# Sex-Specific Transcriptional Changes in Response to Adolescent Social Stress in the Brain's Reward Circuitry

***Supplement 1***

## Supplemental Figures:

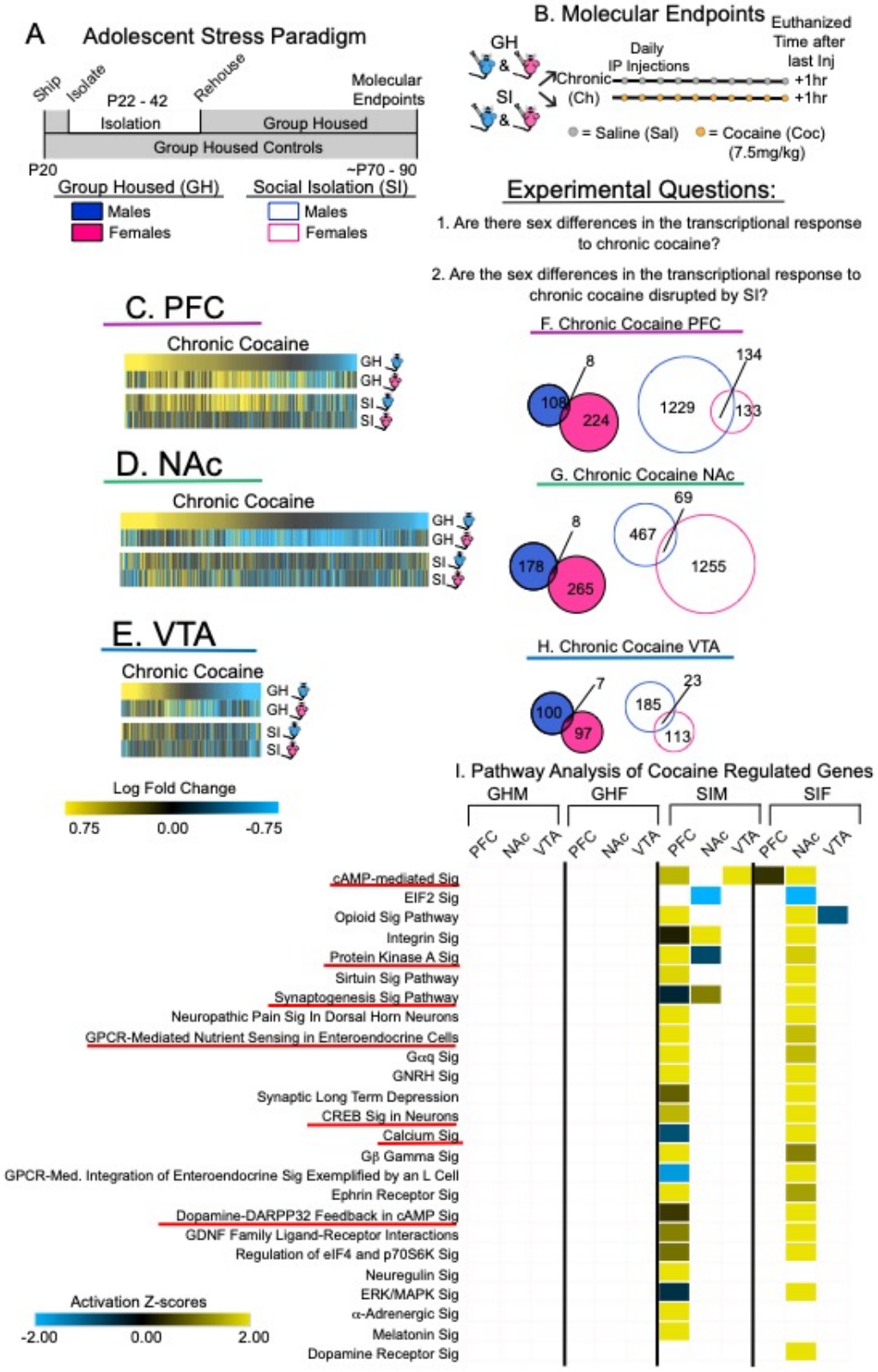


**Figure S1: Sex-specific transcriptional responses to chronic cocaine are disrupted by SI throughout the reward circuitry.**

**(A-B)** Schematic of experimental design. Heatmaps show DEGs with a nominal $p \leq 0.05$ and a $\log_2$ fold change (FC)$\geq 2$ in response to chronic cocaine (CH Coc) in the **(C)** PFC, (**D)** NAc, and (**E)** VTA. Heatmaps are seeded by GHM DEGs. SIM expression resembles GHF expression after chronic cocaine (yellow = upregulated; blue = downregulated). Venn diagrams of DEGs in GH (closed circles) and (transparent circles) M/F after chronic cocaine, each compared to their own saline control in **(F)** PFC, (**G)** NAc, and (**H)** PFC. There was little overlap between GH males and females across all three brain regions, with more overlap seen in SI animals. **(I)** IPA shows intracellular and second messenger signaling, neurotransmitter, and other nervous system signaling pathways that were regulated by chronic cocaine but only in SI animals (activation z-scores $\geq 2.0$; p-value$\leq 0.05$; Yellow = predicted activated; blue = predicted inhibited; black = no predicted activation/inhibition; red lines = overlap with pathways regulated by acute cocaine in Figure 1; Sig = signaling) *p<0.05; **corrected p-value<0.05.

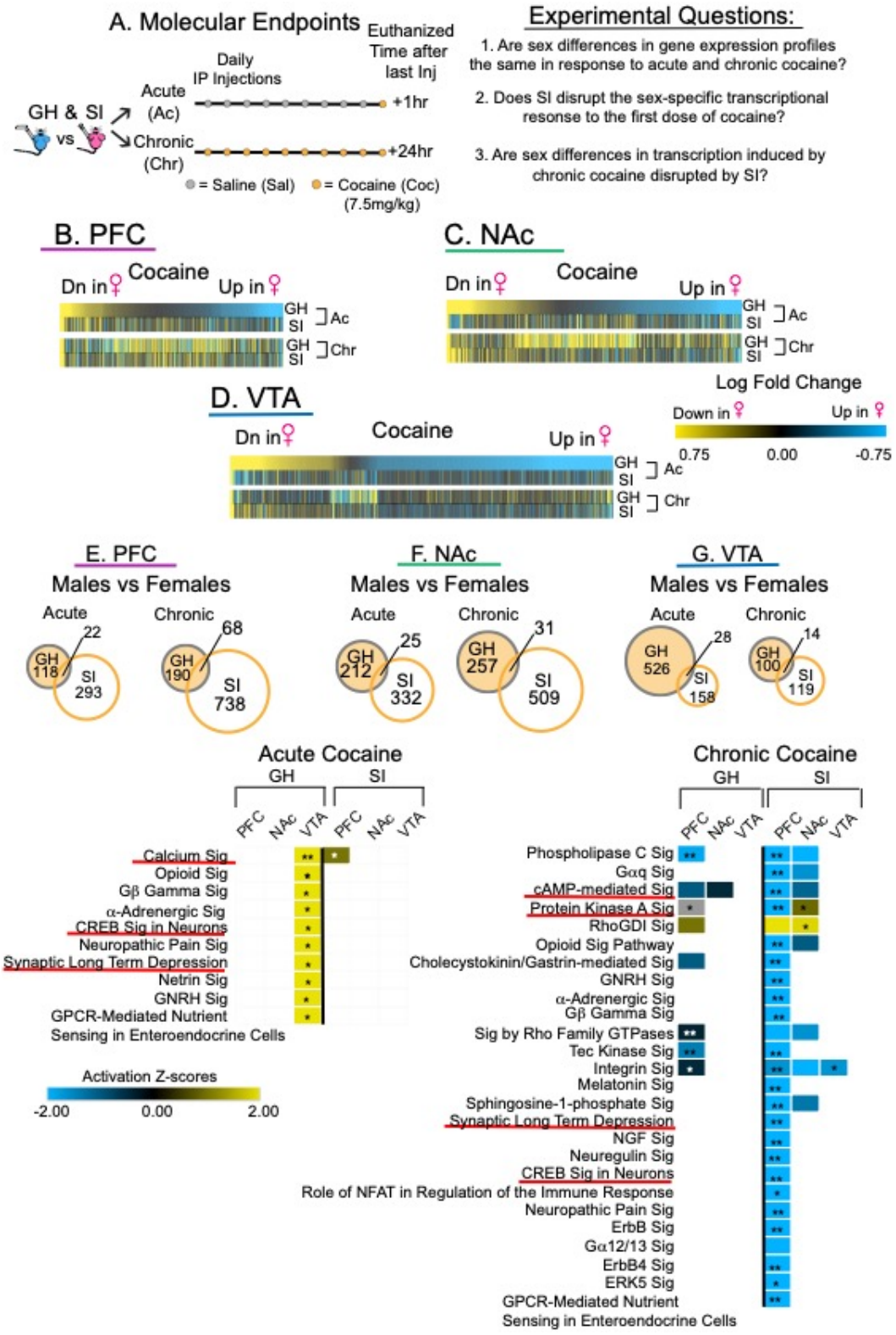



**Figure S2: Adolescent SI disrupts expression of baseline sex differences in the reward circuitry.**

**(A)** Schematic of experimental design. Heatmaps created using RNA-seq data from acute and chronic cocaine exposure of **(B)** PFC, **(C)** NAc, and **(D)** VTA. Blue represents genes upregulated in females, while yellow represents genes downregulated in females. Sex differences are lost with SI in all three brain regions. **(E-G)** Venn diagrams of sex differences in genes in GH (closed circles) and SI (open circles) animals. There is very little overlap of DEGs regulated by cocaine in both GH and SI males vs. females across all brain regions. IPA shows that pathways composed of sex-specific genes overlap with pathways affected by cocaine after experiencing **(H)** acute and **(I)** chronic saline (activation z-scores≥2.0; p-value≤0.05; Yellow = predicted activated; blue = predicted inhibited; black = no predicted activation/inhibition; red lines = overlap with pathways regulated by acute cocaine in Figure 1; Sig = signaling) *p<0.05; **corrected p-value<0.05.

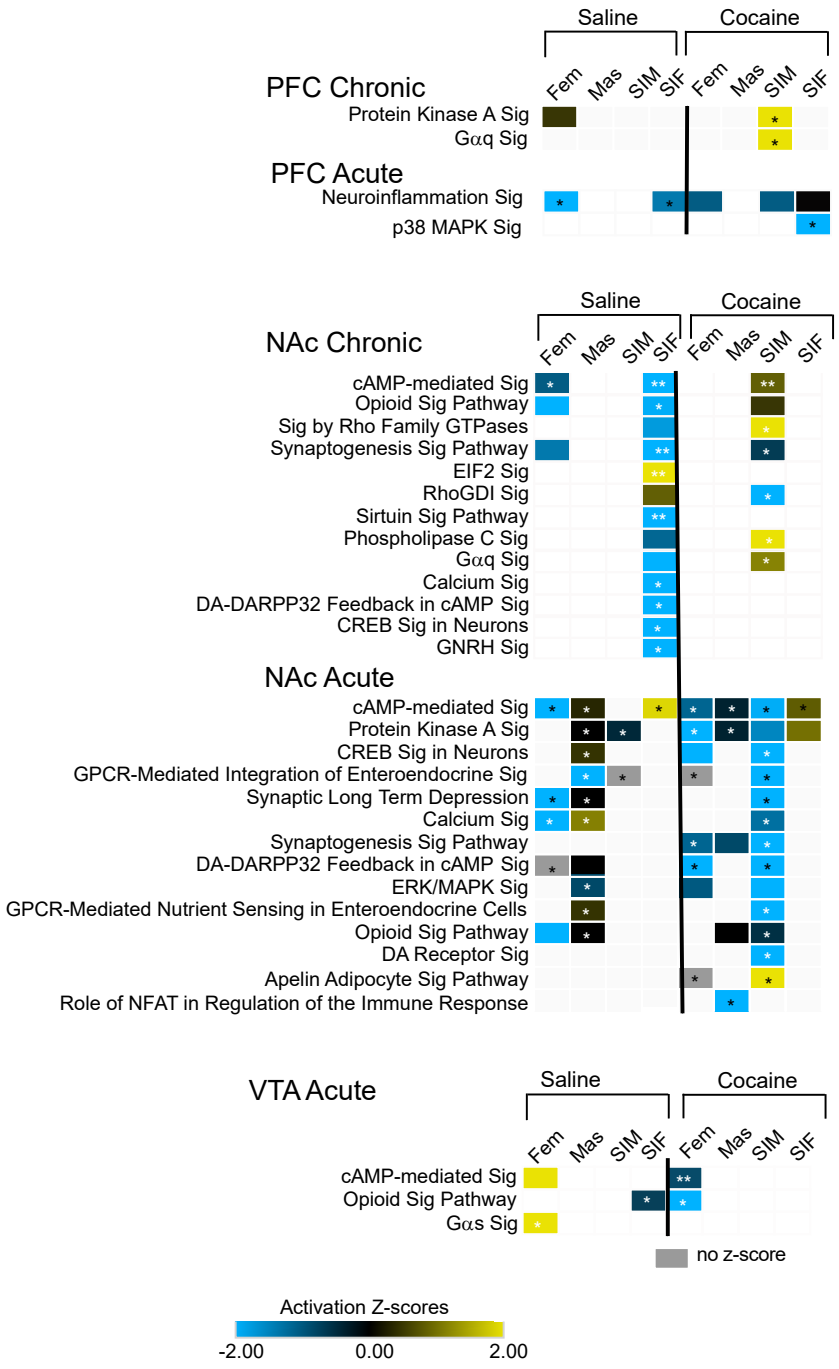



**Figure S3: Neuronal and second messenger signaling pathways associated with genes identified through pattern analysis.**

Complete list of the predicted second messenger and neuronal signaling pathways associated with the genes identified through pattern analysis in (A) PFC, (B) NAC and (C) VTA under all treatment conditions. (activation z-scores≥2.0; p-value≤ 0.05; yellow = predicted activated; blue

= predicted inhibited; black = no predicted activation/inhibition; Sig = signaling) *p<0.05; **corrected p-value<0.05.

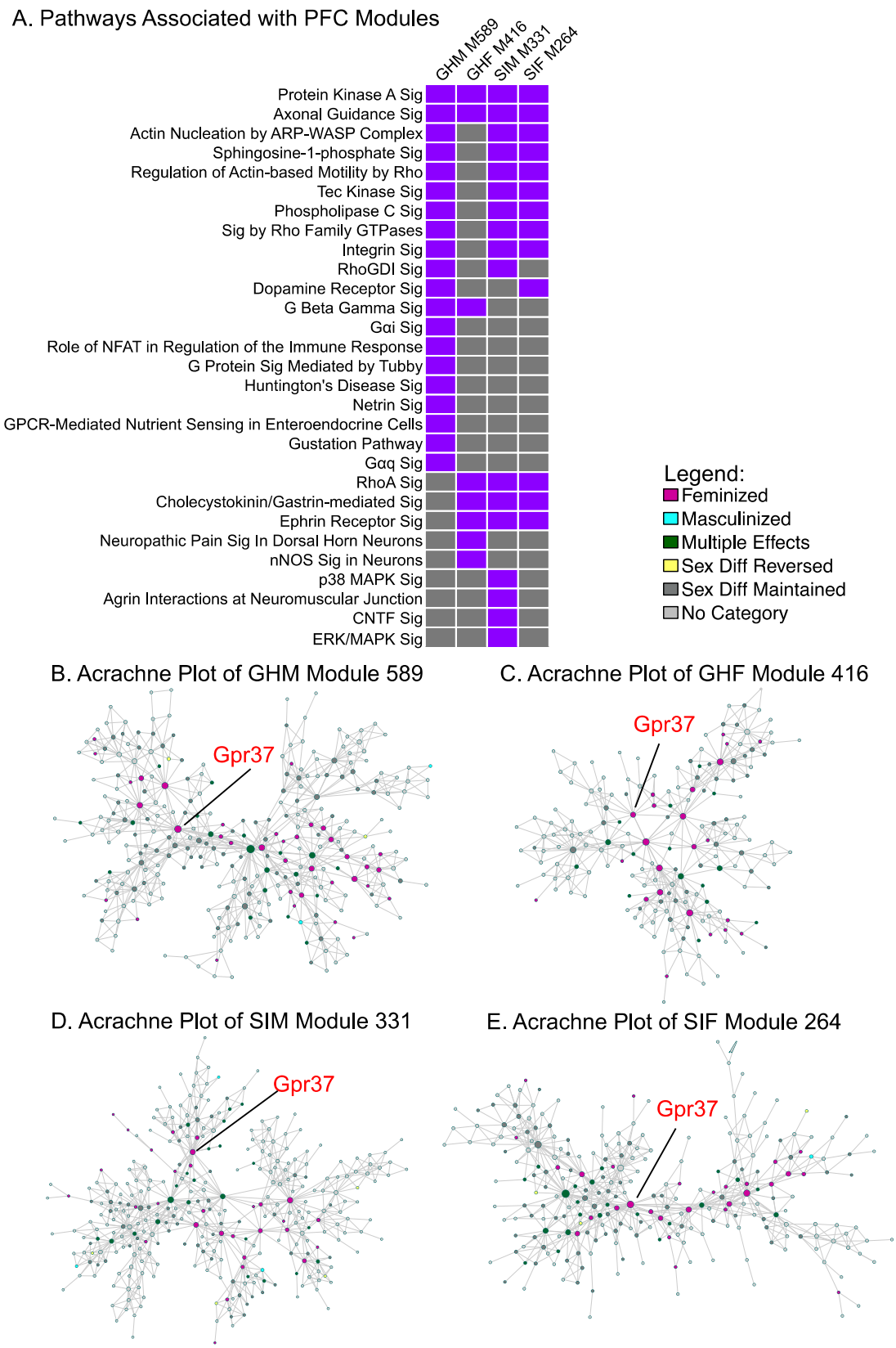



**Figure S4: Co-Expression analysis reveals Gpr37 as sex-specific hub in PFC**

**(A)** Enrichment of all pathways associated with the four gene modules in the PFC with Gpr37 as key driver. Plots show comparison of gene enrichment across modules. Significant enrichment is indicated by purple. (B-E) Arachne plots of gene modules with Gpr37 as a key driver. GHM (B) have more connections within the module than GHF (C) and this effect is lost in SI modules (D-E). Colors = genes defined by pattern analysis, multiple effects = genes that fell under more than one pattern in the 4 different treatment groups.

A. MDC for Children of GHF Parent Module 2

| | GHM | SIM | SIF |
|---|---|---|---|
| M4 | 2.95* | 1.31* | 0.66 |
| M5 | 5.49* | 4.78* | 3.61 |
| M6 | 3.52* | 1.41* | 0.71 |
| M7 | 3.19* | 1.46* | 0.65 |
| M8 | 3.46* | 1.67* | 0.73 |
| M9 | 2.81* | 1.26* | 0.58 |
| M10 | 2.01* | 1.31* | 0.48* |
| M11 | 1.81* | 1.09* | 0.5 |
| M12 | 1.44 | 0.64 | 0.32 |
| M13 | 2.41* | 0.64 | 0.46 |
| M14 | 3.38* | 1.88* | 0.81 |
| M15 | 2.69* | 1.4* | 0.75 |
| M16 | 1.61* | 0.79 | 0.38 |
| M17 | 1.85* | 0.68 | 0.33 |
| M18 | 1.45 | 0.75 | 0.37 |
| M19 | 2.98* | 1.17* | 0.46 |
| M20 | 3.31* | 1.58* | 0.82 |
| M21 | 2.1* | 0.92 | 0.4 |

B. Module Gene Enrichment

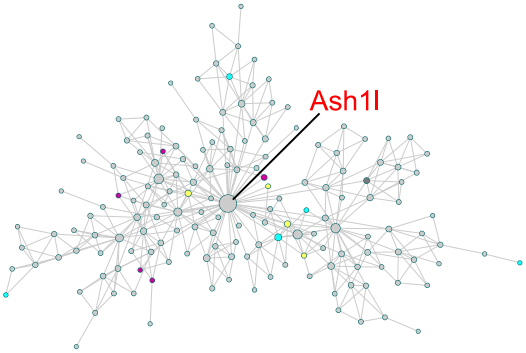
GHM Parent Module:

| GHF Parent Module: M3 (3830) | M2 (395) | M3 (3728) | M4 (1764) | M5 (2953) | M6 (1004) | M7 (637) | M8 (1294) | M9 (756) | M10 (253) |
|---|---|---|---|---|---|---|---|---|---|
| M15 (336) | 155 | 34 | 34 | 22 | 13 | 7 | 17 | 13 | 2 |
| M17 (696) | 7 | 350 | 42 | 83 | 23 | 40 | 24 | 10 | 6 |

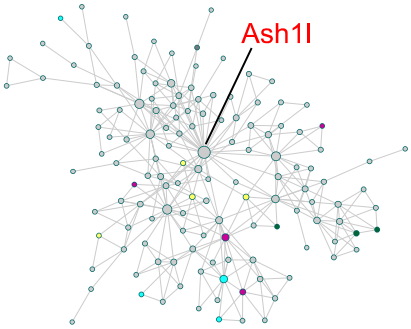
| GHF Parent Module: M3 (3830) | M11 (293) | M12 (199) | M13 (378) | M14 (498) | M16 (412) | M17 (614) | M18 (606) | M19 (365) | M20 (298) |
|---|---|---|---|---|---|---|---|---|---|
| M15 (336) | 3 | 1 | 3 | 3 | 5 | 5 | 4 | 2 | 1 |
| M17 (696) | 11 | 9 | 13 | 5 | 15 | 12 | 12 | 11 | 7 |

Legend:
- Feminized
- Masculinized
- Multiple Effects
- Sex Diff Reversed
- Sex Diff Maintained
- No Category

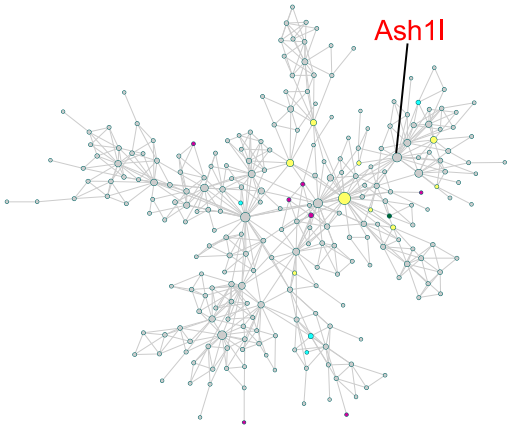
C. Acrachne Plot of GHM Module 28




D. Acrachne Plot of GHF Module 460




E. Acrachne Plot of SIM Module 115




F. Acrachne Plot of SIF Module 318

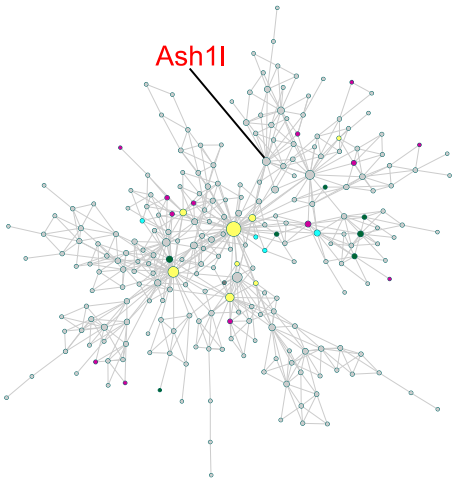


**Figure S5: Co-Expression analysis reveals Ash1l as sex-specific hub in NAc**

**A)** Module Differential Connectivity plots for all submodules of GHF parent module 2 compared to GHM, SIM and SIF. Purple = differentially connected, and gray = not differentially connected. **(B)** Module gene enrichment in all parent modules identified in GHM and GHF. Significant enrichment is indicated by purple, number of genes in each module is in parentheses, and number of overlapping genes is in each box. (C-F) Arachne plots of gene modules with Ash1l as a key driver. GHM (C) have more connections within the module than GHF (D) and this effect is lost in SI modules (E-F). Colors = genes defined by pattern analysis, multiple effects = genes that fell under more than one pattern in the 4 different treatment groups.

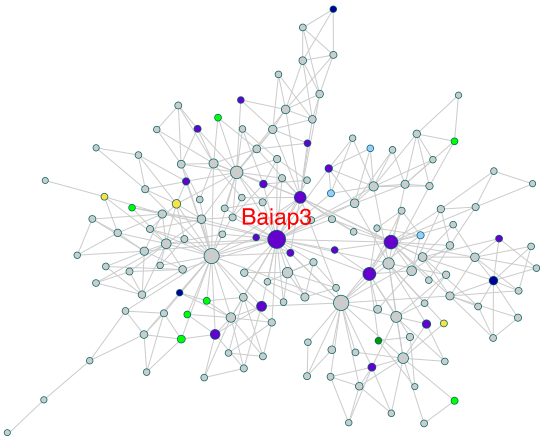
A. Module Differential Connectivity of Submodules of GHM M3

| | GHF | SIM | SIF | |
|---|---|---|---|---|
| M11 | 7.62* | 3.20* | 7.18* | Fem |
| M12 | 8.44* | 3.43* | 7.64* | Fem |
| M13 | 10.71* | 3.86* | 8.05* | Fem |
| M14 | 7.98* | 2.97* | 6.30* | Fem |
| M15 | 9.35* | 3.29* | 7.37* | Fem |
| M16 | 15.1* | 3.98* | 11.0* | Fem |
| M17 | 6.91* | 2.42* | 5.88* | Fem |
| M18 | 12.43* | 4.01* | 10.54 | Fem |
| M19 | 7.43* | 3.27* | 6.08* | Fem |
| M20 | 8.81* | 3.01* | 7.76* | Fem |

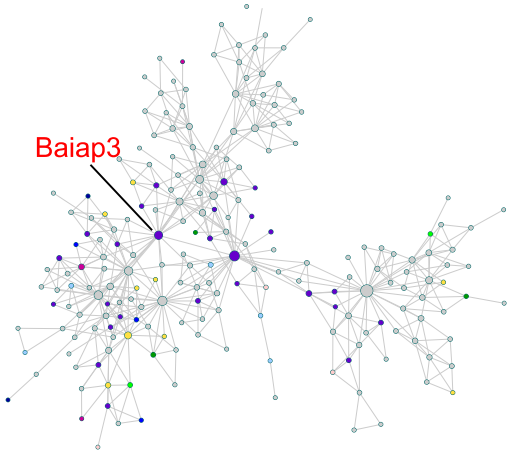
| | GHF | SIM | SIF | |
|---|---|---|---|---|
| M21 | 5.63* | 1.98* | 5.01* | Fem |
| M22 | 7.9* | 3.48* | 7.07* | Fem |
| M23 | 4.89* | 1.44* | 4.37* | Fem |
| M24 | 9.84* | 3.02* | 8.17* | Fem |
| M25 | 10.41* | 3.52* | 6.95* | Fem |
| M26 | 7.63* | 2.55* | 5.93* | Fem |
| M27 | 8.83* | 3.01* | 7.86* | Fem |
| M28 | 5.65* | 2.65* | 5.59* | Fem |
| M29 | 6.33* | 2.57* | 5.11* | Fem |
| M30 | 7.09* | 1.55* | 5.53* | Fem |

| | GHF | SIM | SIF | |
|---|---|---|---|---|
| M31 | 12.02* | 2.58* | 8.31* | Fem |
| M32 | 3.61* | 1.36* | 3.68* | Fem |
| M33 | 4.92* | 1.74* | 4.09* | Fem |
| M34 | 10.0* | 3.68* | 8.93* | Fem |
| M35 | 10.7* | 2.33* | 10.0* | Fem |
| M36 | 6.89* | 2.20* | 5.98* | Fem |
| M37 | 8.92* | 2.05* | 8.63* | Fem |
| M38 | 4.53* | 1.48* | 3.47* | Fem |
| M39 | 3.75* | 2.35* | 4.53* | Fem |

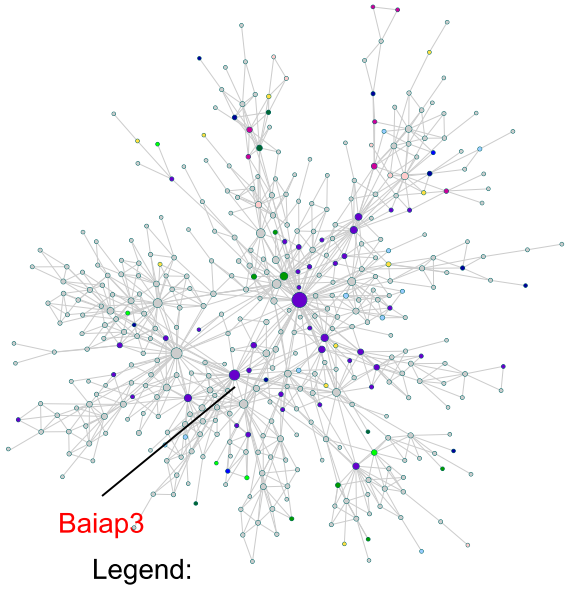
B. Acrachne Plot of GHM Module 190



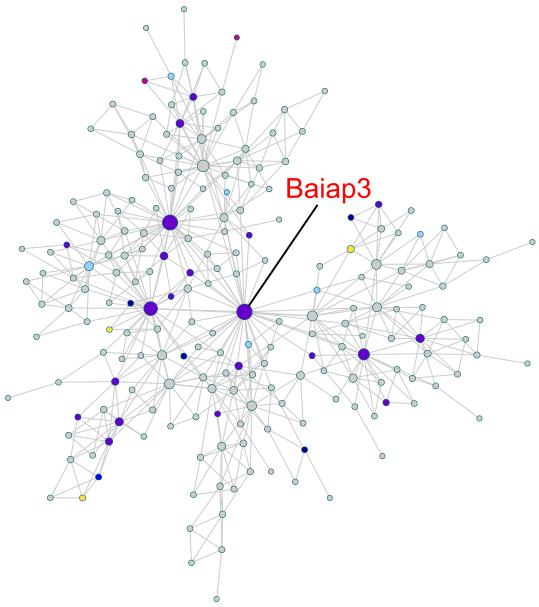

C. Acrachne Plot of GHF Module 113




D. Acrachne Plot of SIM Module 311




E. Acrachne Plot of SIF Module 175




Legend:

- Multiple Effects
- SI DN
- SI No Effect
- SI UP
- No Category
- SIF DN
- SIF No Effect
- SIF UP
- SIM DN
- SIM No Effect
- SIM UP

**Figure S6: Co-Expression analysis reveals Baiap3 as sex-specific hub in VTA**

(A) Module Differential Connectivity plots for all submodules of GHM parent module 3 compared to GHM, SIM and SIF. Purple = differentially connected, and gray = not differentially connected. (B-E) Arachne plots of gene modules with Baiap3 as a key driver. GHM (B) have fewer connections within the module than GHF (C) and this effect is lost in SI modules (D-E). Colors = genes defined by pattern analysis, multiple effects = genes that fell under more than one pattern in the 4 different treatment groups.

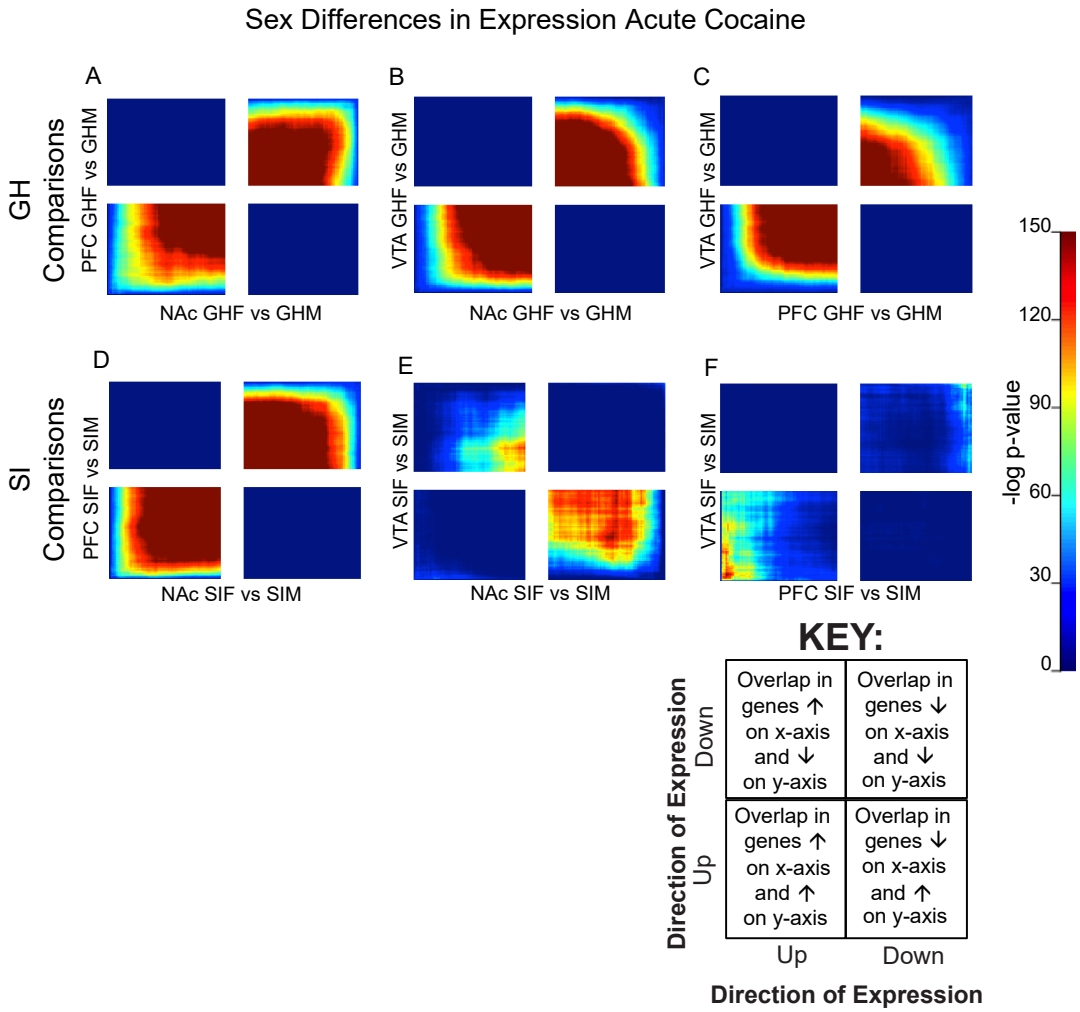



**Figure S7: RRHO analysis of sex-specific expression in each brain region after acute cocaine.**

(**A-C**) Sex differences in expression in all three brain regions align after the first dose of cocaine. (**D-F**) This alignment is maintained, although weakened, between NAc and PFC (**D**). The alignment is completely disrupted between NAc and VTA (**E**) and PFC and VTA (**F**).

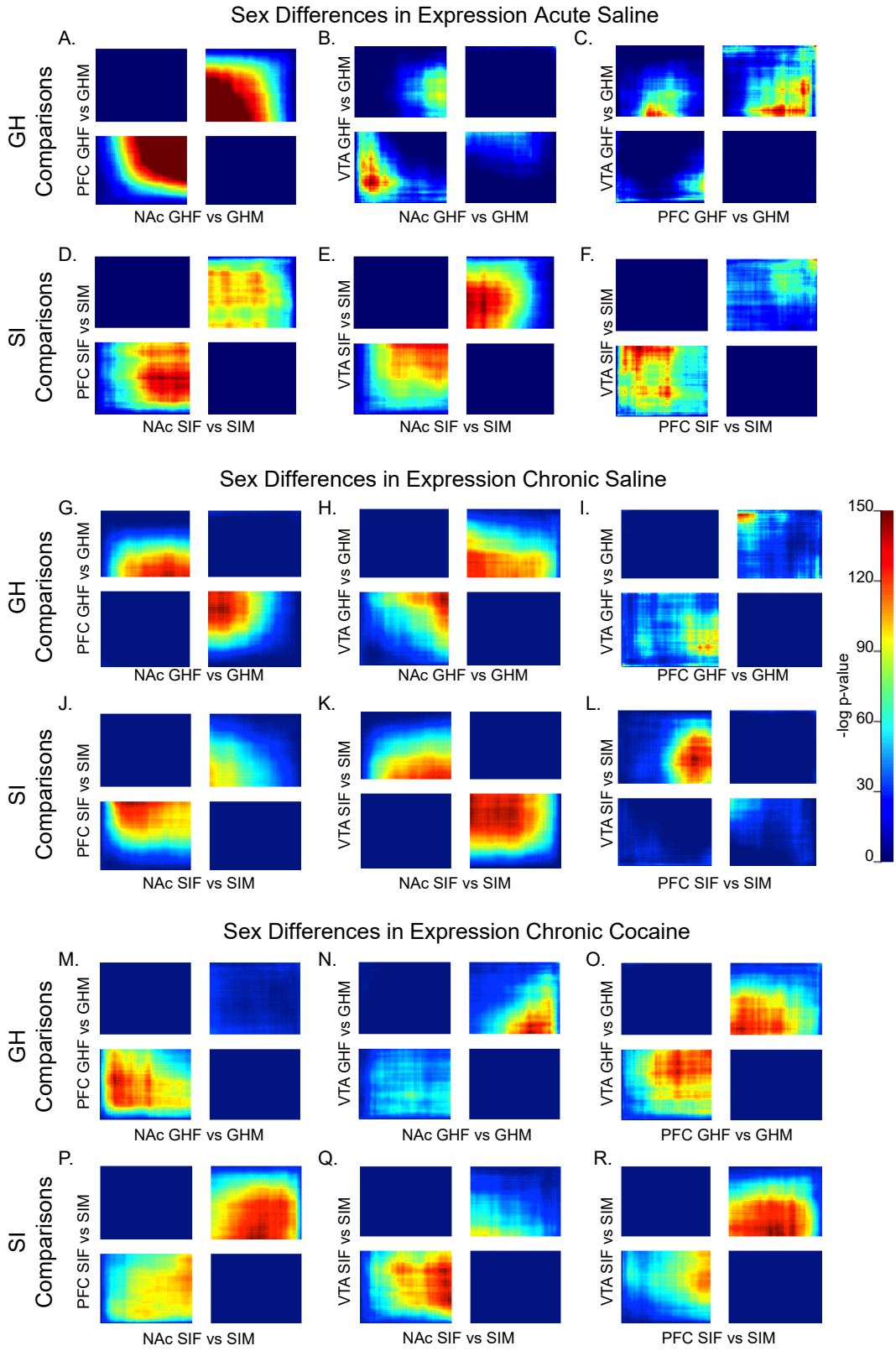



**Figure S8: RRHO analysis of sex-specific expression in each brain region after acute and chronic saline and chronic cocaine.**

**(A-F**) Overlap of sex differences in expression in all three brain regions after an acute dose of saline in GH (A-C) vs SI animals (D-F). Alignment of sex differences in the NAC and PFC (A) is attenuated by SI (D) but the lack of alignment of sex differences (B-C) with the VTA is disrupted by SI (E-F). (**G-L**) Overlap of sex differences in expression in all three brain regions after chronic saline in GH (G-I) vs SI animals (J-L). Alignment of sex differences is between the NAc and other brain regions (G-H) is disrupted by SI (J-K). (**M-R**) Overlap of sex differences in expression in all three brain regions after chronic cocaine in GH(M-O) vs SI animals (P-R). Chronic cocaine results in a alignment of sex-specific expression in the VTA and PFC in both GH (O) and SI animals (R).

**Supplemental Table S1:** Differential expression of cocaine induced genes
**Supplemental Table S2:** Differential expression of sex differences in expression
**Supplemental Table S3:** Pattern analysis of gene expression
**Supplemental Table S4:** Conserved Hub Genes

*See separate Excel files for supplemental tables.*

**SUPPLEMENTAL METHODS**:

## Animals

All experiments were in compliance with guidelines of the Institutional Animal Care and Use Committee at Mount Sinai. Jackson Laboratories (Bar Harbor, ME) shipped age-matched male and female C57BL/6J mice which were received at Mount Sinai on postnatal day (P) 21. Mice had a 24 hr acclimation period before being isolated on P22 by placing one animal in a transparent polycarbonate cage. Mice had olfactory, visual, and auditory interactions with others but no tactile interactions. After three weeks of isolation (P42), mice returned to their original cage mates and were group housed (GH) until ~P90. All animals were collected in one cohort and GH controls were maintained as comparisons. Animals were maintained on a 12 hr light-dark cycle (lights on at 7:00) at 22-25°C with *ad libitum* access to food and water.

## Cocaine Injections and Tissue Collection for RNA-seq

On P80, animals were separated into 8 groups of males and 8 groups of females: GH + chronic cocaine/saline; SI + chronic cocaine/saline; GH + acute cocaine/saline; and SI + acute cocaine/saline (10-15 animals/group). A total of 200 animals were utilized. For chronic injections, animals were given one daily injection (IP) of cocaine (7.5 mg/kg) or saline for 10 days (between 10:00 and 14:00 hr) and were euthanized 24 hr after their final injections (n=6-8 animals/group; total = 116 samples). For acute injections, mice were dosed with saline injections (IP) for 9 days for the purpose of habituating the animals to handling and injection stress. On the 10$^{th}$ day, animals were injected with saline or cocaine (7.5 mg/kg) and euthanized ~1 hr after the injection. All animals were euthanized via cervical dislocation. Brains were extracted and sliced on ice in a brain block (1 mm thick) and micropunches of PFC (12 gauge), NAc (14 gauge), and VTA (16 gauge) were snap-frozen on dry ice and stored at -80°C until use.

Vaginal cytology was monitored throughout the injections and only those females in metestrus/diestrus were used for library preparation.

**RNA Isolation, Library Preparation, and Sequencing**

RNA isolation was conducted as described previously (1, 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 gathered in the flow-through, and total RNA was purified using RNeasy Mini spin columns. Samples were treated with DNAse to purge samples of genomic DNA and run on nanodrop and an Agilent Bioanalyzer 2100 to affirm RNA purity, integrity, and concentration. All samples had RINs>8.

Libraries were prepared using the TruSeq Stranded mRNA HT Sample Prep Kit protocol (Illumina, San Diego, CA). With the use of random hexamers, poly A selection and fragmentation of 300 ng of RNA was converted to cDNA momentarily. 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 were 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 UCLA Sequencing Core (Los Angeles, CA). Each pool was sequenced 3 times with the intention of obtaining ~30 million reads per sample. The number of independent tissue samples incorporated in the final analysis was between 6-8 per group.

**Statistical and Bioinformatic Analyses**

Differential Expression Analysis: Sequencing reads were aligned to the mouse mm10 genome using Tophat2 (3). QC revealed an average of 29 million reads per sample (min = 19 million; Max = 51 million) with an average mapping rate of 90.2%. Read counts were generated using HtSeq-count against the Encode vM4 annotation.

Genes expressed at ≤1 count per million in ≥ 50 samples were excluded from further analysis. Gene expression read counts were normalized using trimmed mean of M-values normalization (TMM) method (4) and followed by voom (5) mean-variance transformation to account for the differences in library size among samples. Differentially expressed genes (DEGs) in comparisons of male vs. female, SI vs. GH, cocaine vs. saline and acute vs. chronic were identified using the Bioconductor package *limma* (6) with the following thresholds: nominal p- value <0.05 and fold-change of 30% (Suppl Table 1).

Pattern Analysis: Pattern analysis was completed as described previously (1, 2). Expression changes from the same baseline were converted into 0s and 1/-1 (0 = no effect; -1 = significantly downregulated; 1 = significantly upregulated) for each condition. The analysis output was a list of combinations of 0s and -1/1s observed in the dataset. In order to avoid bias, two investigators unfamiliar with the experiment defined the patterns of expression. Each gene was then assigned to a defined category. Patterns included genes that were differentially expressed when compared to GHFs injected with chronic saline ($p<0.05$; fold change>30%). The dataset (Suppl Table 2) revealed many patterns; we focused on two: those genes that were differentially connected between GHMs vs. GHFs under each exposure paradigm (sex differences; Figure 3) and those that were differentially expressed in SIMs vs SIFs but not GHMs vs GHFs (SI-only Effects; Figure S3). For example, a gene that was significantly increased in GHMs but not GHFs after acute saline would be categorized as a sex difference in expression; if that gene's expression resembles GHFs in SIMs then it was categorized as

"feminized", whereas if that gene resembles the expression profile of GHMs in SIFs it was categorized as "masculinized." A schematic of the different categories is included in Figure 3A. Importantly, a gene can only be defined as one category under each treatment condition (Ac/Chr Saline and Ac/Chr Cocaine). Thus, we identified genes that are uniquely regulated by each stimulus.

**Multiscale Embedded Gene Co-Expression Network Analysis (MEGENA)**

MEGENA (7) was used to construct gene co-expression networks for GHM, GHF, SIM, and SIF samples. First, MEGENA built a planar filtered network through parallel computation, early termination, and prior quality control. Next, MEGENA performed a multiscale clustering analysis to obtain co-expression modules by introducing compactness of modular structures characterized by a resolution parameter. Lastly, MEGENA completed a multiscale hub analysis to identify highly connected hubs (or driver genes) of each module at each scale. Modules with 50 – 5,000 genes were extracted to perform downstream analyses. DEGs discovered in the differential expression analysis were then laid onto the modules to perform enrichment analysis: ranking modules associated with different treatment conditions according to the adjusted enrichment p-value. In addition, GO-function enrichment analysis (Wang et al., 2012) was applied to the modules to identify enriched biological processes with p-values adjusted by Benjamini-Hochberg correction. Sunburst plots demonstrating the enrichment of sex differences (SD) in expression. SI-induced changes or GO biological processes in individual modules of the networks were visualized using the R package *sunburstR* (8). Top-ranked modules in each network were also visualized in circos plots with the significance of DEG enrichment using the R package *Netweaver* (9). Module subnetworks were visualized by Cytoscape_v3.3 (10). Module differential connectivity (MDC) analysis (11) was used to quantify network module reorganization across different conditions. The significance of MDC was calculated based on the maximum false discovery rate of 100-time permutation of samples and 100-time

permutations of genes in modules. MDC is a measure of the ratio of the average connectivity among genes of a module in one condition to changes among the same set of genes in another condition.

**Heat Maps and Venn Diagrams**

Heat maps were generated using Morpheus from the Broad Institute (https://software.broadinstitute.org/morpheus) to broadly look at changes to the transcriptional profiles of the GH and SI animals across the NAc, PFC, and VTA. Venn Diagrams were then produced using BioVenn (12) to identify changes to the number of genes both expressed in GH/SI males and females. Venn diagrams were also utilized to determine disruption in gene co-expression. Enrichment analysis was conducted using Fisher's exact tests using the Super Exact Test package in R (13).

**Rank Rank Hypergeometric Overlap (RRHO) Analysis**

RRHO analysis was performed using the RRHO2 script in R as described previously (2, 14, 15). RRHO finds overlap between expression profiles in a threshold-free manner to assess the degree and significance of overlap (16). Full differential expression between males and females were ranked by the -log(p-value) multiplied by the sign of the fold change/slope of association. RRHO difference maps were created 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 (10)

**Pathway Analysis**

Predicted molecular pathways were identified through Ingenuity Pathway Analysis (IPA) Software (Qiagen, Fredrick MD). The validity of these outcomes was based on the log fold

change of genes ($p<0.05$; fold change>1.3) analyzed. Pathways were filtered by activation z-score (>2) and p-value (<0.05) as well as for pathways associated with second messenger/intracellular signaling or neuronal/neurotransmitter signaling.

**Supplemental References:**


1. Walker DM, Cates HM, Loh YE, Purushothaman I, Ramakrishnan A, Cahill KM, et al. (2018): Cocaine Self-administration Alters Transcriptome-wide Responses in the Brain's Reward Circuitry. *Biol Psychiatry*.
2. Walker DM, Zhou X, Ramakrishnan A, Cates HM, Cunningham AM, Peña CJ, et al. (2020): Adolescent Social Isolation Reprograms the Medial Amygdala: Transcriptome and Sex Differences in Reward. *bioRxiv*.2020.2002.2018.955187.
3. Trapnell C, Pachter L, Salzberg SL (2009): TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 25:1105-1111.
4. Oshlack A, Robinson MD, Young MD (2010): From RNA-seq reads to differential expression results. *Genome Biol*. 11:220.
5. 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.
6. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. (2015): limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 43:e47.
7. Song WM, Zhang B (2015): Multiscale Embedded Gene Co-expression Network Analysis. *PLoS Comput Biol*. 11:e1004574.
8. Bostock M, Rodden K, Warne K, Russell K, Breitwieser F, Yetman C (2020): Package ‘sunburstR’.
9. Wang M, Zhang B (2019): Package ‘NetWeaver’.
10. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. (2003): Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13:2498-2504.
11. Zhang B, Gaiteri C, Bodea LG, Wang Z, McElwee J, Podtelezhnikov AA, et al. (2013): Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*. 153:707-720.
12. Hulsen T, de Vlieg J, Alkema W (2008): BioVenn - a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*. 9:488.
13. Wang M, Zhao Y, Zhang B (2015): Efficient Test and Visualization of Multi-Set Intersections. *Sci Rep*. 5:16923.
14. Barko K, Paden W, Cahill KM, Seney ML, Logan RW (2019): Sex-Specific Effects of Stress on Mood-Related Gene Expression. *Mol Neuropsychiatry*. 5:162-175.
15. Seney ML, Huo Z, Cahill K, French L, Puralewski R, Zhang J, et al. (2018): Opposite Molecular Signatures of Depression in Men and Women. *Biol Psychiatry*. 84:18-27.
16. 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.