis. All behavioral measurements were first shifted to convert all data to non-negative values followed by log2(x+1) transformation. For “total intake”, an additional variable referred to as “intake or not” was included to indicate whether total intake>0. This accounted for the lack of cocaine intake in the saline groups. A standard factor analysis was performed using the scikit-learn package (7). A 10-fold cross-validation (CV) was utilized to choose the number of factors. We found that the CV log-likelihood was maximized with 8 factors. Therefore, the factor number was set to 8 when factor analysis was then applied to the whole dataset. The transformed data from the analysis was then used as a continuous variable for each factor. Differential analysis was conducted using Voom Limma to determine which factors were associated with gene expression (6).

Factor loading (Supplemental Figure S3) revealed three factors associated with the addicted-like phenotype. Factor 1 was positively associated with intake/infusions and was the Factor that most robustly discriminated between saline versus cocaine SA. Factor 3 was positively associated with active lever pressing and negatively associated with inactive lever pressing,

suggesting that Factor 3 is associated with an animal's ability to identify the reward-paired lever. Factor 4 was positively associated with active lever presses on FR2 and negatively associated with active lever presses on FR1. Factor 2 on the other hand, was positively associated with lever pressing (both inactive and active) and negatively associated with intake. We interpret this as reflecting baseline differences in behavior within our saline groups. Factors 5 – 8 were weakly associated with behaviors and were excluded from further investigation. A full list of transcripts and their associations with each factor are included in Supplemental Table S8.

Generation of an Addiction Index (AI): A composite score, or "addiction index," of the three factors most strongly associated with an addictive phenotype was generated. This allowed us to identify animals with high performance scores across multiple behavioral endpoints associated with addiction and resulted in a continuous variable which could be used to identify genes that were positively or negatively associated with those behavioral endpoints. To calculate the index, factor values were linearly transformed to eliminate negative values. The transformation resulted in values that ranged from 0-1 for each factor: [(individual value – minimum value)/(maximum – minimum value)]. The product of the transformed factors was calculated for each individual. Individual AI values as well as the transformed values for each factor are presented in Figure 4. As indicated, animals with high performance in all three factors have the highest AI but those animals with lower performance on any one factor have a reduced AI.

Enrichment Analysis: Fisher's exact tests were conducted using the Super Exact Test package in R as previously described (8).

Cell-type Enrichment Analysis: Enrichment for cell types were determined as previously described (9). Briefly, we used the Super Exact Test R Package (8) to evaluate statistical overlap between

our differential expression lists and genes expressed at least five times greater in one cell-type than in any other cell type in an established transcriptome study from cortical cells (10).

Rank Rank Hypergeometric Overlap (RRHO) Analysis: We applied an RRHO test to compare gene regulation between the comparisons representing each Pattern (e.g., Pattern A = differential expression between SC vs S24; Pattern B = differential expression between CS vs S24; etc.) and genes associated with the addiction index. RRHO identifies overlap between expression profiles in a threshold free manner to assess the degree and significance of overlap (11). Here we used a modified script that visualizes both positive and negative correlations and illustrates each quadrant separately based on the number of genes in each comparison as previously described (12). Full differential expression or association (Factors) lists were ranked by the -log(p-value) multiplied by the sign of the fold change/slope of association. A one sided version of the test was used to look for over enrichment. RRHO difference maps were produced for each comparison by calculating for each pixel the normal approximation of difference in log odds ratio and standard error of overlap between the comparison representing the Pattern and the Factor. This z-score was then converted to a p-value and corrected for multiple comparisons across pixels (13).

Upstream Regulator and Pathway Analysis: Predicted upstream regulators and molecular pathways were identified using Ingenuity Pathway Analysis (IPA) Software (Qiagen, Fredrick MD). These determinations were based on the log fold change of genes associated with each pattern ($p<0.05$; fold change>1.3) or factor ($p<0.05$) analyzed. Upstream regulators and pathways were filtered by activation z-score (>2) and p-value (<0.001) as well as molecule (genes and proteins).

SUPPLEMENTAL REFERENCES

1. Norman AB, Tsibulsky VL (2006): The compulsion zone: a pharmacological theory of acquired cocaine self-administration. *Brain Res*. 1116:143-152.
2. Labonte B, Engmann O, Purushothaman I, Menard C, Wang J, Tan C, et al. (2017): Sex-specific transcriptional signatures in human depression. *Nat Med*. 23:1102-1111.
3. Schmittgen TD, Livak KJ (2008): Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc*. 3:1101-1108.
4. Trapnell C, Pachter L, Salzberg SL (2009): TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 25:1105-1111.
5. Jeroen Janssens FH, Eric Postma, Jaap van den Herik (2012): Stochastic Outlier Selection. *Tilburg Centre for Creative Computing*.
6. Law CW, Chen Y, Shi W, Smyth GK (2014): voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 15:R29.
7. Pedregosa F, Varoquaux G, Gramfort A, Michel M, Thirion B, Grisel O, et al. (2011): Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*. 12:2825-2830.
8. Wang M, Zhao Y, Zhang B (2015): Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep*. 5:16923.
9. Bagot RC, Cates HM, Purushothaman I, Lorsch ZS, Walker DM, Wang J, et al. (2016): Circuit-wide Transcriptional Profiling Reveals Brain Region-Specific Gene Networks Regulating Depression Susceptibility. *Neuron*. 90:969-983.
10. Zhang Y, Chen K, Sloan SA, Bennett ML, Scholze AR, O'Keeffe S, et al. (2014): An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J Neurosci*. 34:11929-11947.
11. Plaisier S.B. TR, Wong J.A., Graeber T.G. (2010): Rank-rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Res*. 38:169.
12. Seney ML, Huo Z, Cahill K, French L, Puralewski R, Zhang J, et al. Opposite molecular signatures of depression in men and women. *Biological Psychiatry*.
13. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001): Controlling the false discovery rate in behavior genetics research. *Behav Brain Res*. 125:279-284.