# Coding and Non-coding RNA Expression in Rodent Models of Cocaine Craving

by

Annika Vannan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved April 2022 by the
Graduate Supervisory Committee:

Janet L. Neisewander, Co-Chair
Melissa A. Wilson, Co-Chair
Deveroux Ferguson
M. Foster Olive
Nora I. Perrone-Bizzozero

ARIZONA STATE UNIVERSITY

August 2022

ABSTRACT

Substance use disorders (SUDs) are difficult to treat, in part because drug craving can be elicited by exposure to drug-associated environments and cues within the environment. Furthermore, this craving becomes more pronounced as abstinence progresses and it can take months to years for cue-elicited craving to finally wane. This important hallmark of addiction is modeled in rodents by exposing them to light/tone cues associated with the self-administration (SA) of cocaine. Cue exposure results in drug-seeking behavior, an animal analogue for drug craving. The overarching goal of this dissertation was to use the rodent SA model to explore the nucleus accumbens (NAc), a key brain region in the neural pathway of craving, and examine ribonucleic acid (RNA) expression that may underlie cocaine-seeking behavior. This includes messenger RNAs (mRNAs), which encode directly for proteins, and non-coding RNAs, which are important regulators of mRNA expression and cellular function. My first experiment aimed to identify non-coding microRNAs, which directly target and suppress mRNA expression, that are differentially expressed in animals with high or low cocaine-seeking behavior. In the second study, I compared RNA-sequencing (RNA-seq) datasets from rodent models of cocaine abstinence and developed a novel workflow to narrow candidate genes. In the final experiment, I utilized RNA-seq and reverse transcription real-time quantitative polymerase chain reaction (RT-qPCR) to identify and explore non-coding, circular RNAs that may influence gene regulatory networks and impact drug-seeking behavior. Overall, these studies promote our understanding of the neurogenetic mechanisms of craving and they suggest recommendations for improving the experimental design of future neurogenomic studies.

DEDICATION

To my many mentors,

who encouraged me to pursue my interests and challenged me to grow as a scientist.

To my friends,

who ground me and encourage me to grow as a person.

To my parents, Ken and Anke Vannan,

for loving and supporting me through everything in life.

To Aaron Grote, my partner of ten years,

who has always been a source of strength and comfort.

ACKNOWLEDGMENTS

I would like to thank my wonderful advisors Dr. Janet Neisewander and Dr. Melissa Wilson. It is because of their encouragement for my personal passions and interests that I have been able to forge a path between their two fields of expertise. I am also grateful to my committee members Dr. Nora Perrone-Bizzozero, Dr. M. Foster Olive, and Dr. Deveroux Ferguson. I greatly appreciate everyone's support on my unique scientific journey. I would also like to acknowledge the many collaborators who have made these projects possible. In particular, Dr. Michela Dell'Orco has been a continued source of knowledge over the years.

I am also thankful for the many members of the Neisewander and Wilson labs, including Dr. Taleen Der-Ghazarian, Dr. Raul Garcia, Dr. Seema Plaiser, Dr. Kimberly Olney, Samantha Scott, Angela Oill, and many others. I also acknowledge all of the undergraduates who dedicated their time and energy over the years, including Frankie Guerrero, Ayleen Mokbel, Oscar Villareal, Aracely Esquer, Aysha Mahmud, Nafis Eghrari, and Sophia Doyle. I especially thank Dr. Gregory Powell for his many years of mentorship and for being a pioneer in bioinformatics in the Neisewander lab. I also appreciate the support and assistance of Dr. Mark Namba and Michael Johnson. Mark, thank you for our many brainstorming sessions and for helping me grow as a scientist and person. Michael, it has been a pleasure watching you grow from mentee to peer. I am eternally grateful to you both for your friendship. I am also grateful to my friend Tanessa Call for the memories we've shared throughout our time in the Neuroscience program.

These studies were supported by the National Institute of Health grants R21DA048651, R01DA034097, and R21DA048651, and through funding directly from Arizona State University and the School of Life Sciences.

TABLE OF CONTENTS

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

CHAPTER 1

# GENERAL OVERVIEW

## Substance Use Disorders

Substance use disorders (SUDs) are a pervasive global issue with huge financial and social burdens. In the U.S. alone, SUDs cost more than $700 billion annually related to health care, crime, and loss of productivity (National Drug Intelligence Center, 2011; National Institute on Drug Abuse, 2017). Addiction also has devastating personal consequences, both for individuals suffering from SUDs and their communities. Though statistics on SUDs and drug abuse rates have varied somewhat across the past few decades, many trends are worsening. For example, drug overdoses and overdose deaths were on the rise prior to 2019 (Kariisa et al., 2019; National Institute on Drug Abuse, 2022), and they spiked even further during the COVID-19 pandemic as access to SUDs treatments and resources became more limited (Kuehn, 2021; Nguyen & Buxton, 2021). Thus, the need for therapeutic options remains critical.

## SUDs Symptomatology: Drug Craving and Relapse

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM 5) lists 11 diagnostic criteria for SUDs related to a loss of control over drug intake even as personal, health, and psychological consequences of drug use mount (American Psychiatric Association et al., 2013). In response to decades of addiction research, the diagnostic criteria for SUDs were updated in 2013 to include drug craving. Craving, or the desire to seek out and use drug, is a hot topic in SUDs, and there are valid animal models for studying craving (Gawin & Kleber, 1986; Markou et al., 1993; M. A. Smith, 2020; Wolf, 2016). Though craving can be triggered by a variety of stimuli, including

stress (Sinha et al., 1999), it is most often studied in the context of drug-associated cues. Cues such as a crack pipe or lighter that are repeatedly linked with the drug-taking experience can acquire incentive salience, such that presentation of these cues without the drug itself can trigger feelings of craving (Childress et al., 1993; Ehrman et al., 1992; T. E. Robinson & Berridge, 1993). A major difficulty in treating SUDs is that cue-induced craving worsens during abstinence, a phenomenon known as the incubation effect (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). This increased craving often manifests in drug-seeking behavior, which can lead to relapse (Sinha, 2013; Weiss, 2005). Individuals may relapse many times over the course of treatment for SUDs, re-entering the cycle of drug abuse. A better understanding of the neurobiological mechanisms of drug craving is imperative to preventing relapse and developing new SUDs treatments.

**Modeling Drug Abuse and Craving in Animals**

Various behavioral models have been designed to examine SUDs symptomatology in animals. The gold standard of these models is self-administration (SA), which endures as the most highly-regarded preclinical paradigm in the addiction field due to its high face, construct, and predictive validity (Panlilio & Goldberg, 2007; Spanagel, 2017). In the SA model, animals are trained in operant conditioning chambers to voluntarily perform tasks to receive drug (Panlilio & Goldberg, 2007). Over the course of days, weeks, or months, animals learn that a specific response (e.g. pressing a lever, or poking their nose in a hole) results in administration of drug. Delivery of this reinforcer is contingent upon completion of a schedule of reinforcement, where performance requirements can be fixed or varied. For example, with a fixed ratio (FR) 1 schedule, one response (e.g. one lever press) results in one drug delivery. In more complex behavioral

models, the schedule of reinforcement may change between or within sessions (e.g. variable ratio (VR), variable interval (VI), or progressive ratio (PR) schedules).

In animals, incubation of drug craving is robust and replicable using several variations on the SA model, including cue reactivity (Markou et al., 1993; M. A. Smith, 2020; Wolf, 2016). Typically, animals will learn to self-administer drug as the primary reinforcer, with cues presented simultaneously. For example, if a light and tone cue trigger as drug is delivered, the animal will learn to associate those cues with the drug-taking experience. The cue reactivity paradigm harnesses the acquired incentive salience of those cues as a measure of drug craving. With cue reactivity, animals are trained in drug SA over many daily sessions and are then placed in forced abstinence. During this time, they remain in their home cage environment and do not have access to the operant conditioning chamber. After an experimentally-determined time period, animals return to the SA environment. In a post-abstinence test, animals are able to perform the previously trained behavior (e.g. lever press or nose poke), though it results in the delivery of the drug-associated cues without the drug itself. Because animals will readily perform the behavior even in the absence of drug, this measure of drug-"seeking" behavior is considered a proxy for craving (Markou et al., 1993; M. A. Smith, 2020).

**The Genetics of SUDs**

SUDs are marked by high genetic complexity that is difficult to disentangle in human populations (Agrawal et al., 2012; Buckland, 2008; Goldman et al., 2005). Though SUDs are partly heritable (Uhl et al., 2008), they likely involve many genetic susceptibility variants with small effect sizes, which may be overlooked by genome-wide association studies (GWAS) (Buckland, 2008; Deak & Johnson, 2021). Moreover, addictive behaviors are driven by gene-environment interactions and dynamic gene

expression changes difficult to study in human populations (Buckland, 2008; Goldman et al., 2005). While there are many available technologies and tools for human behavioral and genetics research, it is difficult to obtain a fine-tuned understanding of neurobiological mechanisms using human subjects. However, the underlying neurobiological mechanisms of reward and motivation are presumed to be shared between humans and other animals (Martínez-García & Lanuza, 2018; Panksepp et al., 2002; Scaplen & Kaun, 2016). Thus, animal models have been instrumental in developing our understanding of drug abuse and SUDs at the molecular level (Spanagel, 2017).

**Expression of Coding and Non-Coding RNAs**

Under the central dogma of molecular biology, DNA sequences (genes) are transcribed into RNA copies that serve as the template for translation into proteins. In this simplistic model, proteins perform functions in the cell, while "coding" RNAs primarily serve as intermediaries in the process of protein biogenesis. The amount of RNA that corresponds to a specific gene is then a measure of that gene's expression. Gene expression is one of the most widely-used metrics for determining dynamic biological differences between groups. Importantly, gene expression is regulated by many cellular processes. For example, non-coding RNAs, initially considered "junk" because they did not encode for proteins, have now emerged as important regulators of gene expression.

**MicroRNAs**

MicroRNAs (miRNAs) are small, non-coding RNAs approximately 22 nucleotides (NTs) in length that post-transcriptionally regulate gene expression. In their mature,

single-stranded form, miRNAs are incorporated into the RNA-induced silencing complex (RISC) (Filipowicz, 2005). Targets are recognized by RISC through binding of the 6-8 nucleotide "seed sequence" on the miRNA to miRNA response elements (MREs) on the mRNAs. RISC then directs suppression of the mRNA target through translational repression, cleaving of the target, or other mechanisms (Bartel, 2004, 2009). Importantly, the miRNA seed does not need to bind with perfect complementarity to the mRNA in order to regulate its expression (Bartel, 2004). In addition, while MREs typically occur on the 3' untranslated regions (UTRs) of mRNAs, they can also be found on 5'UTRs or in coding regions (Hausser et al., 2013; Lai, 2002; Lee et al., 1993; Lytle et al., 2007). Therefore, a single miRNA may target and suppress hundreds or thousands of mRNAs. Hundreds of miRNAs have now been identified, most of which are conserved across mammals (Friedman et al., 2009). This has led to interest in miRNAs as "master regulators" of gene expression with therapeutic promise in translational medicine.

Many miRNAs have been linked to basic brain function and synaptic plasticity (see (Forero et al., 2010; Gowen et al., 2021; A. C. W. Smith & Kenny, 2018) for reviews). For example, a study in Dicer-deficient zebrafish, which lack the protein necessary to produce mature miRNAs, suggests that miRNAs are critical to healthy development of the central nervous system (Giraldez et al., 2005). miRNAs are widely expressed throughout the brain in both glia and neurons, including at dendrites, axons, and synapses (O'Carroll & Schaefer, 2013). A small number of miRNAs are also enriched in the brain or at synapses (Adlakha & Saini, 2014), but any brain-expressed miRNAs can impact synaptic plasticity genes, and there is increasing evidence that miRNAs regulate drug-induced plasticity. For example, miR-212/miR-132, miR-495, and miR-124 have all been implicated in rodent models of cocaine abuse (Bastle et al., 2018; Chandrasekar & Dreyer, 2009; Im et al., 2010; Jasińska et al., 2016; Sadakierska-Chudy et al., 2017).

These miRNAs are regulated by, or target, known addiction and synaptic plasticity proteins, such as CREB, BDNF, CaMKIIa, and several integrins, which in turn regulate synaptic size and strength (see (A. C. W. Smith & Kenny, 2018) for a review of these miRNAs). However, miRNAs have only been studied in a small number of SUDs models. More research is necessary to understand what other miRNAs may regulate drug-related behaviors, including cocaine-seeking and craving.

**Circular RNAs**

Circular RNAs (circRNAs) are a class of non-coding RNAs (though see (Pamudurti et al., 2017) that have been largely overlooked for decades. The first circRNA, a plant viroid, was discovered in 1976 by (Sanger et al., 1976), and the first eukaryotic circRNA was found just years later (Hsu & Coca-Prados, 1979). Despite this early discovery, circRNAs were not commonly studied until the 2010's, as researchers initially believed they were rare in eukaryotes and arose only due to abnormal or erroneous splicing (Nigro et al., 1991). Now, circRNAs are known to be abundant in eukaryotes and largely conserved across mammalian genomes (Jeck et al., 2013; Rybak-Wolf et al., 2015; Salzman et al., 2013), and thus circRNAs are undergoing a research renaissance.

Though our understanding of circRNAs is still in its infancy, emerging research suggests they are potent regulators of gene expression. circRNAs are generated via backsplicing of linear pre-mRNAs that results in RNA circles (L.-L. Chen & Yang, 2015). Though they may contain introns or span intergenic regions, most known circRNAs are derived from exons of their host genes (Jeck & Sharpless, 2014; Y. Zhang et al., 2013). This close relationship between circRNAs and their corresponding linear mRNA transcripts suggests many circRNAs regulate expression of their host genes through

competitive splicing (Ashwal-Fluss et al., 2014). In addition, circRNAs can interact with RNA-binding proteins (RBPs) (Conn et al., 2015; Dell'Orco et al., 2020; Zang et al., 2020), bind to and sequester miRNAs (Hansen et al., 2013; Memczak et al., 2013), and even bind directly to mRNAs to influence expression (Hafez et al., 2022). Through interactions with multiple protein and RNA classes, circRNAs can act as major players in regulatory networks. For example, circRNAs may compete with linear mRNAs for binding to miRNAs, forming competing endogenous RNA (ceRNA) networks of circRNA-miRNA-mRNA interactions (Salmena et al., 2011). In this way, circRNAs can suppress the action of miRNAs, preventing miRNA repression of target genes.

circRNAs are exciting targets for neurobiological research because they are enriched in the brain, particularly at synapses, and often arise from synaptic plasticity genes (Rybak-Wolf et al., 2015; You et al., 2015). For example, *circHomer1* forms via alternative splicing of the *Homer1b* transcript of *Homer1* (You et al., 2015; Zimmerman et al., 2020), a known synaptic plasticity gene with relevance to drug addiction (Szumlinski et al., 2008, 2006). Early evidence from (You et al., 2015) demonstrated that *circHomer1* is enriched at synaptosomes and rapidly localizes to dendrites during periods of neuronal activity, suggesting that *circHomer1*, like the proteins encoded by its host gene, plays a role in synaptic plasticity. Since then, several studies have implicated this circRNA and others in neurological disorders relating to learning and memory, supporting a role for circRNAs in many aspects of neuronal and brain function (Floris et al., 2017; Hafez et al., 2022; Lukiw, 2013; Mehta et al., 2020; Zimmerman et al., 2020). However, few studies have examined circRNAs and their corresponding ceRNA networks in drug abuse (Dell'Orco et al., 2021, 2020; Floris et al., 2022; J. Li et al., 2020), leaving the role of circRNAs in addiction a wide-open field for discovery.

**Analyzing RNA Expression**

There are many techniques for comparing expression of RNAs. One tried-and-true technique is quantitative real-time polymerase chain reaction (RT-qPCR), which can be used to measure expression of a single RNA (Gibson et al., 1996). To prepare for RT-qPCR, first RNA is isolated from a sample, and then complementary DNA (cDNA) is synthesized and amplified as a proxy for the RNA. To detect and quantify the RNA, PCR is performed with primers that are designed to anneal to the cDNA based on the sequence of interest. As the cDNA is extended, a fluorescent probe or dye added to the reaction emits a signal. The output measure of gene expression is the number of PCR cycles required to reach a threshold level of fluorescence, such that a lower cycle number indicates higher expression. This metric is then compared to expression of a chosen housekeeping gene that has consistent expression across samples. Non-coding RNAs can also be measured with RT-qPCR. In standard cDNA synthesis, an oligo(dT) primer is added which anneals to the poly(A) tail. However, certain non-coding RNAs such as miRNAs and circRNAs do not have poly(A) tails (Bartel, 2004; Jeck & Sharpless, 2014). Thus, a frequent approach for cDNA synthesis of miRNAs is to first add poly(A) tails and adapter sequences (Krepelkova et al., 2019). For circRNAs, random primers can be used in place of oligo(DT) primers to ensure all RNAs are amplified, regardless of poly(A) tailing. For circRNA analysis, primer design is also an important consideration, as the primers should be specific to the circular form and not their linear mRNA counterparts, with which they share high sequence similarity. To do so, primers must target sequences that flank the backsplice junction of the circRNA and face outward with respect to the genomic sequence (Jeck & Sharpless, 2014).

Microarray and other array techniques allow for RNA expression analysis on a larger scale. While only a handful of RNA sequences can be quantified with a single RT-

qPCR run, arrays allow the user to examine many RNAs at once. With microarrays, cDNA is placed on a microchip with hundreds or thousands of wells (Schena et al., 1995; Shalon et al., 1996). Each well corresponds to a specific sequence of interest and contains many DNA probes for that sequence. The cDNA hybridizes to these DNA probes, and RNA expression is then quantified based on the level of fluorescence in each well (Shalon et al., 1996). Importantly, microarrays rely on cDNA and thus amplification, which may introduce bias if some samples are amplified more efficiently than others. Other array techniques, including the nCounter technology developed by Nanostring, bypass this issue by allowing RNA to hybridize to RNA probes with no cDNA synthesis requirement (Eastel et al., 2019; Geiss et al., 2008; Kulkarni, 2011). Whichever technique is used, small modifications to the sample preparation and array can be made to analyze a wide variety of coding and non-coding RNAs.

With RNA-sequencing (RNA-seq), RNA expression can be quantified for the whole transcriptome (Wang et al., 2009). With this technique, cDNA is cut into smaller fragments that are directly sequenced. These sequence reads can then be aligned to a reference genome or transcriptome, and reads that align accurately to a known DNA or RNA sequence can be quantified. Lastly, samples and treatment groups can be compared to find genes that are differentially expressed in the conditions of interest. With this approach, expression can be quantified and assessed for all RNA transcripts, along with some non-coding RNAs. Because of this, RNA-seq has become extremely popular in recent years across many disciplines, particularly as costs continue to decrease. RNA-seq data can also be utilized to identify non-coding RNAs. For example, several tools have been developed to identify circRNAs from RNA-seq data, usually by finding backsplice junctions and adjacent sequences from reads that do not align to the reference (L. Chen et al., 2021; Zeng et al., 2017). As with RT-qPCR, circRNA identification can be

facilitated by enriching for circRNAs and/or depleting mRNAs during sample preparation. However, most RNA-seq data can still be used to measure circRNAs expression, though at lower read counts, as long as it is not poly(A)+ selected (i.e., circRNA-depleted). Though using RNA-seq data that is not mRNA-depleted will reduce the accuracy of circRNA identification (L. Chen et al., 2021), this allows for the assessment of both circRNAs and coding RNAs from the same dataset, and can be appropriate depending on the specific study goal.

**Research Aims and Brief Overview**

In this dissertation, I focus on coding and non-coding RNA expression in the NAc that underlies cocaine craving in rodents. Most commonly, I utilize the cue reactivity model in rats to measure cocaine-seeking behavior post-abstinence. In Chapter 2, I used this model along with environmental enrichment (EE), a therapeutic intervention known to attenuate drug-seeking behavior (Powell, Vannan, et al., 2020; Thiel et al., 2011, 2009; Vannan et al., 2018). Using a Nanostring array, I identified miRNAs in the NAc shell related to craving and the molecular pathways they are likely to regulate. In Chapter 3, I compared several publicly available RNA-seq datasets, including one previously published in our lab, centering around cocaine craving in rodent models. Utilizing these data as a test case, I implemented a new workflow for candidate gene prioritization while providing a critique of standard RNA-seq analysis practices in the field. In Chapter 4, I identified several candidate circRNAs as related to cocaine craving, and constructed theoretical ceRNA networks for these circRNAs to illustrate possible mechanisms of genetic regulation related to cocaine craving. Overall, these works contribute to a greater understanding of coding and non-coding RNAs in rodent models of cocaine craving.

Some of the research in this dissertation has either been previously published or is currently under review. Chapter 2, entitled "microRNA regulation related to the protective effects of environmental enrichment against cocaine-seeking behavior", was previously published in Drug and Alcohol Dependence in April 2021 (Vannan et al., 2021), with Dr. Gregory L. Powell as co-first author, Drs. Michela Dell'Orco and Melissa A. Wilson as co-authors, and Drs. Nora I. Perrone-Bizzozero and Janet L. Neisewander as co-senior authors. I contributed to the bioinformatics analyses with the co-first author and interpreted and curated the results for publication. Chapter 3, "An approach for prioritizing candidate genes from RNA-seq using preclinical cocaine datasets as a test case" is under review for publication. It is co-authored by Drs. Michela Dell'Orco, Nora I. Perrone-Bizzozero, and Janet L. Neisewander, with Dr. Melissa A. Wilson as senior author. I initially conceptualized the questions and approach used in this chapter, and the other authors helped refine the approach and interpretations. Chapter 4, "Alterations in circular RNA expression in the rat nucleus accumbens shell may regulate circRNA-miRNA-mRNA regulatory networks in a model of cocaine craving", has not yet been submitted for publication. Planned authors include Michael C. Johnson as co-first author, Drs. Michela Dell'Orco, Gregory L. Powell, Nikolaos Mellios, Melissa A. Wilson, and Nora I. Perrone-Bizzozero as co-authors, and Dr. Janet L. Neisewander as senior author. I performed the bioinformatics analyses, and collaborated with Michael C. Johnson, a Master's student at Arizona State University, to collect the behavioral and RT-qPCR data.

CHAPTER 2

# MICRORNA REGULATION RELATED TO THE PROTECTIVE EFFECTS OF ENVIRONMENTAL ENRICHMENT AGAINST COCAINE-SEEKING BEHAVIOR

**Previously published as:**

Vannan, A., Powell, G. L., Dell'Orco, M., Wilson, M. A., Perrone-Bizzozero, N. I., & Neisewander, J. L. (2021). MicroRNA regulation related to the protective effects of environmental enrichment against cocaine-seeking behavior. *Drug and Alcohol Dependence*, *221*, 108585. https://doi.org/10.1016/j.drugalcdep.2021.108585

## Abstract

MicroRNAs (miRNAs) are "master post-transcriptional regulators" of gene expression. Here we investigate miRNAs involved in the incentive motivation for cocaine elicited by exposure to cocaine-associated cues. We conducted NanoString nCounter analyses of microRNA expression in the nucleus accumbens shell of male rats that had been tested for cue reactivity in a previous study. These rats had been trained to self-administer cocaine while living in isolate housing, then during a subsequent 21-day forced abstinence period they either stayed under isolate housing or switched to environmental enrichment (EE), as this EE intervention is known to decrease cocaine seeking. This allowed us to create groups of "high" and "low" cocaine seekers using a median split of cocaine-seeking behavior. Differential expression analysis across high- and low-seekers identified 33 microRNAs that were differentially expressed in the nucleus accumbens shell. Predicted mRNA targets of these microRNAs are implicated in

synaptic plasticity, neuronal signaling, and neuroinflammation signaling, and many are known addiction-related genes. Of the 33 differentially-expressed microRNAs, 8 were specifically downregulated in the low-seeking group and another set of 8 had expression levels that were significantly correlated with cocaine-seeking behavior. These findings not only confirm the involvement of previously identified microRNAs (e.g., miR-212, miR-495) but also reveal novel microRNAs (e.g., miR-3557, miR-377) that alter, or are altered by, processes associated with cocaine-seeking behavior. Further research examining the mechanisms involved in these microRNA changes and their effects on signaling may reveal novel therapeutic targets for attenuating drug craving.

## Introduction

Psychostimulant abuse is a significant, ongoing problem in the U.S with devastating economic and social costs to drug abusers and their communities (National Drug Intelligence Center, 2011; National Institute on Drug Abuse, 2017; Pomara et al., 2012). Cocaine use disorders (CUDs) in particular are a serious issue, as cocaine-related deaths have increased substantially over the past few decades, even as general use has declined (Center for Behavioral Health Statistics and Quality, 2015; McCall Jones et al., 2017; National Institute on Drug Abuse, 2020). Unfortunately, there are few treatment options that are effective in promoting long-term abstinence from drug use, especially psychostimulant use. Consequently, 40-60% of drug users relapse within the first year of abstinence (National Institute on Drug Abuse, 2018)). This is in part because drug-associated cues elicit drug craving that strengthens over prolonged abstinence, leaving those with CUDs vulnerable to relapse despite efforts to cease drug use (Dackis & O'Brien, 2001; Gawin & Kleber, 1986; Neisewander et al., 2000). For example, cues such as a crack pipe or crack house acquire conditioned stimulus effects which can trigger craving and relapse (Ciccocioppo et al., 2004; Conklin & Tiffany, 2002; Ehrman et al., 1992; Weiss et al., 2001). Motivational effects of cocaine-conditioned cues persist even after months without drug use in animal models of drug-seeking, a phenomenon referred to as incubation of craving (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). Thus, treatments that reduce cue-elicited craving are needed to promote long-term abstinence.

In both animals and humans, various forms of enrichment are effective in attenuating cocaine-related behaviors throughout the abstinence-relapse cycle (Lynch et al., 2013; Vannan et al., 2018). Typically, environmental enrichment (EE) in animal models consists of social housing in small groups that are given access to novel toys

and/or exercise equipment. Importantly, EE is effective in reducing drug-seeking behavior when given as an intervention during abstinence, as measured in cocaine conditioned place preference and operant behavior animal models (Chauvet et al., 2011; Ma et al., 2016; Thiel et al., 2011, 2010, 2009). Thus, EE can be used as a tool experimentally to create groups of animals with differing levels of cocaine-seeking behavior.

There is growing interest in the epigenetics of drug abuse, including the role of microRNAs (miRNAs). In general, mammalian miRNAs post-transcriptionally silence gene expression by the imperfect base-pairing of nucleotides at positions 2-8 in the 5' end of the miRNA (widely referred to as the "seed sequence") and other miRNA sequences to the 3' untranslated regions (UTRs) of target mRNAs (Bartel, 2009; Schirle et al., 2014). Because hundreds to thousands of genes have miRNA target sequences in the 3'UTRs, miRNAs function as "master regulators" of gene expression (Plotnikova et al., 2019). The capacity of miRNAs to manipulate and alter gene expression has made this class of RNAs an exciting avenue for finding new therapeutic targets for CUDs treatment development. So far, several miRNAs have been shown to play a role in the motivational processes underlying CUDs, including the let-7 family (Chandrasekar & Dreyer, 2009, 2011), miR-212 (Hollander et al., 2010; Im et al., 2010), miR-495 (Bastle et al., 2018), and others (Kenny, 2014). It is likely that many miRNAs that contribute to resilience or susceptibility to CUDs have not yet been identified. Furthermore, many previous studies have examined the NAc as a whole, including both the core and shell, yet these subregions interface differently with corticolimbic inputs and play different roles in cocaine-seeking behavior. For instance, the NAc shell receives input from the basolateral amygdala, and this pathway is involved in processing incentive salience of cocaine-associated cues (Ma et al., 2016; Millan & McNally, 2011).

The present study employed NanoString nCounter analysis to identify miRNAs that are differentially-expressed in the NAc shell in rats with "high" vs. "low" levels of cocaine-seeking behavior. Tissue was harvested from a subset of rats utilized in a previous experiment (Powell et al., 2020) that had confirmed that rats with a history of cocaine self-administration exhibit less operant responding reinforced by cocaine-associated light/tone cues when they were housed in EE for 21 days of abstinence than when housed in isolation (ISO) [90.25 ± 20.96 and 207.4 ± 33.55 mean responses ± SEM, respectively (n=15-16/group)]. Here, we took advantage of the varying degrees of cue-elicited motivation for cocaine across a subset of the animals from this study to explore miRNAs as possible mediators of cue-elicited cocaine-seeking behavior.

## Methods

**Animals and tissue collection.** All experiments were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Institutional Animal care and Use Committee at Arizona State University. Male Sprague-Dawley rats (N=12) used in a previous study (Powell, Vannan, et al., 2020) were sacrificed by isoflurane overdose immediately after a 1-h test for cocaine cue reactivity. In the previous study, single-housed rats had been trained to self-administer cocaine (0.75 mg/kg, IV) delivered response-contingently with light and tone cues. After ≥15, 2-h sessions of training, rats underwent 21 days of forced abstinence, either remaining in single-housing or switching to an enriched environment with 3-5 cage mates, a running wheel, tubes, toys to enhance novelty, and extra nesting materials. Upon completion of forced abstinence, animals were placed back into the self-administration chamber with the cues, but not cocaine, available response-contingently (i.e., cue reactivity test) (Acosta et al., 2008; Kufahl et al., 2009). The number of times

that a rat pressed the lever resulting in cue presentations without cocaine delivery was used as a measure of cocaine-seeking behavior and is thought to reflect the degree of incentive motivation for cocaine elicited by the cues. Within 5 min of completing the test, brains were harvested and rapidly frozen in 2-methylbutane that was placed on dry ice to achieve a temperature of approximately -20°C. Later, 2 mm coronal sections containing the NAc shell were excised using a brain matrix to place razor blades at the appropriate location on the ventral surface of the brain for capturing the NAc in the tissue section. Tissue punches were then taken containing the NAc core and anterior commissure (1 mm diameter). Secondary punches (2 mm diameter) were taken containing the NAc shell using the previously punched location of the core as a landmark. RNA was isolated from the NAc shell samples using the standard Trizol method as performed previously (Bastle et al., 2018). Samples (100-150 ng of RNA) were then analyzed for miRNA expression using the Nanostring® nCounter Rat miRNA Expression Assay Kit v1.5 at the University of Arizona Genetics Core. The panel quantifies expression of 423 rat miRNAs in version v1.5, slightly fewer than the 496 listed in the miRBase Rattus norvegicus miRNA database.

**Bioinformatics analyses.** Nanostring nCounter analysis provided expression levels of miRNAs as raw counts for each miRNA in the sample. 100 ng total RNA was used in a multiplexed reaction to anneal specific miRNA tags followed by ligation and enzymatic purification to remove excess unincorporated tags in the assay, using the manufacturer protocol. Sequence specific fluorescent reporter probes and biotinylated capture probes were hybridized to ligated target nucleic acid complexes overnight at 65°C for >12 h, followed by a series of automated washes and immobilization onto a streptavidin lined cartridge for data collection. Digital images from the cartridges were obtained over 4 h with the nCounter Digital Analyzer (CCD camera and microscope

objective lens) using 555 FOV data resolution. Digital counts were tabulated and exported as comma separated values.

Differential expression analysis was performed using the R package "limma" (Ritchie et al., 2015) to identify differences in miRNA expression between the high- and low-seeking groups. Briefly, normalization factors were calculated to scale the raw library sizes. Raw counts were converted to counts per million reads (CPM) for each miRNA, and miRNAs with very low expression (CPM <30) were filtered out, which removed 68 of 420 miRNAs before the analysis. Weighted least squares were calculated for each miRNA and sample, then a linear model was fit. Contrasts were performed on the fitted linear model to compare expression of each miRNA between high- and low-seeking groups based on log2 fold change values.

TargetScan 7.2, a miRNA target predictor, was used to determine possible mRNA targets of the differentially-expressed miRNAs (http://www.targetscan.org/vert_72) (Agarwal et al., 2015). TargetScan predicts targets based on the miRNA's seed sequence, as well as conserved sites on mRNAs that fully or partially match this sequence. For some miRNAs, the available data on the predicted mRNA targets applied not to a single miRNA, but to a miRNA family. For example, mir-3573-5p is part of the miR-423-5p/3573-5p family and predicted targets for all miRNAs in this family are shared. To obtain accurate TargetScan predictions, MIMAT accession numbers provided by Nanostring were used to determine whether each mature miRNA originated from the 5' or 3' arm of its precursor. Prior literature does not always follow this convention; thus, in this paper, the 3p/5p label is only included when discussing results of the current study exclusive of comparisons to other work, or when prior authors have made the miRNA designations clear.

Because some miRNAs were upregulated, while others were downregulated, predicted mRNA targets were given "impact scores" to signify the levels of up- and downregulation that might result from the miRNA changes. The predicted targets and impact scores were input into IPA (version 51963813; QIAGEN Inc.) to identify significantly regulated pathways (Krämer et al., 2014). We then compared the TargetScan predicted mRNA targets of the differentially-expressed miRNAs to the known addiction-related genes in the Knowledgebase of Addiction-Related Genes (KARG) database (http://karg.cbi.pku.edu.cn) (C.-Y. Li et al., 2008). At the time of analysis, the rat KARG contained 1135 genes, of which 347 had an evidence score of 2 or more (were supported by 2 or more lines of evidence), and these were the only genes included in the analysis.

**miRNA validation with RT-qPCR.** Leftover RNA from the same samples used for the Nanostring analyses was used to validate select miRNAs. For each sample, approximately 45 ng of purified RNA were used to prepare cDNA using the Taqman® Advanced MicroRNA cDNA synthesis kit (Applied Biosystems, Foster City, CA, USA, # A28007), Taqman® Advanced MicroRNA Assay primers (Life Technologies, Grand Island, NY, USA) for miR-376c-3p, miR-107-3p, and miR-212-3p, and Taqman® MicroRNA Assay primer for the control transcript U6. cDNA for each sample was diluted 1:100 with nuclease-free water, then run in triplicate for each miRNA and U6. Relative expression was determined using the comparative 2-ΔCt method (Livak & Schmittgen, 2001).

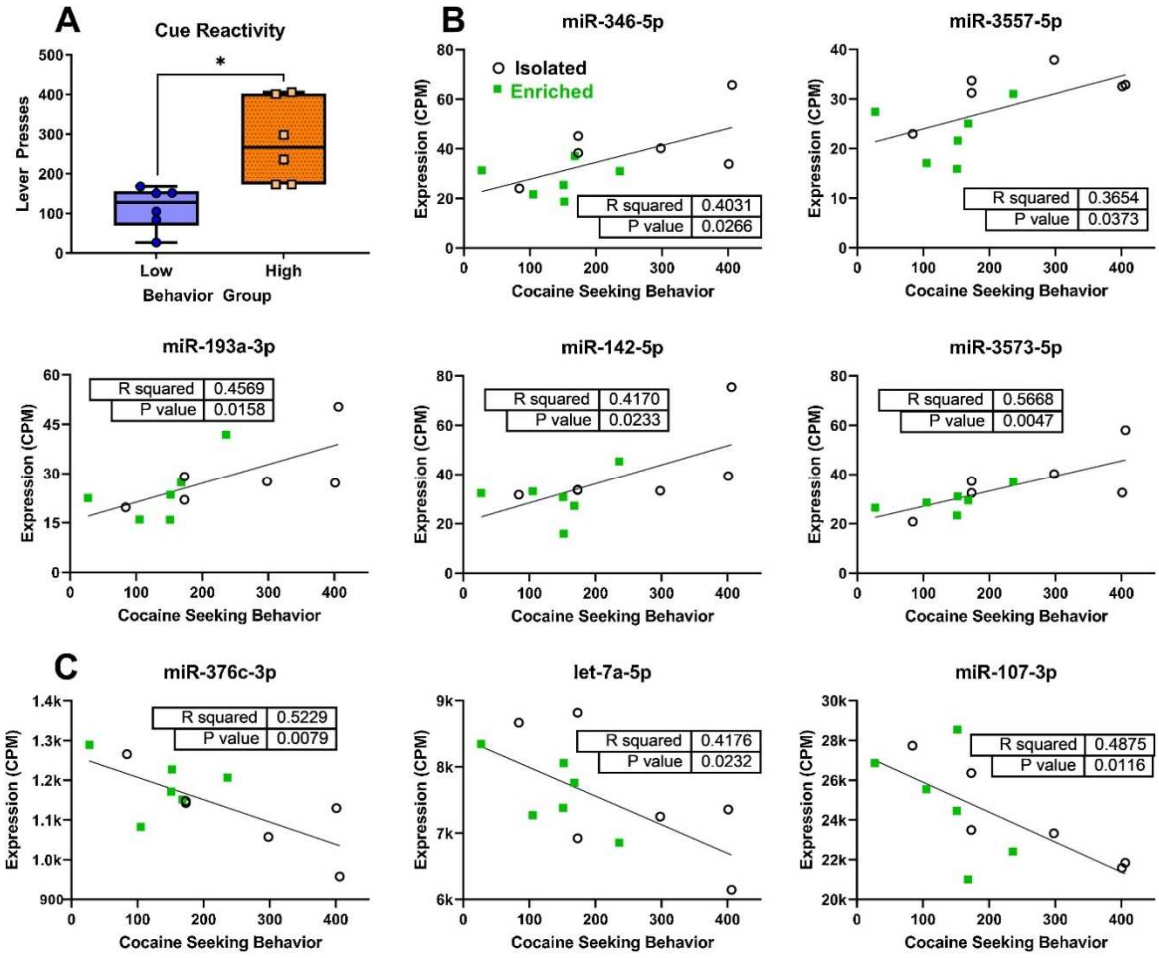
**Statistical analysis.** Rats were divided into groups based on median split of cocaine-seeking behavior. Statistical calculations were performed in GraphPad Prism 8, or R 3.6 (R Core Team, 2014). Linear regressions were used to analyze the correlation between miRNA levels (CPM) and cocaine-seeking behavior. For differential expression

analysis, p-values and false discovery rates (FDR) using the Benjamini-Hochberg method were calculated using the R package "limma" (Ritchie et al., 2015). The statistical threshold for all tests was $p < 0.05$.

**Data availability.** Nanostring data are deposited in the Gene Expression Omnibus (GSE153524). R code used for data analyses are available at https://gitlab.com/neisewander_asu/vannan-powell-2020.

## Results

**Differentially-expressed miRNAs in the NAc shell correlate with cocaine-seeking behavior.** The high and low cocaine-seeking groups (n=6/group) derived from the median split of cocaine seeking values were significantly different in their cocaine-seeking behavior [$t(10) = 3.452$, $p = 0.0062$, Figure 1A]. Cocaine-seeking values aligned well with housing condition: the "high" cocaine-seeking group consisted of 83.3% (5) ISO and 16.7% (1) EE rats and vice versa for the "low" cocaine-seeking group. In total, expression levels of 75 miRNAs were significantly correlated with cocaine-seeking behavior (Supplementary Table 1), although not all of these miRNAs were differentially expressed in the low vs. high seeking groups. Analysis of Nanostring counts using limma identified 33 miRNAs that were differentially expressed in the NAc shell in animals that displayed high versus low cocaine-seeking behavior (Table 1). Of these, 8 were downregulated and 25 were upregulated (Fold Change >1 and <1 on Table 1, respectively) in the low-seeking group relative to the high-seeking group. For Table 1, log2 fold change values have been converted to linear values where Fold Change = 2^(log2 values). In addition, 8 of the 33 miRNAs had expression levels that correlated with cocaine-seeking behavior, 5 positively and 3 negatively (Figure 1C).




***Chapter 2, Figure 1.*** Correlation of miRNA expression levels and the number of active lever presses during the cue reactivity test in each animal. (A). Separation of groups of the low- (blue) and high-seeking (orange) groups determined by median split of active lever presses. Boxes indicate median and quartiles; whiskers indicate minimum and maximum. * indicates difference from low-seeking group, $p < 0.05$, independent samples t-test. Positive (B) and negative (C) Pearson correlations of miRNA expression with cocaine-seeking behavior, as measured by active lever presses during the cue reactivity test. Isolated and enriched animals are depicted with open black circles and green squares, respectively. Only miRNAs that had both significantly different expression between high- and low-seeking groups and that correlated significantly with behavior (table inset on each graph) are shown here.

Chapter 2, Table 1

***Differentially-Expressed miRNAs between the High- and Low-Seeking Groups***

| miRNA | High-Seeking (CPM)[1] | Low-Seeking (CPM) | Fold Change[2] | P-Value | FDR[3] |
|---|---|---|---|---|---|
| miRNAs with higher expression in high cocaine-seeking animals | | | | | |
| miR-463-3p | 37.04 | 23.26 | 1.447 | 0.0188 | 0.3885 |
| miR-346-5p | 42.42 | 26.40 | 1.439 | 0.0114 | 0.3291 |
| miR-483-3p | 45.76 | 29.62 | 1.431 | 0.0036 | 0.3108 |
| miR-3557-5p | 33.24 | 21.68 | 1.417 | 0.0074 | 0.3108 |
| miR-193a-3p | 32.98 | 20.81 | 1.390 | 0.0275 | 0.4400 |
| miR-133a-3p | 68.68 | 46.95 | 1.378 | 0.0133 | 0.3291 |
| miR-142-5p | 43.47 | 28.52 | 1.342 | 0.0385 | 0.4906 |
| miR-3573-5p | 39.58 | 26.59 | 1.331 | 0.0204 | 0.3993 |
| miRNAs with higher expression in low cocaine-seeking animals | | | | | |
| miR-29a-3p | 68099.60 | 70947.29 | 0.878 | 0.0360 | 0.4868 |
| miR-16-5p | 10815.91 | 11290.58 | 0.875 | 0.0465 | 0.5104 |
| miR-93-5p | 306.12 | 324.61 | 0.864 | 0.0326 | 0.4783 |
| miR-495-3p | 6727.51 | 7266.45 | 0.850 | 0.0422 | 0.4906 |
| miR-376c-3p | 1106.98 | 1197.90 | 0.845 | 0.0289 | 0.4429 |
| miR-410-3p[a] | 2722.19 | 2973.03 | 0.841 | 0.0239 | 0.4205 |
| miR-329-3p | 5361.27 | 5863.05 | 0.834 | 0.0432 | 0.4906 |
| let-7a-5p | 7226.04 | 7915.63 | 0.833 | 0.0258 | 0.4320 |
| miR-652-3p | 734.49 | 800.55 | 0.832 | 0.0486 | 0.5104 |
| miR-377-3p | 290.24 | 320.87 | 0.832 | 0.0418 | 0.4906 |
| miR-107-3p | 23171.08 | 25691.54 | 0.828 | 0.0150 | 0.3291 |
| miR-138-5p | 1727.38 | 1915.73 | 0.822 | 0.0225 | 0.4175 |
| miR-487b-3p | 4030.70 | 4628.25 | 0.802 | 0.0066 | 0.3108 |
| miR-344b-1-3p/ miR-344b-2-3p[a] | 357.45 | 411.75 | 0.793 | 0.0026 | 0.3108 |
| miR-128-3p | 1644.39 | 1883.53 | 0.791 | 0.0346 | 0.4868 |
| miR-137-3p | 2451.13 | 2888.88 | 0.790 | 0.0147 | 0.3291 |
| miR-323-3p | 1155.50 | 1336.43 | 0.790 | 0.0088 | 0.3108 |
| miR-337-3p | 317.90 | 377.92 | 0.776 | 0.0051 | 0.3108 |
| miR-125b-3p | 1341.30 | 1589.26 | 0.772 | 0.0125 | 0.3291 |
| miR-409-5p | 326.65 | 389.53 | 0.771 | 0.0049 | 0.3108 |
| miR-218a-5p | 6391.04 | 7740.13 | 0.759 | 0.0393 | 0.4906 |

| | | | | | |
|---|---|---|---|---|---|
| miR-212-3p[b] | 354.77 | 448.44 | 0.728 | 0.0078 | 0.3108 |
| miRNA | High-Seeking (CPM)[1] | Low-Seeking (CPM) | Fold Change[2] | P-Value | FDR[3] |
| miR-130b-3p | 1279.74 | 1616.11 | 0.719 | 0.0104 | 0.3291 |
| miR-221-3p | 862.09 | 1078.44 | 0.718 | 0.0087 | 0.3108 |
| miR-132-3p[b] | 25643.46 | 32814.25 | 0.710 | 0.0188 | 0.3885 |

[a, b]Indicate miRNAs in the same family.

1 miRNA expression levels were calculated as counts per million (CPM) and where available, 3p or 5p designations are included.

2 Fold change is an estimate of the effect derived from the log2 fold changes from "limma", which were then converted to linear values with the equation 2^(log2 value). Values <1 signifies higher expression in high-seeking animals, and values >1 signifies higher expression in low-seeking animals.

3 FDR = false discovery rate.

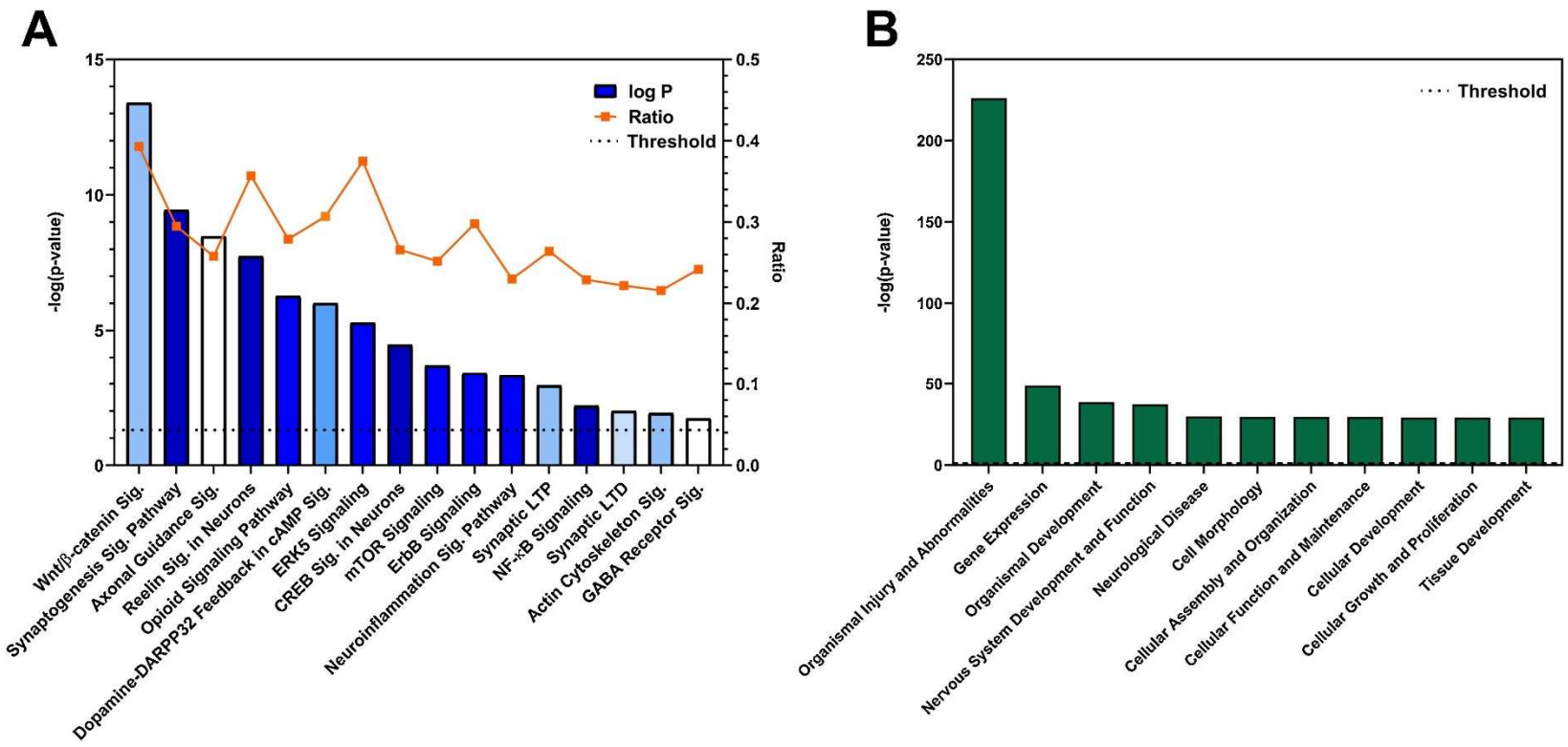
Nanostring results for miR-376c-3p, miR-107-3p, and miR-212-3p, which were all elevated in low-seeking animals, were validated using RT-qPCR (Supplementary Figure 1). Subsequent t-tests including all rats (n=6 of each group) did not show significant differences in expression between high- and low-seeking groups for these miRNAs. However, separating the seeking groups into quartiles (including only the 3 highest and 3 lowest cocaine-seeking animals), revealed significantly higher expression in the low-seeking group for miR-376c-3p [$t(4) = 3.05$, $p = 0.0379$, Supplementary Figure 1A] and miR-107-3p [$t(4) = 2.81$, $p = 0.0481$, Supplementary Figure 1B]. Although the values for miR-212-3p did not quite meet the threshold for significance after separating animals into quartiles [$t(4) = 2.63$, $p = 0.0583$], they did correlate significantly with behavior overall ([$F(1,10) = 6.16$, $p = 0.0324$]) (Supplementary Figure 1C).

**Predicted mRNA targets of the differentially-expressed miRNAs.**

TargetScan 7.2 was utilized to identify predicted mRNA targets of the differentially-

expressed miRNAs for the rat, which were then utilized in subsequent analyses. The number of predicted targets varied for each miRNA and ranged from 11 (miR-487b-3p) to 3,972 (miR-3557-5p).

To create a list for input into IPA, first we determined the overall impact of our miRNAs on the list of predicted targets (mRNAs) relative to the low-seeking condition. If a miRNA was upregulated in low cocaine-seeking animals, its predicted targets were given a score of -1, as they would be downregulated by that miRNA, whereas targets of downregulated miRNAs were given a score of +1. For example, the mRNA Nuclear factor I B (*Nfib*) is a predicted target of 16 differentially-expressed miRNAs, of which 3 were downregulated in the low-seeking group (+3) and 13 of which were upregulated (-13), leading to a total impact score of -10 (Supplementary Table 2). We began with 9,761 predicted targets, which comprised 23.76% of the transcribed genes in the tissue analyzed. Because more miRNAs were upregulated rather than downregulated in the low-seeking condition, predicted targets were more likely to have a negative impact score, and negative impact scores were greater in magnitude: 6,456 targets had a negative impact score (range: -1 to -14); 3,305 had a positive or 0 impact score (range: 0 to +3.) Because IPA analysis is more robust with smaller lists of genes, we prioritized candidates based on impact score. Only mRNA targets that were mapped to IPA and had an impact score of -3 or lower and +1 or higher were included, reducing the list to 3,600 predicted mRNA targets.




***Chapter 2, Figure 2.*** Summary of pathway analysis using IPA. Panels depict several pathways (A) and diseases and functions (B) related to the predicted targets of the differentially-expressed miRNAs. Threshold levels indicate –[log (p = 0.05)]. Bar colors represent the ranges of the z-score calculated by IPA. Darker shades indicate z-scores farther from zero, and blue indicates a negative z-score.

The majority of mRNAs were predicted targets of only 1 or 2 differentially-expressed miRNAs (6,128; 62.8%). However, 9 mRNAs were targets of 15 or more miRNAs: *Zinc finger and BTB domain containing 20* (*Zbtb20*), *Nuclear factor of activated T-cells 5* (*Nfat5*), *Argonaute RNA-induced silencing complex (RISC) component 1* (*Ago1*), *Nfib*, *Protein quaking* (*Qki*), Phosphodiesterase 3A (*Pde3a*), *Retinoic acid receptor-related orphan receptor B* (*Rorb*), *Transcription factor 4* (*Tcf4*), and *Kruppel like factor 7* (*Klf7*), with impact scores ranging from -8 to -14. In total, 136 mRNAs (1.39%) were predicted targets of at least 10 differentially-expressed miRNAs.

IPA revealed many significant pathways, Including *Wnt/ β-catenin Signaling*, *Synaptogenesis Signaling Pathway*, *Axonal Guidance Signaling*, *Dopamine/DARPP34 Feedback in cAMP Signaling*, *CREB Signaling in Neurons*, and *ERK5 Signaling*, among others (Figure 2A, Supplementary Table 3). Significant Diseases and Functions were also provided by IPA (Figure 2B, Supplementary Table 4) and included: *Learning* in the category *Behavior*; *Development of Neurons* and *Morphogenesis of Neurons* in the categories *Cellular Development* and *Cellular Growth and Proliferation*; *Migration of Neurons* in the category *Cellular Movement*; *Transcription of RNA* and *Expression of RNA* in the category *Gene Expression*; and *Neurotransmission* in the category *Nervous System Development and Function* (Supplementary Tables 5-10).

**miRNAs and addiction-related genes.** For each miRNA, the number of predicted targets found in TargetScan was also compiled and then cross-referenced to addiction-related genes in the KARG database (Table 2). Three miRNAs, miR-3557-5p, miR-377-3p, and miR-337-3p, targeted a large percentage of KARG (between 17.00% and 24.21%). Together, the putative targets of significant miRNAs covered 205 (59.1%) of the 347 KARG genes included in our analysis. Several addiction-related mRNAs were predicted to be targets of at least 10 differentially-expressed miRNAs: *Nuclear factor I A*

(*Nfia*), *Nfib*, *Circadian locomotor output cycles protein kaput* (*Clock*), *Ataxin 1* (*Atxn1*), *cyclic AMP (cAMP)-responsive element binding protein 1* (*Creb1*), *Zinc finger and BTB domain containing 16* (*Zbtb16*), *Purine rich element binding protein A* (*Pura*), and *Gamma-aminobutyric acid (GABA) type B receptor subunit 2* (*Gabbr2*), potentially indicative of their key role in cocaine-seeking and drug motivation (Supplementary Table 11). Notably, all these mRNAs were predicted to be downregulated in the low-seeking group compared to high-seeking, with impact scores between -4 (*Zbtb16*) and -11 (*Atxn1*) (Supplementary Table 2).

## Discussion

NanoString nCounter analysis of the NAc shell of male rats with a history of cocaine self-administration that were tested for cocaine-seeking behavior after 21 days of abstinence identified 33 miRNAs displaying differential expression in the "high" and "low" cocaine-seeking groups. Of these, expression of 8 miRNAs correlated significantly with cocaine-seeking behavior. Predicted mRNA targets of the 33 miRNAs were analyzed using IPA, which revealed several significant pathways including *Synaptogenesis* and *Opioid Signaling*. Cross-reference of TargetScan and the KARG database showed that many of these miRNAs were predicted to target mRNAs of addiction-related genes.

Chapter 2, Table 2

***Overlap of the Predicted mRNA Targets for Each miRNA in TargetScan and KARG Databases.***

| miRNA | TargetScan[1] | KARG[1] | % KARG[2] |
|---|---|---|---|
| miRNAs with higher expression in high cocaine-seeking animals | | | |
| miR-463-3p | 1396 | 26 | 7.49 % |
| miR-346-5p | 114 | 6 | 1.73 % |
| miR-483-3p | 79 | 4 | 1.15 % |
| miR-3557-5p | 3972 | 84 | 24.21 % |
| miR-193a-3p | 190 | 5 | 1.44 % |
| miR-133a-3p | 549 | 25 | 7.20 % |
| miR-142-5p | 704 | 21 | 6.05 % |
| miR-3573-5p | 129^ | 6 | 1.73 % |
| miRNAs with higher expression in low cocaine-seeking animals | | | |
| miR-29a-3p | 1013 | 32 | 9.22 % |
| miR-16-5p | 1090^ | 30 | 8.65 % |
| miR-93-5p | 1056^ | 29 | 8.36 % |
| miR-495-3p | 616 | 22 | 6.34 % |
| miR-376c-3p | 188^ | 7 | 2.02 % |
| miR-410-3p[a] | 497^ | 14 | 4.03 % |
| miR-329-3p | 255^ | 7 | 2.02 % |
| let-7a-5p | 1022^ | 19 | 5.48 % |
| miR-652-3p | 14 | 1 | 0.29 % |
| miR-107-3p | 558^ | 22 | 6.34 % |
| miR-138-5p | 537 | 16 | 4.61 % |
| miR-487b-3p | 11 | 1 | 0.29 % |
| miR-344b-1-3p/ miR-344b-2-3p[a] | 497^ | 14 | 4.03 % |
| miR-128-3p | 950 | 25 | 7.20 % |
| miR-137-3p | 1019 | 26 | 7.49 % |
| miR-323-3p | 357 | 10 | 2.88 % |
| miR-337-3p | 3243 | 65 | 18.73 % |
| miR-125b-3p | 443 | 9 | 2.59 % |
| miR-409-5p | 100 | 2 | 0.58 % |
| miR-218a-5p | 869 | 26 | 7.49 % |
| miR-212-3p[b] | 360^ | 14 | 4.03 % |

| | | | |
|---|---|---|---|
| miR-130b-3p | 748^ | 19 | 5.48 % |
| miR-221-3p | 366^ | 14 | 4.03 % |
| miR-132-3p[b] | 360^ | 14 | 4.03 % |

^ Represents miRNAs that were present on TargetScan as miRNA families, not individual miRNAs. miRNAs in the same family share their seed sequence and thus the same predicted miRNA targets.

[a, b] Indicate miRNAs in the same family.

1 Values are the number of predicted mRNA targets found in each respective database.

2 Percentage of KARG that the predicted targets comprise. Colors indicate miRNAs relative to the other group.

Many of the differentially-expressed miRNAs identified in this study have been previously implicated in drug abuse, including miR-29a, miR-16, miR-495, mir-376c, miR-329, miR-138, miR-137, miR-337, miR-125b, miR-212, miR-130b, miR-221, and miR-132 (Bastle et al., 2018; Dave & Khalili, 2010; Eipper-Mains et al., 2011; Hollander et al., 2010; Im & Kenny, 2012; Lippi et al., 2011; Schaefer et al., 2010; Shin et al., 2010). Interestingly, these miRNAs all had greater expression in the low-seeking group compared to the high-seeking group. Of these, miR-212, a CREB-induced activity-dependent miRNA in the same family as miR-132, is perhaps the best-studied in the addiction field. For example, Sadakierska-Chudy et al. (Sadakierska-Chudy et al., 2017) found that 2-h daily access to cocaine increases both miR-212 and miR-132 in the dorsal striatum compared to saline-yoked controls, and this increase is persistent, lasting 10 days into subsequent extinction training. In addition, Hollander et al. (Hollander et al., 2010) demonstrated that miR-212 and miR-132 may be involved in the transition from casual to compulsive drug use, as both are upregulated in the dorsal striatum following extended (6-h) daily cocaine self-administration compared to cocaine-naïve rats. They also found that striatal miR-212 overexpression reduces compulsive-like cocaine-taking behavior during extended access (6-h daily sessions), while knockdown produces more compulsive cocaine consumption. Due to the close relatedness of miR-132 to miR-212, the authors suggest miR-132 may play a similar role in compulsive cocaine-taking. In the present study, the increased expression of miR-212 and miR-132 in the NAc shell of rats with low cocaine-seeking behavior suggests that these miRNAs are protective against motivation for cocaine. Together, it appears that striatal miR-212 and miR-132 expression may shield against two defining phases of CUDs: transition to an addicted-like phenotype as well as craving during protracted abstinence.

Among the miRNAs previously associated with motivation for cocaine is miR-495, a miRNA that we identified because its levels decrease during cocaine self-administration and its over-expression in the NAc shell attenuates motivation for cocaine (Bastle et al., 2018). We have shown that miR-495 regulates expression of multiple addiction-related genes including *Brain derived neurotrophic factor* (*Bdnf*), *Calcium-calmodulin activated protein kinase IIα* (*Camk2a*) and Activity-regulated cytoskeleton-associated protein (*Arc*) among other mRNAs such as *Per2* and *Gria3* (Bastle et al., 2018). Here, we found that miR-495 has higher expression in low-seeking animals, supporting our prior research that suggests upregulation of miR-495 is protective against motivation for drugs of abuse.

To identify novel candidate miRNAs that may regulate cocaine-related behavior in this study we cross-referenced TargetScan and KARG. Of the 33 differentially-expressed miRNAs, miR-3557-5p, miR-377-3p, and miR-337-3p were predicted to target particularly high percentages of the KARG database (>17.00%, up to 24.14%). Of these 3, miR-3557-5p and miR-377-3p have not been studied in substance abuse or psychiatric illness to our knowledge, suggesting they may be novel targets for addiction research. However, miRNAs that were predicted to target a large percentage of KARG also had many TargetScan predicted targets (>2000). Caution is needed when the number of predicted targets is so large, due to an increased likelihood of false positives and limitations of these databases. By contrast, miRNAs with fewer than 200 predicted targets (miR-346-5p, miR-483-3p, miR-193a-3p, miR-3573-5p, miR-376c-3p, miR-652-3p, miR-487b-3p, miR-409-5p) had little overlap with the KARG database (under 2.02%).

Among the miRNAs predicted to target expression of a large number of addiction-related genes, miR-337-3p has been the subject of prior drug abuse research.

Here, we found that miR-337-3p has significantly higher expression in the low-seeking group, suggesting it may be protective against cocaine-seeking behavior. However, in striatal Dopamine Receptor 2 (Drd2)-expressing neurons, miR-337-3p is upregulated after acute cocaine injection in mice (Schaefer et al., 2010), suggesting that this miRNA may be associated with the initial neurobiological changes after drug exposure. Similarly, miR-376c, miR-138, and miR-137, which all have significantly higher expression in the low-seeking group in our study, are also upregulated in striatal Drd2-expressing neurons after acute cocaine injection (Schaefer et al., 2010). Therefore, these miRNAs may either protect against or facilitate addictive behaviors depending on factors such as previous drug experience. Other factors that may contribute to these seemingly contradictory results include differences in the brain region studied and cell-type specificity.

Several addiction studies have identified the importance of let-7 miRNAs, particularly let-7d (Chandrasekar & Dreyer, 2009, 2011; He et al., 2010; Hollander et al., 2010), although these miRNAs have few addiction-related targets according to KARG (1.86%). For example, let-7d expression is decreased in regions of the mesolimbic reward pathway including the NAc core and shell, striatum, and ventral tegmental area after 15 days of daily cocaine injections compared to saline controls (Chandrasekar & Dreyer, 2009), whereas overexpression of let-7d in the NAc attenuates cocaine conditioned place preference (Chandrasekar & Dreyer, 2011). Here, we demonstrate that another let-7 family member, let-7a, has higher expression in low-seeking animals and displays a significant negative correlation with cocaine-seeking. Together, these data suggest let-7 miRNAs may modulate different aspects of cocaine-related behavior. Although TargetScan assumes the same predicted targets for all let-7 miRNAs, differences in their biogenesis and expression patterns may contribute to distinct roles in neuronal function and thus drug abuse (Roush & Slack, 2008).

Many of the miRNAs identified here have been explored primarily in relation to other psychiatric illnesses, such as schizophrenia and depression, which have high comorbidity with substance abuse (Batel, 2000; Kessler et al., 2005; Kosten et al., 1998; Paykel et al., 2005). For example, miR-16, miR-495, miR-652, miR-107, miR-138, and miR-137 have been linked to schizophrenia (Beveridge et al., 2010; Moreau et al., 2011; Ripke et al., 2011; Santarelli et al., 2011; Wright et al., 2013) and were all upregulated in the low-seeking group compared to the high-seeking animals in the present study. Similarly, miR-16 has been implicated in depression, and is believed to underlie the therapeutic effects of the antidepressant fluoextine through targeted downregulation of the serotonin transporter (SERT) mRNA (*Slc6a4*) (Baudry et al., 2010). Importantly, antidepressants that blocks serotonin reuptake through SERT, encoded by the gene Slc6a4, are effective in reducing cocaine-seeking and -taking in some preclinical models (Baker et al., 2001; Burmeister et al., 2003; Harris et al., 2001; Richardson & Roberts, 1991). This suggests that miR-16-5p, which was found here to have higher expression in low-seeking rats, may be therapeutic for treating addiction by reducing depressive symptoms. The overlap of miRNAs implicated in addiction and comorbid psychiatric illnesses may help inform treatments for those suffering from addiction occurring in conjunction with, or exacerbated by, other conditions.

We utilized IPA to uncover pathways that are potentially regulated by the 33 differentially-expressed miRNAs by inputting the miRNAs' predicted mRNA targets and the estimated impact of the miRNAs on their expression (i.e. impact score). This analysis revealed several pathways, including *Axonal Guidance Signaling*, *Opioid Signaling*, and *ERK5 signaling*, that are potentially regulated by the differentially-expressed miRNAs. Many of these pathways have previously been implicated in addiction, including *WNT/β-catenin signaling* (Cuesta & Pacchioni, 2017) and

*Neuroinflammation signaling pathway* (Clark et al., 2013). These results also validate our recent RNA-seq analysis of the NAc shell in animals that underwent the same training and testing procedures as the present study, except that in addition to EE and ISO housing, animals also underwent different lengths of forced abstinence (1 or 21 days) (Powell, Vannan, et al., 2020). We found that contrasting EE and ISO animals given 21 days of abstinence, similar to the present study, implicated several of the pathways found here, including *Synaptogenesis Signaling*, *Reelin Signaling*, *Neuroinflammation Signaling*, *Synaptic Long-Term Potentiation*, and *CREB Signaling in Neurons*, and that *Bdnf*, a widely studied addiction gene (Xuan Li & Wolf, 2015), was a top upstream regulator of this comparison. In the present study, 7 miRNAs are predicted to target *Bdnf* with an impact score of -7. IPA also revealed significant functions of the predicted targets including *Dendritic Growth/Branching* and *Morphology of Dendritic Spines*, which support prior research that several of our miRNAs of interest, including miR-29a, miR-329, miR-137 and miR-132, regulate dendritic spine formation and morphology (Impey et al., 2010; Lippi et al., 2011; Smrt et al., 2010) Indeed, our IPA results both validate and expand on current knowledge by implicating several miRNAs in cue-elicited cocaine-seeking behavior.

Of the pathways identified in the current study, CREB signaling is one of the most highly studied in addiction (William A. Carlezon et al., 2005; Krasnova et al., 2016; Kreibich et al., 2009; Larson et al., 2011; Mattson et al., 2005). As mentioned earlier, increased miR-212 expression is related to reduced compulsive-like cocaine-taking behavior, which is thought to involve increasing CREB signaling and decreasing MeCP2/BDNF signaling (Hollander et al., 2010; Im et al., 2010). Similarly, our prior study suggested *CREB Signaling in Neurons* is an important mechanism underlying cocaine-seeking behavior, and found that Creb1 was a top upstream regulator of the

differentially-expressed RNAs in the contrast between EE and ISO animals with 21 days of abstinence (Powell, Vannan, et al., 2020). Here, we found that 11 of the 33 differentially-expressed miRNAs in high- vs. low-cocaine seeking animals are predicted to target *Creb1* (impact score = -7), suggesting that these miRNAs may be important regulators of the pathways identified in our previous study.

A caveat of the median split used to examine the effects of housing environment during forced abstinence in the present study is that animals at the outer range of the housing groups showed nearly the same lever pressing during the cue reactivity tests, as can be seen in Figure 1A. We considered using a quartile analysis, which includes only the 3 animals with the greatest and least cue reactivity (i.e., 6 total animals representing the first and fourth quartiles of the total dataset), however, a drawback to this approach is the loss of power due to smaller sample size, reducing the capacity to find potentially relevant changes in miRNA expression. Our follow up analyses showing significant correlations of cocaine-seeking behavior with expression of 75 individual miRNAs, including 8 of the 33 miRNAs that are differentially expressed between "low" and "high" reactivity groups, mitigated concern with our approach. Furthermore, a preliminary comparison to the quartile analysis showed 11 differentially-expressed miRNAs, 4 of which are shared with the original analysis (miR-346-5p, miR-193a-3p, miR-3573-3p, and miR-107-3p). These 4 miRNAs, along with 4 of the 7 uniquely identified miRNAs (miR-301b-3p, miR-3561-3p, miR-3558-3p, and miR-208b-3p), are significantly correlated with behavior. We also used the quartile analysis in our follow-up RT-qPCR validation experiments, which confirmed increased expression of miR-276c and miR-107-3p in the low-seeking group relative to high-seeking animals, and additionally showed a significant correlation of miR-212-3p expression with cocaine-seeking behavior with all 12 animals. Finally, Nanostring nCounter employs dual probes and hybridization

to directly measure target molecules without the bias of amplification-dependent techniques (e.g. RT-qPCR) (Eastel et al., 2019), and though not as sensitive as small RNA-seq, Nanostring displays improved detection of lowly-expressed miRNAs that might be missed by techniques such as microarrays (Eastel et al., 2019; Kulkarni, 2011). This is especially important for detecting disease biomarkers, which may be present in low abundance (Foye et al., 2017). Indeed, for the present study, all 8 differentially-expressed miRNAs that were elevated in the high-seeking group displayed low CPM values. Still, Nanostring is a proprietary platform, which somewhat limits flexibility and future use. In addition, the specificity of the assay depends largely on the design of the probes (Eastel et al., 2019); here, however, we used a pre-designed assay from Nanostring that has been widely used by other researchers (Chaudhuri et al., 2018; Mellios et al., 2018; Murphy et al., 2014), bolstering confidence in our approach.

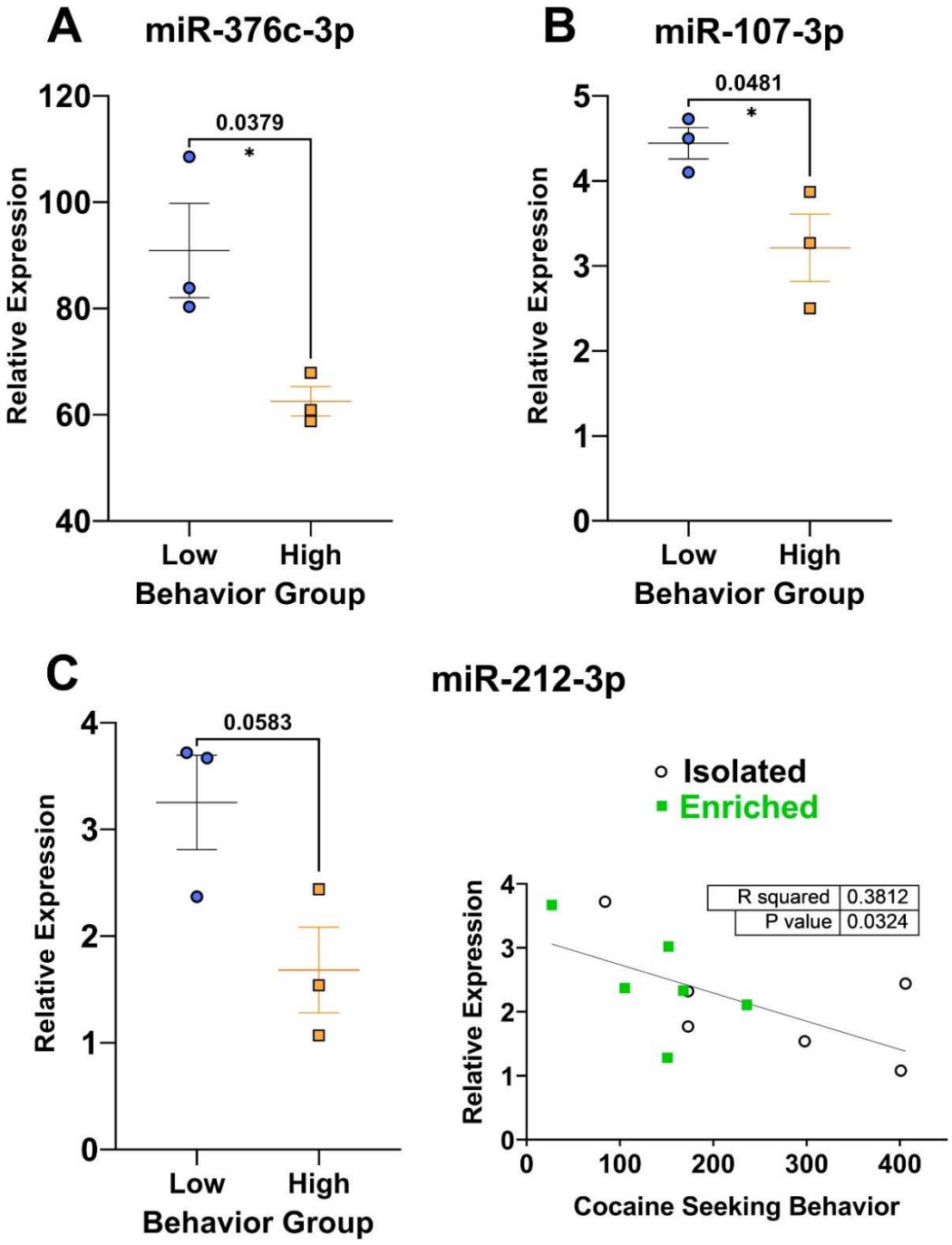
**Conclusion**

In this study, we identified 33 miRNAs that are differentially-expressed in rats displaying high versus low cocaine-seeking behavior. Although this study focused on motivation for cocaine during abstinence, it is possible that the miRNAs identified may be relevant to other aspects of CUDs, as others have linked miRNAs to acute cocaine exposure (Bastle et al., 2018), escalation of cocaine self-administration (Hollander et al., 2010), and cocaine CPP (Chandrasekar & Dreyer, 2011; Viola et al., 2016). Understanding the role of these miRNAs in motivation for cocaine may lead to novel treatments as currently pioneered in the cancer field, where some miRNA therapeutics have even advanced to clinical trials (Bonneau et al., 2019; Wahid et al., 2010). Further research on the role of miRNAs in CUDs will aid in understanding the underlying

mechanisms involved and position the field to capitalize on the knowledge for development of treatments.

## Supplementary information

Supplementary Figure 1 is included at the end of this chapter. Supplementary Tables 1-11 are provided on GitHub (https://github.com/SexChrLab/Vannan-Dissertation-Supplements).




***Chapter 2, Supplementary Figure 1.*** Select differentially-expressed miRNAs from the Nanostring analysis were validated with RT-qPCR. Animals were separated into quartiles by cocaine-seeking behavior, and the 3 lowest (blue) and 3 highest (orange) seekers are depicted. Low-seeking animals had higher expression of miR-376c-3p (A) and miR-107-3p (B) compared to high seekers. For miR-212-3p (C), this comparison was not significant, though there was a significant correlation of relative expression derived from RT-qPCR and cocaine-seeking behavior. For the correlation, isolated and enriched animals are depicted with open black circles and green squares, respectively.

CHAPTER 3

# AN APPROACH FOR PRIORITIZING CANDIDATE GENES FROM RNA-SEQ USING PRECLINICAL COCAINE CRAVING DATASETS AS A TEST CASE

**Submitted for publication as:**

Vannan, A., Dell'Orco, M., Perrone-Bizzozero, N. I., Neisewander, J. L., Wilson M. A. An approach for prioritizing candidate genes from RNA-seq using preclinical cocaine craving datasets as a test case.

## Abstract

The addiction field relies heavily on rodent models to probe the complex molecular mechanisms underlying drug motivation and reward. However, findings in animal research often fail to be reproduced or translated into clinical treatments. Here, we developed and applied standardized workflows to three RNA-sequencing (RNA-seq) studies in rodent self-administration models to assess reproducibility and translational viability of potential gene targets. Generally, we found low overlap of differentially-expressed genes (DEGs) across datasets using a significance level of $P < 0.05$, and only 1 DEG in any dataset after correcting for multiple testing (FDR < 0.05 or FDR < 0.1), raising concerns about false positives. When using a nominal P value, which is a standard used in the field, we found potential genomic targets for drug craving and withdrawal that overlap across datasets. Several of these shared genes are conserved in humans and/or are highly expressed in reward regions of the brain, but the sample sizes in all datasets limit the power to distinguish true positives from false positives. Thus, we advocate for improved RNA-seq experimental design, data collection, statistical testing,

and metadata reporting that will bolster the field's ability to communicate across research groups and advance addiction science in a more reproducible manner.

## Introduction

Substance use disorders (SUDs) are a critical health issue resulting in substantial financial and social costs in the U.S. and worldwide (Cartwright, 2008). SUDs are driven by gene-environment interactions and manifest in complex behavioral phenotypes that are difficult to disentangle in clinical populations (Buckland, 2008; Goldman et al., 2005). Because of this, few studies in human cocaine abusers are able to incorporate both behavioral information and an unbiased evaluation of genetic data, e.g. through a genome-wide association study (GWAS) or next-generation sequencing (though see (Marees et al., 2020). Thus, animal models have been critical to our understanding of the behavioral and genetic correlates of drug abuse. The self-administration (SA) paradigm, in which animals learn to repeatedly perform a behavior (e.g. lever press or nose poke) to receive a drug reinforcer, has long been considered the gold standard of preclinical addiction research (Mello & Negus, 1996). Variations on the SA model have been used to study many aspects of substance abuse, including the pharmacological effects of drugs, initiation and maintenance of drug use, and escalation of drug intake, among others (see (Belin-Rauscent et al., 2016; Sanchis-Segura & Spanagel, 2006; M. A. Smith, 2020) for reviews and critiques). Many preclinical paradigms have shown strong predictive validity for human behavior (Spanagel, 2017), but insights from rodent models have often failed to translate to effective clinical treatments (Kalant, 2010; Venniro et al., 2020).

Drug craving increases or “incubates" after prolonged withdrawal, which promotes drug-seeking behavior in animals and is thought to precede relapse in human drug abusers (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). Because animal models allow researchers to probe biological mechanisms with much greater detail than is viable in human subjects, the translational relevance of SA relies not just on behavioral similarity between humans and animals, but on convergent

neurobiology. While the neural circuitry of reward is broadly shared between mammals (Martínez-García & Lanuza, 2018; Panksepp et al., 2002; Scaplen & Kaun, 2016), gene expression variation throughout development between humans and rodents (Cardoso-Moreira et al., 2019, 2020), may have implications for developing treatments based on rodent data. Therapeutic development remains a critical goal of preclinical research, particularly for CUDs, which have no FDA-approved therapeutics despite successful medications development for other SUDs (Lerman et al., 2007; Soyka & Müller, 2017; Veilleux et al., 2010) and increased cocaine-related overdose deaths in recent years (Cano et al., 2020). Even more challenging for studying CUDs is that there exist clear, and sometimes stark, differences in behavioral and neurobiological responses to drugs, both between rodent species (Ary & Szumlinski, 2007; Cunningham et al., 1993; Ellenbroek & Youn, 2016; Spanagel, 2017; H.-Y. Zhang et al., 2015) and even between strains of the same species (Crabbe, 2002; McClearn & Rodgers, 1959). Some researchers have begun to tackle these issues by leveraging data across rodents and humans to hone in on convergent pathways and genes in drug abuse that may lead to more successful therapeutic targets (Forero & González-Giraldo, 2020; Huggett et al., 2020; Huggett & Stallings, 2020).

Here, we developed and implemented a pipeline to use multiple lines of evidence to identify candidate genes involved in craving and abstinence. We used three publicly-available RNA-sequencing (RNA-seq) datasets from rodents that underwent cocaine SA and prolonged abstinence to identify reproducible signals with convergent preclinical evidence. Though these studies had some methodological variation, each had the key, common feature of prolonged cocaine abstinence, an important hallmark of the incubation effect that suggests all studies shared the underlying phenomenon of cocaine craving. However, we found low overlap of craving-related genes across datasets using a

nominal P value ($P < 0.05$) and only 1 significant gene in a single dataset after correcting for multiple testing (FDR < 0.05 or 0.1). Utilizing the shared genes ($P < 0.05$) as a test case, we demonstrate the need to consider evolutionary conservation and expression across central nervous system (CNS) tissues when selecting candidate genes for further study. We propose that using this methodology can narrow down candidate genes that are potential therapeutic targets due to likelihood of functional conservation with humans or expression levels in reward-related brain regions. However, low sample sizes in all datasets limit our ability to detect true positives. Therefore, we make specific recommendations for collecting and reporting RNA-seq data, which will facilitate communication across labs and allow us to leverage our collective knowledge to advance addiction science.

## Methods

**Data selection, access, and availability.** Three publicly available RNA-seq datasets were selected based on similarity in their experimental designs and the inclusion of treatment groups relevant to cocaine craving (Carpenter et al., 2020; Powell, Vannan, et al., 2020; Walker et al., 2018). Because the datasets have been published previously, we hereafter refer to them by their first authors (Carpenter, Powell, and Walker). Raw RNA-seq files were downloaded from NCBI's Sequencing Read Archive (Carpenter: SRP234876, Powell: SRP246331; Walker: SRP132477). Summaries of relevant design parameters and differences between studies are described in Table 1. Additional information about these files, detailed descriptions of analyses, and scripts are provided on GitHub (https://github.com/SexChrLab/RodentAddiction).

Chapter 3, Table 1

***Description of Cocaine Self-Administration Paradigms Used in Each Study***

| Study | Species | Sex | Surgical Anesthesia | Drug | SA Sessions | Schedule(s) of Reinforcement | Abstinence | Post-Abstinence |
|---|---|---|---|---|---|---|---|---|
| Carpenter et al. (2020) | Mouse | Male | Ketamine, xylazine | Cocaine (0.7 mg/kg) or Saline, IV | 2 hr/d, 21d | FR1 | 1 or 28 days | Extracted whole nucleus accumbens* |
| Walker et al. (2018) | Mouse | Male | Ketamine, xylazine | Cocaine (0.5 mg/kg) or Saline, IV | 2 hr/d, 10-15d | Progression from FR1 up to FR2 | 1 or 30 days | 1d abstinence: None<br>30d abstinence: Saline or Cocaine challenge injection and SA context re-exposure<br>Extracted whole nucleus accumbens* |
| Powell et al. (2020) | Rat | Male | Isoflurane | Cocaine (0.75 mg/kg), IV | 2 hr/d, ≥15d | Progression from FR1 up to VR5 | 1 or 21 days^ | Cue reactivity test<br>Extracted nucleus accumbens shell |

^Animals in this study were either maintained in their original housing (isolation) or moved to environmental enrichment conditions during abstinence. Only animals that underwent abstinence in isolation were considered for further analysis (see Table 2).

*Other tissues were extracted for these experiments, but only the nucleus accumbens was included in this analysis.

We focused on the nucleus accumbens (NAc), a mesolimbic region fundamental to drug motivation and reward (Floresco, 2015). Treatment groups were selected to create pairwise contrasts within an experiment that were similar to an experiment by another research group. In all three studies, rodents underwent cocaine SA and forced abstinence for a short (1d) or long (21-30d) interval. The Carpenter and Walker studies additionally included saline SA, allowing for contrasts between long-term abstinence from saline and long-term abstinence from cocaine. All three studies compare a treatment group likely to show low cocaine craving (either short term abstinence from cocaine SA or long term abstinence from saline SA) to a group likely to show higher craving (long term abstinence from cocaine SA) based on previous literature (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). Full descriptions of the treatment groups and contrasts performed for each study, along with abbreviated names for each contrast, are available in Table 2. Two contrasts were excluded from the analysis due to extreme batch effects or substantial differences between groups beyond the phenomenon of interest (abstinence-induced craving).

Chapter 3, Table 2

***Description of Tissues and Treatment Groups in Each Study That Were Considered in the Present Analysis***

| Study | Group Self-Administration (SA) and Abstinence (ABS) Conditions | Included Comparisons | Excluded Comparisons | Reason for Exclusion |
|---|---|---|---|---|
| Carpenter et al. (2020) | S1: Saline SA, 1d ABS<br>S28: Saline SA, 28d ABS<br>C1: Cocaine SA, 1d ABS<br>C28: Cocaine SA, 28d ABS | S28 vs. *C28* | 1. S1 vs. S28<br>2. S1 vs. *C1*<br>3. C1 vs. *C28* | 1. No cocaine groups<br>2. Both groups received 1d ABS; No incubation of cocaine craving<br>3. Extreme batch effects; Samples in each treatment group were analyzed in separate sequencing runs, with no overlap |
| Walker et al. (2018) | S30: Saline SA, 30d ABS, Saline challenge injection and context re-exposure<br>C1: Cocaine SA, 1d ABS<br>C30: Cocaine SA, 30d ABS, Saline challenge injection and context re-exposure | S30 vs. *C30* | 1. S1 vs. S30<br>2. S1 vs. *C1*<br>3. C1 vs. *C30* | 1. No cocaine groups<br>2. Both groups received 1d ABS; No incubation of cocaine craving<br>3. Groups are different not just in ABS length, but in post-abstinence testing |
| Powell et al. (2020) ^ | C1: Cocaine SA, 1d ABS, cue reactivity test<br>C21: Cocaine SA, 21d ABS, cue reactivity test | C1 vs. *C21* | N/A | N/A |

Only the treatment groups considered for the present analysis are described here. Where applicable, treatment groups that display relatively higher levels of cocaine craving or cocaine-seeking behavior, either measured directly within a study or predicted by the literature, are underlined and italicized.

^ For the Powell dataset, only animals that underwent abstinence in isolation were listed here (see Table 1).

We focused on the nucleus accumbens (NAc), a mesolimbic region fundamental to drug motivation and reward (Floresco, 2015). Treatment groups were selected to create pairwise contrasts within an experiment that were similar to an experiment by another research group. In all three studies, rodents underwent cocaine SA and forced abstinence for a short (1d) or long (21-30d) interval. The Carpenter and Walker studies additionally included saline SA, allowing for contrasts between long-term abstinence from saline and long-term abstinence from cocaine. All three studies compare a treatment group likely to show low cocaine craving (either short term abstinence from cocaine SA or long term abstinence from saline SA) to a group likely to show higher craving (long term abstinence from cocaine SA) based on previous literature (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). Full descriptions of the treatment groups and contrasts performed for each study, along with abbreviated names for each contrast, are available in Table 2. Two contrasts were excluded from the analysis due to extreme batch effects or substantial differences between groups beyond the phenomenon of interest (abstinence-induced craving).

**RNA-seq workflow and differential expression analysis.** The project workflow from downloading sequencing files to obtaining differentially expressed genes (DEGs) is described in Supplementary Figure 1. Original sequencing files were assessed for quality using FastQC (Andrews, 2010) to generate reports and MultiQC (Ewels et al., 2016) to visualize reports in aggregate. BBDuk (part of the BBTools suite; http://jgi.doe.gov/data-and-tools/bbtools) was used to trim reads from all experiments using the same basic parameters ("ktrim=r k=21 mink=11 hdist=2 qtrim=rl trimq=30 maq=20"). The minlen parameter was set to half of the original read length, which varied between experiments. Adapters and overrepresented sequences were additionally trimmed where applicable. Before alignment, library type (e.g. unstranded, reverse-

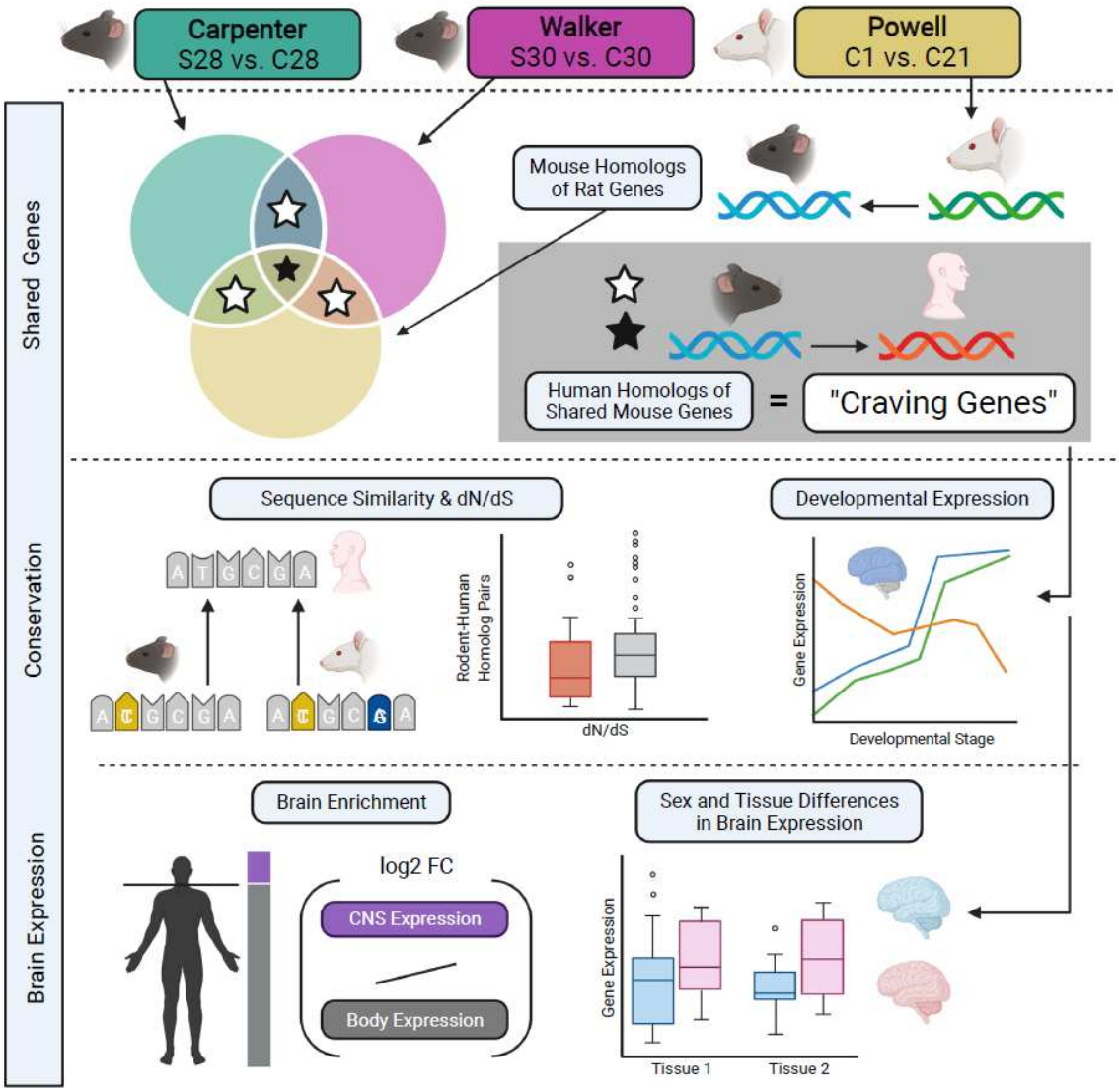
stranded) was determined with Salmon using the Ensembl GRCm39 mouse (Carpenter, Walker) and Rn6 rat transcriptome indexes from Refgenie (Stolarczyk et al., 2020). Splice-aware HISAT2 genome indexes were created for the Ensembl GRCm39 mouse (Carpenter, Walker) or mRatBN7.2 rat (Powell), and then trimmed reads were aligned using the library type identified by Salmon for the --rna-strandness parameter and other settings at default. Gene read counts were quantified for primary alignments (--primary) with featureCounts (Liao et al., 2014) using the -s parameter to specify library type.

For each of the three datasets, edgeR (M. D. Robinson et al., 2010) was used to transform and normalize read counts (e.g. from raw counts to fragments per kilobase per million mapped reads (FPKM)) and to filter out lowly-expressed genes. Genes were retained if their mean expression was greater than 0.5 FPKM for at least one of the two treatment groups in a study and had a raw read count of at least 6 in greater than or equal to the lowest sample size for that comparison. Raw counts were transformed to $\log_2$ counts per million (CPM), and outlier samples were removed based on multidimensional scaling (MDS) plots of the top 100 genes with the highest variance shared between groups. Genes were filtered and transformed again after outlier removal, and gene expression values from the remaining samples were normalized using the trimmed-means method (TMM) (M. D. Robinson & Oshlack, 2010). Next, limma voom (Law et al., 2014; Ritchie et al., 2015) was used to build linear models that accounted for technical sources of variation (sequencing lane or batch) identified by the package variancePartition (M. D. Robinson et al., 2010). Linear models were fitted using the least squares method, and empirical Bayes statistics were generated for each contrast.

**Common craving genes between studies.** To test our proposed workflow, DEGs in each dataset were identified with a cutoff of P value < 0.05, as only 1 DEG remained in any dataset after FDR correction (FDR > 0.05 or FDR > 0.1). The data

analysis process beginning with identified DEGs is visualized in Figure 1. Bioconductor's biomaRt (Durinck et al., 2005, 2009) and genomes from Ensembl release 105 (December 2021) were used to identify common DEGs between mouse and rat by identifying homologous genes. Next, DEGs shared between at least 2 contrasts were designated Craving genes and included in subsequent analyses. To further increase confidence, DEGs were only considered "shared" if they were regulated in the same direction for both contrasts. The final set of Craving genes were those shared genes that also had human homologs when compared to the human genome GRCh38.p13. For example, *Cartpt* was downregulated in C28 relative to S28 for Carpenter, downregulated in C30 relative to S30 in Walker, and had a human homolog (*CARTPT*), so was included as a Craving gene.

**RNA-seq power analysis.** After filtering genes and removing outlier samples according to the above criteria, power analyses were conducted separately for each dataset with the goal of 0.8 power using a FDR > 0.1 threshold. Using the R package *ssizeRNA* (Bi & Liu, 2016), the sample size required to achieve appropriate power was calculated based on values from each dataset, including the total number of genes (nGenes), the mean read count of the control group (mu), and tagwise dispersions estimated by edgeR (disp). Several values were used for the expected proportion of DEGs (0.05, 0.1, 0.2) and fold change for each DEG (1.15, 1.25, 1.5, and 2).




***Chapter 3, Figure 1.*** Workflow after differential expression analysis. After obtaining differentially-expressed genes (DEGs) from each study, DEGs were compared between datasets. Because the Powell dataset used rats and the others used mice, the DEGs for Powell were first converted to their mouse homologs. Genes were considered "shared" if they were called as DEGs in 2 or all 3 studies. Human homologs were found for these shared "Craving" genes, which were then analyzed for conservation across species by comparing sequence similarity, dN/dS, and developmental expression in the forebrain between human, mouse, and rat. Sequence similarity and dN/dS data were obtained from ENSEMBL through the R package biomaRt. Developmental expression data were obtained from Cardoso-Moreira et al. (2020). Next, Craving genes were assessed for their specificity to the brain and expression across brain tissues, including possible sex differences, using human data from the Gene-Tissue Expression (GTEx) database.

**Evolutionary conservation of craving-related genes.** To determine the evolutionary conservation of Craving genes across human, mouse, and rat, we incorporated developmental gene expression data, sequence similarity, and the ratio of divergence at nonsynonymous and synonymous sites (dN/dS), which has been used as an indicator of evolution's selective pressure acting on protein coding genes (Yang & Bielawski, 2000). For sequence similarity and dN/dS comparisons, Craving genes were analyzed separately for human-mouse and human-rat homologs, using only one-to-one homologs. Sequence similarity, scored as a percentage of matched nucleotides of a rodent gene compared to its human homolog, was obtained from the Ensembl database using biomaRt (Durinck et al., 2005, 2009). To obtain dN and dS values, an older Ensembl release (99; January 2020) was used due to availability of these metrics. dN/dS values were calculated for each Craving gene compared to its human homolog. Mann-Whitney *U* tests were used for both sequence similarity and dN/dS comparisons.

As part of a comprehensive project on species differences in developmental gene expression, Cardoso-Moreira et al. (Cardoso-Moreira et al., 2020) utilized previously collected RNA-sequencing data from several tissues of human, mouse, rat, and other species across developmental time points (Cardoso-Moreira et al., 2019) to determine which genes diverged in their temporal expression patterns across species. Utilizing their forebrain data, we categorized Craving genes as developmentally conserved between human, mouse, and rat (HMR), only human and mouse (HM), only human and rat (HR), or not conserved between humans and either rodent species (H), and compared them to all other genes in the dataset using Fisher's exact test.

**Brain specificity and expression across central nervous system (CNS) tissues.** Human homologs for the Craving genes were examined for expression in healthy human tissues using the publicly available Genotype-Tissue Expression (GTEx)

Project database (GTEx Consortium, 2013; Melé et al., 2015) (https://www.gtexportal.org/home/). Expression data in transcripts per million (TPM) was obtained for 51 human tissues, including 12 CNS tissues: amygdala, anterior cingulate cortex BA24, caudate, cerebellum, frontal cortex BA9, hippocampus, hypothalamus, nucleus accumbens, putamen, pituitary gland, spinal cord cervical C1, and substantia nigra. All CNS tissues were collected by the Miami Brain Bank, with the exception of 2 cortical and cerebellar tissues that were excluded from the present analysis, as they were replicates not collected by the same research group.

For each gene, the mean expression in TPM across individuals was summed for both CNS and non-CNS tissues, excluding brain replicates and cell lines. From this, a brain specificity score was calculated as the $\log_2$ fold change of the ratio of total CNS expression over total expression in other tissues. Specificity of Craving genes was compared to all other genes using a Mann-Whitney *U* test.

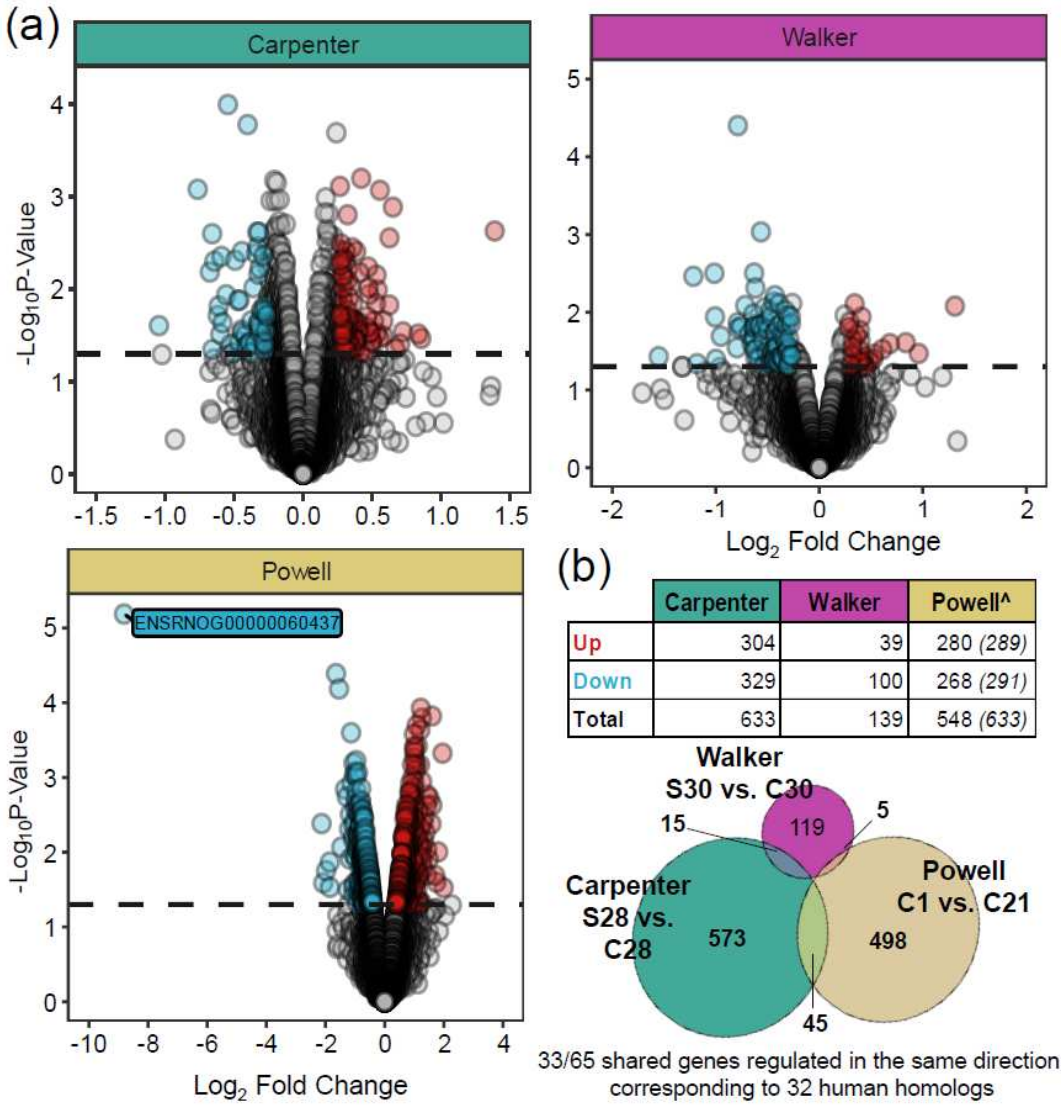
To assess for potential sex differences in CNS expression of human homologs of Craving genes, mean and median gene expression was obtained only for samples that were age-matched between the sexes for individuals ≥ 55 years of age. Two-way Type III ANOVAs were used to analyze effects of Tissue and Sex as well as Sex:Tissue interactions. Post-hoc Tukey's tests were used to further assess main effects or interactions.

## Results

**Differential expression analysis reveals few common genes and regulated pathways between craving-related datasets.** After filtering out genes with low expression, the datasets contained 10,026 (Carpenter), 9,948 (Walker), and 7,104 (Powell) genes each. The Supplementary Materials contain summaries of the

processing for each sample (Supplementary Table 1) and final statistics for each dataset (Supplementary Table 2). Two outlier samples were removed from the Powell C21 group because they did not cluster with other samples on the MDS plot. MDS and variancePartition plots showed that the experimental variables sequencing batch (Walker) and lane (Powell) explained a larger portion of the variation in gene expression than the treatment groups, so these variables were incorporated into the linear models before differential expression analysis. Supplementary Figures 2-4 contain MDS and variancePartition plots, along with voom (Law et al., 2014) results before and after linear modeling. Final sample sizes were 12 (6/group; Carpenter), 11 (5-6/group; Walker) and 4 (2/group; Powell).

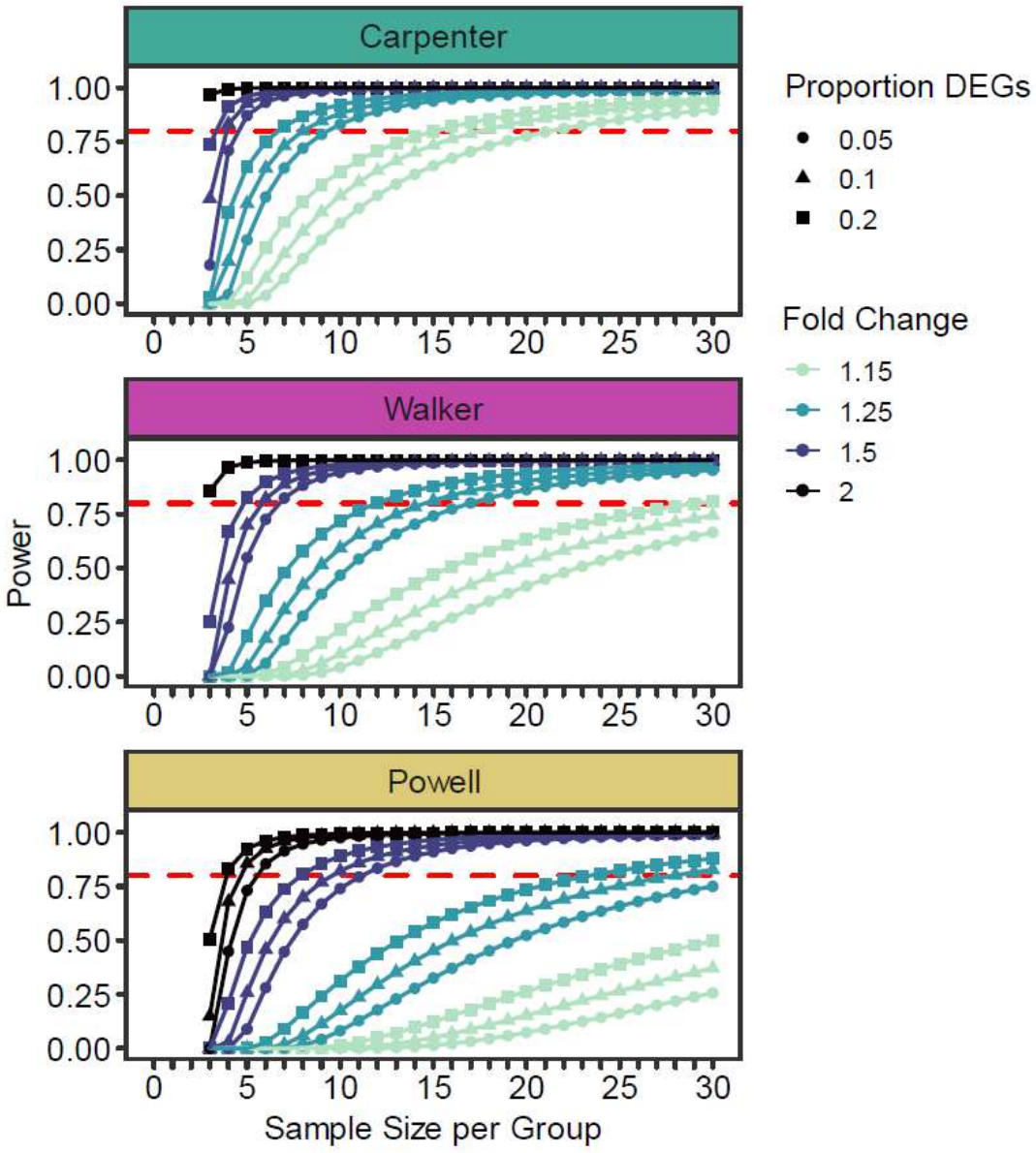
To demonstrate how to use evolutionary and human medical data to prioritize RNAseq results, we proceeded with using a P value < 0.05 (Figure 2, Supplementary Tables 2-4). For the Carpenter contrast, there were 633 DEGs, with 304 up- and 329 downregulated in C28 relative to S28. A similar number of DEGs were identified for the Walker contrast (139 DEGs; 39 up- and 100 downregulated in C30 relative to S30), while the Powell contrast had 633 DEGs (289 up- and 291 downregulated in C21 relative to C1). In contrast, only 1 DEG was identified using an FDR threshold of 0.05 or 0.1 across all three datasets. This gene, *AC239701.1* (ENSRNOG00000060437, adjusted P = 0.046), was significantly downregulated in the Powell dataset and is thought to be a pseudogene and does not yet have a known function, though a recent study suggests it may be sexually dimorphic in microglia (Ewald et al., 2020).




| | Carpenter | Walker | Powell^ |
|---|---|---|---|
| Up | 304 | 39 | 280 *(289)* |
| Down | 329 | 100 | 268 *(291)* |
| Total | 633 | 139 | 548 *(633)* |

***Chapter 3, Figure 2.*** Differentially-expressed genes (DEGs) within and across datasets. (A) Volcano plots depicting the spread of gene expression data for the Carpenter (teal), Walker (purple), and Powell (yellow) comparisons. Dashed lines indicate an uncorrected P value cutoff of < 0.05 (horizontal). Genes that met this criterion are colored to indicate that they are downregulated (blue) or upregulated (red) in the high craving group relative to the low craving group of a given dataset. (B) The number of DEGs according to the listed criteria. Note that only 1 DEG (for Powell) was found using stricter statistical thresholds corrected for False Discovery Rate (FDR), either FDR > 0.05 or FDR > 0.1. A 3-way Venn diagram shows the overlapping DEGs between studies, with 65 genes shared between two studies and none shared between all three. Comparisons on the Venn diagram are listed as "low" vs. "high" craving groups. After accounting for the direction of the overlapping genes (e.g. a gene upregulated in "high" craving for one study that is also upregulated in the "high" craving group for another study is regulated in the same direction), 33 genes were considered "shared" between studies. This corresponded to 31 human homologs. ^For Powell, the number of genes is listed as mouse homologs, with the original rat genes indicated in parentheses.

We further used sequence homology to identify comparable genes across studies. To compare DEGs between the Powell contrast and the others, Ensembl IDs for rat genes were first converted to a list of mouse homologs. This reduced the Powell dataset to 534 rat DEGs (270 up- and 264 downregulated) corresponding to 548 mouse homologs (280 up- and 268 downregulated). Some rat genes had more than one ortholog in mouse, resulting in a higher number of mouse orthologs than initial rat genes. No genes were differentially expressed between all 3 contrasts. There were 15 shared genes between Carpenter and Walker, 45 between Carpenter and Powell, and 5 between Walker and Powell, of which 10, 21, and 3 were regulated in the same direction. These correspond to 9, 21, and 3 human homologs, respectively, for a total of 33 Craving genes when taking the union across all three comparisons. Of these 33 genes, one was excluded from the analyses because it was not annotated/present in the majority of the downstream datasets, in human it is annotated as a novel transcript (*AC009690.3* or ENSG00000273025), and because it is annotated as a novel paralog of another Craving gene, *CELF6*. The final 32 genes are as follows: *AGK*, *AMZ1*, *B2M*, *BCAS1*, *BTG1*, *CACYBP*, *CARTPT*, *CCDC88C*, *CELF6*, *EGR2*, *FABP7*, *FKBP4*, *FTH1*, *GPD1*, *GUCY1A3*, *HAPLN2*, *HSPA8*, *IRS2*, *KIF5A*, *LYPD1*, *MBP*, *MOBP*, *NTS*, *PHLDA1*, *PITPNM3*, *RGS5*, *RPS6KA2*, *SOX17*, *TIPARP*, *TTLL1*, *USP46*, and *VIM*. For full names of all discussed genes, see the Supplementary Material.



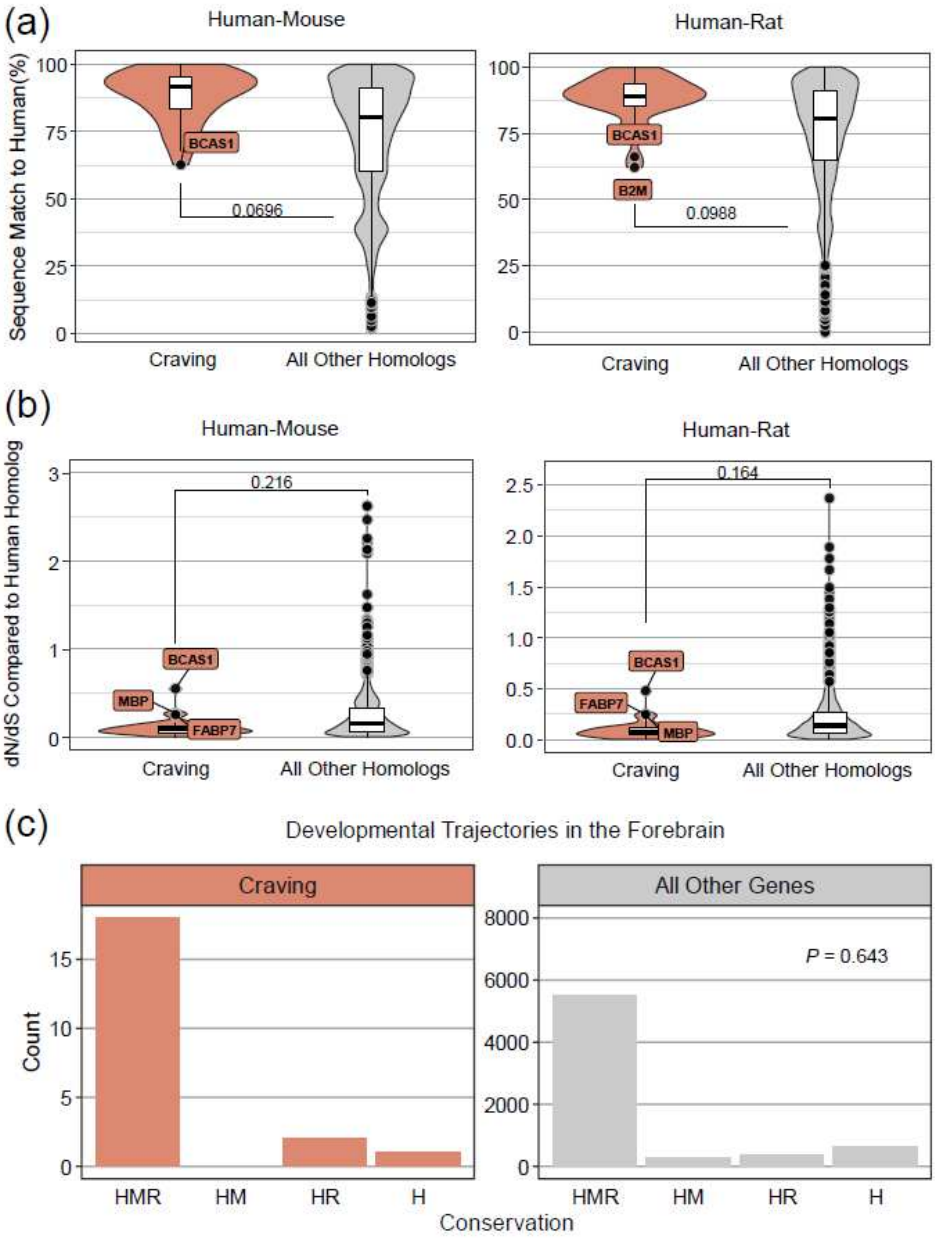

***Chapter 3, Figure 3.*** Estimated necessary sample size per group using the R package *ssizeRNA*. The number of samples, up to 30, that are necessary to reach varying levels of power are shown for the Carpenter (top panel; teal), Walker (middle panel; purple), and Powell (bottom panel; yellow) datasets. Dashed red lines indicate 0.8 power with FDR < 0.05. Power varies with the proportion of differentially-expressed genes (DEGs) and their fold change (FC). Because these metrics were not directly attainable for these datasets, varying proportions of DEGs and fold changes were tested and are indicated with different shapes and colors, respectively. For these theoretical calculations, FC is relative to all DEGs (e.g. if the proportion of DEGs is 0.2 and FC is 2, it is expected that 20% of genes are DEGs, all with FC of at least 2). For clarity on the Carpenter (teal) and Walker (purple) plots, the sample size estimations for FC of 2 are shown only for a proportion of DEGs equaling 0.2, though similar sample sizes are required for 0.05 or 0.1.

**RNA-seq power analysis.** Because almost no DEGs were identified for each study using an FDR cutoff, we calculated the sample size per group required to reach 0.8 power for each dataset with FDR < 0.05. Estimated sample size varied for each dataset based on the expected proportion of DEGs and their fold change values (Figure 3). For example, assuming a small effect size for each DEG (fold change of 1.25) but a high proportion of DEGs (0.2, or 20%), the datasets were estimated to need 7 (Carpenter), 13 (Walker), or about 25 (Powell) samples per group. Calculating for both a large effect size (fold change of 2) and high proportion of DEGs (0.2, or 20%), all three datasets were estimated to need only about 3 samples. In a more realistic scenario, where fewer genes are differentially expressed (5%, corresponding to 500 genes for every 10,000), each with a small fold change difference (1.25), the estimated necessary sample size varies more between datasets, with just 8 for Carpenter, but 16 for Walker and > 30 for Powell. This is unsurprising, since power analysis tools for RNA-seq are more accurate in estimating necessary sample size when the pilot or input data has more samples (Poplawski & Binder, 2018), and the Carpenter and Powell datasets contained the most and fewest samples per group, respectively. The Powell dataset also shows the highest mean dispersion across all genes (0.056, vs. 0.015 for Carpenter and 0.028 for Walker; Supplementary Table 2).

**Prioritizing Craving genes using evolutionarily conservation.** We propose that therapeutically relevant Craving genes are conserved between rodents and humans, and as such evolutionary conservation is a measure to prioritize RNA-seq candidates. We find that overall, Craving genes showed somewhat higher sequence similarity for both human-mouse (Mann–Whitney $U = 307576$, $n_1 = 31$, $n_2 = 16699$, $\text{median}_1 = 91.7$, $\text{median}_2 = 87.1$, P = 0.0696) and human-rat (Mann–Whitney $U = 270762$, $n_1 = 29$, $n_2 = 15863$, $\text{median}_1 = 89.1$, $\text{median}_2 = 86.2$, P = 0.0988) homologs

when compared to all other homologous pairs in the genome (Figure 4A), though results did not reach statistical significance. Craving genes did not have significantly different dN/dS values for either human-mouse (Mann–Whitney $U$ = 182458, $n_1$ = 28, $n_2$ = 15067, $\text{median}_1$ = 0.093, $\text{median}_2$ = 0.109, P = 0.216) or human-rat (Mann–Whitney $U$ = 162708, $n_1$ = 27, $n_2$ = 14260, $\text{median}_1$ = 0.088, $\text{median}_2$ = 0.110, P = 0.164) homologs compared to other homologs (Figure 4B). Only one-to-one homologs were included in the statistical analyses, but where available, sequence similarity and dN/dS ranges for all homologs are available in Supplementary Table 6.

We suggest developmental trajectory as another measure of conservation for prioritizing candidate genes. A Fisher's exact test indicated no difference in the evolutionary conservation of Craving gene homolog developmental trajectories in the forebrain compared to all other genes ($n_1$ = 21, $n_2$ = 9824, P = 0.643, Figure 4C). This analysis excluded genes that had no data available (*CELF6*, *FKBP4*, *FTH1*, *HSPA8*, *LYPD1*, *TIPARP*, and *USP46*), or that had data only for human-mouse (*GPD1*, *PHLDA1*, *RPS6KA2*) or human-rat (*SOX17*) comparisons and not the other rodent species (Cardoso-Moreira et al., 2020). All Craving genes with complete available data showed conservation in their developmental expression in the forebrain of human, mouse, and rat, with the exception of *CACYBP*, *NTS*, and *RGS5*. Individual conservation results for each gene are presented in Supplementary Table 6.




***Chapter 3, Figure 4.*** Conservation of Craving genes. (A) Sequence similarity, measured as percentage similarity of the rodent gene to its human homolog. (B) dN/dS values for human-mouse and human-rat orthologs. (C) Conservation of gene expression patterns across development in the forebrain, utilizing data from Cardoso-Moreira et al. (Cardoso-Moreira et al., 2020). HMR = conserved expression across development in human, mouse, and rat; HM = conserved expression in only human and mouse; HR = conserved expression only in human and rat; H = developmental expression not conserved between humans and either rodent species. For all conservation analyses, only homologs that mapped one-to-one from rodent to human were included in the analysis. *P*-values are listed on each panel, indicating no significant difference between Craving and all other homologs/genes utilizing Mann-Whitney U test (sequence similarity, dN/dS) or Fisher's exact test (developmental expression).

For each Conservation measurement (sequence similarity, dN/dS, developmental conservation), genes were categorized as having High, Medium, or Low conservation (Table 3). For sequence similarity, genes were categorized as High, Medium, and Low if they had similarity ≥ 90%, between 80 and 90%, or < 80%, respectively. For dN/dS scores, genes were considered to have High conservation if they had values below the Craving median (0.11), Medium conservation at any other non-outlier value (≤ 0.2), and Low if they were outliers (> 0.2). dN/dS values were categorized as "High" if they were missing due to 100% sequence similarity, and not categorized in any other cases of missing data. Developmental conservation was categorized as High if the gene shared its developmental expression pattern in human, mouse and rat (HMR), Medium if its temporal expression in humans was shared with either mouse (HM) or rat (HR), and Low for no conservation with either rodent species (H). Genes with no data for both the human-mouse and human-rat comparisons were not categorized. For genes where data was available for one rodent species and not the other, they were categorized as Medium for shared developmental expression in the forebrain with that species, and Low if they did not.

Chapter 3, Table 3

***Prioritization of Candidate Craving Genes***

| | Conservation Prioritization | | | | | Expression Prioritization | | |
|---|---|---|---|---|---|---|---|---|
| Human Gene Symbol | Human-Mouse Sequence Similarity | Human-Mouse dN/dS | Human-Rat Sequence Similarity | Human-Rat dN/dS | Developmental Conservation in Forebrain | Brain Specificity Score | Brain Region Enrichment | Mean Brain Expression (TPM) |
| CARTPT | 96.55 | 0.0523 | 94.83 | 0.0654 | HMR | 1.44 | BRN-HYP | 49.13 |
| LYPD1* | 93.62 | 0.0644 | 92.91 | 0.0573 | *NA* | 1.97 | BRN-NAC | 17.48 |
| MOBP | 75.96 | 0.1524 | 76.50 | 0.1363 | HMR | 7.82 | BRN-SPC | 253.85 |
| HAPLN2 | 91.76 | 0.0484 | 91.47 | 0.0567 | HMR | 4.00 | BRN-SPC | 134.58 |
| TIPARP* | 92.69 | 0.0983 | 92.39 | 0.0999 | *NA* | -3.00 | BRN-PTRY, BRN-SPC | 10.39 |
| IRS2 | 85.58 | 0.0517 | 86.25 | 0.0517 | HMR | -1.86 | BRN-PTRY | 18.62 |
| VIM | 97.42 | 0.0239 | 88.84 | 0.0223 | HMR | -4.31 | BRN-PTRY | 126.47 |
| CELF6$+ | 95.43% | 0.0470 | 96.47% | 0.0375 | *NA* | -1.32 | BRN-PTRY | 19.13 |
| GUCY1A3 | 89.86 | 0.0697 | 89.13 | 0.0763 | HMR | -2.93 | BRN-NAC | 4.66 |
| CCDC88C | 82.20 | 0.1051 | 84.02 | 0.0881 | HMR | -2.30 | BRN-CAU, BRN-NAC | 5.24 |
| EGR2 | 88.87 | 0.1078 | 89.08 | 0.1010 | HMR | -2.82 | BRN-PTRY | 3.24 |

| | Conservation Prioritization | | | | | Expression Prioritization | | |
|---|---|---|---|---|---|---|---|---|
| Human Gene Symbol | Human-Mouse Sequence Similarity | Human-Mouse dN/dS | Human-Rat Sequence Similarity | Human-Rat dN/dS | Developmental Conservation in Forebrain | Brain Specificity Score | Brain Region Enrichment | Mean Brain Expression (TPM) |
| HSPA8$ | 99.85 | 0.0011 | 86.53% to 99.85% | 0.0035 to 0.024 | *NA* | -1.64 | BRN-CB-a, BRN-PTRY | 389.02 |
| RGS5 | 90.06 | 0.1354 | 85.08 | 0.1260 | HR | -2.83 | BRN-CAU, BRN-CTX-a, BRN-PUT, BRN-SN | 73.24 |
| NTS | 81.18 | 0.1602 | 77.06 | 0.1942 | HR | 0.85 | BRN-HYP, BRN-PTRY | 12.82 |
| PITPNM3 | 94.46 | 0.0431 | 94.76 | 0.0416 | HMR | -0.63 | BRN-CB-a | 24.02 |
| AGK | 91.71 | 0.0889 | 90.76 | 0.0965 | HMR | -1.54 | BRN-CB-a | 14.41 |
| BTG1$ | 100.00 | *NA* | 99.42 | 0.0180 | HMR | -3.55 | BRN-CB-a | 31.06 |
| TTLL1 | 96.93 | *NA* | 97.40 | *NA* | HMR | -1.04 | BRN-CB-a | 13.73 |
| USP46* | 100.00 | *NA* | 96.99 | 0.0035 | *NA* | -0.30 | BRN-CB-a | 17.55 |
| FKBP4 | 89.76 | 0.0962 | 89.98 | 0.1169 | *NA* | -1.73 | BRN-CB-a | 47.41 |
| RPS6KA2* | 95.63 | 0.0518 | 91.95 | 0.0488 | HM* | -1.33 | BRN-SPC | 31.66 |
| GPD1 | 93.98 | 0.0561 | 94.27 | 0.0586 | HM* | -3.15 | BRN-SPC | 20.05 |
| FTH1^ | 91.80 | 0.1156 | *NA*^ | 0.0800 | *NA* | -2.05 | BRN-SPC | 1135.39 |

| | Conservation Prioritization | | | | | Expression Prioritization | | |
|---|---|---|---|---|---|---|---|---|
| Human Gene Symbol | Human-Mouse Sequence Similarity | Human-Mouse dN/dS | Human-Rat Sequence Similarity | Human-Rat dN/dS | Developmental Conservation in Forebrain | Brain Specificity Score | Brain Region Enrichment | Mean Brain Expression (TPM) |
| AMZ1 | 73.69 | 0.1273 | 75.50 | 0.1067 | HMR | 0.98 | BRN-ACC, BRN-AMY, BRN-NAC | 2.09 |
| MBP | 71.71 | 0.2698 | 88.16 | 0.2377 | HMR | 4.92 | BRN-SPC | 1830.85 |
| FABP7 | 79.55 | 0.2535 | 88.64 | 0.2462 | HMR | 2.80 | BRN-CB-a | 32.50 |
| B2M | 68.07 | 0.1612 | 62.18 | 0.1328 | HMR | -3.49 | BRN-PTRY, BRN-SPC | 379.43 |
| BCAS1 | 62.64 | 0.5496 | 66.14 | 0.4783 | HMR | 3.29 | BRN-SPC | 204.70 |
| CACYBP | 93.42 | 0.0623 | 90.35 | 0.0732 | H | -0.75 | BRN-CB-a, BRN-CTX | 55.87 |
| SOX17* | 84.54 | 0.0666 | 84.30 | 0.0658 | H* | -4.67 | BRN-SN, BRN-SPC | 2.42 |
| PHLDA1* | 85.54 | 0.1058 | 85.79 | 0.1127 | H* | -2.56 | BRN-AMY, BRN-CTX-a, BRN-HYP, BRN-PTRY, BRN-SN | 10.80 |

Genes were categorized as High (green), Medium (yellow), and Low (red) priority for each conservation and brain expression metric. Complete data are listed here, even for homologs that do not map one-to-one and were not included in the statistical analyses. Sequence similarity was measured as the percentage match between a rodent gene and its human homolog. Data from Cardoso-Moreira et al. (Cardoso-Moreira et al., 2020) was used to determine conservation of gene expression patterns across

development in the forebrain. HMR = conserved expression across development in human, mouse, and rat; HM = conserved expression in only human and mouse; HR = conserved expression only in human and rat; H = developmental expression not conserved between humans and either rodent species. Developmental conservation assignments listed in this table are based on this available data only, assuming developmental expression is not conserved with the untested species. For example, GPD1 is "HM" because its developmental expression is conserved between human and mouse, but data is not available for human vs. rat. Brain specificity was calculated as the $\log_2$ fold change of the total expression in central nervous system (CNS) tissues compared to non-CNS tissues, using data from the GTEx database. The brain region(s) with the highest expression were determined based on post-hoc Tukey's tests (Supplementary Table 9). See Supplementary Material for brain region abbreviations. Mean brain expression was calculated by obtaining the mean expression in TPM for each brain tissue across all individuals regardless of age, then calculating the grand mean.

* Indicates genes that do not have complete developmental expression data available, i.e. genes that only had data comparing human and mouse (*GPD1*, *PHLDA1*, *RPS6KA2*) or human and rat (*SOX17*).

$ Indicates homologs with one-to-many or many-to-one mapping. For some of these homologs, sequence similarity and/or dN/dS are given as a range.

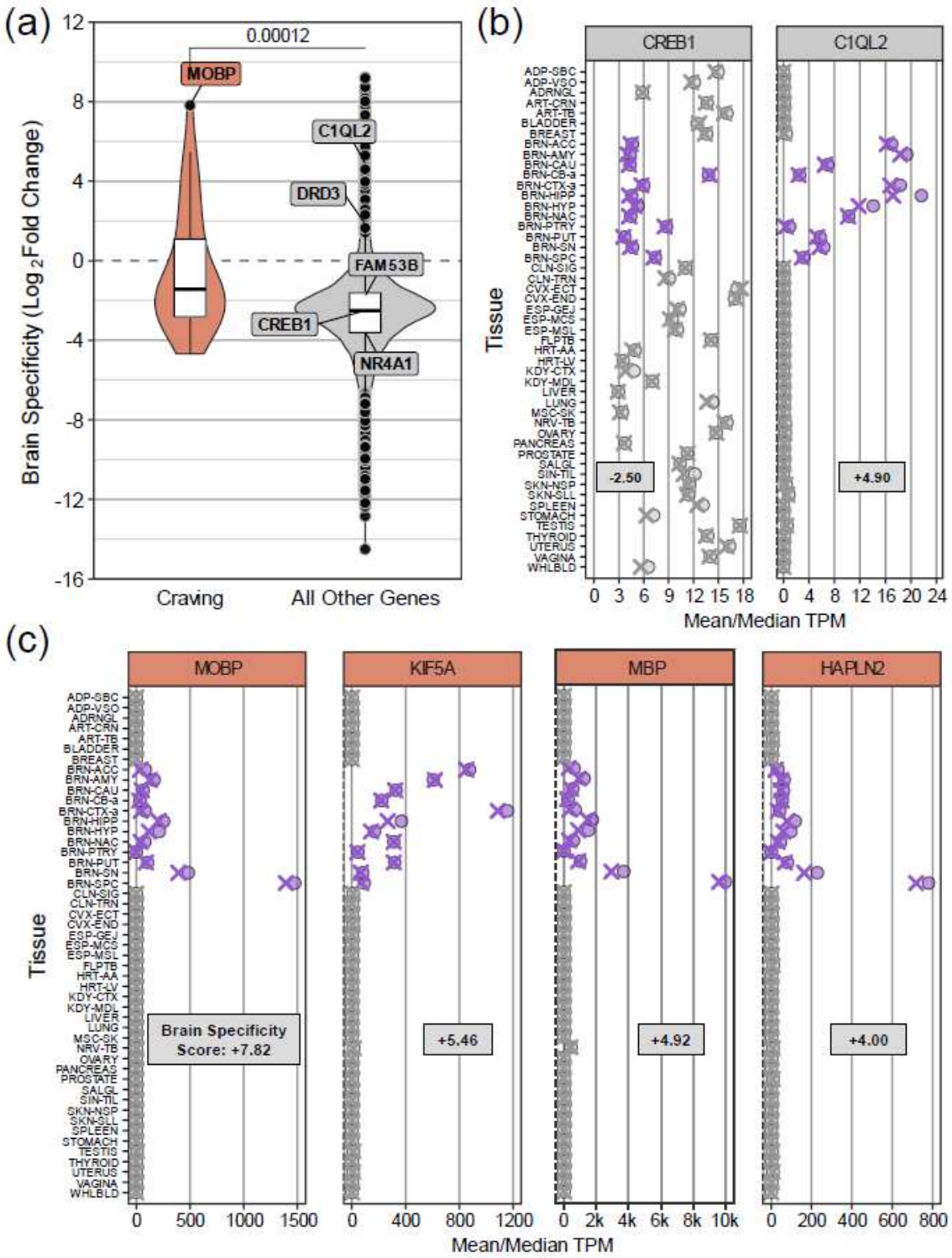
^ FTH1 has a rat ortholog according to the January 2020 but not December 2021 Ensembl release, and thus a dN/dS value is listed but not sequence similarity.

+ CELF6 is a paralog of another Craving gene that corresponds to a poorly annotated novel transcript, ENSG00000273025, which was excluded from the analyses.

Though the overall set of Craving genes is not more or less evolutionarily conserved than other genes, there are individual genes that would make poor candidates based on their conservation. For example, human *FTH1* has a homolog in mouse but not rat, so studies of this gene would need to be confined only to mouse models, limiting translational utility. Some Craving genes were also outliers with very low sequence similarity and/or high dN/dS values, including *BCAS1*, *B2M*, *FABP7*, and *MBP* (Figure 5A, B, Table 3); however, all of these genes are conserved in their developmental expression in the forebrain across the three species. Importantly, several other genes are not developmentally conserved, including *NTS* and *RGS5*, which have conserved expression between human and rat only. *CACYBP,* which has high sequence similarity between human and mouse (93.4%) and human and rat (90.3%), is not conserved between human and either rodent species for developmental expression in the forebrain. This emphasizes the importance of considering functional conservation when selecting a candidate gene.

**Narrowing candidate Craving genes by brain specificity and brain tissue- and sex-specific expression.** Utilizing the GTEx dataset, we found that Craving genes are expressed with higher brain specificity than other genes after CNS- and body-exclusive genes were removed (Mann-Whitney $U = 1233008$, $n_1 = 32$, $n_2 = 55305$, $\text{median}_1 = -1.44$, $\text{median}_2 = -2.51$, $P = 0.000116$, Figure 5A). This result remains significant even after filtering out lowly-expressed genes with a grand mean of ≤ 1 (Mann–Whitney $U = 428911$, $n_1 = 31$, $n_2 = 20624$, $\text{median}_1 = -1.54$, $\text{median}_2 = -2.36$, $P = 0.000992$; excluding the lowly-expressed Craving gene *AMZ1*) or ≤ 4 TPM (Mann–Whitney $U = 299236$, $n_1 = 31$, $n_2 = 14404$, $\text{median}_1 = -1.54$, $\text{median}_2 = -2.35$, $P = 0.00105$; also excluding *AMZ1*) across all tissues. Figure 5B and 5C depict the overall

gene expression across tissues for 2 previously identified addiction genes, *CREB1* (W. A. Carlezon et al., 1998; McClung & Nestler, 2003) and *C1QL2* (Gelernter et al., 2014; Huggett & Stallings, 2020), and the 4 Craving genes with the highest CNS specificity: *MOBP*, *KIF5A*, *MBP*, and *HAPLN2*.




***Chapter 3, Figure 5.*** Brain specificity of Craving genes. (A) Brain specificity, as measured by log2 fold change of expression in CNS tissues vs. all other tissues, of Craving genes (orange) compared to all other genes (gray) in GTEx V8. Select genes are labeled, including 2 outliers for Craving genes, and several established addiction genes -

*CREB1* (W. A. Carlezon et al., 1998; McClung & Nestler, 2003), *C1QL2* (Gelernter et al., 2014; Huggett & Stallings, 2020), *DRD3* (Neisewander et al., 2014; Powell et al., 2018; Powell, Namba, et al., 2020), *FAM53B* (Gelernter et al., 2014), and *NR4A1* (Carpenter et al., 2020). P = 0.00012 indicates a significant difference between Craving and all other genes with Mann-Whitney *U* test. Mean and median expression in transcripts per million (TPM), indicated by X's and circles, respectively, are shown for (B) 2 known addiction genes and (c) the 4 Craving genes with the highest brain specificity, using all available samples from GTEx. For (B) and (C), the overall brain specificity score for each gene is also listed. Purple and gray indicate CNS and non-CNS tissues, respectively. Full names for abbreviated genes and tissues are listed in the Supplementary Material.

Next, we probed for possible sex and tissue differences in brain expression of Craving genes, which is important when developing additional experiments and considering translation to humans. Using age-matched data for males and females ≥ 55 years, we performed two-way ANOVAs by Sex and Tissue for the 12 CNS tissues. ANOVA and post-hoc Tukey statistics for Sex and Sex:Tissue interactions are reported in Supplementary Tables 7-10. All 32 genes showed a significant main effect of Tissue. Post-hoc Tukey's tests were used to determine which brain tissue(s) showed enriched expression of a gene relative to others and to generate compact letter displays (CLDs) to cluster tissues with similar expression levels (e.g., "a" has the highest expression, "b" has the second-highest expression, and so on) (Supplementary Table 9). For example, *KIF5A* shows higher expression in the cortex relative to all other tissues and is denoted with the only "a"; however, for *NTS*, the pituitary gland shows significantly higher expression than all other tissues except the hypothalamus, so they are designated as "a" and "ab", respectively; Supplementary Figure 5). We then categorized genes as having "enriched" expression in specific brain tissue(s) if the CLD included the letter "a" (Table 3).

Of the 5 genes that had a significant main effect of Sex, 4 had higher expression in males (*AMZ1*, *NTS*, *PITPNM3*, *TTLL1*) and 1 was more highly expressed in females (*B2M*). Several genes also showed a significant Sex:Tissue interaction, including *CARTPT*, which has higher expression in males in BRN-CAU, as well as *NTS* and *RPS6KA2*, which both have higher expression in males in BRN-PTRY. *HAPLN2* also showed a significant Sex:Tissue interaction, but none of the post-hoc tests for this interaction were significant after Tukey's correction. Summary brain expression results for each gene are presented in Table 3. Supplementary Figure 5 shows expression of each Craving gene across Sex and Tissue using the age-matched samples.

Craving genes were categorized as High, Medium, or Low priority based on the brain expression data (Table 3). The following criteria were used: High - brain specificity > 0 and enriched expression in reward regions, but not cerebellum or spinal cord, according to post-hoc Tukey's tests (Supplementary Tables 9 & 11); Medium - brain specificity > 0 but enriched expression only in non-reward reward regions, or brain specificity < 0 but enriched expression in reward regions; Low - brain specificity ≤ 0 and has enriched expression only in cerebellum and/or spinal cord. The following tissues were considered reward-related: anterior cingulate cortex, amygdala, caudate, frontal cortex, hippocampus, hypothalamus, nucleus accumbens, pituitary gland, putamen, and substantia nigra. The grand mean of expression (TPM) for all brain tissues was also calculated for each gene, and genes were categorized as High, Medium, and Low if their expression was ≥ 20, between 10 and 20, or < 10 TPM, respectively.

Using this approach, we found genes that would make exciting candidates to explore for a role in craving and reward. Though most of these genes do not show sex differences in brain expression for healthy human adults, for those that do (e.g. *CARTPT*, *NTS*), extra consideration should be made to design any preclinical experiments with adequate power to detect possible sex differences. Ten genes have higher expression in the brain than the rest of the body: *MOBP*, *KIF5A*, *MBP*, *HAPLN2*, *BCAS1*, *FABP7*, *LYPD1*, *CARTPT*, *AMZ1*, and *NTS* (Supplementary Table 11). Post-hoc Tukey's tests indicate that five of these genes (*MOBP*, *MBP*, *HAPLN2*, *BCAS1*, and *FABP7*) are preferentially expressed in the spinal cord or cerebellum, which are tissues not conventionally associated with addiction. Rather, creating therapeutics that target genes expressed highly in these tissues may lead to negative side effects. However, other genes are highly expressed in reward-related regions. For example, *KIF5A* and *LYPD1* are most expressed in the cortex and nucleus accumbens, respectively. *AMZ1* has equally high

expression in the anterior cingulate cortex, amygdala, and nucleus accumbens, which are all involved in emotional and reward processing. Many other Craving genes do not have high brain specificity, but within the brain are most expressed in reward-related regions. This includes *IRS2* (highest expression in pituitary gland), *CCDC88C* (highest in caudate and nucleus accumbens), and *GUCY1A3* (highest in nucleus accumbens).

**Discussion**

We analyzed three RNA-seq datasets from rodents that underwent cocaine SA and prolonged abstinence, with the goal to identify craving-related genes with convergent evidence across preclinical models and research labs. Using the same, reproducible RNA-seq analysis pipeline for each dataset, we found that only 1 DEG survived corrections for false discovery rate in any dataset, using FDR < 0.05 or < 0.1. Using a less stringent threshold (uncorrected P value < 0.05), hundreds of DEGs were identified for each study. However, no DEGs overlapped between all 3 datasets, and only 33 genes were shared between 2 datasets and also regulated in the same direction for both. This low overlap may suggest large differences in the neurogenomic outcomes of each behavioral model, but is more likely caused by low sample sizes and low power to detect true DEGs. Using this set of shared "Craving" genes as a test case, we suggest a next step toward further refining lists of candidate genes is prioritizing those with evolutionary and functional conservation and those with brain specificity. Within each conservation or brain expression metric, we categorized genes as having High, Medium, or Low priority to determine which would make better candidates for translational research and validation.

Using this approach, we found several candidate genes of particular interest because they are both highly conserved and are enriched in brain tissues involved in

reward (*CARTPT*, *KIF5A*, *LYPD1*). Several other candidates are well-conserved and show high brain specificity, though they have enriched expression in non-reward regions (*MOBP*, *HAPLN2*). Using this prioritization strategy, we also identify several genes that are poorer candidates because they are less conserved (*BCAS1*, *MBP*, *FTH1*, *PHLDA1*), are not brain-specific and have higher expression in non-reward tissues (*AGK*, *BTG1*, *TTLL1*, *USP46*, *PITPNM3*) or all three (*B2M*, *CACYBP*, *FKBP4*, *GPD1*, *RPS6KA2*).

We initially expected a large overlap in DEGs across the studies reflective of cocaine craving. While there are differences in each study's design (Table 1), each dataset examined gene expression patterns in the NAc and included treatment groups that vary in post-abstinence cocaine craving, whether this was measured directly (Powell) or can be predicted based on prior literature on the incubation effect (Carpenter, Walker) (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). For the Carpenter and Walker data sets, we performed pairwise comparisons between mice that underwent saline or cocaine SA followed by prolonged abstinence (28 or 30d). For the Powell experiment, incubation was modeled more directly, as rats underwent cocaine SA and either short (1d) or long-term abstinence (21d). For all three datasets, we compared a group likely to demonstrate low cocaine craving (long-term abstinence from saline SA, or short-term abstinence from cocaine SA) to a group with higher craving (long-term abstinence from cocaine SA). Thus, we expected to identify potential Craving genes from a high degree of overlap between studies. Instead, we found low overlap of DEGs between the studies. Notably, there are some differences between the experiments that might contribute to a reduced overlap in DEGs, such as the rodent species (mouse, rat), tissue studied (whole NAc, NAc shell), or experimental manipulations post-abstinence (saline injections, context re-exposure, cue-reactivity) (Tables 1 and 2). For example, in both the Carpenter and Walker studies, craving was not measured directly, e.g. through a

cue reactivity test. Though the Carpenter and Walker studies are more similar to each other than to the Powell study, they still have low overlap in DEGs (Figure 2). Meanwhile, the Powell dataset shows more DEGs than either Carpenter or Walker. This may be due to the cue reactivity test, which assesses general learning and memory in addition to cocaine-seeking behavior. Many learning and memory genes are also implicated in the SUDs literature, and the addition of a cue reactivity test may lead to a greater number of DEGs as a reflection of this additional psychological process.

Use of uncorrected P values may also result in low overlap, even between similar datasets. This may in part be because DEGs identified with this lenient threshold are likely to include many false positives, creating the appearance of large differences between datasets (Mukamel, 2021; Noble, 2009). In response to a recent critique about using uncorrected P values (Mukamel, 2021), Walker et al. (Walker et al., 2021) note that even using a nominal P value, potential genomic targets can be validated by comparisons to prior literature. Similarly, individual genes or transcripts may be validated with approaches such as RT-qPCR, though this is typically used to spot check the validity of only a small subset of DEGs. However, it is not always clear whether authors have reported those DEGs, if any, that failed to validate with RT-qPCR. In either case, confirming RNA-seq results for a single gene or a small set of genes does not automatically indicate that all identified DEGs are true positives.

Even after running all datasets through the same pipeline to avoid technical biases that could result in differences in detecting DEGs, the datasets used here are underpowered to detect true DEGs after correcting for multiple testing. We performed power analyses for each dataset and found that appropriate statistical power (0.8) is only achieved at low sample sizes when the expected number of DEGs and their fold changes are both high (Figure 3). The true number of DEGs and their fold changes are unknown,

but based on volcano plots showing the spread of gene expression differences, it is likely that the actual fold change for most DEGs is low, around 1.15 to 1.5, with fewer DEGs having a fold change of 2 or higher (Figure 2, Supplementary Table 2). Notably, the issue of low power is not limited to the datasets selected here, and instead reflects a common trend in the addiction field and neuroscience at large, where tissue-level gene expression changes are usually small. Differential expression and pathway analyses are regularly published using uncorrected P values, perhaps because correcting for multiple testing would severely reduce or entirely eliminate any DEGs (for example, see (Mukamel, 2021)). Because of this, we can neither detect true DEGs within the datasets nor determine the true overlap between datasets. However, with greater sample sizes, this method of comparing datasets could reveal not just shared DEGs, but gene expression patterns that are exclusively altered by a specific experimental design. This would enhance our understanding of different preclinical models and be a step toward unraveling the complexity of addictive behaviors.

There has been extensive discussion about necessary sample sizes for RNA-seq analysis, which have evolved from early rules of thumb suggesting only 3 samples per group is sufficient. A highly-cited analysis by Schurch et al. (Schurch et al., 2016) determined that a sample size of 6 to 12 should be sufficient for detecting the majority of DEGs using most differential expression tools. However, this was based on RNA-seq data from yeast, which have simple, small transcriptomes, and by comparing wild-types to mutants with a robust gene expression phenotype. In contrast, SUDs are likely to be driven by complex interactions of many genes with low effect sizes (Geschwind & Flint, 2015; Prom-Wormley et al., 2017) that are difficult to detect transcriptome-wide without an adequately large sample size. Rather than assume a specific sample size will be adequate, many tools have been developed to estimate the necessary sample size for a

given RNA-seq experiment based on pilot data. While some of these are more accurate than others, all tools drop off in reliability when the fold changes of true DEGs are small (Poplawski & Binder, 2018). This is partly because these tools rely on simplified assumptions about the data (e.g. that the fold change is fixed for all DEGs, as in Figure 3) that may be too complex to holistically model (Z. Wu & Wu, 2016). While sample size estimation tools are a useful starting point, it is likely that the number of samples required for an appropriately powered RNA-seq study of behavior or addiction is higher than for simpler experimental designs. In general, the necessary sample size is likely to be at least the number needed for RT-qPCR validation or behavioral testing, if not more, due to the many variables in sequencing that cannot be accounted for directly.

Under the assumption that overlapping DEGs between studies generated from uncorrected P values will still contain false positives, we moved forward using the shared "Craving" genes ($P < 0.05$) as a test case for illustrating how to prioritize candidate genes that are likely to translate from rodents to humans. We reasoned that such genes should have shared evolutionary conservation between mouse, rat, and human. Indeed, we found that Craving genes had generally high sequence similarity and low dN/dS values between their rodent and human homologs (Figure 5A, 5B), though they were not significantly different from all other genes. Overall, it appears that these genes are evolving under strong purifying selection, and that there are evolutionary pressures to retain the same amino acid encoding even where the exact sequences diverge. While all of our Craving genes have low dN/dS values (all < 1, median = 0.11 for both human-mouse and human-rat orthologs), there are several outliers with higher dN/dS values (> 0.2; Low priority), including *BCAS1*, *FABP7*, and *MBP* (Figure 5A, Table 3). Though these also have Low sequence similarity (62.6 to 71.2% for both human-mouse and human-rat homologs, with the exception of 88.6% for the *FABP7* human-rat

comparison), several other genes (*B2M*, *MOBP*) show Low (> 80%) or Medium (between 80% and 90%) sequence similarity but still have low dN/dS values (> 0.17; Low or Medium priority). Thus, despite some changes in their coding sequences, the genes are still likely to be under evolutionary selection that maintains their protein conformation and perhaps their molecular function as well. By considering both sequence similarity and dN/dS values, we can discard genes as poor candidates due to low sequence similarity (*BCAS1*, *FABP7*, *MBP*) while tentatively retaining genes with low sequence similarity but low dN/dS values (*B2M*, *MOBP*).

Similarly, our analysis revealed that most Craving genes have conserved expression across matched developmental stages in the forebrain of human, mouse, and rat (Figure 5C), though again they are not conserved more than other genes. However, by adding this with dN/dS we can identify genes with relatively high sequence similarity (High or Medium priority) and low dN/dS (High priority), but which show divergent developmental expression (*CACYBP*, *PHLDA1*, *SOX17*), suggesting their function may not be conserved across species, at least in the context of development. Given that many known addiction-related genes are strongly implicated in brain development, and abnormal neurodevelopment can increase risk of SUDs in adulthood (McCrory & Mayes, 2015), genes that are not conserved across development are likely to have less translational utility. In addition, genes that are highly conserved across all metrics, such as *AGK*, *BTG1*, *CARTPT*, *HAPLN2*, *KIF5A*, and *PITPNM3*, may be even better-suited for study in preclinical rodent models.

We further propose to narrow translational rodent targets based on whether their homologs in humans have enriched expression across human brain tissues. Within the brain, genes that are most highly expressed in reward regions such as the amygdala, nucleus accumbens, and the dorsal striatum are appealing candidates for future study.

For example, *CCDC88C*, *PHLDA1*, *RGS5*, and *GUCY1A3* show elevated expression in these regions (Table 3). Of these, *GUCY1A3* is of particular interest as it encodes an isoform of guanylate cyclase that can form heterodimers that are activated by nitric oxide (NO) (Koesling et al., 2004). NO signaling has been implicated in SUDs (Alexander C. W. Smith et al., 2017; Vleeming et al., 2002), and *GUCY1A3* is specifically implicated in nonhuman primate models of early-life stress (Sabatini et al., 2007), ADHD, Tourette Syndrome (Tsetsos et al., 2016), and glioma (Saino et al., 2004). However, these genes, and Craving genes overall, do not have higher brain specificity than other genes, and thus we have categorized them as Medium priority based on brain expression (Figure 6A, Table 3). Still, several candidate Craving genes do show elevated expression in the brain compared to non-CNS tissues (*MOBP*, *KIF5A*, *MBP*, *HAPLN2*, *BCAS1*, *FABP7*, *LYPD1*, *CARTPT*, *AMZ1*, *NTS*). While a gene does not need to be brain-specific to have an impact on reward processing and behavior (e.g. the well-studied addiction gene *CREB1* is not brain specific, Figure 6B), ubiquitous expression may make these genes difficult to target with therapeutics without negative off-target effects in other tissues. For example, we find the myelin oligodendrocyte gene *MOBP* has 8-fold higher expression in the brain relative to all other tissues, suggesting targeting its expression may have a stronger effect in the brain than the rest of the body. *MOBP* has already been implicated in cocaine abuse (Albertson et al., 2004, 2006). However, *MOBP* is most highly expressed in spinal cord tissue, and therapeutics that target this gene may provoke motor side effects. Other Craving genes score as High for brain expression prioritization because they show brain-specificity and are highly expressed in reward regions (*CARTPT*, *KIF5A*, *LYPD1*). Of these, *CARTPT* is a prepropeptide related to the addiction gene CART (Cocaine and Amphetamine Regulated Transcript) (Kuhar et al., 2005; Ong & McNally, 2020), and both *KIF5A* and *LYPD1* are involved in neuroplasticity. *KIF5A* is a neuron-specific

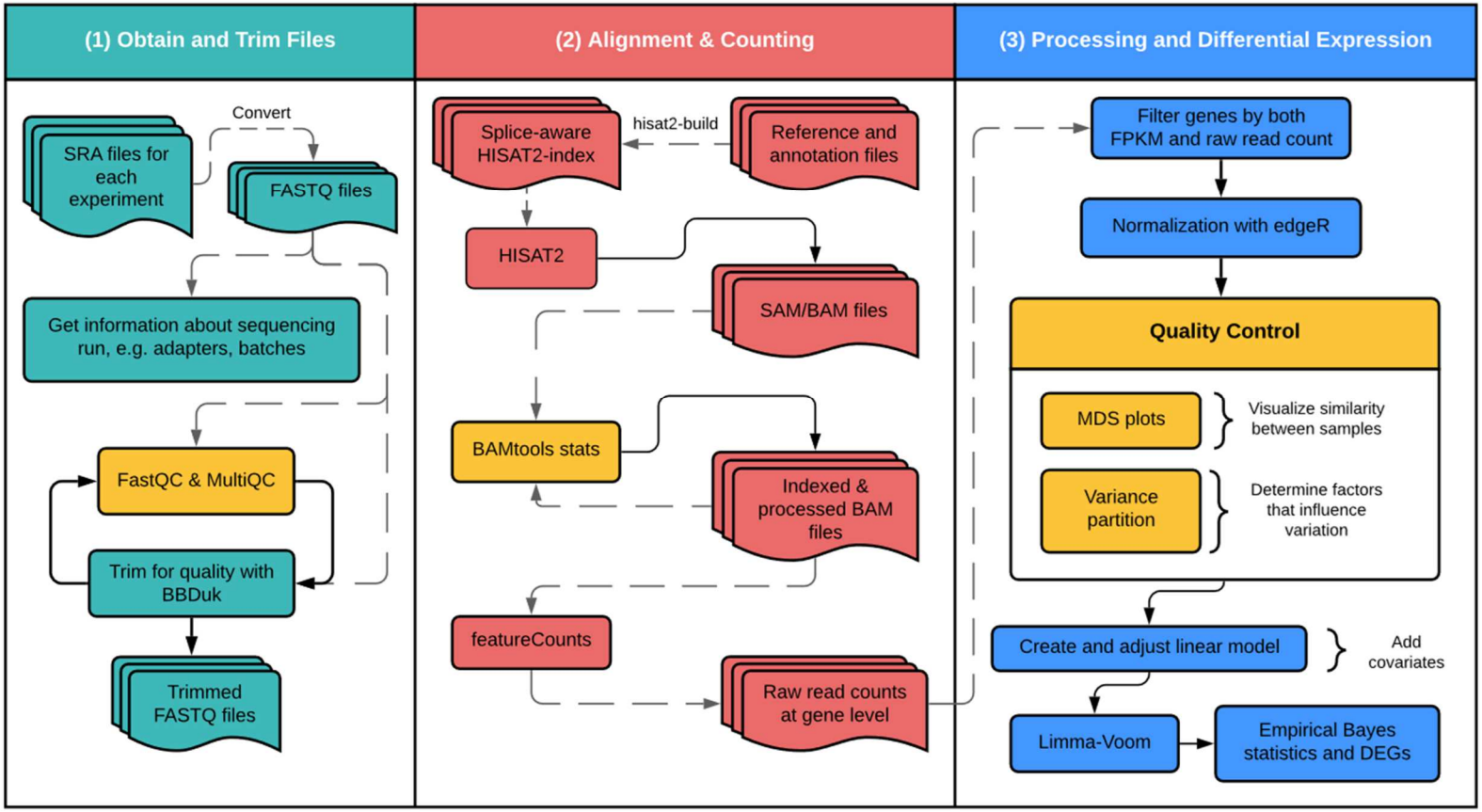
kinesin that plays a role in trafficking of neuronal vesicles (Matsuzaki et al., 2011), including maintenance of axons via transport of mitochondria (Campbell et al., 2014). Meanwhile, *LYPD1*, also called *LYNX2*, modulates nicotinic acetylcholine receptors and is associated with increased fear and anxiety-like behavior in LYPD1-knockout mice (Dessaud et al., 2006; Tekinay et al., 2009). With the exception of missing developmental expression data for *LYPD1*, all three of these genes are also highly conserved. Genes that are not brain specific and show enriched expressed in non-reward regions of the brain (e.g. cerebellum and/or spinal cord) are Low priority candidate genes (*AGK*, *BTG1*, *FKBP4*, *FTH1*, *GPD1*, *PITPNM3*, *RPS6KA2*, *TTLL1*). Still, when selecting candidate genes for preclinical rodent research, conservation is of greater importance than brain expression. Some of these genes are already Low priority for one or more conservation metrics (*FTH1*), while others are highly conserved (*AGK*, *BTG1*, *PITPNM3*).

At its best, RNA-seq allows for unbiased candidate gene discovery across the transcriptome, but the standard common practice fails to capitalize on its potential. Though it may be tempting to reduce sequencing and animal costs by lowering sample size, this is particularly problematic in RNA-seq, because gene expression trends identified with a nominal P value will not necessarily hold with higher sample sizes and more robust statistics. Though financial costs were once a major limiter of sample size, sequencing has become less expensive in recent years, and there remains the opportunity for large-scale collaboration between labs, such as through NIH U or P grants (National Institutes of Health, 2019). With more funding, along with better reporting of methods, results, and code by researchers, RNA-seq can become an important tool in discovery of candidate genes rather than suspect for generating unreliable DEGs (Simoneau et al., 2021). Further, we show here that additional steps taken after differential expression

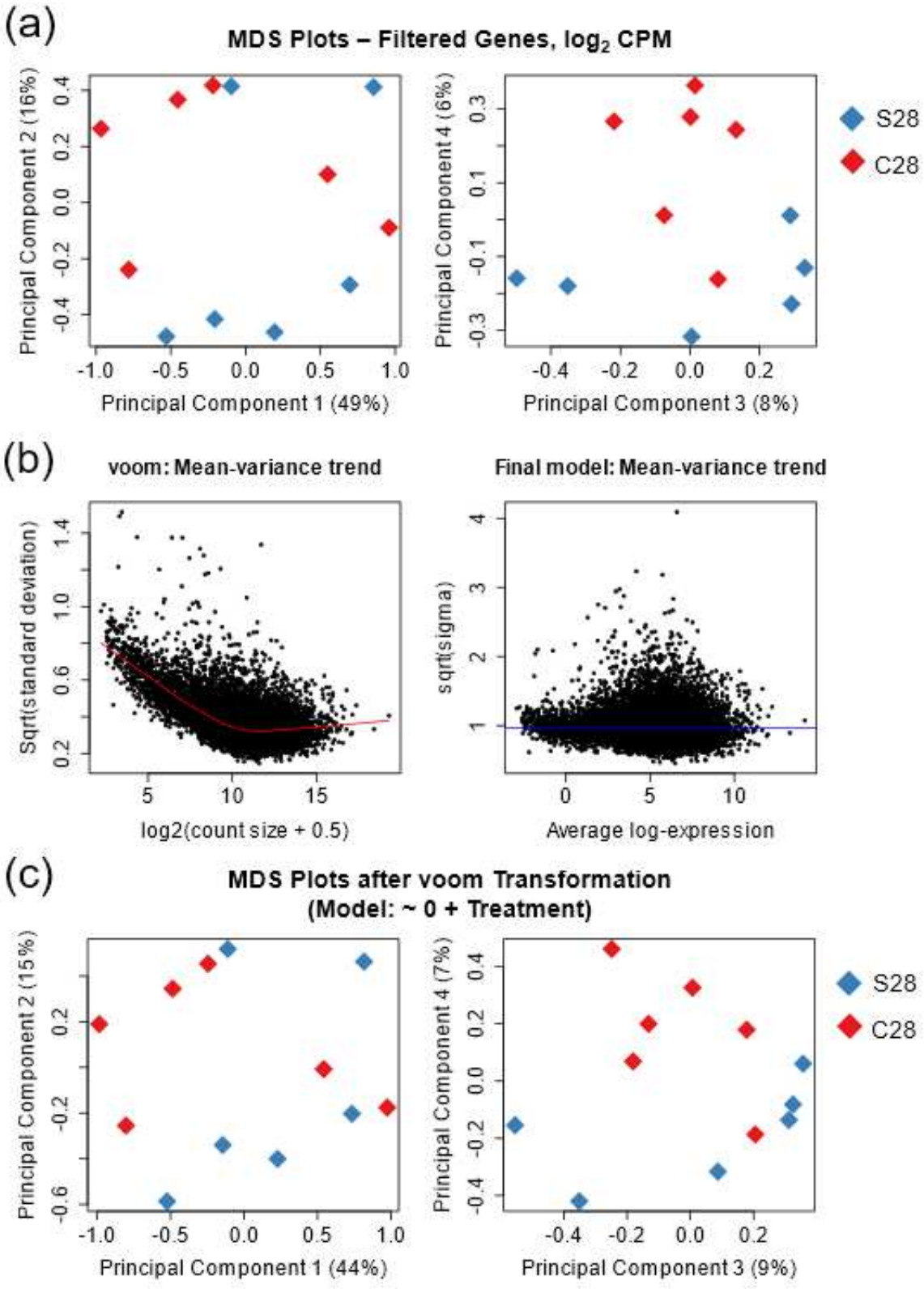
analysis, such as determining the evolutionary conservation and regional expression of candidate genes, can potentially narrow down future targets of study. Utilizing this approach, along with ensuring proper power, will both bolster the reliability of individual RNA-seq studies and provide a stronger foundation for preclinical research to bridge the translational gap.

**Supplementary information**

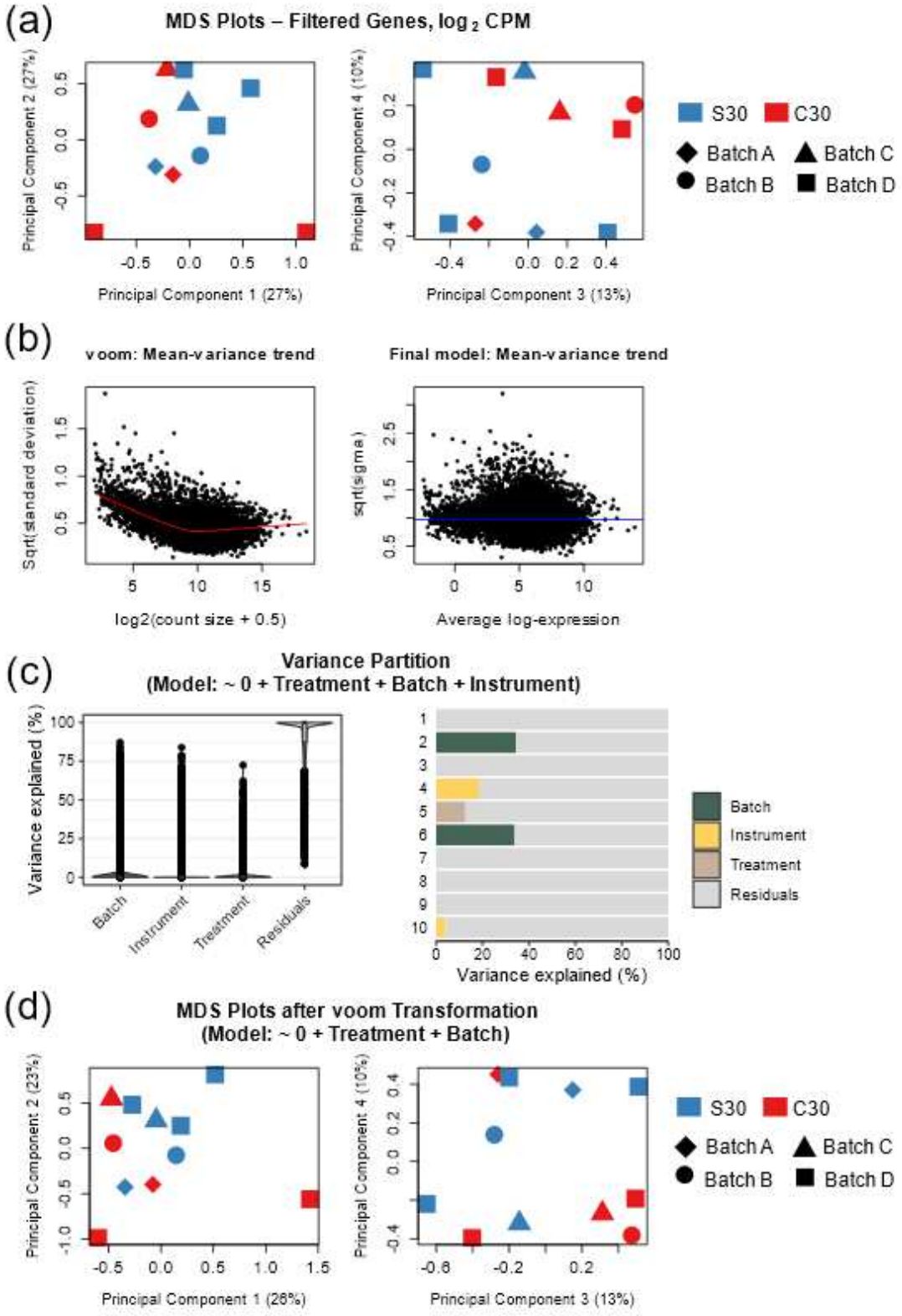
Full names for all discussed genes can be found in Appendix A: Chapter 3 Gene and Tissue Abbreviations. Supplementary Figures 1-4 are included at the end of this chapter. Supplementary Tables 1-11 and Supplementary Figure 5 are provided on GitHub (https://github.com/SexChrLab/Vannan-Dissertation-Supplements).



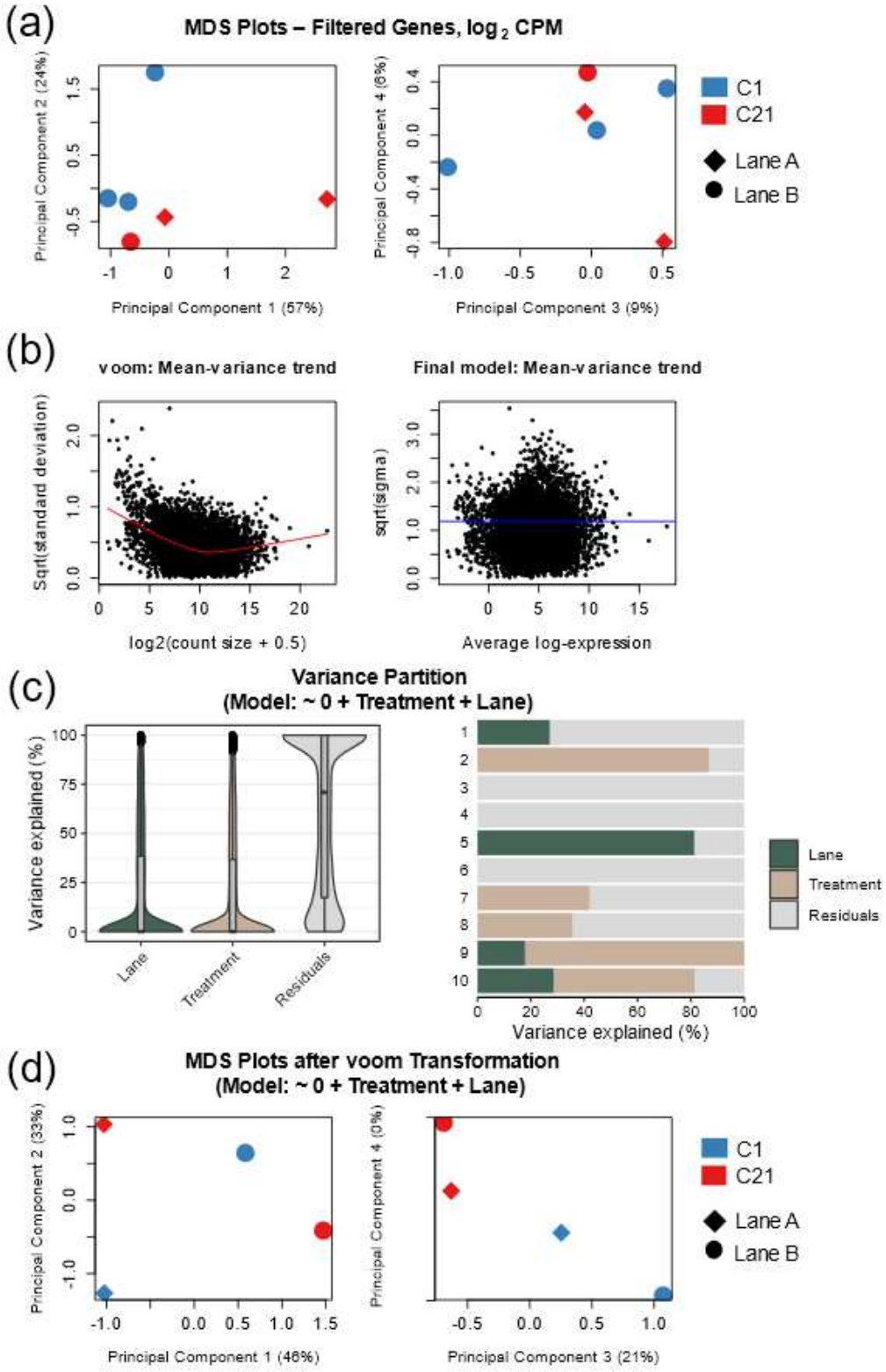

***Chapter 3, Supplementary Figure 1.*** RNA-seq pipeline for differential expression analysis. All 3 datasets were analyzed separately using the above pipeline.


(a)
MDS Plots – Filtered Genes, log2 CPM
Principal Component 2 (16%)
Principal Component 1 (49%)
Principal Component 4 (6%)
Principal Component 3 (8%)
S28
C28
(b)
voom: Mean-variance trend
Sqrt(standard deviation)
log2(count size + 0.5)
Final model: Mean-variance trend
sqrt(sigma)
Average log-expression
(c)
MDS Plots after voom Transformation
(Model: ~ 0 + Treatment)
Principal Component 2 (15%)
Principal Component 1 (44%)
Principal Component 4 (7%)
Principal Component 3 (9%)
S28
C28

***Chapter 3, Supplementary Figure 2.*** RNA-seq processing for Carpenter S28 vs. C28 dataset. (A, C) Multidimensional scaling (MDS) plots for the top 100 genes with the largest fold change differences between groups before (A) and after (C) voom transformation. Principal components 1 and 2 (left) and 3 and 4 (right) are shown, along with the percentage of variation captured by each component. Blue and red signify the S28 and C28 groups, respectively. (B) limma voom was used to apply a linear model to the data. Gene expression was measured in CPM with (A, C) and without (B) log2 transformation.


(a)
MDS Plots – Filtered Genes, log 2 CPM
Principal Component 2 (27%)
Principal Component 1 (27%)
Principal Component 4 (10%)
Principal Component 3 (13%)
S30
C30
Batch A
Batch C
Batch B
Batch D
(b)
voom: Mean-variance trend
Final model: Mean-variance trend
Sqrt(standard deviation)
log2(count size + 0.5)
sqrt(sigma)
Average log-expression
(c)
Variance Partition
(Model: ~ 0 + Treatment + Batch + Instrument)
Variance explained (%)
Batch
Instrument
Treatment
Residuals
Variance explained (%)
(d)
MDS Plots after voom Transformation
(Model: ~ 0 + Treatment + Batch)
Principal Component 2 (23%)
Principal Component 1 (26%)
Principal Component 4 (10%)
Principal Component 3 (13%)
S30
C30
Batch A
Batch C
Batch B
Batch D

***Chapter 3, Supplementary Figure 3.*** RNA-seq processing for Walker S30 vs. C30 dataset. (A, D) Multidimensional scaling (MDS) plots for the top 100 genes with the largest fold change differences between groups before (A) and after (D) voom transformation. Principal components 1 and 2 (left) and 3 and 4 (right) are shown, along with the percentage of variation captured by each component. Blue and red signify the S30 and C30 groups, respectively. Shapes represent different sequencing batches. (B) limma voom was used to apply a linear model to the data, including both Treatment group and Batch, based on plots in (C) which indicate the highest percentage of variance was explained by Batch. (C) Variance partition plots showing the percentage of the gene expression variance explained by the variables sequencing Batch (green), sequencing Instrument (yellow), and Treatment (brown), along with Residuals (gray). Total percentages are shown on the left and are broken down by the first 10 dimensions on the right. Gene expression was measured in CPM with (A, D) and without (B, C) log2 transformation.


(a)
MDS Plots – Filtered Genes, log2 CPM
Principal Component 2 (24%)
Principal Component 1 (57%)
Principal Component 4 (6%)
Principal Component 3 (9%)
C1
C21
Lane A
Lane B
(b)
voom: Mean-variance trend
Sqrt(standard deviation)
log2(count size + 0.5)
Final model: Mean-variance trend
sqrt(sigma)
Average log-expression
(c)
Variance Partition
(Model: ~ 0 + Treatment + Lane)
Variance explained (%)
Lane
Treatment
Residuals
Variance explained (%)
(d)
MDS Plots after voom Transformation
(Model: ~ 0 + Treatment + Lane)
Principal Component 2 (33%)
Principal Component 1 (46%)
Principal Component 4 (0%)
Principal Component 3 (21%)
C1
C21
Lane A
Lane B

***Chapter 3, Supplementary Figure 4.*** RNA-seq processing for Powell C1 vs. C21 dataset. (A, D) Multidimensional scaling (MDS) plots for the top 100 genes with the largest fold change differences between groups before (A) and after (D) voom transformation and removal of 2 outliers. Principal components 1 and 2 (left) and 3 and 4 (right) are shown, along with the percentage of variation captured by each component. Blue and red signify the C1 and C21 groups, respectively. Shapes represent different sequencing lanes. (B) limma voom was used to apply a linear model to the data, including both Treatment group and Lane, based on plots (c) which indicate the highest percentage of variance was explained by Lane. (C) Variance partition plots showing the percentage of the gene expression variance explained by the variables sequencing Lane (green) and Treatment (brown), along with Residuals (gray). Total percentages are shown on the left and are broken down by the first 10 dimensions on the right. Gene expression was measured in CPM with (A, D) and without (B, C) log2 transformation.

CHAPTER 4

# ALTERATIONS IN CIRCULAR RNA EXPRESSION IN THE RAT NUCLEUS ACCUMBENS SHELL MAY REGULATE circRNA-miRNA-mRNA REGULATORY NETWORKS IN A MODEL OF COCAINE CRAVING

**In preparation for publication as:**

Vannan, A., Johnson, M. C., Dell'Orco, M., Powell, G. L., Mellios, N., Wilson, M. A., Perrone-Bizzozero, N. I., Neisewander, J. L. Alterations in circular RNA expression in the rat nucleus accumbens shell may regulate circRNA-miRNA-mRNA regulatory networks in a model of cocaine craving.

## Abstract

Emerging evidence indicates that circular RNAs (circRNAs) regulate complex RNA networks in the brain that underlie synaptic plasticity and psychiatric disease. Here we used bioinformatics tools to analyze publicly available RNA-sequencing (RNA-seq) data from our group, which allowed us to identify candidate circRNAs associated with cocaine craving in the nucleus accumbens (NAc) shell. A new set of male and female rats was then used to validate these findings. Among the candidate craving circRNAs is *circHomer1*, which is widely implicated in learning, memory, and cognition, along with *circArid1a* and *circMapkap1*. In particular, we find that *circAnkrd12*, *circArid1a*, *circHomer1*, and *circMapkap1* are expressed more highly in the NAc shell of rats that show lower cocaine-seeking behavior in a post-abstinence cue-reactivity test using RNA-seq data. Follow-up RT-qPCR suggests that *circHomer1* is specifically downregulated after long term abstinence from cocaine compared to saline controls, but that this effect is specific to bulk RNA and not synaptosomal fractions. As well, *circMapkap1* appears to

be downregulated after short abstinence from cocaine compared to saline. By constructing theoretical competing endogenous RNA (ceRNA) networks, our findings suggest these circRNAs may regulate expression of synaptic plasticity and neuroinflammatory genes through binding and sequestering microRNAs (miRNAs) that typically suppress expression of target mRNAs. These miRNAs include miR-330-5p and miR-760-3p among others. These findings support ongoing work implicating *circHomer1* in psychiatric disorders and encourage the exploration of novel circRNAs such as *circArid1a* and *circMapkap1* in addiction and drug craving research.

## Introduction

Substance use disorders (SUDs) are both financially costly (Cartwright, 2008; National Drug Intelligence Center, 2011) and socially devastating, but effective therapeutic options are limited. A major complication in treating SUDs is that addiction is a chronic, relapsing disorder, in part because sensory cues associated with the drug-taking experience (e.g. a crack pipe or lighter) can take on conditioned stimulus effects with long-lasting salience (Childress et al., 1993, 1988; Ehrman et al., 1992; Grant et al., 1996). Subsequently, exposure to drug-associated cues can drive craving and lead to drug-seeking behavior and relapse (Grimm et al., 2001; Namba et al., 2018; Neisewander et al., 2000; Tran-Nguyen et al., 1998). Worse yet, craving is known to increase during periods of prolonged abstinence (known as the "incubation effect") in both humans (Gawin & Kleber, 1986; Parvaz et al., 2016) and animal models (Grimm et al., 2001; Neisewander et al., 2000; Tran-Nguyen et al., 1998). This is due in part to drug-induced, maladaptive neuroplasticity that persists or worsens during abstinence (Koob & Le Moal, 2005; Russo et al., 2010; A. C. W. Smith & Kenny, 2018). As a result, there is a profound need to understand the neural mechanisms of neuroplasticity and cue-elicited drug craving after abstinence.

Neural plasticity is regulated in part by circular RNAs (circRNAs), a relatively novel class of highly stable, non-coding RNAs (Memczak et al., 2013; You et al., 2015). circRNAs typically form via post-transcriptional back-splicing to produce a circular form of linear pre-mRNA from their host genes (L.-L. Chen & Yang, 2015; Jeck et al., 2013). Though thousands of circRNAs have been identified in humans and rodents, their specific functions are still the subject of ongoing research (W. Chen & Schuman, 2016; Xiang Li et al., 2018; Qu et al., 2015). Importantly, many circRNAs are enriched at synapses and arise from genes that are implicated in synaptic plasticity (You et al., 2015).

Several studies show that circRNAs act as sponges for microRNAs (miRNAs) by binding to and sequestering them, thereby preventing them from suppressing their targets (Hansen et al., 2013). Because circRNAs can compete with other RNAs, particularly mRNAs, for miRNA binding, these RNAs form competing endogenous RNA (ceRNA) networks of circRNA-miRNA-mRNA interactions that regulate gene expression(Salmena et al., 2011). circRNAs can also modulate transcription of their linear mRNA counterparts through competitive alternative splicing (Ashwal-Fluss et al., 2014; Hafez et al., 2022). Overall, these highly conserved RNAs (Rybak-Wolf et al., 2015) have great potential as translational targets for regulating mechanisms of drug-induced neural plasticity and drug craving.

In this study, we utilize prior RNA-sequencing (RNA-seq) data to identify candidate circRNAs predicted to regulate gene networks in a model of cocaine craving. In this previous experiment (Powell, Vannan, et al., 2020), male rats were trained in cocaine self-administration (SA), and then placed in environmental enrichment (EE) during abstinence. In line with prior work (Thiel et al., 2010, 2009; Vannan et al., 2018), we found that after prolonged (21d) abstinence, rats had increased cocaine-seeking behavior in a cue-reactivity test, and that this incubation of craving was attenuated by EE. Using circRNA prediction tools (CIRI2, CIRCExplorer2) on RNA-seq data conducted on the nucleus accumbens (NAc) shell of these rats revealed three circRNAs that may underlie cocaine craving, including *circHomer1*, which has previously been studied in addiction and reversal learning (Dell'Orco et al., 2020; Zimmerman et al., 2020), and a novel candidate, *circMapkap1*. To build on this foundation, I validated expression of these circRNAs using tissue obtained from my collaborator, Michael Johnson, who tested the hypothesis that male and female rats would exhibit down-regulation of *circHomer1* after protracted abstinence for his Master's thesis research. I then

constructed ceRNA networks using bioinformatics analyses of predicted circRNA-miRNA and miRNA-target interactions. Overall, these studies provide evidence for several candidate circRNAs in the modulation of cue-elicited cocaine craving, and suggest potential regulatory networks for these circRNAs in the NAc shell.

## Methods

**RNA-seq and data availability.** RNA-seq data collected for a prior experiment (Powell, Vannan, et al., 2020) were used to identify circRNAs involved in cocaine craving. Briefly, RNA was prepared using the Nugen Ovation RNA-seq system (#7102-32; Tecan Genomics, Inc., Redwood City, California USA) and had not been treated to deplete linear RNAs or enrich circRNAs. Libraries were created using the Kapa Biosystem library, and sequencing was performed on a 1 × 75 flow cell using an Illumina NextSeq500 instrument. Raw RNA-seq files can be downloaded from NCBI's Sequencing Read Archive (SRP246331) and are deposited in the Gene Expression Omnibus (GEO; GSE144606). The raw circRNA count files will be uploaded to GEO upon submission of this work for publication. Descriptions of analyses and associated code are provided on GitHub (https://github.com/SexChrLab/Craving-circRNAs).

**Animals.** Adult male and female Sprague-Dawley rats (200-225 g on arrival; Charles River Laboratories, Wilmington, Massachusetts, USA) were individually housed on arrival (21.6 x 45.7 x 17.8 cm) on a 14:10 light:dark cycle with access to food and water ad libitum. All experiments were conducted in accordance with the National Institutes of Health Guide for the Care and Use of Laboratory Animals and approved by the Institutional Animal care and Use Committee at Arizona State University.

**Surgical procedures.** For rats used to generate the prior RNA-seq data, see surgical procedures in (Powell, Vannan, et al., 2020). The new set of rats underwent

surgery under 2-3% isoflurane anesthesia to implant a catheter into the jugular vein as previously described (Powell et al., 2019). A catheter was implanted into the jugular vein and tunneled subcutaneously to a small incision between the shoulder blades and secured there with a backpack-style cannula (PlasticsOne®, Roanoke, VA). After surgery, rats were given meloxicam (1 mg/kg/mL, SC) and buprenorphine (0.05 mg/kg/mL, SC) as analgesics and the antibiotic cefazolin (100 mg/kg/mL, IV). Meloxicam was administered for at least 5 days after surgery, and cefazolin was given for at least 2 days after surgery. Thereafter, catheters were flushed daily with heparinized saline (70 U/mL, IV) to maintain patency. Catheter function was assessed periodically with methohexital sodium (16.67 mg/mL, IV) (Pentkowski et al., 2010).

**Self-administration and cue reactivity testing.** Separate sets of rats were tested using similar methods. For (Powell, Vannan, et al., 2020), male rats (N = 62) were trained to self-administer cocaine (0.75 mg/kg/0.1 mL infusion, IV) in daily 2h sessions 6d/wk. Self-administration (SA) sessions took place in operant conditioning chambers (30 × 24 × 21 cm; Med Associates Inc, St. Albans, VT). Completion of an operant schedule resulted in a 6s drug infusion and contingent presentation of light and tone cues after a 1-s delay. A second inactive lever was available, and presses were recorded though they produced no consequences. Infusions were followed by a 20s timeout during which infusions and cues were not available. Rats began SA training on a fixed ratio (FR) 1 schedule, with a single active lever press required for a drug infusion. Within the session, rats advanced to variable ratio (VR) 2, 3, and 5 schedules if 7 infusions were received in 1 hr. After three consecutive sessions ending on a VR5 schedule, the starting schedule was advanced, such that rats progressed through FR1, VR2, VR3, and VR5 schedules over the course of ≥ 15 sessions. After each session, rats were fed to maintain them at 90% of their free-feeding body weight. Upon completion of three consecutive

sessions on a VR5 starting schedule, food was given ad libitum thereafter throughout the rest of the experiment. Rats that achieved stable (< 25% variability) cocaine infusions across three consecutive sessions after free feeding were placed into abstinence for 1 or 21 d, either in the same isolated housing (ISO) as before or in environmental enrichment (EE), which is known to attenuate drug craving in rats when used as a therapeutic intervention during abstinence (Thiel et al., 2012, 2009; Vannan et al., 2018). In EE, rats were housed in large tubs (74 × 91 × 36 cm) in groups of 3-6 and were given access to toys, such as a running wheel or PVC tubes, that were exchanged every 3 days to enhance novelty. Treatment groups were counterbalanced based on cocaine consumption. Rats were handled daily during both SA and abstinence. After 1 or 21d of forced abstinence, rats returned to their SA chambers for a 1 h cue-reactivity test. During this time, cocaine was unavailable, but the light and tone cues that were previously paired with cocaine infusions were conferred on an FR1 schedule. Cocaine-seeking behavior, a proxy for drug craving, was measured by active lever presses.

An additional set of rats was used to validate craving-related circRNAs of interest. Male and female rats underwent behavioral testing as described above, with some differences. For this set, all rats were maintained in isolated housing throughout the experiment, including during the 1 or 21 d abstinence period, and additional rats were added as saline-yoked controls. For saline animals, cues were presented and infusion of saline (0.1 mL) was delivered only when a partnered rat received a cocaine infusion, and thus these rats did not develop a cue-drug association as with cocaine SA. This second set of animals was run in three separate behavioral cohorts (n = 12-21/cohort) to achieve the complete sample size (N = 48, 24 males and 24 females).

**Identification, expression, and species conservation of circRNAs.** Raw sequencing reads were assessed for quality using FastQC (Andrews, 2010) and MultiQC

(Ewels et al., 2016) and were subsequently trimmed of adapters and overrepresented sequences using BBDuk from the BBTools suite (http://jgi.doe.gov/data-and-tools/bbtools). Reads were trimmed using the following parameters: ref=adapters.fa ktrim=r k=21 mink=11 hdist=2 trimpolya=10 trimpolyg=10 qtrim=rl trimq=30 minlen=38 maq=20, where minlen was set to approximately half of the original read length (75 bp). Trimmed reads were aligned to the Ensembl mRatBN7.2 rat genome release 105 using BWA (H. Li, 2013) using the threshold parameter -T 19 as recommended by the CIRI2 developers. The BamTools (Barnett et al., 2011) and Samtools (H. Li et al., 2009) packages were used to assess the number of mapped reads after alignment.

To increase accuracy in identifying circRNAs from a standard RNA-seq preparation, two separate tools were used to verify the presence of circRNAs. Aligned reads from BWA were run through the basic CIRI2 (Gao et al., 2018) pipeline and through CIRCexplorer2 (X.-O. Zhang et al., 2016) annotate and parse functions, all using default settings. As part of these pipelines, back-splice junctions were quantified and annotated to their corresponding host genes or introns. To increase confidence in the analysis, only those circRNAs identified in a given sample by both tools with at least 2 reads were selected for further analysis. CIRI2 and CIRCexplorer2 use 1-based and 0-based indexing, respectively, so genomic locations from CIRCexplorer2 were first converted to 1-based indexing before comparing between tools. To compare circRNA expression across treatment groups, read counts were first converted into counts per million (CPM) based on the number of BWA-mapped reads per sample.

To determine whether the identified rat circRNAs are conserved with human, we utilized NCBI's standard nucleotide BLAST (blastn) to compare rat circRNA sequences to their closest human circRNAs obtained from circRNADB (X. Chen et al., 2016). We

also obtained the sequence similarity of the linear genes for rat compared to human using biomaRt in R (Durinck et al., 2005) and extracting data from the human genome from Ensembl release 105 (December 2021).

For simplification, circRNAs are hereafter referred to by "*circ*" and their host gene (e.g. *circHomer1*), though other circRNAs may arise from the same host genes. For circRNAs discussed in this text, genomic locations are listed in Table 2. For all other circRNAs, please see the CIRI2 and CIRCExplorer2 output files deposited in GEO.

**Synaptosome and RNA isolation, cDNA synthesis, and RT-qPCR.** Within 10 minutes of completing the cue reactivity test, rats were sacrificed by isoflurane overdose and brains were harvested. The fresh tissue was sliced into 2 mm coronal sections containing the NAc. To isolate the shell, first the core was excised with a 1 mm tissue punch, and then a secondary 2 mm punch was taken using the removed core as a landmark. Next, crude synaptosomes (SYN) were isolated as previously described (Boese et al., 2016; Dell'Orco et al., 2020; Rao & Steward, 1991). Shell tissue was homogenized in 200 mL of ice cold homogenization buffer (0.1 mM EDTA, 0.25 mM DTT, 2 mM HEPES in a 7.5 pH, 0.32 M sucrose solution) which was supplemented with 200 U/mL RNAseOUT™ (Invitrogen). Samples were centrifuged at 2,000 x g for 2 min, resulting in a pellet containing nuclei and cellular debris (P1) and a supernatant (S1a). P1 was resuspended in 100 µL homogenization buffer and centrifuged again at 2,000 x g for 2 min. The resulting supernatant (S1b) was combined with S1a and then centrifuged at 17,000 x g for 10 min. The final supernatant (S2) was removed, and the final pellet (P2) constituted the crude synaptosomal fraction. A subset of samples were validated for purification efficiency by measuring the brain-specific and SYN-enriched gene *BC1* in S1 and S2 supernatant fractions and the P2 pellet (Boese et al., 2016). For each sample, a

portion of homogenized tissue 50 µL was stored for isolation of total RNA without fractionation. RNA was stored at -80C until further processing.

RNA isolation was performed on both SYN and total RNA fractions using the standard Trizol® (Invitrogen) method. RNA samples (35 ng), were used to prepare circRNA-enriched and circRNA-depleted cDNA using the SuperScript™ III First-Strand Synthesis System (Invitrogen, # 18080051). cDNA was prepared with random hexamers or oligo-dT to aid in circRNA or linear mRNA detection, respectively. cDNA for each sample was run in triplicate for each circRNA or mRNA in conjunction with the and iTaq™ Universal SYBR® Green Supermix (BioRad, #1725121) kit. Relative expression from RT-qPCR data was calculated using the comparative $2^{-\Delta CT}$ method using GAPDH as a control (Livak & Schmittgen, 2001). To account for cohort effects $2^{-\Delta CT} * 100$ values were divided by the average expression values for the appropriate saline controls. Primers for both circRNAs and mRNAs are listed in Table 1. circRNA primers were designed to target the circRNA without amplifying the linear mRNAs of the host gene.

Chapter 4, Table 1

***circRNA and mRNA Primers Tested With RT-qPCR***

| Target | Forward Primer | Reverse Primer |
|---|---|---|
| *circ-Homer1* | TCAATGGGACAGATGATG | TTGTGTTTGGGTCGATCTGG |
| *circ-Mapkap1* | CCCTCTGCAGCTGAATAACC | TTCTGTGGGCCAGTCTCTTTA |
| *Gapdh* | CTCTCTGCTCCTCCCTGTTC | TACGGCCAAATCCGTTCACA |

**Constructing ceRNA networks.** circRNA sequences were obtained using the R package FcircSEC (Hossain et al., 2020) and the data provided by CIRCexplorer2,

including the circRNA chromosome, start and end sites, strand, and exon number and lengths where applicable. Through FcircSEC, transcript information was extracted from the CIRCexplorer2 files, and then circRNA sequences were pulled from the rat mRatBN7.2 genome using biomaRt in R (Durinck et al., 2005).

To determine likely circRNA-miRNA interactions, the 764 rat miRNA sequences from miRBase release 22.1 (Griffiths-Jones, 2004; Kozomara et al., 2019) were used as queries against 1125 potential circRNA target sequences using miRanda v3.3a (Enright et al., 2003) using default settings. To increase confidence, predicted circRNA-miRNA interactions were narrowed to those with a match score ≥ 150 and a free energy ≤ -20. If a circRNA was predicted to bind to a miRNA at more than one site, its match score and free energy scores were summed across all predicted binding sites.

To find miRNA interactions with potential targets, all the summary counts and predictions for 759 rat miRNAs were downloaded from TargetScanMouse 8.0 (Agarwal et al., 2015; Friedman et al., 2009). Direct predictions between rat miRNA and rat mRNA sequences are not available on TargetScan, so the TargetScanMouse was used to obtain predictions for rat miRNAs and 3'UTR sequences that are homologous in mouse. This provided information for 759 of the 764 rat miRNAs. Most of the target sequences were from mRNAs, but some were long non-coding RNAs or other RNA targets; collectively, these are referred to as "mRNAs" throughout this paper. When referring to mRNAs in these networks (e.g. *Rgp1*), we use the gene names from rat (e.g. *RGP1 homolog, RAB6A GEF complex partner 1*) rather than their mouse orthologs (e.g. *RAB6A GEF complex partner 1*), for clarity of reading. miRNA-mRNA interactions were labeled as either "High", "Moderate", or "Low" confidence based if they had a cumulative weighted context++ score (CWCS) ≤ -0.4, > -0.4 and ≤ -0.2, or > -0.2, respectively. For reference, CWCS of -0.2 and -0.4 indicate the miRNA is predicted to suppress the

mRNA's expression by 13% and 25%, respectively (Agarwal et al., 2015; QIAGEN, n.d.). miRNA-mRNA interactions were only included if they had "High" or "Moderate" CWCS and at least 1 binding site predicted to be conserved across mammals. Finally, the list of miRNAs was narrowed to those that are known to be expressed in the NAc shell based on our recent Nanostring array of 423 rat miRNAs (Vannan et al., 2021) and/or expressed in the cerebrum according to the RATEmiRs database of 856 rat miRNAs using the data collected from Maastricht University (Bushel et al., 2018, 2020). After filtering to keep miRNAs expressed at least 2 CPM in our prior array and/or 2 TMM in the RATEmiRs array, 478 miRNAs remained.

The final dataset consisted of 1105 circRNAs, 410 miRNAs belonging to 319 miRNA families, and 9544 mRNAs. To assess potential indirect interactions of circRNAs and mRNAs (i.e., potential ceRNA pairs) through miRNA intermediaries, hypergeometric tests were performed on each ceRNA pair with similar methodology to prior publications (Dell'Orco et al., 2021; J.-H. Li et al., 2014). Briefly, the hypergeometric tests used 4 parameters, determined for each ceRNA pair: 1) the total number of miRNAs with target predictions across the dataset, including those miRNAs that did not have any potential circRNA interactions, 2) the number of miRNAs that interact with a given mRNA, 3) the number of miRNAs that interact with a given circRNA, and 4) the number of miRNAs predicted to interact with both the circRNA and the mRNA, i.e. both members of the ceRNA pair. Each miRNA was considered individually for this analysis, though miRNAs in the same family share the same seed sequence and are predicted to bind to the same target mRNAs. Though some miRNAs are predicted to bind to a given circRNA at more than one binding site, each unique circRNA-miRNA pair was only considered once in the analysis. Hypergeometric tests for 2,168,937 potential ceRNA pairs were subjected to FDR correction (Benjamini &

Hochberg, 1995). ceRNA-pairs with FDR < 0.05 with at least 4 shared miRNA interactions were considered significant. ceRNA networks were visualized using Cytoscape v3.9.1 (Su et al., 2014).

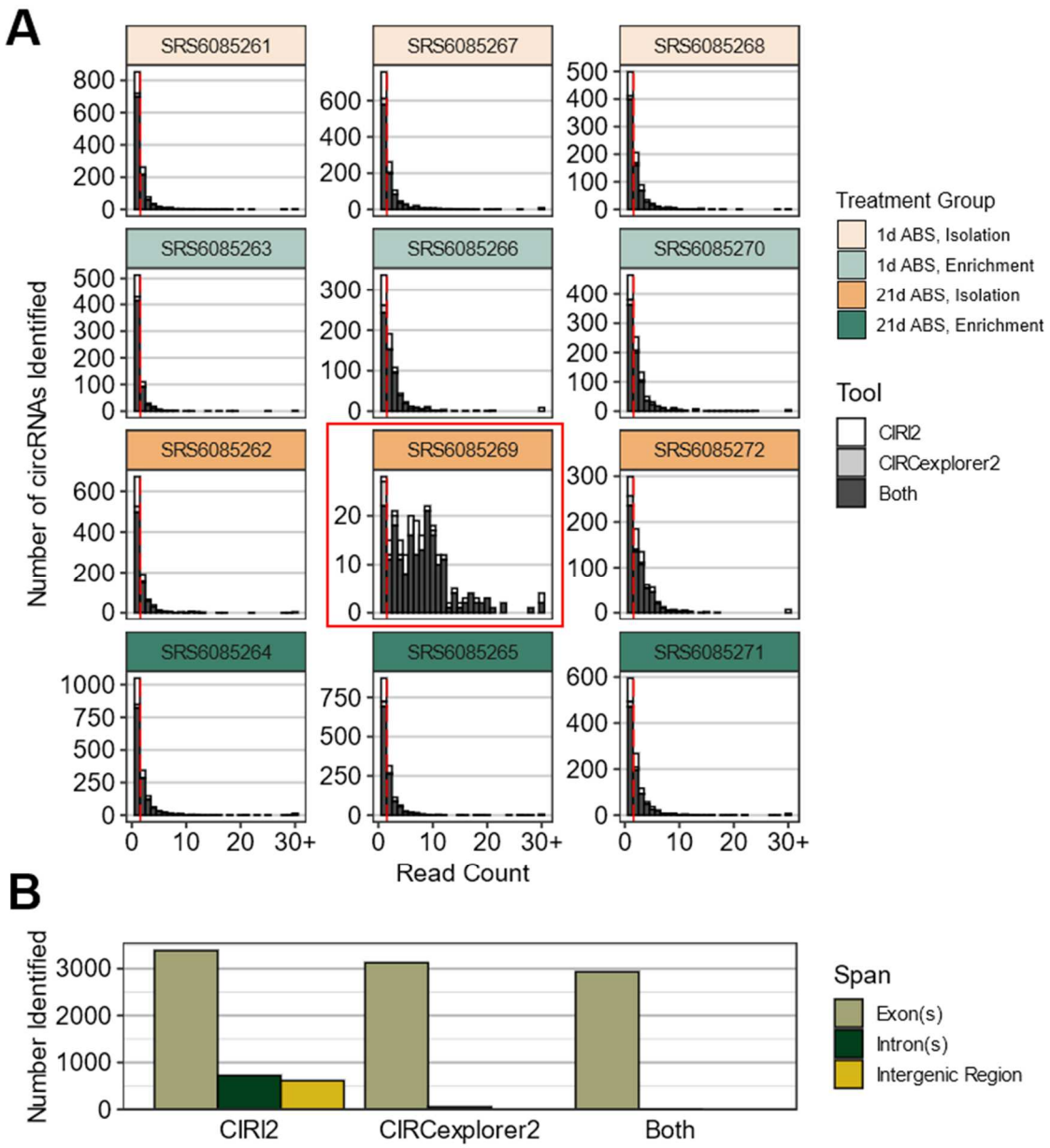
**Statistical analyses.** Statistical analyses were performed in R 4.1.2 and Graphpad Prism 8. For RNA-seq data, the circRNA identification Tool (CIRI2, CIRCexplorer2) and Sequencing Lane were included in a linear model assessing the effects of Housing condition (Isolation, Enrichment) by Abstinence Length (1, 21d), which was subsequently analyzed using two-way ANOVAs. For behavioral and RT-qPCR data, saline-yoked animals were combined into one control group as there was no difference among them. Data were then compared using two-way ANOVAs of Sex (male, female) by Group (saline-yoked, cocaine 1d abstinence, cocaine 21d abstinence). Values higher than the third quartile (Q3) + 1.5x the interquartile (IQR) range or lower than the first quartile (Q1) – 1.5x IQR were removed as outliers, with a maximum of one outlier per group. If two samples in the same group were called as outliers, neither value was removed. For RT-qPCR, expression values were divided by the average for the appropriate saline control animals in the same cohort. We tested for outliers both before and after controlling for saline expression. Outliers were removed prior to controlling for saline, though if there were no outliers prior to normalization, we checked for outliers after controlling for saline. Type III ANOVAs were used where there was a significant interaction effect, and Type II ANOVAs were used in the absence of interaction effects. Specific differences between groups were assessed using post-hoc Tukey's tests. Hypergeometric tests were performed as described above.

## Results

**Identification of candidate circRNAs related to cocaine craving.**

Hundreds of circRNAs were identified in each sample, with the majority of circRNAs expressed at low levels (≤ 30 reads, Figure 1A). CIRI2 identified more circRNAs than CIRCexplorer2 overall, though most circRNAs were identified by both tools. Collapsing across all samples, most of the circRNAs identified by both tools are exonic (2931 circRNAs, 99.73%) though some were intronic (8 circRNAs, 2.72%) (Figure 1B). CIRI2 identified circRNAs spanning intergenic regions, though these were not identified by CIRCexplorer2. One sample from the isolation, 21d group had unusually high sequence duplication levels according to FastQC (68.3%, while other samples ranged from 24.7 to 45.9%) and low identification of circRNAs (Figure 1A), and thus it was excluded from subsequent analyses.

In the search for candidate circRNAs related to cocaine craving, we selected for circRNAs that were identified by both tools, but in only 1, 2, or 3 treatment groups (Table 2). *circBbx* and *circGlce* were identified exclusively in the enrichment 21d and isolation 21d groups, respectively. Three other circRNAs were exclusively identified in one treatment group, but only after filtering low-confidence circRNAs (read count < 2): *circPpil4* (isolation 21d), along with *circAABR07049060.1* and *circPbrm1* (enrichment 21d). However, all 5 of these circRNAs were identified at very low read counts (1-5 reads per sample), reducing our interest in them as candidate circRNAs in cocaine craving.




***Chapter 4, Figure 1.*** circRNA identification from RNA-seq data. (A) circRNAs were identified by each tool at various expression levels, though most were expressed at low read counts (< 30). circRNAs with a read count ≥ 30 were combined for this plot. Across all samples, hundreds of circRNAs were identified, though one outlier sample (marked with a red box) had far fewer circRNAs identified by either tool. Red dashed lines indicate the filtering threshold, and all circRNAs below this threshold (expressed with > 2 reads) were removed prior to statistical analyses. (B) The type and genomic span of circRNAs identified by either or both tools before filtering. The majority of identified were exonic, with some intronic or spanning intergenic regions.

Chapter 4, Table 2

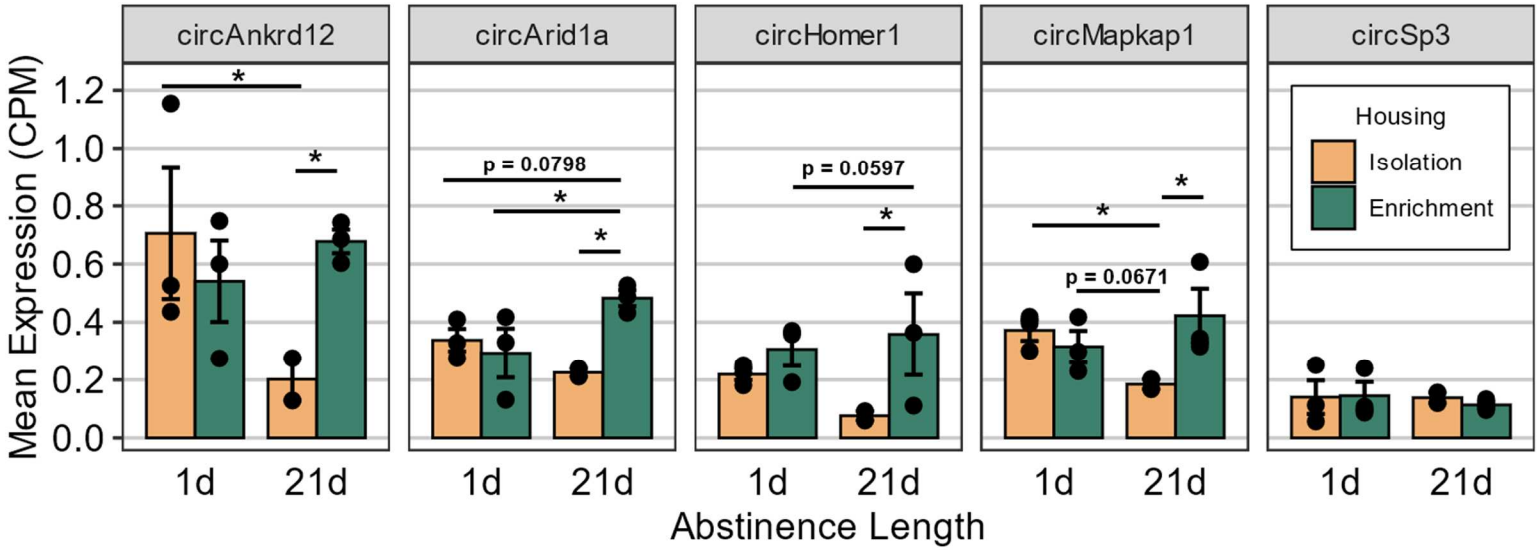
***circRNA Composition and Conservation***

| circRNA and Corresponding Gene ID | Genomic Location | Treatment Group(s) | circRNA Length (bp) | Exon Lengths (bp) | Homologous Human circRNA | Rat-Human Sequence Similarity (circRNA) | Rat-Human Sequence Similarity (Gene) |
|---|---|---|---|---|---|---|---|
| *circAnkrd12,* ENSRNOG00000012733 | 9:105629568-105660854 | All | 994 | 148, 143, 201, 147, 69, 148, 138 | hsa_circ_31497 | 89.84% | 80.31% |
| *circArid1a,* ENSRNOG00000006137 | 5:145948264-145951212 | All | 786 | 117, 456, 213 | hsa_circ_17116 | 92.10% | 96.50% |
| *circHomer1,* ENSRNOG00000047014 | 2:24583888-24598599 | All | 522 | 157, 132, 93, 140 | hsa_circ_21738 | 94.25% | 97.46% |
| *circMapkap1,* ENSRNOG00000017583 | 3:17741317-17753437 | All | 567 | 328, 90, 149 | hsa_circ_29649 | 93.83% | 97.13% |
| *circSp3,* ENSRNOG00000060479 | 3:57821407-57847776 | All | 1553 | 93, 1360 | hsa_circ_00233 | 95.88% | 96.80% |
| *circBbx*,* ENSRNOG00000001971 | 11:50592225-50597061 | Enrichment, 21d ABS | 1291 | 1006, 136, 149 | Unknown | N/A | 87.46% |
| *circPbrm1*,* ENSRNOG00000028227 | 16:6256270-6268891 | Enrichment, 21d ABS | 1241 | 280, 106, 643, 212 | Unknown | N/A | 97.45% |

| circRNA and Corresponding Gene ID | Genomic Location | Treatment Group(s) | circRNA Length (bp) | Exon Lengths (bp) | Homologous Human circRNA | Rat-Human Sequence Similarity (circRNA) | Rat-Human Sequence Similarity (Gene) |
|---|---|---|---|---|---|---|---|
| *circPpil4,* ENSRNOG00000015552 | 1:2277718-2282190 | Isolation, 21d ABS | 602 | 62, 65, 118, 143, 97, 117 | hsa_circ_28189 | 95.73% | 97.15% |
| *circGlce,* ENSRNOG00000025372 | 8:62559509-62579508 | Isolation, 21d ABS | 939 | 246, 599, 94 | hsa_circ_24529 | 90.74% | 96.11% |

Information on circRNAs of interest, including the genomic location and host genes. Where possible, sequence similarity was calculated for circRNAs using BLAST for homologous human sequences obtained from circRNADB, and for genes using data from Ensembl and biomaRt. Included circRNAs were either expressed in all 11 samples or expressed exclusively in one of the 4 treatment groups: Isolation 1d abstinence (ABS), Enrichment 1d ABS, Isolation 21d ABS, or Enrichment 21d ABS. * Indicates circRNAs that had no human ortholog listed in circRNADB. In all three cases, circRNADB listed circRNAs originating from the orthologous human gene, though the circRNAs were not themselves orthologous. ^ This circRNA originates from the rat gene *AABR07049060.1*, which is a low-confidence ortholog of the human gene *CCDC171* according to Ensembl.

We then selected for differential expression across groups for circRNAs that were confidently expressed (read count > 2) in all 11 samples. Using this approach, we found that 5 circRNAs were identified in all samples: *circAnkrd12*, *circArid1a*, *circHomer1*, *circMapkap1*, and *circSp3* (Table 2). For these circRNAs, expression (CPM) was compared using a linear model of the Housing x Abstinence Length interaction while accounting for circRNA identification Tool (CIRI2, CIRCExplorer2) and sequencing Lane (2 different lanes) as covariates (Figure 2). One circRNA, *circSp3*, showed consistent expression across all 4 groups, and did not have a significant main effect of Housing ($F_{1,16}$ = 0.0535, p = 0.81998), Abstinence Length ($F_{1,16}$ = 0.3821, p = 0.5452), or Housing x Abstinence Length interaction ($F_{1,16}$ = 0.1806, p = 0.6765). Three circRNAs showed a significant effect of Tool: *circAnkrd12* ($F_{1,16}$ = 5.225, p = 0.036), *circArid1a* ($F_{1,16}$ = 28.877, p < 0.0001), and *circMapkap1* ($F_{1,16}$ = 22.701, p = 0.0002), which also had a significant effect of Lane ($F_{1,16}$ = 9.526, p = 0.0071). Three circRNAs had a significant main effect of Housing: *circArid1a* ($F_{1,16}$ = 5.236, p = 0.0361), *circHomer1* ($F_{1,16}$ = 9.4968, p = 0.0071), and *circMapkap1* ($F_{1,16}$ = 7.313, p = 0.0156), all of which showed higher expression for enrichment compared to isolation. Lastly, three circRNAs showed a significant HousingxAbstinence Length interaction: *circAnkrd12* ($F_{1,16}$ = 8.663, p = 0.0095), *circArid1a* ($F_{1,16}$ = 14.382, p = 0.0016), and *circMapkap1* ($F_{1,16}$ = 17.920, p = 0.0006). Though not statistically significant, *circHomer1* showed a trend for a Housing x Abstinence Length interaction ($F_{1,16}$ = 3.116, p = 0.0966). With only 2-3 rats per group, we may be underpowered to detect a significant difference in the samples, especially since circRNAs are enriched specifically at synapses and we measured total tissue RNA in the RNA-seq experiment. However, we moved forward with *circHomer1* as a candidate circRNA because it was trending toward significance and has been implicated in other addiction models (Bu et al., 2019; J. Li et al., 2020).

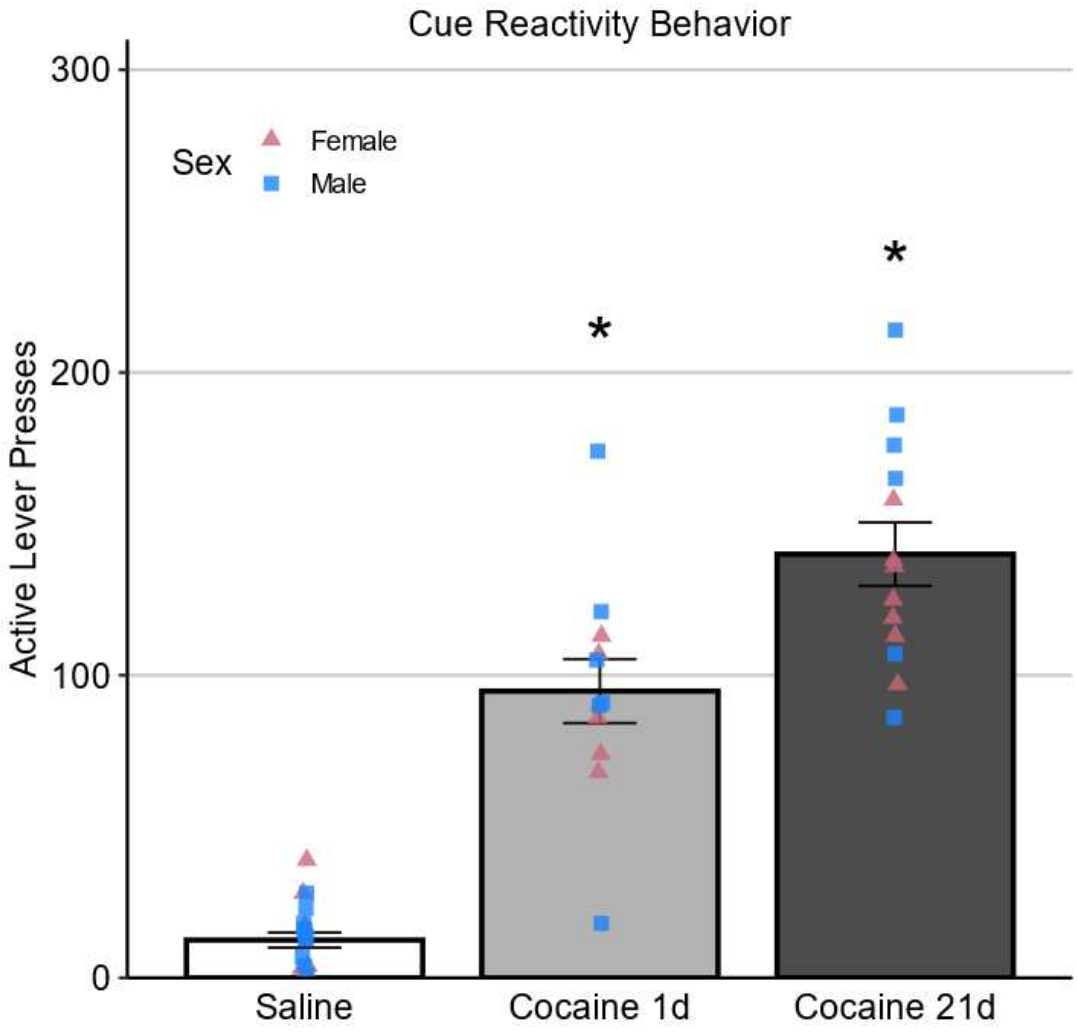





**Chapter 4, Figure 2.** circRNA expression from RNA-seq data. Read counts were first converted to counts per million mapped reads (CPM) for each of the 5 circRNAs expressed in all 11 samples. Expression differences between housing condition and abstinence length were assessed using a linear model accounting for circRNA identification tool and sequencing lane. Expression is depicted here as the mean for each sample across both tools (CIRI2 and CIRCexplorer2). * indicates a significant difference with Tukey's test ($p < 0.05$). P-values for non-significant trends ($p < 0.1$) are also listed.

For those circRNAs showing a Housing effect (*circHomer1*) or Housing x Abstinence Length interaction (*circAnkrd12*, *circArid1a*, and *circMapkap1*), we further probed for differences between groups with post-hoc Tukey's tests. In all cases, it appeared that the group with the highest cocaine-seeking, isolation 21d (Powell, Vannan, et al., 2020), showed lower expression relative to other treatment groups. For *circAnkrd12*, there was a significantly lower expression in isolation 21d compared to both isolation 1d ($t_{16}$ = -3.098, p = 0.0316) or enrichment 21d ($t_{16}$ = -2.927, p = 0.0440). *circArid1a* showed significantly lower expression in the isolation 21d group compared to enrichment 21d ($t_{16}$ = -4.355, p = 0.0025) and enrichment 1d ($t_{16}$ = -3.421, p = 0.0166), and a trend for higher expression compared to isolation 1d ($t_{16}$ = 2.613, p = 0.0798). *circHomer1* showed significantly lower expression in isolation 21d compared enrichment 21d ($t_{16}$ = -2.768, p = 0.0185), and a trend for higher expression in enrichment 21d compared to enrichment 1d ($t_{16}$ = 3.366, p = 0.0597). For *circMapkap1*, there was significantly lower expression in the isolation 21d group compared to enrichment 21d ($t_{16}$ = -4.640, p = 0.0014) and isolation 1d ($t_{16}$ = -3.729, p = 0.00089), along with a trend for higher expression in enrichment 1d compared to isolation 21d ($t_{16}$ = 2.706, p = 0.0671). Overall, the expression patterns of these 4 circRNAs suggest the potential for a protective role in attenuating cocaine craving. We moved forward with exploring these circRNAs, narrowing the focus to those that are highly conserved with human (< 90% sequence similarity for both the circRNA and host gene): *circArid1a*, *circHomer1* and *circMapkap1* (Table 2). Additionally, the latter two circRNAs were validated with RT-qPCR.

**Male and female rats show elevated cue-reactivity after prolonged abstinence, indicative of incubated cocaine craving.** For the validation experiment, behavior data was primarily collected by my collaborator, Michael Johnson. Cocaine-seeking behavior, measured by active lever presses in a post-abstinence cue reactivity test, was compared between cocaine SA and saline-yoked groups. Rats were assigned to either 1 or 21d of abstinence, counterbalancing based on average cocaine intake during SA. For cue reactivity, 2-way ANOVAs of Sex and Group were performed after removing one outlier each from the male saline, male cocaine 21d, and female cocaine 1d groups. The ANOVA indicated no significant effect of Sex ($F_{2, 39} = 2.142$, $p = 0.1513$) nor Treatment x Sex interaction ($F_{2, 39} = 0.8631$, $p = 0.4298$) (Figure 3). However, there was a significant main effect of Group ($F_{2, 39} = 86.451$, $p < 0.0001$). Post-hoc Tukey's tests for the Group effect indicate that cue reactivity was lower in the saline group than cocaine 1d abstinence ($t_{39} = 7.978$, $p < 0.001$) or cocaine 21d abstinence groups ($t_{39} = 12.763$, $p < 0.001$) Additionally, cocaine 1d abstinence showed lower active lever presses than cocaine 21d abstinence ($t_{39} = 4.096$, $p = 0.0006$), consistent with prior research on the incubation of cocaine craving over prolonged abstinence.



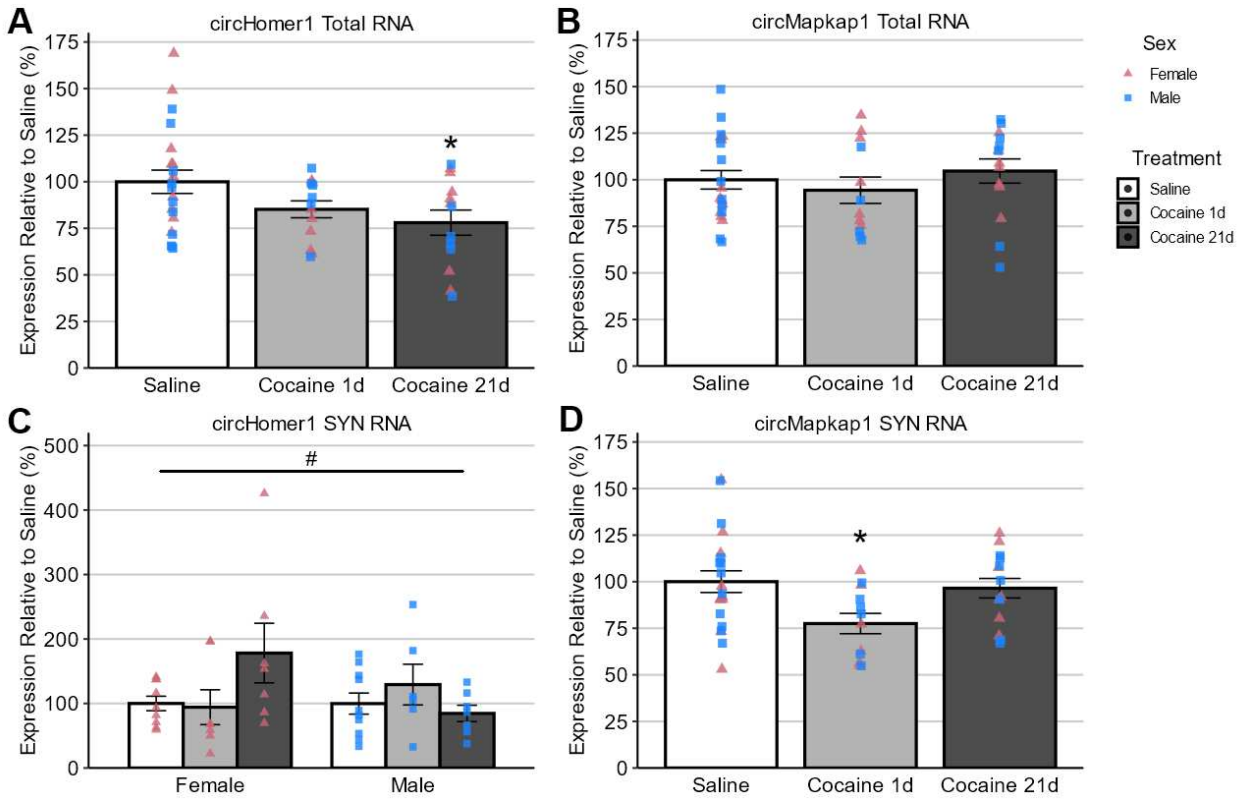

***Chapter 4, Figure 3.*** Active lever presses (±SEM) during the post-abstinence cue reactivity test in a second set of male and female rats. Rats underwent cocaine SA or were saline-yoked, and then underwent either 1 or 21d abstinence (N = 47 after removing 3 outliers). Saline animals were combined into one control group. * indicates significant difference from all other groups in post-hoc Tukey' test after a significant main effect of Treatment. Significance level is $p < 0.05$.

**Select candidate circRNAs were validated with RT-qPCR.** RT-qPCR was used to validate expression of the candidate circRNAs in an expanded set of male and female rats that underwent cocaine SA or were saline-yoked, with a cue-reactivity test after 1 or 21d of abstinence. To account for cohort effects, expression values were divided by the average for saline animals in the same cohort as described in the Methods.

We first analyzed circRNA expression in total RNA fractions, as this was most comparable to the RNA-seq analysis, which also used bulk RNA. For *circHomer1*, 2-way ANOVAs revealed a significant main effect of Group ($F_{2, 41} = 3.685$, $p = 0.0338$), though no effect of Sex ($F_{1, 41} = 1.084$, $p = 0.3043$) or Group x Sex interaction ($F_{2, 41} = 1.236$, $p = 0.3012$) after removing 1 outlier (Figure 4A). Post-hoc Tukey's tests indicated that circHomer1 expression was higher in cocaine 21d abstinence compared to saline ($t_{41} = 2.619$, $p = 0.0323$). However, there was no difference between saline and cocaine 1d ($t_{41} = 1.705$, $p = 0.2154$) or cocaine 1d compared to cocaine 21d ($t_{41} = 0.821$, $p = 0.6922$). This is consistent with the RNA-seq data, which also suggested *circHomer1* expression is lower in animals with higher cocaine-seeking behavior. For *circMapkap1*, 2-way ANOVAs suggested that expression was consistent across groups and sexes, as there was no effect of Group ($F_{2, 41} = 0.616$, $p = 0.545$), Sex ($F_{1, 41} = 0.078$, $p = 0.782$), or Group x Sex interaction ($F_{2, 41} = 1.043$, $p = 0.361$) after removing 1 outlier (Figure 4B).

Because many circRNAs are enriched at synapses, we also analyzed expression in SYN fractions. For *circHomer1*, 2-way ANOVAs indicated no significant main effect of Group ($F_{2, 40} = 0.827$, $p = 0.4447$) or Sex ($F_{1, 40} = 0.850$, $p = 0.3620$), although there was a strong trend for a Group x Sex interaction ($F_{2, 40} = 3.218$, $p = 0.0506$) after removing 2 outliers (Figure 4C). This appears to be because females showed high expression of SYN

*circHomer1* in cocaine 21d abstinence, while expression in males is more consistent across treatment groups. For SYN *circMapkap1*, the 2-way ANOVA indicated a significant main effect of Group ($F_{2, 41} = 3.623$, $p = 0.0356$) with no Sex ($F_{1, 41} = 0.0227$, $p = 0.8809$) effect or Group x Sex interaction ($F_{2, 41} = 0.105$, $p = 0.902$) after removing 1 outlier (Figure 4B). In the RNA-seq data from bulk RNA, we found that *circMapkap1* expression was lower in animals with 21d abstinence than 1d abstinence. By contrast, post-hoc Tukey's test showed higher *circMapkap1* expression in SYN for saline compared to cocaine 1d abstinence ($t_{41} = 2.614$, $p = 0.0327$). However, there was no difference between saline and cocaine 21d abstinence ($t_{41} = 0.421$, $p = 0.9070$) or between cocaine 1d and 21d ($t_{41} = 2.035$, $p = 0.1165$).




***Chapter 4, Figure 4.*** RT-qPCR for candidate circRNAs in the NAc shell. Expression (±SEM) was assessed in total RNA samples for (A) *circHomer1* and (B) *circMapkap1* and in SYN fractions for (C) *circHomer1* and (D) *circMapkap1* in male (blue squares) and female (pink triangles) rats. Expression values were quantified relative to the housekeeping gene GAPDH using the comparative $2^{-\Delta CT}$ method (Livak & Schmittgen, 2001). Expression values were normalized to the average saline expression in each cohort and sex (*circHomer1* SYN) or each cohort overall (*circHomer1* Total RNA, *circMapkap1* Total and SYN) and are depicted here as a percentage relative to saline. * indicates significant difference from saline with post-hoc Tukey's tests ($p < 0.05$). # indicates trending Group x Sex interaction ($p = 0.051$).

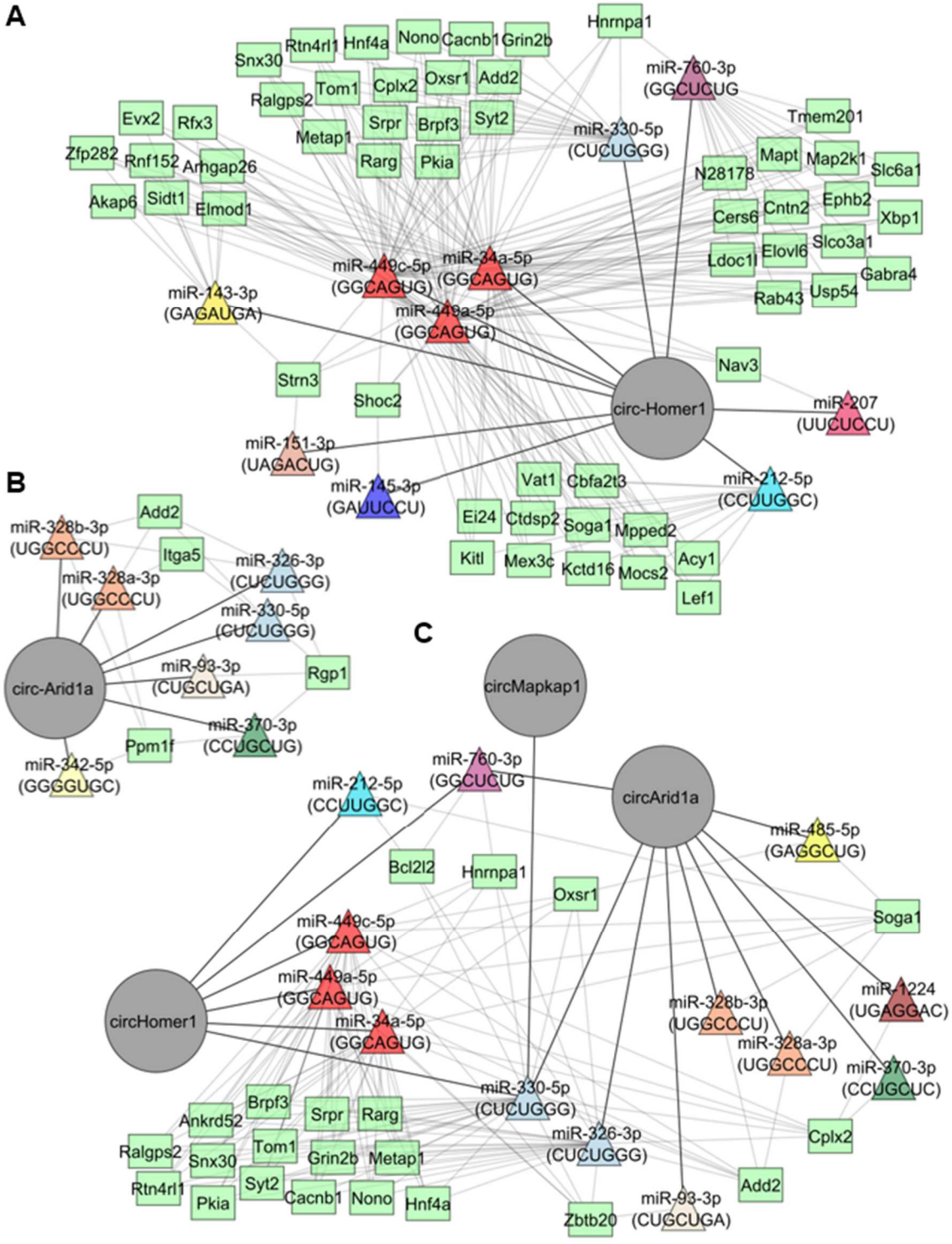
**circRNA-miRNA interactions and ceRNA networks.** ceRNA networks were visualized based on predicted circRNA-miRNA and miRNA-mRNA binding along with hypothetical circRNA regulation of mRNA targets through hypergeometric tests (Figure 5). This analysis was performed for the 3 highly-conserved circRNAs (*circArid1a*, *circHomer1*, and *circMapkap1*). Of these circRNAs, *circArid1a*, *circHomer1*, and *circMapkap1* are predicted to bind to 26, 14, and 8 miRNAs with a match score ≥150 and free energy ≤ -20 across all predicted binding sites, respectively. After filtering for only those miRNAs that were expressed in cerebrum or NAc shell, this was reduced to 16, 10, and 4 miRNAs each. Though most of the circRNA-miRNA interactions are predicted to occur at only one binding site, some circRNAs are likely to bind to the same miRNA at many sites. Of note, *circHomer1* has 2 predicted binding sites each for miR-143-3p, miR-151-3p, and miR-873-3p, and *circMapkap1* has 2 predicted binding sites each for miR-330-5p, miR-345-3p, and miR-6314. Nearly all of the predicted miRNA interactions of *circArid1a* occur at 2 or more binding sites, including one miRNA with 7 sites (miR-3575) and another with 10 sites (miR-370-3p). Both *circArid1a* and *circHomer1* are predicted to bind rno-miR-760-3p, each 1 site, and *circArid1a*, *circHomer1*, and *circMapkap1* are all predicted to bind rno-miR-330-5p, with 3, 1, and 2 site(s) each. Interestingly, *circArid1a* is also predicted to bind rno-miR-326-3p, which has the same seed sequence as rno-miR-330-5p. Importantly, most miRNAs predicted to bind to *circArid1a*, *circHomer1*, and/or *circMapkap1* are conserved in mouse (35/45, 78%) and human (35/45, 78%), including miR-760-3p and miR-330-5p. Predicted miRNA interactions for the 3 validated circRNAs are listed in Supplementary Table 2.

Hypergeometric tests suggested that *circArid1a* and *circHomer1* are predicted to significantly regulate 4 and 56 mRNAs, respectively (FDR < 0.05, common miRNAs predicted to interact with both the circRNA and target ≥ 4; Supplementary Tables 3 & 4). While there were significant predicted interactions for *circMapkap1* (FDR < 0.05), the highest number of common miRNAs for any interaction was just 2 (*Tomm34*; miR-330-5p and miR-421-3p). Predicted ceRNA networks are depicted for *circHomer1* (Figure 5A) and *circArid1a* (Figure 5B). One limitation of hypergeometric tests is the simplified assumption that each circRNA is independent. However, our RT-qPCR data for *circHomer1* and *circMapkap1* suggest that multiple circRNAs are co-regulated in cocaine craving. Thus, we also noted a predicted ceRNA network for *circArid1a*, *circHomer1*, and *circMapkap1* together, including mRNAs that are targeted by ≥ 7 unique circRNA-miRNA interactions, regardless of the significance of any individual interaction (Figure 5C).

**Discussion**

In this study, we identified several circRNAs of potential interest in cocaine-seeking behavior utilizing RNA-seq data from a prior experiment (Powell, Vannan, et al., 2020), and then further probed these candidate circRNAs in a set of both male and female rats. Two highly-conserved circRNAs (*circHomer1* and *circMapkap1*) were also assessed using RT-qPCR for their role in incubation of cocaine craving. Using the miRanda and TargetScan algorithms to predict circRNA-miRNA and miRNA-mRNA interactions, respectively, we then constructed predicted ceRNA networks for *circHomer1*, *circMapkap1*, and another conserved circRNA, *circArid1a*. We found

miRNAs predicted to bind to all three (miR-330-5p) or two of the three circRNAs (miR-760-3p), suggesting that *circArid1a*, *circHomer1*, and *circMapkap1* may co-regulate gene expression in a model of cocaine-seeking behavior.


A
Hnrnpa1
Rtn4rl1
Hnf4a
Nono
Cacnb1
Grin2b
Snx30
Tom1
Cplx2
Oxsr1
Add2
Ralgps2
Srpr
Brpf3
Syt2
Metap1
Rarg
Pkia
Evx2
Rfx3
Zfp282
Rnf152
Arhgap26
Akap6
Sidt1
Elmod1
miR-760-3p
(GGCUCUG
miR-330-5p
(CUCUGGG)
Tmem201
Mapt
Map2k1
Slc6a1
N28178
Cers6
Cntn2
Ephb2
Xbp1
Ldoc1l
Elovl6
Slco3a1
Gabra4
Rab43
Usp54
miR-449c-5p
(GGCAGUG)
miR-34a-5p
(GGCAGUG)
miR-143-3p
(GAGAUGA)
miR-449a-5p
(GGCAGUG)
Strn3
Shoc2
Nav3
circ-Homer1
miR-207
(UUCUCCU)
miR-151-3p
(UAGACUG)
miR-145-3p
(GAUUCCU)
Vat1
Cbfa2t3
miR-212-5p
(CCUUGGC)
Ei24
Ctdsp2
Soga1
Mpped2
Kitl
Mex3c
Kctd16
Mocs2
Acy1
Lef1
B
Add2
miR-328b-3p
(UGGCCCU)
Itga5
miR-326-3p
(CUCUGGG)
miR-328a-3p
(UGGCCCU)
miR-330-5p
(CUCUGGG)
circ-Arid1a
miR-93-3p
(CUGCUGA)
Rgp1
miR-370-3p
(CCUGCUG)
Ppm1f
miR-342-5p
(GGGGUGC)
C
circMapkap1
miR-760-3p
(GGCUCUG
circArid1a
miR-212-5p
(CCUUGGC)
miR-485-5p
(GAGGCUG)
Bcl2l2
Hnrnpa1
Oxsr1
Soga1
miR-449c-5p
(GGCAGUG)
miR-449a-5p
(GGCAGUG)
circHomer1
miR-34a-5p
(GGCAGUG)
miR-1224
(UGAGGAC)
miR-328b-3p
(UGGCCCU)
miR-328a-3p
(UGGCCCU)
miR-370-3p
(CCUGCUC)
miR-330-5p
(CUCUGGG)
Brpf3
Ankrd52
Srpr
Rarg
miR-326-3p
(CUCUGGG)
Cplx2
Ralgps2
Snx30
Tom1
Grin2b
Metap1
Rtn4rl1
Syt2
Pkia
Cacnb1
Nono
Hnf4a
Add2
miR-93-3p
(CUGCUGA)
Zbtb20

**Chapter 4, Figure 5.** ceRNA networks depicting predicted regulation of target genes by circRNAs through miRNA interactions. Networks for (A) *circHomer1* and (B) *circArid1a* include only significant interactions (FDR < 0.05, hypergeometric test) with at least 4 common miRNAs between the circRNA and targets. (C) ceRNA network for all three circRNAs that were validated with RT-qPCR, showing all targets of at least 7 unique circRNA-miRNA interactions, regardless of statistical significance in the hypergeometric tests. Note that *circArid1a*, *circHomer1*, and *circMapkap1* are all predicted to bind to rno-miR-330-5p, and both *circArid1a* and *circHomer1* are predicted to bind to rno-miR-760-3p.

For the initial RNA-seq analysis, we found that *circArid1a*, *circHomer1*, and *circMapkap1* showed lower expression in groups that demonstrated higher cocaine-seeking behavior post-abstinence (Figure 2). In subsequent RT-qPCR analyses, we observed that *circHomer1* is specifically downregulated after cocaine 21d abstinence relative to saline controls (Figure 4). However, this effect is specific to total RNA samples. In SYN fractions, *circHomer1* shows a strong trend for a Group x Sex interaction in which *circHomer1* is upregulated in cocaine 21d compared to saline controls, specifically in female rats. For *circMapkap1*, the RT-qPCR data suggest that *circMapkap1* is specifically dysregulated at synapses, in contrast with the original RNA-seq data which suggested differential expression in total RNA samples. For SYN fractions, we found that *circMapkap1* had decreased expression in cocaine 1d animals compared to saline controls. While this does not directly replicate the RNA-seq findings, this does suggest *circMapkap1* may play a role in cocaine-motivated behavior after short term abstinence.

Of the three candidate circRNAs identified in this study, only *circHomer1* has been the subject of prior drug addiction research. *circHomer1* was first described in detail by (You et al., 2015), who demonstrated that it localizes to dendrites during periods of neuronal activity. Based on this finding, it is believed that *circHomer1* plays a role in synaptic plasticity akin to its host gene, *Homer scaffold protein 1 (Homer1)* (Luo et al., 2012; Szumlinski et al., 2006), which has been implicated in several models of cocaine addiction (Ary & Szumlinski, 2007; Ghasemzadeh et al., 2003; Ghasemzadeh, Vasudevan, et al., 2009; Ghasemzadeh, Windham, et al., 2009; Swanson et al., 2001; Szumlinski et al., 2004). Similarly, recent research suggests a role for *circHomer1* in

learning, memory, and cognition (Dube et al., 2018, 2020; Hafez et al., 2022; Zimmerman et al., 2020), and in cocaine and methamphetamine reinforcement (Dell'Orco et al., 2020; J. Li et al., 2020). Additionally, we find that *circHomer1* is predicted to bind to several miRNAs in the same family (miR-34a-5p, miR-449a-5p, and miR-449c-5p) that are broadly involved in brain development (Fededa et al., 2016; Nowakowski et al., 2018; J. Wu et al., 2014). This miRNA cluster targets many synaptic plasticity genes, including *Adducin 2 (Add2), Rho GTPase activating protein 26 (Arhgap26)*, *Complexin 2 (Cplx2)*, *Contactin 2 (Cntn2)*, *Gamma-aminobutyric acid type A receptor subunit alpha 4 (Gabra4)*, *Glutamate ionotropic receptor NMDA type subunit 2B (Grin2b)*, *Potassium channel tetramerization domain containing 16 (Kctd16)*, *Mitogen activated protein kinase kinase 1 (Map2k1)*, *Microtubule-associated protein tau (Mapt*), *Neuron navigator 3 (Nav3)*, *Oxidative stress response kinase 1 (Oxsr1)*, *RAB43, member of RAS oncogene family (Rab43)*, *Reticulon 4 receptor-like 1 (Rtn4rl1)*, *Solute carrier family 6 member 1 (Slc6a1*; encodes for the GABA transporter), *Striatin 3 (Strn3)*, *Synaptotagmin 2 (Syt2)*, and *Vesicle amine transport 1 (Vat1)* (Figure 5, Supplementary Tables 2-4). In addition, *circHomer1* may regulate neuroinflammatory pathways. In their recent methamphetamine study, (J. Li et al., 2020) found that *circHomer1* binds to multiple miRNAs that target *Bcl2-binding component 3 (Bb3)*, which is also downregulated after *circHomer1* knockdown in HT-22 cells. Interestingly, *Bb3* is a modulator of p53-induced apoptosis (Nakano & Vousden, 2001) and is also a predicted target of miR-34a-5p, miR-449a-5p, miR-449c-5p, which themselves are implicated in this process (Chang et al., 2007; Lizé et al., 2009; Raver-Shapira et al., 2007; Rokavec et al., 2014; Yamakuchi et al., 2008) (Supplementary

Tables 2-4). Thus, *circHomer1* may impact motivation for cocaine through regulation of neuroplasticity and neuroinflammatory mechanisms.

The other two candidate circRNAs, *circArid1a* and *circMapkap1*, may similarly regulate synaptic plasticity and neuroinflammatory networks. *circArid1a* is encoded by the gene *AT-rich interaction domain 1A (Arid1a)*, which is involved in neurogenesis (J. Liu et al., 2020; X. Liu et al., 2021). In addition, a study using postmortem tissue from subjects with autism spectrum disorder (ASD) suggests that human *circARID1A* regulates the synaptic remodeling gene Neuroligin-1 (*NGLN1*) by binding and sequestering miR-204-3p. Though we also found that rat *circArid1a* is predicted to bind to miR-204-3p (Supplementary Table 2), this miRNA was excluded from the ceRNA network analysis because it did not pass our expression filter. However, our analyses indicate *circArid1a* targets other miRNAs that are involved in brain function. For example, *circArid1a* has 2 predicted binding sites each for miR-204-3p and miR-377-5p, which have been implicated in neurodegenerative and/or psychiatric disorders (Cammaerts et al., 2015; Garbett et al., 2015; Tao et al., 2021), as well as rodent models of cocaine craving and addiction vulnerability (Dell'Orco et al., 2021; Ni et al., 2019; Vannan et al., 2021). Our data also suggest that *circArid1a* regulates immune response genes, including *Integrin subunit alpha 5 (Itga5)*, *Protein phosphatase, Mg2+/Mn2+ dependent, 1F (Ppm1f),* and *RGP1 homolog, RAB6A GEF complex partner 1* (*Rgp1*) (Figure 5B, Supplementary Tables 3 & 4). Similarly, *circMapkap1* is encoded by the gene *MAPK associated protein 1*, which is involved in immune response through stress-activated MAPK signaling (Wilkinson et al., 1999). Though *circMapkap1* had few significant targets (FDR < 0.05), and none with 4 common miRNAs, this may be due to

the limitations of hypergeometric tests, which only account for the impact of any single circRNA. However, *circMapkap1* may regulate ceRNA networks and neuroinflammatory mechanisms in conjunction with *circArid1a* and *circHomer1*.

By considering the candidate circRNAs together, we find evidence for co-regulation of ceRNA networks. For example, all three circRNAs are predicted to bind to one common miRNA, miR-330-5p (Figure 5, Supplementary Tables 2-4). miR-330-5p is a brain-enriched miRNA that has also been linked to Parkinson's disease, Alzheimer's disease, and bipolar disorder (Cai et al., 2017; Maffioletti et al., 2016; Ravanidis, Bougea, Papagiannakis, Koros, et al., 2020; Ravanidis, Bougea, Papagiannakis, Maniati, et al., 2020) and is predicted to suppress expression of several neural plasticity, synaptic vesicle, and actin cytoskeleton genes, including many that are also targeted by the miR-34 cluster: *Add2*, *Cplx2*, *Grin2b*, *Oxsr1*, *Rtn4rl1*, and *Syt2*. Notably, miR-330-5p has also been shown to modulate microglial response to traumatic brain injury (Y. Li et al., 2020) and impact microglial differentiation (Feng et al., 2021), and is in the same family as miR-326-3p, which is also predicted to bind *circArid1a*. Though we did not validate the candidate cocaine-craving circRNA *circAnkrd12* due to lower conservation with human compared to the other circRNAs, this circRNA is also predicted to bind to miR-330-5p using the miRanda algorithm (match score = 284, free energy = -43.47, 2 binding sites). We also find that both *circArid1a* and *circHomer1* are predicted to bind to miR-760-3p. This miRNA targets several plasticity and neuronal development genes, including some regulated by the miR-34 cluster: *Cntn2*, *Gabra4*, *Map2k1*, *Mapt*, *Rab43*, and *Slc6a1*. Thus, our data suggest that *circArid1a*, *circHomer1*, and *circMapkap1* may interact with

miR-326-3p/miR-330-5p and miR-760-3p to modulate synaptic plasticity and neuroinflammatory genes.

Few studies have focused circRNA-miRNA-mRNA interactions in any model of SUDs (Dell'Orco et al., 2021, 2020; Floris et al., 2022; J. Li et al., 2020), and this is the first investigation of ceRNA networks in the NAc shell in a behavioral model of cocaine craving. Here, we find that three circRNAs, *circArid1a*, *circHomer1*, and *circMapkap1*, may underlie motivation for cocaine in the NAc shell of rats. While *circHomer1* has been widely studied in neuroplasticity, cognition and psychiatric disease (Dube et al., 2018, 2020; Hafez et al., 2022; You et al., 2015; Zimmerman et al., 2020), including addiction (Dell'Orco et al., 2021, 2020; J. Li et al., 2020), *circArid1a* and *circMapkap1* are potentially novel circRNAs in drug abuse research. Together, these circRNAs may regulate synaptic plasticity and neuroinflammatory pathways that are often dysregulated in models of SUDs (Bachtell et al., 2017; Namba et al., 2021; Russo et al., 2010). However, more work needs to be done to investigate these and other circRNAs in addiction, with particular attention to how circRNAs vary across behavioral models and drugs of abuse. Further insight into these complex RNA interactions will help build a strong foundation for basic synaptic plasticity and translational SUDs research.

**Supplementary information**

Supplemental Tables 1-4 are provided on GitHub (https://github.com/SexChrLab/Vannan-Dissertation-Supplements).

CHAPTER 5

# CONCLUDING REMARKS

## Major Contributions of Dissertation

**Chapter 2. microRNA regulation related to the protective effects of environmental enrichment against cocaine-seeking behavior.** Here, we collected samples from the NAc shell of 12 male rats that self-administered cocaine and underwent 21d abstinence in either EE or standard, isolated housing. Rats were split into high and low cocaine-seeking groups based on active lever presses in a post-abstinence cue reactivity test. Using a Nanostring array, we found 33 miRNAs that were differentially-expressed between these groups ($p < 0.05$), including miR-212 and miR-495, which have previously been implicated in addiction, and novel miRNAs such as miR-377. Of the 33 miRNAs, 8 miRNAs are of particular interest because their expression correlates directly with cocaine-seeking behavior: miR-3557, miR-142-5p, miR-3573-5p, miR-346, miR-193a-3p, miR-107, let-7a, and miR-376c. Further, we find that some miRNAs (e.g. miR-107, miR-3557) are predicted to target a large number of addiction-related genes, on par with or higher than miRNAs previously implicated in SUDs models. As a whole, the predicted mRNA targets of the 33 differentially-expressed miRNAs are involved in synaptic plasticity pathways, including synaptogenesis and regulation of the actin cytoskeleton, that are dysregulated during cocaine abstinence. We also highlight candidate miRNAs, including several previously studied (e.g. miR-212, miR-495) and novel miRNAs (e.g. miR-107, miR-376c, miR-3557, miR-377), which may be exciting targets for future SUDs research. Raw data files are deposited in the Gene

Expression Omnibus (GEO), accession GSE153524. Code is available at: https://gitlab.com/neisewander_asu/vannan-powell-2020.

**Chapter 3. An approach for prioritizing candidate genes from RNA-seq using preclinical cocaine craving datasets as a test case.** Using a standardized workflow, we analyzed three publicly available RNA-seq datasets from rodent cocaine self-administration and abstinence studies to identify differentially-expressed genes (DEGs) with multiple lines of evidence. Overall, we found that there were almost no DEGs in any dataset after multiple testing correction (FDR < 0.5 or FDR < 0.1), though there were many more DEGs when using a less stringent, nominal p-value threshold ($p < 0.05$). However, with the latter approach we saw very low overlap in these DEGs ($p < 0.05$) between datasets in contrast to our original assumptions. Both the lack of significant genes after FDR correction and low number of shared DEGs between datasets may be related to limited sample size and thus reduced ability to distinguish true positive DEGs. Still, we moved forward using the shared DEGs as a test case to apply a novel pipeline for narrowing candidate genes from RNA-seq. Using conservation data (dN/dS, sequence similarity, and developmental expression for rat, mouse, and human) along with expression data from the human GTEx database, we found that several of the shared DEGs are highly conserved and/or expressed preferentially in the brain. Thus, we promote our pipeline as a useful tool for narrowing down candidate genes and translational targets from RNA-seq analysis. We conclude by advocating for improved transparency and experimental design in RNA-seq studies to promote more replicable, and thus more translational, research. Raw data files are deposited in GEO with the following accession numbers: GSE141520 (Carpenter et al., 2020), GSE110344

(Walker et al., 2018), GSE144606 (Powell, Vannan, et al., 2020). Code is available at: https://github.com/SexChrLab/RodentAddiction

**Chapter 4. Alterations in circular RNA expression in the rat nucleus accumbens shell may regulate circRNA-miRNA-mRNA regulatory networks in a model of cocaine craving.** In this project, circRNAs were initially quantified from a prior RNA-seq study of rats that underwent cocaine self-administration and abstinence. Several circRNAs, including *circArid1a*, *circHomer1*, *circMapkap1*, had expression patterns suggesting a role in cocaine craving. A separate set of male and female rats was then used to validate two of these highly conserved candidate circRNAs with RT-qPCR. These analyses revealed potential dysregulation of circHomer1 expression after long term (21d) abstinence and of *circMapkap1* after short-term (1d) abstinence. Theoretical competing endogenous (ceRNA) networks of circRNA-miRNA-mRNA interactions were then constructed to understand how these circRNAs may regulate gene expression. Generally, the 3 candidate circRNAs appear to regulate genes involved in synaptic plasticity and neuroinflammation through interactions with miRNAs such as miR-760-5p and miR-760-3p. These results support a role for a previously-studied circRNA (*circHomer1*) and a novel circRNA (*circMapkap1*) in animal models of SUDs. Raw RNA-seq files were deposited in GEO previously (GSE144606). circRNA identification files will be deposited in GEO upon submission of the manuscript. Code is available at: https://github.com/SexChrLab/Craving-circRNAs

**Limitations of Bioinformatics Tools**

**Comparison of RNA expression analysis techniques.** Throughout this dissertation, expression of coding and non-coding RNAs was quantified using 3 different methods: RT-qPCR (Chapters 2, 3, and 4), Nanostring array (Chapter 2), and RNA-seq (Chapters 3 and 4). Each technique has different strengths and pitfalls. While RT-qPCR can only be used to measure expression of one RNA at a time, Nanostring arrays and RNA-seq quantify hundreds or thousands of RNAs in just one experiment. Of these techniques, Nanostring arrays have superior accuracy, as they do not require cDNA synthesis, and therefore avoid bias introduced by amplification. While analysis of Nanostring arrays requires more background knowledge than RT-qPCR, it is not as intensive as RNA-seq. RNA-seq has the most potential of any other RNA analysis technique, as it can be used to quantify all known RNAs for a species in addition to novel or alternative splicing events (e.g. circRNAs). However, RNA-seq has the most difficult learning curve, with many steps to learn and a variety of analysis tools one can use, and it is by far the most costly. Often, RNA-seq or Nanostring arrays are used to assess a large number of RNAs, and then RT-qPCR is used to verify the results. In Chapters 2 and 4, we used RT-qPCR to specifically validate expression of a small subset of RNAs, either in the same samples or in samples from a new cohort. However, there is some disagreement among researchers about how to validate results with RT-qPCR and in what cases this is necessary (Dallas et al., 2005; Fang & Cui, 2011). Overall, choosing which RNA analysis technique to use is an imperfect process, and requires a balance of financial costs, study goals, and researcher knowledge.

**Tools for predicted RNA binding.** We used the TargetScan (Chapters 2 and 4) and miRanda (Chapter 4) algorithms to predict miRNA interactions with mRNAs and

circRNAs. These tools have become more accurate as current understanding of miRNA binding has improved, though they still have limitations. First, any algorithm will have false positives or negatives that cannot be accounted for without more experimental data. For this reason, prediction tools generally include binding scores or other estimations of accuracy for each predicted interaction. For example, TargetScan provides a cumulative weighted context++ score (CWCS) (Agarwal et al., 2015), which incorporates all possible binding interactions between a miRNA-mRNA pair to estimate how much a miRNA will suppress its target. Likewise, miRanda gives a match score, which describes how complementary two RNA sequences are, and the free energy, which estimates the thermodynamic probability of binding (Enright et al., 2003; John et al., 2004). In Chapter 4, we increased confidence in our results by using TargetScan's CWCS and miRanda's match score and free energy to filter predicted miRNA-mRNA and circRNA-miRNA interactions.

We also used the conservation data incorporated in TargetScan to improve accuracy. Prior work from the TargetScan researchers suggests most miRNA-mRNA interactions are likely to be conserved across mammals (Friedman et al., 2009). While TargetScan has no official rat database, it has versions for human and mouse which can also provide target predictions for miRNAs in other species based on homologous 3'UTR sequences of target mRNAs. In both Chapters 2 and 4, we filtered our results to miRNA target predictions with at least 1 conserved binding site. While using TargetScan for rat miRNAs has some limitations, target predictions are likely to be conserved across multiple species, which is an important consideration in translational research.

**Identifying circRNAs from RNA-seq data.** Various tools have been developed to identify circRNAs, including CIRI2 and CIRCExplorer2. Though there are some differences between tools, they broadly follow the same principles. As part of standard RNA-seq analysis, reads are aligned to a reference genome or transcriptome. However, not all reads will map to known sequences. For example, *circHomer1* is formed by backsplicing of *Homer1b* linear mRNA at exons 2 and 5 (Zimmerman et al., 2020). If exon 2 had the sequence GGAA and exon 5 had the sequence ATAG, then the backsplice junction for *circHomer1* would have the sequence GGAAGATA or ATAGAAGG, i.e. one exon would be reversed relative to the linear sequence. RNA-seq reads that contain this backsplice junction would not map to known features of the genome during alignment. circRNA identification tools take advantage of this to identify possible circRNAs from unmapped reads.

circRNA identification tools can vary in accuracy based on factors such as the original RNA preparation. In Chapter 4, we utilized prior RNA-seq data to identify circRNAs in a model of cocaine craving. The previous experiment was originally designed to compare mRNA expression, and therefore the RNA samples were not treated with RNAse R to degrade linear RNAs prior to cDNA synthesis (Suzuki et al., 2006). Because linear mRNAs were not depleted from our dataset, there is some level of noise in our data as circRNAs can contain sequences that exactly match their linear counterparts, and thus the small cDNA fragments (~50-150 bp) that are sequenced may generate reads that appear to map to linear mRNAs during alignment. Additionally, circRNA detection tools vary in their precision and specificity (Zeng et al., 2017). Therefore, regardless of the RNA preparation, multiple tools should be utilized to improve accuracy of circRNA

detection. Accordingly, in Chapter 4 we utilized two tools and only analyzed circRNAs that were identified with both, as recommended by other researchers (Hansen et al., 2016). We also followed up with RT-qPCR validation of candidate circRNAs in a separate set of animals. This approach allowed us to bolster confidence in our results in light of these known concerns with circRNA detection.

**Constructing ceRNA networks.** In Chapter 4, multiple tools were used to construct ceRNA networks of interacting circRNAs, miRNAs, and mRNAs. We used miRanda to determine possible circRNA-miRNA interactions and TargetScan to understand predicted targets of miRNAs. As described above, we narrowed these interactions to those with high confidence based on miRanda and TargetScan scores. To understand how circRNAs may indirectly regulate mRNAs through miRNAs, we then performed hypergeometric tests on all theoretical circRNA-mRNA pairs. This led to the construction of potential circRNA-miRNA-mRNA interaction networks. This approach is similar to a proprietary workflow developed by Arraystar, Inc. (Rockville, MD, USA; https://www.arraystar.com) based on methodology from the creators of the Starbase ceRNA database (J.-H. Li et al., 2013); http://starbase.sysu.edu.cn/).

There are several limitations to this ceRNA network analysis. As briefly described in Chapter 4, performing hypergeometric tests for theoretical circRNA-mRNA pairs ignores the potential for multiple circRNAs to influence the same mRNA. For example, miRanda predictions suggest that *circArid1a*, *circHomer1*, and *circMapkap1* are all likely to interact with miR-330-5p. Therefore, all three of these circRNAs should influence expression of miR-330-5p's target genes, such as *Grin2b*. However, because hypergeometric tests consider each circRNA-mRNA pair individually, only *circHomer1*,

which is predicted to interact with 4 total miRNAs that target *Grin2b*, showed a significant (FDR < 0.05) predicted relationship with *Grin2b*. By contrast, the other circRNAs interact with fewer miRNAs that are predicted to target *Grin2b* (*circArid1a*: 2 miRNAs; *circMapkap1*: 1 miRNA), and thus the hypergeometric tests were not significant for the *circArid1a-Grin2b* or *circMapkap1-Grin2b* pairs. Still, this does not preclude the possibility that all three circRNAs influence *Grin2b* expression cooperatively.

Creation of accurate ceRNA networks is also limited by the accuracy of the underlying tools, i.e. miRanda and TargetScan. As described above, we increased confidence in our results by applying thresholds to the miRanda and TargetScan predictions. However, using different parameters or tools can lead to different results. For example, (J. Li et al., 2020) developed a ceRNA network for *circHomer1* using Starbase and TargetScan to inform their mouse model of METH reinforcement. This network included 6 miRNAs (miR-101b-3p, miR-138-5p, miR-143-3p, miR-217-5p, miR-455-3p, miR-802-5p, and miR-883-3p) and 68 mRNAs. In our study, we found that rat *circHomer1* is predicted to bind to most of these miRNAs (5/6; excluding miR-101b-3p), though all were filtered out of our analysis due to low match scores (< 150) and/or high free energy scores (> 20) except miR-143-3p. Unfortunately, (J. Li et al., 2020) do not specify the criteria they used to exclude or include miRNAs and mRNAs in their network, so it is not possible to directly compare our results. In any case, construction of theoretical ceRNA networks requires both specificity and precision. In other words, the RNA binding predictions that influence network construction must be narrowed to reduce false positives while retaining true positives. However, there is no scientific

consensus on what parameters and thresholds should be used to balance these two competing needs. We believe the thresholds we selected for miRanda, TargetScan, and the hypergeometric test significance are appropriately stringent, though further testing of our networks is necessary to determine this with certainty.

Another important consideration in creating realistic networks is whether the RNAs that are predicted to bind to one another are located in the same tissues and parts of the cell. For this reason, we narrowed our testing to miRNAs that are known to be expressed in the cerebrum (Bushel et al., 2018, 2020), which contains the NAc, or are specifically expressed in the NAc shell according to our data from the Nanostring array in Chapter 2. Thus, our approach improves accuracy by considering actual potential of RNAs to interact based on tissue localization. As our understanding of circRNAs and miRNAs improves, we may be able to further hone ceRNA network analysis by only testing for RNAs that are present at the same cellular location, e.g. expressed together at the postsynaptic density.

**Considerations for Translational Research**

**Sample size, power, and other statistical concerns.** Low sample size limited the interpretations of the projects outlined in this dissertation. For example, in Chapter 2, a Nanostring array was used to quantify the expression of 423 of the 496 rat miRNAs known at the time. Just as with RNA-seq, robust analysis of array data requires correcting for multiple testing. However, with only 12 samples total (6/group), our study was not powered to detect differences after FDR correction. Instead, we utilized an uncorrected p-value ($p < 0.05$) for differential expression and the subsequent correlation

and predicted pathway analyses. We also looked at expression of three miRNAs (miR-107, miR-212, and miR-376c) using RT-qPCR from the same samples. With this data, animals were split by quartile rather than by median, such that only the ¼ highest and lowest cocaine-seeking rats were included (i.e., 3 high and 3 low seekers). Using this approach, significant differences ($p < 0.05$) were seen for miR-376c and miR-107, with a trend for miR-212 ($p = 0.0583$). While we did not examine all 33 differentially-expressed miRNAs with RT-qPCR, this may not be necessary since Nanostring arrays measure expression more accurately. In either case, this project functions as an important exploratory examination of miRNAs in cocaine craving, though further experiments are needed to verify the results.

Similarly, low sample size limited the detection of candidate circRNAs in Chapter 4. For the initial circRNA identification, we had only 11 samples (2-3/group) after removing an outlier, and we had not treated the RNA with RNAse R. Because of this, we took a conservative approach for selecting candidate circRNAs to increase accuracy. Only those circRNAs that were identified by 2 tools and were additionally expressed in all samples or exclusively in a specific treatment group were considered potential candidates. Still, our subsequent ANOVAs and post-hoc Tukey's tests may not have been fully powered to detect true differences in expression. Therefore, our additional experiment analyzing expression in a well-powered set of both male and female rats was important for increasing our confidence in the circRNA candidates.

Chapter 3 explores in detail how an insufficient sample size can impact RNA-seq analysis and prevent researchers from making the most of this type of data. For example, researchers may erroneously use RNA-seq data that is underpowered to detect

differences using robust statistical methods, i.e. multiple testing corrections such as FDR (Benjamini & Hochberg, 1995). However, we did not delve further into some of the other statistical issues affecting RNA-seq analysis. RNA-seq counts follow a non-normal distribution that some have compared to a Poisson distribution, but there are disagreements on how to best model this type of data. Accordingly, differential expression analysis tools vary in their approach to analyzing RNA-seq data. For example, three of the most popular differential expression analysis software are DESeq2 (Love et al., 2014), edgeR (M. D. Robinson et al., 2010), and limma voom (Law et al., 2014; Ritchie et al., 2015) which are based on a negative binomial distribution, Poisson distribution, and linear model, respectively (see (Costa-Silva et al., 2017) for discussion of RNA-seq statistics and comparison of analysis tools). Of these, limma voom (utilized in Chapters 2 and 4) and DESeq2 are generally considered among the most accurate (Costa-Silva et al., 2017; Schurch et al., 2016). However, no one tool consistently outperforms all others, and it is likely that all tools perform poorly in some scenarios. For this reason, new differential expression tools are continuously being developed to address the full complexity of RNA-seq data.

**Conservation between rodents and humans.** One of the most important considerations for preclinical research is whether the findings are applicable to humans. Though rodent models have provided strong insights into SUDs, not all findings translate well to clinical populations (Kalant, 2010; Venniro et al., 2020). One issue is that, while neural reward circuitry is shared across mammals, (Martínez-García & Lanuza, 2018; Panksepp et al., 2002; Scaplen & Kaun, 2016), there are still neurobiological and behavioral differences between species. However, where both

human and rodent data are available, researchers can use this information to increase the translational relevance of animal models. For example, in Chapters 2 and 4, we utilized the conservation data available with TargetScan to filter predicted targets of miRNAs. Likewise, in Chapter 4 candidate circRNAs in rat were assessed for sequence similarity in human. In Chapter 3, we went a step further and used conservation data as part of a pipeline for narrowing down candidate genes from RNA-seq data. To do so, we incorporated data from multiple sources, including sequence similarity and dN/dS data from Ensembl and biomaRt (Durinck et al., 2005) and data from (Cardoso-Moreira et al., 2020) on temporal expression of genes across mammalian development. Using this approach on a test set of genes, we saw that some genes are highly conserved and may make better candidate genes for future research. Though conservation is not often discussed in the preclinical literature, it is an essential component for translating animal data to the human condition.

**Utility of animal models.** Though animal models have been critical to our understanding of SUDs, they are not without drawbacks. For example, the "gold-standard" of preclinical addiction research is self-administration (SA), in which animals perform learned responses to receive infusions of drug. Because this corresponds well to human SUDs on a surface level, the SA model is considered to have high face validity, particularly when compared to procedures involving experimenter-delivered drug. However, many animal studies do not lead directly to therapeutic development, and those therapies that do get tested in human populations are often not effective. While genetic and biological differences between humans and model organisms may be one reason for this, there are other complicating factors. For one, behavioral experiments in

animals are often simple in their design, and thus they fail to recapitulate the complex symptomatology of human SUDs. For example, throughout this dissertation we performed rodent SA experiments with 1) a simple, though increasingly difficult, schedule of reinforcement, 2) abstinence for a set amount of time (1 or 21 days, nothing between or longer), and 3) a cue reactivity test for measuring animals' post-abstinence response to cocaine-paired cues. We used this to model cocaine craving, though this design fails to capture other aspects of human SUDs, such as loss of control and escalated drug intake. Another drawback of animal models is communicate their emotional or psychological state. Throughout this dissertation we have referred to the cue reactivity test as measuring "craving". While craving often leads to drug-seeking and relapse in humans, relapse can also occur without craving (e.g. (Miller & Gold, 1994)) However, we cannot ask rats about their subjective feeling of craving, and thus we are unable to fully understand this phenomenon using animal models alone.

However, animal models have many strengths, which has led to their continued use, development, and refinement over the last few decades. To fully understand the neurological mechanisms of SUDs, researchers need to probe the brain directly and manipulate its biology. Most importantly, this needs to be coupled with measurements of behavior. For the most part, this is not possible in human subjects. While humans are likely the best models of human behavior, many of the most advanced tools and techniques in neuroscience require either brain extraction or direct neural manipulation. Post-mortem human brain tissue does not fully meet this need, as it is uncoupled from behavior. For this reason, animal models are likely to be a mainstay in SUDs research for

the foreseeable future. However, more work is needed to develop and refine animal models to better recapitulate symptoms of human SUDs.

**Transparency and reproducibility.** In all areas of science, it is important that the methods, approaches, and results are clearly disseminated. If the methodology is not described with a sufficient level of detail, it can be difficult for others to replicate the work. In addition, it can lead to issues in interpreting research findings, and more importantly, determining whether there are any missteps or errors that may impact interpretation. However, for research projects involving bioinformatics and computation, transparency can be aided by providing access to any scripts involved in the analysis. While many journals have made efforts to increase transparency in other ways, e.g. by requiring that raw data files are deposited in a public database, making scripts publicly available is not yet a standard in neuroscience. Of course, this does require scientists to do the additional labor of making their data organized for public consumption. Moreover, many scientists may not want to invite this additional scrutiny on their work, even if their analyses have been performed accurately and ethically. This is unfortunate, because providing scripts will improve the translational value of neuroscience research by allowing others to both evaluate its accuracy and learn from prior work.

Another hurdle for reproducibility is that many widely-used software are not available for all researchers. For example, in Chapter 2, we used Ingenuity Pathway Analysis (IPA; QIAGEN Inc.) to understand what pathways and biological mechanisms might be regulated by our miRNAs of interest. However, because IPA is proprietary, not all researchers have access to this software. Other pathway analysis tools can be used for

free, such as Reactome (G. Wu & Haw, 2017) or Gene Ontology (GO) (The Gene Ontology Consortium, 2019). Still, there are benefits to using IPA over other tools, as it includes additional features such as visualizations and analysis of predicted upstream regulators of the input genes. In Chapter 2, we provide a list of the genes input into IPA in the supplementary material. Yet while our pathway analysis process is transparent and described in detail, it is only reproducible by researchers with access to IPA. Importantly, many of the other software, algorithms, and databases used throughout this dissertation are freely available, including miRanda and TargetScan (Chapters 2 and 4), as well as GTEx (GTEx Consortium, 2013) (Chapter 3).

## Future Directions

**Using the proposed RNA-seq workflow.** In Chapter 2, we proposed a new pipeline for narrowing candidate genes identified from RNA-seq data based on translational potential. In this workflow, we refine candidate gene lists by incorporating additional information on genetic conservation and tissue expression from publicly available databases. The scripts associated with this project are available online for researchers to use, and we hope that others will take advantage of this resource. Additionally, we look forward to continuing the dialogue on RNA-seq best practices and other improvements that can lead to more translational outcomes.

**Validation of candidate RNAs.** A major next step for these works is to further validate candidate miRNAs (Chapters 2 and 4) and circRNAs (Chapter 4). For both projects, our data show that expression of non-coding RNAs is altered in the NAc shell in rats that display low or high cocaine-seeking behavior. For example, several RNAs, such

as miR-376c and miR-107-3p (Chapter 2) as well as *circHomer1* (Chapter 4), show lower expression in rats with higher cocaine motivation. This suggests that increasing levels of these RNAs may be protective against drug-seeking behavior, though this has not yet been tested directly. Ideally, future experiments will specifically manipulate levels of these RNAs in the NAc shell. For example, an ongoing project in the Neisewander lab is to use a lentiviral vector designed to overexpress *circHomer1* and test whether rats with increased *circHomer1* show reduced cocaine-seeking behavior relative to controls. Similar experiments can be designed for any of the candidate miRNAs and circRNAs described in this dissertation. Moreover, further work needs to be done to determine whether these RNAs modulate behavior in other models of cocaine motivation and reinforcement, or across other drugs of abuse.

In addition to validating candidate RNAs, a future goal is to verify their predicted targets. Algorithms such as TargetScan and miRanda are generally considered accurate, but their predictions are not perfect. Therefore, further experiments are needed to truly verify whether specific RNAs bind to one another. This could be achieved by manipulating levels of an RNA *in vivo* (e.g. *circHomer1* in the NAc shell) and then measuring expression of predicted downstream targets (e.g. miR-330-5p, *Homer1b*), along with luciferase assays in cell culture to directly assess binding *in vitro* (see (Bastle et al., 2018) for an example with miR-495). In either case, fully elucidating these regulatory gene networks will advance knowledge in both the addiction field and non-coding RNA research overall.

## Conclusions

The major goal of this dissertation was to identify coding and non-coding RNAs involved in cocaine craving using rodent models, and a variety of tools, techniques, and resources. While our findings are partly limited by low sample size and other factors, we were able to hone in on several candidate RNAs of interest. Further research is necessary to fully understand these RNAs and their role in other cocaine- and drug-related behaviors. In future endeavors, it is important to continue striving for improved methodology and transparency in characterizing these and other RNAs.

REFERENCES


Acosta, J. I., Thiel, K. J., Sanabria, F., Browning, J. R., & Neisewander, J. L. (2008). Effect of schedule of reinforcement on cue-elicited reinstatement of cocaine-seeking behavior. *Behavioural Pharmacology*, *19*(2), 129–136.

Adlakha, Y. K., & Saini, N. (2014). Brain microRNAs and insights into biological functions and therapeutic potential of brain enriched miRNA-128. *Molecular Cancer*, *13*, 33.

Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *ELife*, *4*. https://doi.org/10.7554/eLife.05005

Agrawal, A., Verweij, K. J. H., Gillespie, N. A., Heath, A. C., Lessov-Schlaggar, C. N., Martin, N. G., Nelson, E. C., Slutske, W. S., Whitfield, J. B., & Lynskey, M. T. (2012). The genetics of addiction-a translational perspective. *Translational Psychiatry*, *2*, e140.

Albertson, D. N., Pruetz, B., Schmidt, C. J., Kuhn, D. M., Kapatos, G., & Bannon, M. J. (2004). Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin. *Journal of Neurochemistry*, *88*(5), 1211–1219.

Albertson, D. N., Schmidt, C. J., Kapatos, G., & Bannon, M. J. (2006). Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *31*(10), 2304–2312.

American Psychiatric Association, D. S., Association, A. P., & Others. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5). American psychiatric association Washington, DC.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. *URL Http://Www. Bioinformatics. Babraham. Ac. Uk/Projects/Fastqc*.

Ary, A. W., & Szumlinski, K. K. (2007). Regional differences in the effects of withdrawal from repeated cocaine upon Homer and glutamate receptor expression: A two-species comparison. *Brain Research*, *1184*, 295–305.

Ashwal-Fluss, R., Meyer, M., Pamudurti, N. R., Ivanov, A., Bartok, O., Hanan, M., Evantal, N., Memczak, S., Rajewsky, N., & Kadener, S. (2014). circRNA biogenesis competes with pre-mRNA splicing. *Molecular Cell*, *56*(1), 55–66.

Bachtell, R. K., Jones, J. D., Heinzerling, K. G., Beardsley, P. M., & Comer, S. D. (2017). Glial and neuroinflammatory targets for treating substance use disorders. *Drug and Alcohol Dependence*, *180*, 156–170.

Baker, D. A., Tran-Nguyen, L. T. L., Fuchs, R. A., & Neisewander, J. L. (2001). Influence of individual differences and chronic fluoxetine treatment on cocaine-seeking behavior in rats. *Psychopharmacology*, *155*(1), 18–26.

Barnett, D. W., Garrison, E. K., Quinlan, A. R., Strömberg, M. P., & Marth, G. T. (2011). BamTools: a C++ API and toolkit for analyzing and managing BAM files. *Bioinformatics* , *27*(12), 1691–1692.

Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*, *116*(2), 281–297.

Bartel, D. P. (2009). MicroRNAs: Target recognition and regulatory functions. *Cell*, *136*(2), 215–233.

Bastle, R. M., Oliver, R. J., Gardiner, A. S., Pentkowski, N. S., Bolognani, F., Allan, A. M., Chaudhury, T., St. Peter, M., Galles, N., Smith, C., Neisewander, J. L., & Perrone-Bizzozero, N. I. (2018). *In silico* identification and *in vivo* validation of miR-495 as a novel regulator of motivation for cocaine that targets multiple addiction-related networks in the nucleus accumbens. *Molecular Psychiatry*, *23*(2), 434–443.

Batel, P. (2000). Addiction and schizophrenia. *European Psychiatry: The Journal of the Association of European Psychiatrists*, *15*(2), 115–122.

Baudry, A., Mouillet-Richard, S., Schneider, B., Launay, J.-M., & Kellermann, O. (2010). miR-16 targets the serotonin transporter: a new facet for adaptive responses to antidepressants. *Science* , *329*(5998), 1537–1541.

Belin-Rauscent, A., Fouyssac, M., Bonci, A., & Belin, D. (2016). How Preclinical Models Evolved to Resemble the Diagnostic Criteria of Drug Addiction. *Biological Psychiatry*, *79*(1), 39–46.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Beveridge, N. J., Gardiner, E., Carroll, A. P., Tooney, P. A., & Cairns, M. J. (2010). Schizophrenia is associated with an increase in cortical microRNA biogenesis. *Molecular Psychiatry*, *15*(12), 1176–1189.

Bi, R., & Liu, P. (2016). Sample size calculation while controlling false discovery rate for differential expression analysis with RNA-sequencing experiments. *BMC Bioinformatics*, *17*, 146.

Boese, A. S., Saba, R., Campbell, K., Majer, A., Medina, S., Burton, L., Booth, T. F., Chong, P., Westmacott, G., Dutta, S. M., Saba, J. A., & Booth, S. A. (2016). MicroRNA abundance is altered in synaptoneurosomes during prion disease. *Molecular and Cellular Neurosciences*, *71*, 13–24.

Bonneau, E., Neveu, B., Kostantin, E., Tsongalis, G. J., & De Guire, V. (2019). How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market. In *Electronic Journal of the International Federation of Clinical Chemistry and Laboratory Medicine* (Vol. 30, Issue 2, pp. 114–127). International Federation of Clinical Chemistry and Laboratory Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6599191/

Bu, Q., Long, H., Shao, X., Gu, H., Kong, J., Luo, L., Liu, B., Guo, W., Wang, H., Tian, J., Zhao, Y., & Cen, X. (2019). Cocaine induces differential circular RNA expression in striatum. *Translational Psychiatry*, *9*(1), 199.

Buckland, P. R. (2008). Will we ever find the genes for addiction? *Addiction* , *103*(11), 1768–1776.

Burmeister, J. J., Lungren, E. M., & Neisewander, J. L. (2003). Effects of fluoxetine and d-fenfluramine on cocaine-seeking behavior in rats. *Psychopharmacology*, *168*(1–2), 146–154.

Bushel, P. R., Caiment, F., Wu, H., O'Lone, R., Day, F., Calley, J., Smith, A., & Li, J. (2018). RATEmiRs: the rat atlas of tissue-specific and enriched miRNAs database. *BMC Genomics*, *19*(1), 825.

Bushel, P. R., Caiment, F., Wu, H., O'Lone, R., Day, F., Calley, J., Smith, A., Li, J., & Harrill, A. H. (2020). RATEmiRs: the rat atlas of tissue-specific and enriched miRNAs for discerning baseline expression exclusivity of candidate biomarkers. *RNA Biology*, *17*(5), 630–636.

Cai, Y., Sun, Z., Jia, H., Luo, H., Ye, X., Wu, Q., Xiong, Y., Zhang, W., & Wan, J. (2017). Rpph1 Upregulates CDC42 Expression and Promotes Hippocampal Neuron Dendritic Spine Formation by Competing with miR-330-5p. *Frontiers in Molecular Neuroscience*, *10*, 27.

Campbell, P. D., Shen, K., Sapio, M. R., Glenn, T. D., Talbot, W. S., & Marlow, F. L. (2014). Unique function of Kinesin Kif5A in localization of mitochondria in axons. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(44), 14717–14732.

Cano, M., Oh, S., Salas-Wright, C. P., & Vaughn, M. G. (2020). Cocaine use and overdose mortality in the United States: Evidence from two national data sources, 2002-2018. *Drug and Alcohol Dependence*, *214*, 108148.

Cardoso-Moreira, M., Halbert, J., Valloton, D., Velten, B., Chen, C., Shao, Y., Liechti, A., Ascenção, K., Rummel, C., Ovchinnikova, S., Mazin, P. V., Xenarios, I., Harshman, K., Mort, M., Cooper, D. N., Sandi, C., Soares, M. J., Ferreira, P. G., Afonso, S., … Kaessmann, H. (2019). Gene expression across mammalian organ development. *Nature*, *571*(7766), 505–509.

Cardoso-Moreira, M., Sarropoulos, I., Velten, B., Mort, M., Cooper, D. N., Huber, W., & Kaessmann, H. (2020). Developmental Gene Expression Differences between Humans and Mammalian Models. *Cell Reports*, *33*(4), 108308.

Carlezon, W. A., Thome, J., Olson, V. G., Lane-Ladd, S. B., Brodkin, E. S., Hiroi, N., Duman, R. S., Neve, R. L., & Nestler, E. J. (1998). Regulation of cocaine reward by CREB. *Science* , *282*(5397), 2272–2275.

Carlezon, William A., Duman, R. S., & Nestler, E. J. (2005). The many faces of CREB. In *Trends in Neurosciences* (Vol. 28, Issue 8, pp. 436–445). https://doi.org/10.1016/j.tins.2005.06.005

Carpenter, M. D., Hu, Q., Bond, A. M., Lombroso, S. I., Czarnecki, K. S., Lim, C. J., Song, H., Wimmer, M. E., Pierce, R. C., & Heller, E. A. (2020). Nr4a1 suppresses cocaine-induced behavior via epigenetic regulation of homeostatic target genes. *Nature Communications*, *11*(1), 1–14.

Cartwright, W. S. (2008). Economic costs of drug abuse: financial, cost of illness, and services. *Journal of Substance Abuse Treatment*, *34*(2), 224–233.

Center for Behavioral Health Statistics and Quality. (2015). *Results from the 2014 National Survey on Drug Use and Health: Detailed Tables*.

Chandrasekar, V., & Dreyer, J.-L. (2009). microRNAs miR-124, let-7d and miR-181a regulate Cocaine-induced Plasticity. *Molecular and Cellular Neurosciences*, *42*(4), 350–362.

Chandrasekar, V., & Dreyer, J.-L. (2011). Regulation of MiR-124, Let-7d, and MiR-181a in the accumbens affects the expression, extinction, and reinstatement of cocaine-induced conditioned place preference. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *36*(6), 1149–1164.

Chang, T.-C., Wentzel, E. A., Kent, O. A., Ramachandran, K., Mullendore, M., Lee, K. H., Feldmann, G., Yamakuchi, M., Ferlito, M., Lowenstein, C. J., Arking, D. E., Beer, M. A., Maitra, A., & Mendell, J. T. (2007). Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Molecular Cell*, *26*(5), 745–752.

Chaudhuri, A. D., Dastgheyb, R. M., Yoo, S. W., Trout, A., Talbot, C. C., Hao, H., Witwer, K. W., & Haughey, N. J. (2018). TNFα and IL-1β modify the miRNA cargo of astrocyte shed extracellular vesicles to regulate neurotrophic signaling in neurons article. *Cell Death & Disease*, *9*(3). https://doi.org/10.1038/s41419-018-0369-4

Chauvet, C., Lardeux, V., Jaber, M., & Solinas, M. (2011). Brain regions associated with the reversal of cocaine conditioned place preference by environmental enrichment. *Neuroscience*, *184*, 88–96.

Chen, L., Wang, C., Sun, H., Wang, J., Liang, Y., Wang, Y., & Wong, G. (2021). The bioinformatics toolbox for circRNA discovery and analysis. *Briefings in Bioinformatics*, *22*(2), 1706–1728.

Chen, L.-L., & Yang, L. (2015). Regulation of circRNA biogenesis. *RNA Biology*, *12*(4), 381–388.

Chen, W., & Schuman, E. (2016). Circular RNAs in Brain and Other Tissues: A Functional Enigma. *Trends in Neurosciences*, *39*(9), 597–604.

Chen, X., Han, P., Zhou, T., Guo, X., Song, X., & Li, Y. (2016). circRNADb: A comprehensive database for human circular RNAs with protein-coding annotations. *Scientific Reports*, *6*, 34985.

Childress, A. R., Hole, A. V., Ehrman, R. N., Robbins, S. J., McLellan, A. T., & O'Brien, C. P. (1993). Cue reactivity and cue reactivity interventions in drug dependence. *NIDA Research Monograph*, *137*, 73–95.

Childress, A. R., McLellan, A. T., Ehrman, R., & O'Brien, C. P. (1988). Classically conditioned responses in opioid and cocaine dependence: a role in relapse? *NIDA Research Monograph*, *84*, 25–43.

Ciccocioppo, R., Martin-Fardon, R., & Weiss, F. (2004). Stimuli associated with a single cocaine experience elicit long-lasting cocaine-seeking. *Nature Neuroscience*, *7*(5), 495–496.

Clark, K. H., Wiley, C. A., & Bradberry, C. W. (2013). Psychostimulant abuse and neuroinflammation: Emerging evidence of their interconnection. *Neurotoxicity Research*, *23*(2), 174–188.

Conklin, C. A., & Tiffany, S. T. (2002). Applying extinction research and theory to cue-exposure addiction treatments. *Addiction* , *97*(2), 155–167.

Conn, S. J., Pillman, K. A., Toubia, J., Conn, V. M., Salmanidis, M., Phillips, C. A., Roslan, S., Schreiber, A. W., Gregory, P. A., & Goodall, G. J. (2015). The RNA binding protein quaking regulates formation of circRNAs. *Cell*, *160*(6), 1125–1134.

Costa-Silva, J., Domingues, D., & Lopes, F. M. (2017). RNA-Seq differential expression analysis: An extended review and a software tool. *PloS One*, *12*(12), e0190152.

Crabbe, J. C. (2002). Genetic contributions to addiction. *Annual Review of Psychology*, *53*, 435–462.

Cuesta, S., & Pacchioni, A. M. (2017). Are changes in the Wnt/β-catenin pathway involved in cocaine and stress-induced long-term neuroadaptations? *Journal of Addiction and Preventive Medicine*, *2*(2), 112.

Cunningham, C. L., Niehus, J. S., & Noble, D. (1993). Species difference in sensitivity to ethanol's hedonic effects. *Alcohol* , *10*(2), 97–102.

Dackis, C. A., & O'Brien, C. P. (2001). Cocaine dependence: a disease of the brain's reward centers. *Journal of Substance Abuse Treatment*, *21*(3), 111–117.

Dallas, P. B., Gottardo, N. G., Firth, M. J., Beesley, A. H., Hoffmann, K., Terry, P. A., Freitas, J. R., Boag, J. M., Cummings, A. J., & Kees, U. R. (2005). Gene expression levels assessed by oligonucleotide microarray analysis and quantitative real-time RT-PCR -- how well do they correlate? *BMC Genomics*, *6*, 59.

Dave, R. S., & Khalili, K. (2010). Morphine treatment of human monocyte-derived macrophages induces differential miRNA and protein expression: Impact on inflammation and oxidative stress in the central nervous system. *Journal of Cellular Biochemistry*, *110*(4), 834–845.

Deak, J. D., & Johnson, E. C. (2021). Genetics of substance use disorders: a review. *Psychological Medicine*, 1–12.

Dell'Orco, M., Elyaderani, A., Vannan, A., Sekar, S., Powell, G., Liang, W. S., Neisewander, J. L., & Perrone-Bizzozero, N. I. (2021). HuD Regulates mRNA-circRNA-miRNA Networks in the Mouse Striatum Linked to Neuronal Development and Drug Addiction. *Biology*, *10*(9). https://doi.org/10.3390/biology10090939

Dell'Orco, M., Oliver, R. J., & Perrone-Bizzozero, N. (2020). HuD Binds to and Regulates Circular RNAs Derived From Neuronal Development- and Synaptic Plasticity-Associated Genes. *Frontiers in Genetics*, *11*, 790.

Dessaud, E., Salaün, D., Gayet, O., Chabbert, M., & deLapeyrière, O. (2006). Identification of lynx2, a novel member of the ly-6/neurotoxin superfamily, expressed in neuronal subpopulations during mouse development. *Molecular and Cellular Neurosciences*, *31*(2), 232–242.

Dube, U., Del-Aguila, J. L., Li, Z., Budde, J. P., Jiang, S. H. A. N., Harari, O., & Cruchaga, C. (2018). Circular RNAs in Alzheimer's Disease Brain Tissues. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *14*(7), P1401.

Dube, U., Harari, O., Zimmerman, A., Hafez, A., Brigman, J., Cruchaga, C., Mellios, N., & Benitez, B. A. (2020). Alzheimer□associated circular RNA circHOMER1 regulates synaptic gene expression and cognition. *Alzheimer's & Dementia: The Journal of the Alzheimer's Association*, *16*(S2). https://doi.org/10.1002/alz.042335

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., De Moor, B., Brazma, A., & Huber, W. (2005). BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* , *21*(16), 3439–3440.

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–1191.

Eastel, J. M., Lam, K. W., Lee, N. L., Lok, W. Y., Tsang, A. H. F., Pei, X. M., Chan, A. K. C., Cho, W. C. S., & Wong, S. C. C. (2019). Application of NanoString technologies in companion diagnostic development. *Expert Review of Molecular Diagnostics*, *19*(7), 591–598.

Ehrman, R. N., Robbins, S. J., Childress, A. R., & O'Brien, C. P. (1992). Conditioned responses to cocaine-related stimuli in cocaine abuse patients. *Psychopharmacology*, *107*(4), 523–529.

Eipper-Mains, J. E., Kiraly, D. D., Palakodeti, D., Mains, R. E., Eipper, B. A., & Graveley, B. R. (2011). microRNA-Seq reveals cocaine-regulated expression of striatal microRNAs. *RNA* , *17*(8), 1529–1543.

Ellenbroek, B., & Youn, J. (2016). Rodent models in neuroscience research: is it a rat race? *Disease Models & Mechanisms*, *9*(10), 1079–1087.

Enright, A., John, B., Gaul, U., Tuschl, T., Sander, C., & Marks, D. (2003). MicroRNA Targets in Drosophila. *Genome Biology*, *4*(11), P8.

Ewald, A. C., Kiernan, E. A., Roopra, A. S., Radcliff, A. B., Timko, R. R., Baker, T. L., & Watters, J. J. (2020). Sex- and Region-Specific Differences in the Transcriptomes of Rat Microglia from the Brainstem and Cervical Spinal Cord. *The Journal of Pharmacology and Experimental Therapeutics*, *375*(1), 210–222.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* , *32*(19), 3047–3048.

Fang, Z., & Cui, X. (2011). Design and validation issues in RNA-seq experiments. *Briefings in Bioinformatics*, *12*(3), 280–287.

Fededa, J. P., Esk, C., Mierzwa, B., Stanyte, R., Yuan, S., Zheng, H., Ebnet, K., Yan, W., Knoblich, J. A., & Gerlich, D. W. (2016). MicroRNA-34/449 controls mitotic spindle orientation during mammalian cortex development. *The EMBO Journal*, *35*(22), 2386–2398.

Feng, Y., Li, T., Xing, C., Wang, C., Duan, Y., Yuan, L., & Zhang, Y. (2021). Effective inhibition of miR-330/SHIP1/NF-κB signaling pathway via miR-330 sponge repolarizes microglia differentiation. *Cell Biology International*, *45*(4), 785–794.

Filipowicz, W. (2005). RNAi: the nuts and bolts of the RISC machine. *Cell*, *122*(1), 17–20.

Floresco, S. B. (2015). The nucleus accumbens: an interface between cognition, emotion, and action. *Annual Review of Psychology*, *66*, 25–52.

Floris, G., Gillespie, A., Zanda, M. T., Dabrowski, K. R., & Sillivan, S. E. (2022). Heroin Regulates Orbitofrontal Circular RNAs. *International Journal of Molecular Sciences*, *23*(3). https://doi.org/10.3390/ijms23031453

Floris, G., Zhang, L., Follesa, P., & Sun, T. (2017). Regulatory role of circular RNAs and neurological disorders. *Molecular Neurobiology*, *54*(7), 5156–5165.

Forero, D. A., & González-Giraldo, Y. (2020). Convergent functional genomics of cocaine misuse in humans and animal models. *The American Journal of Drug and Alcohol Abuse*, *46*(1), 22–30.

Forero, D. A., van der Ven, K., Callaerts, P., & Del-Favero, J. (2010). miRNA genes and the brain: implications for psychiatric disorders. *Human Mutation*, *31*(11), 1195–1204.

Foye, C., Yan, I. K., David, W., Shukla, N., Habboush, Y., Chase, L., Ryland, K., Kesari, V., & Patel, T. (2017). Comparison of miRNA quantitation by Nanostring in serum and plasma samples. *PloS One*, *12*(12), e0189165.

Friedman, R. C., Farh, K. K.-H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92–105.

Gao, Y., Zhang, J., & Zhao, F. (2018). Circular RNA identification based on multiple seed matching. *Briefings in Bioinformatics*, *19*(5), 803–810.

Gawin, F. H., & Kleber, H. D. (1986). Abstinence symptomatology and psychiatric diagnosis in cocaine abusers. Clinical observations. *Archives of General Psychiatry*, *43*(2), 107–113.

Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., … Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, *26*(3), 317–325.

Gelernter, J., Sherva, R., Koesterer, R., Almasy, L., Zhao, H., Kranzler, H. R., & Farrer, L. (2014). Genome-wide association study of cocaine dependence and related traits: FAM53B identified as a risk gene. *Molecular Psychiatry*, *19*(6), 717–723.

Geschwind, D. H., & Flint, J. (2015). Genetics and genomics of psychiatric disease. *Science*, *349*(6255), 1489–1494.

Ghasemzadeh, M. B., Permenter, L. K., Lake, R., Worley, P. F., & Kalivas, P. W. (2003). Homer1 proteins and AMPA receptors modulate cocaine-induced behavioural plasticity. *The European Journal of Neuroscience*, *18*(6), 1645–1651.

Ghasemzadeh, M. B., Vasudevan, P., Mueller, C., Seubert, C., & Mantsch, J. R. (2009). Neuroadaptations in the cellular and postsynaptic group 1 metabotropic glutamate receptor mGluR5 and Homer proteins following extinction of cocaine self-administration. *Neuroscience Letters*, *452*(2), 167–171.

Ghasemzadeh, M. B., Windham, L. K., Lake, R. W., Acker, C. J., & Kalivas, P. W. (2009). Cocaine activates Homer1 immediate early gene transcription in the mesocorticolimbic circuit: differential regulation by dopamine and glutamate signaling. *Synapse* , *63*(1), 42–53.

Gibson, U. E., Heid, C. A., & Williams, P. M. (1996). A novel method for real time quantitative RT-PCR. *Genome Research*, *6*(10), 995–1001.

Giraldez, A. J., Cinalli, R. M., Glasner, M. E., Enright, A. J., Thomson, J. M., Baskerville, S., Hammond, S. M., Bartel, D. P., & Schier, A. F. (2005). MicroRNAs regulate brain morphogenesis in zebrafish. *Science*, *308*(5723), 833–838.

Goldman, D., Oroszi, G., & Ducci, F. (2005). The genetics of addictions: uncovering the genes. *Nature Reviews. Genetics*, *6*(7), 521–532.

Gowen, A. M., Odegaard, K. E., Hernandez, J., Chand, S., Koul, S., Pendyala, G., & Yelamanchili, S. V. (2021). Role of microRNAs in the pathophysiology of addiction. *Wiley Interdisciplinary Reviews. RNA*, *12*(3), e1637.

Grant, S., London, E. D., Newlin, D. B., Villemagne, V. L., Liu, X., Contoreggi, C., Phillips, R. L., Kimes, A. S., & Margolint, A. (1996). *Activation of memory circuits during cue-elicited cocaine craving* (Vol. 93, pp. 12040–12045). https://www.pnas.org/content/pnas/93/21/12040.full.pdf

Griffiths-Jones, S. (2004). The microRNA Registry. *Nucleic Acids Research*, *32*(Database issue), D109-11.

Grimm, J. W., Hope, B. T., Wise, R. A., & Shaham, Y. (2001). Neuroadaptation. Incubation of cocaine craving after withdrawal. *Nature*, *412*(6843), 141–142.

GTEx Consortium. (2013). The Genotype-Tissue Expression (GTEx) project. *Nature Genetics*, *45*(6), 580–585.

Hafez, A. K., Zimmerman, A. J., Papageorgiou, G., Chandrasekaran, J., Amoah, S. K., Lin, R., Lozano, E., Pierotti, C., Dell'Orco, M., Hartley, B. J., Alural, B., Lalonde, J., Esposito, J. M., Berretta, S., Squassina, A., Chillotti, C., Voloudakis, G., Shao, Z., Fullard, J. F., … Mellios, N. (2022). A bidirectional competitive interaction between circHomer1 and Homer1b within the orbitofrontal cortex regulates reversal learning. *Cell Reports*, *38*(3), 110282.

Hansen, T. B., Jensen, T. I., Clausen, B. H., Bramsen, J. B., Finsen, B., Damgaard, C. K., & Kjems, J. (2013). Natural RNA circles function as efficient microRNA sponges. *Nature*, *495*(7441), 384–391.

Hansen, T. B., Venø, M. T., Damgaard, C. K., & Kjems, J. (2016). Comparison of circular RNA prediction tools. *Nucleic Acids Research*, *44*(6), e58.

Harris, G. C., Altomare, K., & Aston-Jones, G. (2001). Preference for a cocaine-associated environment is attenuated by augmented accumbal serotonin in cocaine withdrawn rats. *Psychopharmacology*, *156*(1), 14–22.

Hausser, J., Syed, A. P., Bilen, B., & Zavolan, M. (2013). Analysis of CDS-located miRNA target sites suggests that they can effectively inhibit translation. *Genome Research*, *23*(4), 604–615.

He, Y., Yang, C., Kirkmire, C. M., & Wang, Z. J. (2010). *Behavioral/Systems/Cognitive Regulation of Opioid Tolerance by let-7 Family MicroRNA Targeting the Opioid Receptor*. https://doi.org/10.1523/JNEUROSCI.2419-10.2010

Hollander, J. A., Im, H.-I. I., Amelio, A. L., Kocerha, J., Bali, P., Lu, Q., Willoughby, D., Wahlestedt, C., Conkright, M. D., & Kenny, P. J. (2010). Striatal microRNA controls cocaine intake through CREB signalling. *Nature*, *466*(7303), 197–202.

Hossain, M. T., Peng, Y., Feng, S., & Wei, Y. (2020). FcircSEC: An R Package for Full Length circRNA Sequence Extraction and Classification. *International Journal of Genomics and Proteomics*, *2020*, 9084901.

Hsu, M. T., & Coca-Prados, M. (1979). Electron microscopic evidence for the circular form of RNA in the cytoplasm of eukaryotic cells. *Nature*, *280*(5720), 339–340.

Huggett, S. B., Bubier, J. A., Chesler, E. J., & Palmer, R. H. C. (2020). Do gene expression findings from mouse models of cocaine use recapitulate human

cocaine use disorder in reward circuitry? *Genes, Brain, and Behavior*. https://doi.org/10.1111/gbb.12689

Huggett, S. B., & Stallings, M. C. (2020). Cocaine'omics: Genome□wide and transcriptome□wide analyses provide biological insight into cocaine use and dependence. *Addiction Biology*, *25*(2). https://doi.org/10.1111/adb.12719

Im, H. I., Hollander, J. A., Bali, P., & Kenny, P. J. (2010). MeCP2 controls BDNF expression and cocaine intake through homeostatic interactions with microRNA-212. *Nature Neuroscience*, *13*(9), 1120–1127.

Im, H. I., & Kenny, P. J. (2012). MicroRNAs in neuronal function and dysfunction. In *Trends in Neurosciences* (Vol. 35, Issue 5, pp. 325–334). https://doi.org/10.1016/j.tins.2012.01.004

Impey, S., Davare, M., Lasiek, A., Fortin, D., Ando, H., Varlamova, O., Obrietan, K., Soderling, T. R., Goodman, R. H., & Wayman, G. A. (2010). An activity-induced microRNA controls dendritic spine formation by regulating Rac1-PAK signaling. *Molecular and Cellular Neurosciences*, *43*(1), 146–156.

Jasińska, M., Miłek, J., Cymerman, I. A., Łęski, S., Kaczmarek, L., & Dziembowska, M. (2016). miR-132 Regulates Dendritic Spine Structure by Direct Targeting of Matrix Metalloproteinase 9 mRNA. *Molecular Neurobiology*, *53*(7), 4701–4712.

Jeck, W. R., & Sharpless, N. E. (2014). Detecting and characterizing circular RNAs. *Nature Biotechnology*, *32*(5), 453–461.

Jeck, W. R., Sorrentino, J. A., Wang, K., Slevin, M. K., Burd, C. E., Liu, J., Marzluff, W. F., & Sharpless, N. E. (2013). Circular RNAs are abundant, conserved, and associated with ALU repeats. *RNA* , *19*(2), 141–157.

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., & Marks, D. S. (2004). Human MicroRNA targets. *PLoS Biology*, *2*(11), e363.

Kalant, H. (2010). What neurobiology cannot tell us about addiction. *Addiction* , *105*(5), 780–789.

Kariisa, M., Scholl, L., Wilson, N., Seth, P., & Hoots, B. (2019). Drug Overdose Deaths Involving Cocaine and Psychostimulants with Abuse Potential - United States, 2003-2017. *MMWR. Morbidity and Mortality Weekly Report*, *68*(17), 388–395.

Kenny, P. J. (2014). Epigenetics, microRNA, and addiction. *Dialogues in Clinical Neuroscience*, *16*(3), 335–344.

Kessler, R. C., Wai, T. C., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey

Replication. In *Archives of General Psychiatry* (Vol. 62, Issue 6, pp. 617–627). American Medical Association. https://doi.org/10.1001/archpsyc.62.6.617

Koesling, D., Russwurm, M., Mergia, E., Mullershausen, F., & Friebe, A. (2004). Nitric oxide-sensitive guanylyl cyclase: structure and regulation. *Neurochemistry International*, *45*(6), 813–819.

Koob, G. F., & Le Moal, M. (2005). Plasticity of reward neurocircuitry and the "dark side" of drug addiction. *Nature Neuroscience*, *8*(11), 1442–1444.

Kosten, T. R., Markou, A., & Koob, G. F. (1998). Depression and stimulant dependence: Neurobiology and pharmacotherapy. *The Journal of Nervous and Mental Disease*, *186*(12), 737–745.

Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162.

Krämer, A., Green, J., Pollard, J., & Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* , *30*(4), 523–530.

Krasnova, I. N., Justinova, Z., & Cadet, J. L. (2016). Methamphetamine addiction: involvement of CREB and neuroinflammatory signaling pathways. In *Psychopharmacology* (Vol. 233, Issue 10, pp. 1945–1962). Springer Verlag. https://doi.org/10.1007/s00213-016-4235-8

Kreibich, A. S., Briand, L., Cleck, J. N., Ecke, L., Rice, K. C., & Blendy, J. A. (2009). Stress-induced potentiation of cocaine reward: A role for CRF R1 and CREB. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *34*(12), 2609–2617.

Krepelkova, I., Mrackova, T., Izakova, J., Dvorakova, B., Chalupova, L., Mikulik, R., Slaby, O., Bartos, M., & Ruzicka, V. (2019). Evaluation of miRNA detection methods for the analytical characteristic necessary for clinical utilization. *BioTechniques*, *66*(6), 277–284.

Kuehn, B. M. (2021). Accelerated Overdose Deaths Linked With COVID-19. *JAMA: The Journal of the American Medical Association*, *325*(6), 523.

Kufahl, P. R., Zavala, A. R., Singh, A., Thiel, K. J., Dickey, E. D., Joyce, J. N., & Neisewander, J. L. (2009). c-Fos expression associated with reinstatement of cocaine-seeking behavior by response-contingent conditioned cues. *Synapse* , *63*(10), 823–835.

Kuhar, M. J., Jaworski, J. N., Hubert, G. W., Philpot, K. B., & Dominguez, G. (2005). Cocaine- and amphetamine-regulated transcript peptides play a role in drug abuse and are potential therapeutic targets. *The AAPS Journal*, *7*(1), E259-65.

Kulkarni, M. M. (2011). Digital multiplexed gene expression analysis using the nanostring ncounter system. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]*, *94*(SUPPL.94), 25B.10.1-25B.10.17.

Lai, E. C. (2002). Micro RNAs are complementary to 3' UTR sequence motifs that mediate negative post-transcriptional regulation. *Nature Genetics*, *30*(4), 363–364.

Larson, E. B., Graham, D. L., Arzaga, R. R., Buzin, N., Webb, J., Green, T. A., Bass, C. E., Neve, R. L., Terwilliger, E. F., Nestler, E. J., & Self, D. W. (2011). Overexpression of CREB in the nucleus accumbens shell increases cocaine reinforcement in self-administering rats. *Journal of Neuroscience*, *31*(45), 16447–16457.

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). Voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29.

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell*, *75*(5), 843–854.

Lerman, C., LeSage, M. G., Perkins, K. A., O'Malley, S. S., Siegel, S. J., Benowitz, N. L., & Corrigall, W. A. (2007). Translational research in medication development for nicotine dependence. *Nature Reviews. Drug Discovery*, *6*(9), 746–762.

Li, C.-Y., Mao, X., & Wei, L. (2008). Genes and (common) pathways underlying drug addiction. *PLoS Computational Biology*, *4*(1), e2.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. In *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* , *25*(16), 2078–2079.

Li, J., Sun, Q., Zhu, S., Xi, K., Shi, Q., Pang, K., Liu, X., Li, M., Zhang, Y., & Sun, J. (2020). Knockdown of circHomer1 ameliorates METH-induced neuronal injury through inhibiting Bbc3 expression. *Neuroscience Letters*, *732*, 135050.

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., & Yang, J.-H. (2013). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, *42*(D1), D92–D97.

Li, J.-H., Liu, S., Zhou, H., Qu, L.-H., & Yang, J.-H. (2014). starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic Acids Research*, *42*(Database issue), D92-7.

Li, Xiang, Yang, L., & Chen, L.-L. (2018). The Biogenesis, Functions, and Challenges of Circular RNAs. *Molecular Cell*, *71*(3), 428–442.

Li, Xuan, & Wolf, M. E. (2015). Multiple faces of BDNF in cocaine addiction. In *Behavioural Brain Research* (Vol. 279, pp. 240–254). Elsevier. https://doi.org/10.1016/j.bbr.2014.11.018

Li, Y., Wang, X., Cao, X., Li, N., Meng, S., & Li, A. (2020). SHED-derived exosomes regulate microglial polarization in the treatment of traumatic brain injury in rats via miR-330-5p. In *Research Square*. https://doi.org/10.21203/rs.3.rs-65521/v1

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* , *30*(7), 923–930.

Lippi, G., Steinert, J. R., Marczylo, E. L., D'Oro, S., Fiore, R., Forsythe, I. D., Schratt, G., Zoli, M., Nicotera, P., & Young, K. W. (2011). Targeting of the Arpc3 actin nucleation factor by miR-29a/b regulates dendritic spine morphology. *The Journal of Cell Biology*, *194*(6), 889–904.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* , *25*(4), 402–408.

Lizé, M., Pilarski, S., & Dobbelstein, M. (2009). E2F1-inducible microRNA 449a/b suppresses cell proliferation and promotes apoptosis. *Cell Death and Differentiation*, *17*(3), 452–458.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550.

Lukiw, W. J. (2013). Circular RNA (circRNA) in Alzheimer's disease (AD). *Frontiers in Genetics*, *4*, 1–2.

Luo, P., Li, X., Fei, Z., & Poon, W. (2012). Scaffold protein Homer 1: implications for neurological diseases. *Neurochemistry International*, *61*(5), 731–738.

Lynch, W. J., Peterson, A. B., Sanchez, V., Abel, J., & Smith, M. A. (2013). Exercise as a novel treatment for drug addiction: a neurobiological and stage-dependent hypothesis. *Neuroscience and Biobehavioral Reviews*, *37*(8), 1622–1644.

Lytle, J. R., Yario, T. A., & Steitz, J. A. (2007). Target mRNAs are repressed as efficiently by microRNA-binding sites in the 5′ UTR as in the 3′ UTR. *Proc. Natl. Acad. Sci. U. S. A.* https://www.pnas.org/content/pnas/104/23/9667

Ma, Y.-Y., Wang, X., Huang, Y., Marie, H., Nestler, E. J., Schlüter, O. M., & Dong, Y. (2016). Re-silencing of silent synapses unmasks anti-relapse effects of environmental enrichment. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(18), 5089–5094.

Maffioletti, E., Cattaneo, A., Rosso, G., Maina, G., Maj, C., Gennarelli, M., Tardito, D., & Bocchio-Chiavetto, L. (2016). Peripheral whole blood microRNA alterations in major depression and bipolar disorder. *Journal of Affective Disorders*, *200*, 250–258.

Marees, A. T., Gamazon, E. R., Gerring, Z., Vorspan, F., Fingal, J., van den Brink, W., Smit, D. J. A., Verweij, K. J. H., Kranzler, H. R., Sherva, R., Farrer, L., International Cannabis Consortium, Gelernter, J., & Derks, E. M. (2020). Post-GWAS analysis of six substance use traits improves the identification and functional interpretation of genetic risk loci. *Drug and Alcohol Dependence*, *206*, 107703.

Markou, A., Weiss, F., Gold, L. H., Caine, S. B., Schulteis, G., & Koob, G. F. (1993). Animal models of drug craving. *Psychopharmacology*, *112*(2–3), 163–182.

Martínez-García, F., & Lanuza, E. (2018). Evolution of vertebrate survival circuits. *Current Opinion in Behavioral Sciences*, *24*, 113–123.

Matsuzaki, F., Shirane, M., Matsumoto, M., & Nakayama, K. I. (2011). Protrudin serves as an adaptor molecule that connects KIF5 and its cargoes in vesicular transport during process formation. *Molecular Biology of the Cell*, *22*(23), 4602–4620.

Mattson, B. J., Bossert, J. M., Simmons, D. E., Nozaki, N., Nagarkar, D., Kreuter, J. D., & Hope, B. T. (2005). Cocaine-induced CREB phosphorylation in nucleus accumbens of cocaine-sensitized rats is enabled by enhanced activation of extracellular signal-related kinase, but not protein kinase A. *Journal of Neurochemistry*, *95*(5), 1481–1494.

McCall Jones, C., Baldwin, G. T., & Compton, W. M. (2017). Recent Increases in Cocaine-Related Overdose Deaths and the Role of Opioids. *American Journal of Public Health*, *107*(3), 430–432.

McClearn, G. E., & Rodgers, D. A. (1959). Differences in Alcohol Preference among Inbred Strains of Mice. *Quarterly Journal of Studies on Alcohol*, *20*, 691–695.

McClung, C. A., & Nestler, E. J. (2003). Regulation of gene expression and cocaine reward by CREB and ΔFosB. *Nature Neuroscience*, *6*(11), 1208–1215.

McCrory, E. J., & Mayes, L. (2015). Understanding Addiction as a Developmental Disorder: An Argument for a Developmentally Informed Multilevel Approach. *Current Addiction Reports*, *2*(4), 326–330.

Mehta, S. L., Dempsey, R. J., & Vemuganti, R. (2020). Role of circular RNAs in brain development and CNS diseases. *Progress in Neurobiology*, *186*, 101746.

Melé, M., Ferreira, P. G., Reverter, F., DeLuca, D. S., Monlong, J., Sammeth, M., Young, T. R., Goldmann, J. M., Pervouchine, D. D., Sullivan, T. J., Johnson, R., Segrè, A. V., Djebali, S., Niarchou, A., GTEx Consortium, Wright, F. A., Lappalainen, T., Calvo, M., Getz, G., … Guigó, R. (2015). Human genomics. The human transcriptome across tissues and individuals. *Science*, *348*(6235), 660–665.

Mellios, N., Feldman, D. A., Sheridan, S. D., Ip, J. P. K., Kwok, S., Amoah, S. K., Rosen, B., Rodriguez, B. A., Crawford, B., Swaminathan, R., Chou, S., Li, Y., Ziats, M., Ernst, C., Jaenisch, R., Haggarty, S. J., & Sur, M. (2018). MeCP2-regulated miRNAs control early human neurogenesis through differential effects on ERK and AKT signaling. *Molecular Psychiatry*, *23*(4), 1051–1065.

Mello, N. K., & Negus, S. S. (1996). Preclinical evaluation of pharmacotherapies for treatment of cocaine and opioid abuse using drug self-administration procedures. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *14*(6), 375–424.

Memczak, S., Jens, M., Elefsinioti, A., Torti, F., Krueger, J., Rybak, A., Maier, L., Mackowiak, S. D., Gregersen, L. H., Munschauer, M., Loewer, A., Ziebold, U., Landthaler, M., Kocks, C., le Noble, F., & Rajewsky, N. (2013). Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature*, *495*(7441), 333–338.

Millan, E. Z., & McNally, G. P. (2011). Accumbens shell AMPA receptors mediate expression of extinguished reward seeking through interactions with basolateral amygdala. *Learning and Memory*, *18*(7), 414–421.

Miller, N. S., & Gold, M. S. (1994). Dissociation of "Conscious Desire" (Craving) from and Relapse in Alcohol and Cocaine Dependence. *Annals of Clinical Psychiatry: Official Journal of the American Academy of Clinical Psychiatrists*, *6*(2), 99–106.

Moreau, M. P., Bruse, S. E., David-Rus, R., Buyske, S., & Brzustowicz, L. M. (2011). Altered MicroRNA expression profiles in postmortem brain samples from individuals with schizophrenia and bipolar disorder. *Biological Psychiatry*, *69*(2), 188–193.

Mukamel, E. A. (2021). Multiple Comparisons and Inappropriate Statistical Testing Lead to Spurious Sex Differences in Gene Expression. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2021.06.026

Murphy, S. J., Lusardi, T. A., Phillips, J. I., & Saugstad, J. A. (2014). Sex differences in microRNA expression during developmentin rat cortex. *Neurochemistry International*, *77*, 24–32.

Nakano, K., & Vousden, K. H. (2001). PUMA, a novel proapoptotic gene, is induced by p53. *Molecular Cell*, *7*(3), 683–694.

Namba, M. D., Leyrer-Jackson, J. M., Nagy, E. K., Olive, M. F., & Neisewander, J. L. (2021). Neuroimmune Mechanisms as Novel Treatment Targets for Substance Use Disorders and Associated Comorbidities. *Frontiers in Neuroscience*, *15*, 650785.

Namba, M. D., Tomek, S. E., Olive, M. F., Beckmann, J. S., & Gipson, C. D. (2018). The Winding Road to Relapse: Forging a New Understanding of Cue-Induced Reinstatement Models and Their Associated Neural Mechanisms. *Frontiers in Behavioral Neuroscience*, *12*, 17.

National Drug Intelligence Center. (2011). *The economic impact of illicit drug use on American society*. Department of Justice.

National Institute on Drug Abuse. (2017). *Costs of Substance Abuse*. https://archives.drugabuse.gov/trends-statistics/costs-substance-abuse

National Institute on Drug Abuse. (2018). *The Science of Drug Use and Addiction: The Basics | National Institute on Drug Abuse (NIDA)*. https://www.drugabuse.gov/publications/media-guide/science-drug-use-addiction-basics

National Institute on Drug Abuse. (2020). *DrugFacts: Nationwide Trends*. https://www.drugabuse.gov/publications/drugfacts/nationwide-trends

National Institute on Drug Abuse. (2022, January 20). *Overdose Death Rates*. National Institute on Drug Abuse. https://nida.nih.gov/drug-topics/trends-statistics/overdose-death-rates

National Institutes of Health. (2019, February 27). *Types of Grant Programs*. https://grants.nih.gov/grants/funding/funding_program.htm

Neisewander, J. L., Baker, D. A., Fuchs, R. a., Tran-Nguyen, L. T., Palmer, A., & Marshall, J. F. (2000). Fos protein expression and cocaine-seeking behavior in rats after exposure to a cocaine self-administration environment. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *20*(2), 798–805.

Neisewander, J. L., Cheung, T. H. C., & Pentkowski, N. S. (2014). Dopamine D3 and 5-HT1B receptor dysregulation as a result of psychostimulant intake and forced abstinence: Implications for medications development. *Neuropharmacology*, *76 Pt B*, 301–319.

Nguyen, T., & Buxton, J. A. (2021). Pathways between COVID-19 public health responses and increasing overdose risks: A rapid review and conceptual framework. *The International Journal on Drug Policy*, *93*, 103236.

Nigro, J. M., Cho, K. R., Fearon, E. R., Kern, S. E., Ruppert, J. M., Oliner, J. D., Kinzler, K. W., & Vogelstein, B. (1991). Scrambled exons. *Cell*, *64*(3), 607–613.

Noble, W. S. (2009). How does multiple testing correction work? *Nature Biotechnology*, *27*(12), 1135–1137.

Nowakowski, T. J., Rani, N., Golkaram, M., Zhou, H. R., Alvarado, B., Huch, K., West, J. A., Leyrat, A., Pollen, A. A., Kriegstein, A. R., Petzold, L. R., & Kosik, K. S. (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nature Neuroscience*, *21*(12), 1784–1792.

O'Carroll, D., & Schaefer, A. (2013). General principals of miRNA biogenesis and regulation in the brain. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *38*(1), 39–54.

Ong, Z. Y., & McNally, G. P. (2020). CART in energy balance and drug addiction: Current insights and mechanisms. *Brain Research*, *1740*, 146852.

Pamudurti, N. R., Bartok, O., Jens, M., Ashwal-Fluss, R., Stottmeister, C., Ruhe, L., Hanan, M., Wyler, E., Perez-Hernandez, D., Ramberger, E., Shenzis, S., Samson, M., Dittmar, G., Landthaler, M., Chekulaeva, M., Rajewsky, N., & Kadener, S. (2017). Translation of CircRNAs. *Molecular Cell*, *66*(1), 9-21.e7.

Panksepp, J., Knutson, B., & Burgdorf, J. (2002). The role of brain emotional systems in addictions: a neuro-evolutionary perspective and new "self-report" animal model. *Addiction* , *97*(4), 459–469.

Panlilio, L. V., & Goldberg, S. R. (2007). Self-administration of drugs in animals and humans as a model and an investigative tool. *Addiction* , *102*(12), 1863–1870.

Parvaz, M. A., Moeller, S. J., & Goldstein, R. Z. (2016). Incubation of Cue-Induced Craving in Adults Addicted to Cocaine Measured by Electroencephalography. *JAMA Psychiatry* , *73*(11), 1127–1134.

Paykel, E. S., Brugha, T., & Fryers, T. (2005). Size and burden of depressive disorders in Europe. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *15*(4), 411–423.

Pentkowski, N. S., Duke, F. D., Weber, S. M., Pockros, L. A., Teer, A. P., Hamilton, E. C., Thiel, K. J., & Neisewander, J. L. (2010). Stimulation of Medial Prefrontal Cortex Serotonin 2C (5-HT2C) Receptors Attenuates Cocaine-Seeking Behavior. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *35*(10), 2037–2048.

Plotnikova, O., Baranova, A., & Skoblov, M. (2019). Comprehensive analysis of human microRNA–mRNA interactome. *Frontiers in Genetics*, *10*, 933.

Pomara, C., Cassano, T., D'Errico, S., Bello, S., Romano, A. D., Riezzo, I., Serviddio, G., & Riezzo, I. (2012). Data available on the extent of cocaine use and dependence: biochemistry, pharmacologic effects and global burden of disease of cocaine abusers. *Current Medicinal Chemistry*, *19*(33), 5647–5657.

Poplawski, A., & Binder, H. (2018). Feasibility of sample size calculation for RNA-seq studies. *Briefings in Bioinformatics*, *19*(4), 713–720.

Powell, G. L., Bonadonna, J. P., Vannan, A., Xu, K., Mach, R. H., Luedtke, R. R., & Neisewander, J. L. (2018). Dopamine D3 receptor partial agonist LS-3-134 attenuates cocaine-motivated behaviors. *Pharmacology, Biochemistry, and Behavior*, *175*, 123–129.

Powell, G. L., Leyrer-Jackson, J. M., Goenaga, J., Namba, M. D., Piña, J., Spencer, S., Stankeviciute, N., Schwartz, D., Allen, N. P., Del Franco, A. P., McClure, E. A., Olive, M. F., & Gipson, C. D. (2019). Chronic treatment with N-acetylcysteine decreases extinction responding and reduces cue-induced nicotine-seeking. *Physiological Reports*, *7*(1), e13958.

Powell, G. L., Namba, M. D., Vannan, A., Bonadonna, J. P., Carlson, A., Mendoza, R., Chen, P.-J., Luetdke, R. R., Blass, B. E., & Neisewander, J. L. (2020). The Long-Acting D3 Partial Agonist MC-25-41 Attenuates Motivation for Cocaine in Sprague-Dawley Rats. *Biomolecules*, *10*(7). https://doi.org/10.3390/biom10071076

Powell, G. L., Vannan, A., Bastle, R. M., Wilson, M. A., Dell'Orco, M., Perrone-Bizzozero, N. I., & Neisewander, J. L. (2020). Environmental enrichment during forced abstinence from cocaine self-administration opposes gene network expression changes associated with the incubation effect. *Scientific Reports*, *10*(1), 11291.

Prom-Wormley, E. C., Ebejer, J., Dick, D. M., & Bowers, M. S. (2017). The genetic epidemiology of substance use disorder: A review. *Drug and Alcohol Dependence*, *180*, 241–259.

QIAGEN. (n.d.). *Confidence Level Filter - microRNA*. Retrieved January 30, 2022, from https://qiagen.secure.force.com/KnowledgeBase/articles/Knowledge/Confidence-level-filter

Qu, S., Yang, X., Li, X., Wang, J., Gao, Y., Shang, R., Sun, W., Dou, K., & Li, H. (2015). Circular RNA: A new star of noncoding RNAs. *Cancer Letters*, *365*(2), 141–148.

R Core Team. (2014). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Rao, A., & Steward, O. (1991). Evidence that protein constituents of postsynaptic membrane specializations are locally synthesized: analysis of proteins synthesized within synaptosomes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *11*(9), 2881–2895.

Ravanidis, S., Bougea, A., Papagiannakis, N., Koros, C., Simitsi, A. M., Pachi, I., Breza, M., Stefanis, L., & Doxakis, E. (2020). Validation of differentially expressed brain-enriched microRNAs in the plasma of PD patients. *Annals of Clinical and Translational Neurology*, *7*(9), 1594–1607.

Ravanidis, S., Bougea, A., Papagiannakis, N., Maniati, M., Koros, C., Simitsi, A.-M., Bozi, M., Pachi, I., Stamelou, M., Paraskevas, G. P., Kapaki, E., Moraitou, M., Michelakakis, H., Stefanis, L., & Doxakis, E. (2020). Circulating Brain-enriched MicroRNAs for detection and discrimination of idiopathic and genetic Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, *35*(3), 457–467.

Raver-Shapira, N., Marciano, E., Meiri, E., Spector, Y., Rosenfeld, N., Moskovits, N., Bentwich, Z., & Oren, M. (2007). Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Molecular Cell*, *26*(5), 731–743.

Richardson, N. R., & Roberts, D. C. S. (1991). Fluoxetine pretreatment reduces breaking points on a progressive ratio schedule reinforced by intravenous cocaine self-administration in the rat. *Life Sciences*, *49*(11), 833–840.

Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., Lin, D. Y., Duan, J., Ophoff, R. A., Andreassen, O. A., Scolnick, E., Cichon, S., St. Clair, D., Corvin, A., Gurling, H., Werge, T., Rujescu, D., Blackwood, D. H. R., Pato, C. N., … Gejman, P. V. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nature Genetics*, *43*(10), 969–978.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47–e47.

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* , *26*(1), 139–140.

Robinson, M. D., & Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*, *11*(3), 1–9.

Robinson, T. E., & Berridge, K. C. (1993). The neural basis of drug craving: an incentive-sensitization theory of addiction. *Brain Research. Brain Research Reviews*, *18*(3), 247–291.

Rokavec, M., Li, H., Jiang, L., & Hermeking, H. (2014). The p53/miR-34 axis in development and disease. *Journal of Molecular Cell Biology*, *6*(3), 214–230.

Roush, S., & Slack, F. J. (2008). The let-7 family of microRNAs. In *Trends in Cell Biology* (Vol. 18, Issue 10, pp. 505–516). https://doi.org/10.1016/j.tcb.2008.07.007

Russo, S. J., Dietz, D. M., Dumitriu, D., Morrison, J. H., Malenka, R. C., & Nestler, E. J. (2010). The addicted synapse: mechanisms of synaptic and structural plasticity in nucleus accumbens. *Trends in Neurosciences*, *33*(6), 267–276.

Rybak-Wolf, A., Stottmeister, C., Glažar, P., Jens, M., Pino, N., Giusti, S., Hanan, M., Behm, M., Bartok, O., Ashwal-Fluss, R., Herzog, M., Schreyer, L., Papavasileiou, P., Ivanov, A., Öhman, M., Refojo, D., Kadener, S., & Rajewsky, N. (2015). Circular RNAs in the Mammalian Brain Are Highly Abundant, Conserved, and Dynamically Expressed. *Molecular Cell*, *58*(5), 870–885.

Sabatini, M. J., Ebert, P., Lewis, D. A., Levitt, P., Cameron, J. L., & Mirnics, K. (2007). Amygdala gene expression correlates of social behavior in monkeys experiencing maternal separation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(12), 3295–3304.

Sadakierska-Chudy, A., Frankowska, M., Miszkiel, J., Wydra, K., Jastrzębska, J., & Filip, M. (2017). Prolonged induction of miR-212/132 and REST expression in rat striatum following cocaine self-administration. *Molecular Neurobiology*, *54*(3), 2241–2254.

Saino, M., Maruyama, T., Sekiya, T., Kayama, T., & Murakami, Y. (2004). Inhibition of angiogenesis in human glioma cell lines by antisense RNA from the soluble guanylate cyclase genes, GUCY1A3 and GUCY1B3. *Oncology Reports*, *12*(1), 47–52.

Salmena, L., Poliseno, L., Tay, Y., Kats, L., & Pandolfi, P. P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell*, *146*(3), 353–358.

Salzman, J., Chen, R. E., Olsen, M. N., Wang, P. L., & Brown, P. O. (2013). Cell-type specific features of circular RNA expression. *PLoS Genetics*, *9*(9), e1003777.

Sanchis-Segura, C., & Spanagel, R. (2006). Behavioural assessment of drug reinforcement and addictive features in rodents: an overview. *Addiction Biology*, *11*(1), 2–38.

Sanger, H. L., Klotz, G., Riesner, D., Gross, H. J., & Kleinschmidt, A. K. (1976). Viroids are single-stranded covalently closed circular RNA molecules existing as highly base-paired rod-like structures. *Proceedings of the National Academy of Sciences of the United States of America*, *73*(11), 3852–3856.

Santarelli, D. M., Beveridge, N. J., Tooney, P. A., & Cairns, M. J. (2011). Upregulation of dicer and MicroRNA expression in the dorsolateral prefrontal cortex Brodmann area 46 in schizophrenia. *Biological Psychiatry*, *69*(2), 180–187.

Scaplen, K. M., & Kaun, K. R. (2016). Reward from bugs to bipeds: a comparative approach to understanding how reward circuits function. *Journal of Neurogenetics*, *30*(2), 133–148.

Schaefer, A., Im, H. I., Venø, M. T., Fowler, C. D., Min, A., Intrator, A., Kjems, J., Kenny, P. J., O'Carroll, D., & Greengard, P. (2010). Argonaute 2 in dopamine 2 receptor-expressing neurons regulates cocaine addiction. *The Journal of Experimental Medicine*, *207*(9), 1843–1851.

Schena, M., Shalon, D., Davis, R. W., & Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, *270*(5235), 467–470.

Schirle, N. T., Sheu-Gruttadauria, J., & MacRae, I. J. (2014). Structural basis for microRNA targeting. *Science*, *346*(6209), 608–613.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., Blaxter, M., & Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA* , *22*(6), 839–851.

Shalon, D., Smith, S. J., & Brown, P. O. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Research*, *6*(7), 639–645.

Shin, V. Y., Jin, H., Ng, E. K., Ng, E. K., Cheng, A. S., Chong, W. W., Wong, C. Y., Leung, W. K., Sung, J. J., & Chu, K. M. (2010). NF-κB targets miR-16 and miR-21 in gastric cancer: involvement of prostaglandin E receptors. *Carcinogenesis*, *32*(3), 240–245.

Simoneau, J., Dumontier, S., Gosselin, R., & Scott, M. S. (2021). Current RNA-seq methodology reporting limits reproducibility. *Briefings in Bioinformatics*, *22*(1), 140–145.

Sinha, R. (2013). The clinical neurobiology of drug craving. *Current Opinion in Neurobiology*, *23*(4), 649–654.

Sinha, R., Catapano, D., & O'Malley, S. (1999). Stress-induced craving and stress response in cocaine dependent individuals. *Psychopharmacology*, *142*(4), 343–351.

Smith, A. C. W., & Kenny, P. J. (2018). MicroRNAs regulate synaptic plasticity underlying drug addiction. *Genes, Brain, and Behavior*, *17*(3), e12424.

Smith, Alexander C. W., Scofield, M. D., Heinsbroek, J. A., Gipson, C. D., Neuhofer, D., Roberts-Wolfe, D. J., Spencer, S., Garcia-Keller, C., Stankeviciute, N. M., Smith, R. J., Allen, N. P., Lorang, M. R., Griffin, W. C., 3rd, Boger, H. A., & Kalivas, P. W. (2017). Accumbens nNOS Interneurons Regulate Cocaine Relapse. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *37*(4), 742–756.

Smith, M. A. (2020). Nonhuman animal models of substance use disorders: Translational value and utility to basic science. *Drug and Alcohol Dependence*, *206*, 107733.

Smrt, R. D., Szulwach, K. E., Pfeiffer, R. L., Li, X., Guo, W., Pathania, M., Teng, Z. Q., Luo, Y., Peng, J., Bordey, A., Jin, P., & Zhao, X. (2010). MicroRNA miR-137 regulates neuronal maturation by targeting ubiquitin ligase mind bomb-1. *Stem Cells* , *28*(6), 1060–1070.

Soyka, M., & Müller, C. A. (2017). Pharmacotherapy of alcoholism - an update on approved and off-label medications. *Expert Opinion on Pharmacotherapy*, *18*(12), 1187–1199.

Spanagel, R. (2017). Animal models of addiction. *Dialogues in Clinical Neuroscience*, *19*(3), 247–258.

Stolarczyk, M., Reuter, V. P., Smith, J. P., Magee, N. E., & Sheffield, N. C. (2020). Refgenie: a reference genome resource manager. *GigaScience*, *9*(2). https://doi.org/10.1093/gigascience/giz149

Su, G., Morris, J. H., Demchak, B., & Bader, G. D. (2014). Biological network exploration with Cytoscape 3. *Current Protocols in Bioinformatics / Editoral Board, Andreas D. Baxevanis ... [et Al.]*, *47*, 8.13.1-24.

Suzuki, H., Zuo, Y., Wang, J., Zhang, M. Q., Malhotra, A., & Mayeda, A. (2006). Characterization of RNase R-digested cellular RNA source that consists of lariat and circular RNAs from pre-mRNA splicing. *Nucleic Acids Research*, *34*(8), e63.

Swanson, C. J., Baker, D. A., Carson, D., Worley, P. F., & Kalivas, P. W. (2001). *Repeated Cocaine Administration Attenuates Group I Metabotropic Glutamate Receptor-Mediated Glutamate Release and Behavioral Activation: A Potential Role for Homer*. http://www.jneurosci.org/content/jneuro/21/22/9043.full.pdf

Szumlinski, K. K., Ary, A. W., & Lominac, K. D. (2008). Homers regulate drug-induced neuroplasticity: Implications for addiction. *Biochemical Pharmacology*, *75*(1), 112–133.

Szumlinski, K. K., Dehoff, M. H., Kang, S. H., Frys, K. A., Lominac, K. D., Klugmann, M., Rohrer, J., Griffin, W., Toda, S., Champtiaux, N. P., Berry, T., Tu, J. C., Shealy, S. E., During, M. J., Middaugh, L. D., Worley, P. F., & Kalivas, P. W. (2004). Homer proteins regulate sensitivity to cocaine. *Neuron*, *43*(3), 401–413.

Szumlinski, K. K., Kalivas, P. W., & Worley, P. F. (2006). Homer proteins: implications for neuropsychiatric disorders. *Current Opinion in Neurobiology*, *16*(3), 251–257.

Tekinay, A. B., Nong, Y., Miwa, J. M., Lieberam, I., Ibanez-Tallon, I., Greengard, P., & Heintz, N. (2009). A role for LYNX2 in anxiety-related behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4477–4482.

The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338.

Thiel, K. J., Engelhardt, B., Hood, L. E., Peartree, N. A., & Neisewander, J. L. (2011). The interactive effects of environmental enrichment and extinction interventions in attenuating cue-elicited cocaine-seeking behavior in rats. *Pharmacology, Biochemistry, and Behavior*, *97*(3), 595–602.

Thiel, K. J., Painter, M. R., Pentkowski, N. S., Mitroi, D., Crawford, C. A., & Neisewander, J. L. (2012). Environmental enrichment counters cocaine abstinence-induced stress and brain reactivity to cocaine cues but fails to prevent the incubation effect. *Addiction Biology*, *17*(2), 365–377.

Thiel, K. J., Pentkowski, N. S., Peartree, N. A., Painter, M. R., & Neisewander, J. L. (2010). Environmental living conditions introduced during forced abstinence alter cocaine-seeking behavior and Fos protein expression. *Neuroscience*, *171*(4), 1187–1196.

Thiel, K. J., Sanabria, F., Pentkowski, N. S., & Neisewander, J. L. (2009). *Anti-craving effects of environmental enrichment*. https://doi.org/10.1017/S1461145709990472

Tran-Nguyen, L. T., Fuchs, R. A., Coffey, G. P., Baker, D. A., O'Dell, L. E., & Neisewander, J. L. (1998). Time-dependent changes in cocaine-seeking behavior and extracellular dopamine levels in the amygdala during cocaine withdrawal. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *19*(1), 48–59.

Tsetsos, F., Padmanabhuni, S. S., Alexander, J., Karagiannidis, I., Tsifintaris, M., Topaloudi, A., Mantzaris, D., Georgitsi, M., Drineas, P., & Paschou, P. (2016). Meta-Analysis of Tourette Syndrome and Attention Deficit Hyperactivity Disorder Provides Support for a Shared Genetic Basis. *Frontiers in Neuroscience*, *10*, 340.

Uhl, G. R., Drgon, T., Johnson, C., Li, C.-Y., Contoreggi, C., Hess, J., Naiman, D., & Liu, Q.-R. (2008). Molecular genetics of addiction and related heritable phenotypes: genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Annals of the New York Academy of Sciences*, *1141*, 318–381.

Vannan, A., Powell, G. L., Dell'Orco, M., Wilson, M. A., Perrone-Bizzozero, N. I., & Neisewander, J. L. (2021). microRNA regulation related to the protective effects of environmental enrichment against cocaine-seeking behavior. *Drug and Alcohol Dependence*, *221*, 108585.

Vannan, A., Powell, G. L., Scott, S. N., Pagni, B. A., & Neisewander, J. L. (2018). Animal models of the impact of social stress on cocaine use disorders. In M. F. Olive & S. E. Tomek (Eds.), *Animal models for examining social influences on drug addiction*. (Vol. 140, pp. 131–169). International Review of Neurobiology, Academic Press.

Veilleux, J. C., Colvin, P. J., Anderson, J., York, C., & Heinz, A. J. (2010). A review of opioid dependence treatment: pharmacological and psychosocial interventions to treat opioid addiction. *Clinical Psychology Review*, *30*(2), 155–166.

Venniro, M., Banks, M. L., Heilig, M., Epstein, D. H., & Shaham, Y. (2020). Improving translation of animal models of addiction and relapse by reverse translation. *Nature Reviews. Neuroscience*, *21*(11), 625–643.

Viola, T. W., Wearick-Silva, L. E., De Azeredo, L. A., Centeno-Silva, A., Murphy, C., Marshall, P., Li, X., Singewald, N., Garcia, F., Bredy, T. W., & Grassi-Oliveira, R. (2016). Increased cocaine-induced conditioned place preference during periadolescence in maternally separated male BALB/c mice: the role of cortical BDNF, microRNA-212, and MeCP2. *Psychopharmacology*, *233*(17), 3279–3288.

Vleeming, W., Rambali, B., & Opperhuizen, A. (2002). The role of nitric oxide in cigarette smoking and nicotine addiction. *Nicotine & Tobacco Research: Official Journal of the Society for Research on Nicotine and Tobacco*, *4*(3), 341–348.

Wahid, F., Shehzad, A., Khan, T., & Kim, Y. Y. (2010). MicroRNAs: Synthesis, mechanism, function, and recent clinical trials. In *Biochimica et Biophysica Acta - Molecular Cell Research* (Vol. 1803, Issue 11, pp. 1231–1243). https://doi.org/10.1016/j.bbamcr.2010.06.013

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., … Nestler, E. J. (2018). Cocaine Self-administration Alters Transcriptome-wide Responses in the Brain's Reward Circuitry. *Biological Psychiatry*, *84*(12), 867–880.

Walker, D. M., Cunningham, A. M., & Nestler, E. J. (2021). Reply to: Multiple Comparisons and Inappropriate Statistical Testing Lead to Spurious Sex Differences in Gene Expression. *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2021.08.006

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews. Genetics*, *10*(1), 57–63.

Weiss, F. (2005). Neurobiology of craving, conditioned reward and relapse. *Current Opinion in Pharmacology*, *5*(1), 9–19.

Weiss, F., Martin-Fardon, R., Ciccocioppo, R., Kerr, T. M., Smith, D. L., & Ben-Shahar, O. (2001). Enduring resistance to extinction of cocaine-seeking behavior induced by drug-related cues. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *25*(3), 361–372.

Wilkinson, M. G., Pino, T. S., Tournier, S., Buck, V., Martin, H., Christiansen, J., Wilkinson, D. G., & Millar, J. B. (1999). Sin1: an evolutionarily conserved component of the eukaryotic SAPK pathway. *The EMBO Journal*, *18*(15), 4210–4221.

Wolf, M. E. (2016). Synaptic mechanisms underlying persistent cocaine craving. *Nature Reviews. Neuroscience*, *17*(6), 351–365.

Wright, C., Turner, J. A., Calhoun, V. D., & Perrone-Bizzozero, N. (2013). Potential impact of miR-137 and its targets in schizophrenia. *Frontiers in Genetics*, *4*. https://doi.org/10.3389/fgene.2013.00058

Wu, G., & Haw, R. (2017). Functional Interaction Network Construction and Analysis for Disease Discovery. *Methods in Molecular Biology* , *1558*, 235–253.

Wu, J., Bao, J., Kim, M., Yuan, S., Tang, C., Zheng, H., Mastick, G. S., Xu, C., & Yan, W. (2014). Two miRNA clusters, miR-34b/c and miR-449, are essential for normal brain development, motile ciliogenesis, and spermatogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(28), E2851-7.

Wu, Z., & Wu, H. (2016). Experimental Design and Power Calculation for RNA-seq Experiments. *Methods in Molecular Biology* , *1418*, 379–390.

Yamakuchi, M., Ferlito, M., & Lowenstein, C. J. (2008). miR-34a repression of SIRT1 regulates apoptosis. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(36), 13421–13426.

Yang, Z., & Bielawski, J. P. (2000). Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution*, *15*(12), 496–503.

You, X., Vlatkovic, I., Babic, A., Will, T., Epstein, I., Tushev, G., Akbalik, G., Wang, M., Glock, C., Quedenau, C., Wang, X., Hou, J., Liu, H., Sun, W., Sambandan, S., Chen, T., Schuman, E. M., & Chen, W. (2015). Neural circular RNAs are derived from synaptic genes and regulated by development and plasticity. *Nature Neuroscience*, *18*(4), 603–610.

Zang, J., Lu, D., & Xu, A. (2020). The interaction of circRNAs and RNA binding proteins: An important part of circRNA maintenance and function. *Journal of Neuroscience Research*, *98*(1), 87–97.

Zeng, X., Lin, W., Guo, M., & Zou, Q. (2017). A comprehensive overview and evaluation of circular RNA detection tools. *PLoS Computational Biology*, *13*(6), e1005420.

Zhang, H.-Y., Bi, G.-H., Li, X., Li, J., Qu, H., Zhang, S.-J., Li, C.-Y., Onaivi, E. S., Gardner, E. L., Xi, Z.-X., & Liu, Q.-R. (2015). Species differences in cannabinoid receptor 2 and receptor responses to cocaine self-administration in mice and rats. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, *40*(4), 1037–1051.

Zhang, X.-O., Dong, R., Zhang, Y., Zhang, J.-L., Luo, Z., Zhang, J., Chen, L.-L., & Yang, L. (2016). Diverse alternative back-splicing and alternative splicing landscape of circular RNAs. *Genome Research*, *26*(9), 1277–1287.

Zhang, Y., Zhang, X.-O., Chen, T., Xiang, J.-F., Yin, Q.-F., Xing, Y.-H., Zhu, S., Yang, L., & Chen, L.-L. (2013). Circular intronic long noncoding RNAs. *Molecular Cell*, *51*(6), 792–806.

Zimmerman, A. J., Hafez, A. K., Amoah, S. K., Rodriguez, B. A., Dell'Orco, M., Lozano, E., Hartley, B. J., Alural, B., Lalonde, J., Chander, P., Webster, M. J., Perlis, R. H., Brennand, K. J., Haggarty, S. J., Weick, J., Perrone-Bizzozero, N., Brigman, J.

L., & Mellios, N. (2020). A psychiatric disease-related circular RNA controls synaptic gene expression and cognition. *Molecular Psychiatry*, *25*(11), 2712–2727.

# APPENDIX A

## CHAPTER 3 GENE AND TISSUE ABBREVIATIONS

**Gene abbreviations**
Gene abbreviations are listed for human orthologs.

*Shared "Craving" genes*

| | |
|---|---|
| AGK | Acylglycerol Kinase |
| AMZ1 | Archaelysin Family Metallopeptidase 1 |
| B2M | Beta-2-Microglobulin |
| BCAS1 | Brain Enriched Myelin Associated Protein 1 |
| BTG1 | B-Cell Translocation Gene (BTG) Anti-Proliferation Factor 1 |
| CACYBP | Calcyclin Binding Protein |
| CARTPT | Cocaine- And Amphetamine Regulated Transcript (CART) Prepropeptide |
| CCDC88C | Coiled-Coil Domain Containing 88C |
| EGR2 | Early Growth Response 2 |
| FABP7 | Fatty Acid Binding Protein 7 |
| FKBP4 | FKBP Prolyl Isomerase 4 |
| FTH1 | Ferritin Heavy Chain 1 |
| GPD1 | Glycerol-3-Phosphate Dehydrogenase 1 |
| GUCY1A3 | Guanylate Cyclase 1 Soluble Subunit Alpha 1 |
| HAPLN2 | Hyaluronan And Proteoglycan Link Protein 2 |
| HSPA8 | Heat Shock Protein Family A (Hsp70) Member 8 |
| IRS2 | Insulin Receptor Substrate 2 |
| KIF5A | Kinesin Family Member 5A |
| LYPD1 | LY6/PLAUR Domain Containing 1 |
| MBP | Myelin Basic Protein |
| MOBP | Myelin Associated Oligodendrocyte Basic Protein |
| NTS | Neurotensin |
| PHLDA1 | Pleckstrin Homology Like Domain Family A Member 1 |
| PITPNM3 | Phosphatidylinositol Transfer Protein, Membrane Associated (PITPNM) Family Member 3 |
| RGS5 | Regulator Of G Protein Signaling 5 |
| RPS6KA2 | Ribosomal Protein S6 Kinase A2 |
| SOX17 | Sex Determining Region Y (SRY)-Box Transcription Factor 17 |
| TIPARP | TCDD Inducible Poly(ADP-Ribose) Polymerase |
| TTLL1 | Tubulin Tyrosine Ligase (TTL) Family Tubulin Polyglutamylase Complex Subunit L1 |
| USP46 | Ubiquitin Specific Peptidase 46 |
| VIM | Vimentin |

*Other genes*

| | |
|---|---|
| ARC | Activity Regulated Cytoskeleton Associated Protein |
| BDNF | Brain Derived Neurotrophic Factor |
| C1QL2 | Complement C1q Like 2 |
| CART | Cocaine- and Amphetamine-Related Transcript |
| CREB1 | cAMP Responsive Element Binding Protein 1 |
| DRD3 | Dopamine receptor D3 |
| FAM53B | Family With Sequence Similarity 53 Member B |

**Tissue abbreviations**
From Gene-Tissue Expression (GTEx) V8 database. Tissues with the prefix "BRN" (brain) are located in the central nervous system (CNS).

| | |
|---|---|
| ADP-SBC | Adipose – subcutaneous |
| ADP-VSO | Adipose - visceral omentum |
| ADRNLGL | Adrenal gland |
| ART-CRN | Artery – coronary |
| ART-TB | Artery – tibial |
| BLADDER | Bladder |
| BRN-AMY | Amygdala |
| BRN-ACC | Anterior cingulate cortex (BA24) |
| BRN-CAU | Caudate, basal ganglia |
| BRN-CB-a | Cerebellar hemisphere |
| BRN-CB-b | Cerebellum |
| BRN-CTX-a | Frontal cortex (BA9) |
| BRN-CTX-b | Cortex |
| BRN-HIPP | Hippocampus |
| BRN-HYP | Hypothalamus |
| BRN-NAC | Nucleus accumbens, basal ganglia |
| BRN-PTRY | Pituitary gland |
| BRN-PUT | Putamen, basal ganglia |
| BRN-SN | Substantia nigra |
| BRN-SPC | Spinal cord (cervical C1) |
| BREAST | Breast |
| CELL-FB | Cells - cultured fibroblasts |
| CELL-LYM | Cells - EBV-transformed lymphocytes |
| CVX-ECT | Cervix - ectocervix |
| CVX-END | Cervix - endocervix |
| CLN-SIG | Colon - sigmoid |
| CLN-TRN | Colon - transverse |
| ESP-GEJ | Esophagus - gastroesophageal junction |
| ESP-MCS | Esophagus mucosa |
| ESP-MSL | Esophagus mucularis |
| FLPTB | Fallopian tube |
| HRT-AA | Heart - atrial appendage |
| HRT-LV | Heart - left ventricle |
| KDY-CTX | Kidney - cortex |
| KDY-MDL | Kidney - medulla |
| LIVER | Liver |
| LUNG | Lung |
| SALGL | Salivary gland |
| MSC-SK | Skeletal muscle |
| NRV-TB | Nerve - tibial |
| OVARY | Ovary |
| PANCREAS | Pancreas |

| | |
|---|---|
| PROSTATE | Prostate |
| SKN-NSP | Skin - not sun exposed, subrapubic |
| SKN-SLL | Skin - sun exposed, lower leg |
| SIN-TIL | Small intestine - terminal ileum |
| SPLEEN | Spleen |
| STOMACH | Stomach |
| TESTIS | Testis |
| THYROID | Thyroid |
| UTERUS | Uterus |
| VAGINA | Vagina |
| WHLBLD | Whole blood |

# APPENDIX B

## PERMISSION FROM CO-AUTHORS

**Chapter 2. microRNA Regulation Related to the Protective Effects of Environmental Enrichment Against Cocaine-Seeking Behavior.**

This work was previously published (Vannan et al., 2021). Co-first author Dr. Gregory L. Powell, co-authors Drs. Michela Dell'Orco and Melissa A. Wilson, and co-senior authors Drs. Nora I. Perrone-Bizzozero and Janet L. Neisewander have all consented to inclusion of the publication in this dissertation.

**Chapter 3. An Approach for Prioritizing Candidate Genes from RNA-seq Using Preclinical Cocaine Craving Datasets as a Test Case.**

This chapter has been submitted for publication. It is co-authored by Drs. Michela Dell'Orco, Nora I. Perrone-Bizzozero, and Janet L. Neisewander, with Dr. Melissa A. Wilson as senior author, who have all consented to use of this work in the dissertation.

**Chapter 4. Alterations in Circular RNA Expression in the Rat Nucleus Accumbens Shell May Regulate circRNA-miRNA-mRNA Networks in a Model of Cocaine Craving.**

This work has not yet been submitted for publication. The planned authors, including Michael C. Johnson as co-first author, Drs. Michela Dell'Orco, Gregory L. Powell, Nikolaos Mellios, Melissa A. Wilson, and Nora I. Perrone-Bizzozero as co-authors, and Dr. Janet L. Neisewander as senior author, have consented to use of this work in the dissertation.