UNIVERSITY OF CALIFORNIA SAN DIEGO

Genome-wide Association Discovery, Replication, and Validation in Mouse Models

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Biomedical Sciences

by

Xinzhu Zhou

Committee in charge:

Professor Abraham A. Palmer, Chair
Professor Kelly A. Frazer
Professor Ralph J. Greenspan
Professor John Kelsoe
Professor Francesca Telese

2020

The Dissertation of Xinzhu Zhou is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2020

DEDICATION

To family and friends with whom I have shared my excitement, frustration, discovery, and strain during this momentous passage.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL FILES

- Zhou_Chapter1_FigureS1.pdf
- Zhou_Chapter1_FigureS2.pdf
- Zhou_Chapter1_FigureS3.pdf
- Zhou_Chapter1_FigureS4.pdf
- Zhou_Chapter1_FigureS5.pdf
- Zhou_Chapter1_FigureS6.pdf
- Zhou_Chapter1_FigureS7.pdf
- Zhou_Chapter1_FigureS8.pdf
- Zhou_Chapter1_FileS1.txt
- Zhou_Chapter1_TableS1.csv
- Zhou_Chapter1_TableS2.csv
- Zhou_Chapter1_TableS3.csv
- Zhou_Chapter1_TableS4.csv
- Zhou_Chapter1_TableS5.csv
- Zhou_Chapter1_TableS6.csv
- Zhou_Chapter1_TableS7.csv
- Zhou_Chapter1_TableS8.csv
- Zhou_Chapter1_TableS9.csv
- Zhou_Chapter1_TableS10.csv
- Zhou_Chapter1_TableS11.csv
- Zhou_Chapter1_TableS12.csv

- Zhou_Chapter1_TableS13.csv
- Zhou_Chapter2_FigureS1.pdf
- Zhou_Chapter2_FigureS2.pdf
- Zhou_Chapter2_FigureS3.pdf
- Zhou_Chapter2_FigureS4.pdf
- Zhou_Chapter2_FigureS5.pdf
- Zhou_Chapter2_FigureS6.pdf
- Zhou_Chapter2_FigureS7.pdf
- Zhou_Chapter2_FigureS8.pdf
- Zhou_Chapter2_TableS1.csv
- Zhou_Chapter2_TableS2.csv
- Zhou_Chapter2_TableS3.csv
- Zhou_Chapter2_TableS4.csv
- Zhou_Chapter2_TableS5.csv
- Zhou_Chapter2_TableS6.csv
- Zhou_Chapter2_TableS7.csv

ACKNOWLEDGMENTS

I would like to thank my adviser Abraham Palmer for his extensive training in basic research, communication skills, scientific writing, team management and much more. Not only did Abe guided me through developing project ideas, problem shooting roadblocks, synthesizing insights, and revising manuscripts, he also exemplified a conscientious scientist, colleague, citizen, and father. I am very grateful for the hard and soft skills that I have garnered through the years under his mentorship, mostly through his living example.

I would like to thank the rest of my dissertation committee – Kelly Frazer, John Kelsoe, Ralph J. Greenspan, and Francesca Telese – for their continuing support and advice. Francesca Telese provided me unreserved advice and resource in RNA-Seq library preparations.

I would like to thank my colleagues, Celine Pierre, Natalia Gonzales, Jennifer Zou, Riyan Cheng, Apurva Chitre, and Patricia Montilla-Perez for their close collaborations and deep insights across several projects. I would like to thank all past and present members of the Palmer Lab, with special thanks to Alex Gileta, Jianjun Gao, Laura Sittig, Sandra Sanchez-Roige, Amelie Baud, Oksana Polesskaya, Amanda Barkley-Levenson, Hannah Bimschleger, and Ariana Page. Although they have not collaborated with me on my dissertation projects directly, they have assisted me in various aspects of running experiments and performing computational analysis.

I would like to thank my family for their unconditional love. My mom, Xiaojuan Pan, and my dad, Bo Zhou, have always showered me with kindness and care through video chats, text messages, and occasional international care packages. Due to complications

of the visa where I would have to leave the country for at least six weeks to renew it, I was only able to visit home once during the entirety of my PhD. If not my parents' understanding, I would not be able pursue my PhD career far away from home without feeling the guilt of neglect.

I would like to thank my boyfriend Rongxin Fang for his unwavering love and support through the daily undulations of life. I find the PhD experience much more enjoyable (and bearable at times) to share it with my best friend. As a PhD himself, he provided me a window to another discipline, community, and perspective. I am most fond of our physical explorations in and outside of San Diego and our intellectual adventures in science, art, music, history, politics, pop culture, and last but not least, the treasure trove that is the internet.

I would like to thank my close friends. Jinyi Yang and Wei Ren, my closest friends since middle school, who have listened to many of my PhD stories; Sissi Lei, Hongguo Chen, Zhuo Yang, Yifei Zhang, Weini Yu, my childhood and college close friends, who have supported me through deep conservations and food; many, many PhD friends in the Biomedical Sciences program, the Bioinformatics program, the Biology program, and Physics program at UCSD, who have enlightened me with their research findings and experiences; all of the students in the Genetics, Genomics, and Systems Biology program and the Human Genetics program at University of Chicago, with whom I have shared the first year of PhD.

I would like to thank my college mentor in science, Laura Kats, whose passion about the pursuit of truth inspired me to join a PhD program and prepared me with skills to succeed in whatever ventures I undertake in life.

Chapter 1, in full, is a reprint of the material as it appears in G3 2020. Xinzhu Zhou, Celine L. St. Pierre, Natalia M. Gonzales, Jennifer Zou, Riyan Cheng, Apurva S. Chitre, Greta Sokoloff, Abraham A. Palmer. Genome-wide Association Study in Two Cohorts from a Multi-generational Mouse Advanced Intercross Line Highlights the Difficulty of Replication Due to Study-specific Heterogeneity. G3: GENES, GENOMES, GENETICS, 2020, 10(3):951-965; https://doi.org/10.1534/g3.119.400763. The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Xinzhu Zhou, Amanda Barkley-Levenson, Patricia Montilla-Perez, Francesca Telese, Abraham A. Palmer. The dissertation author was the primary investigator and author of this material.

VITA

EDUCATION

2014 Bachelor of Arts, Smith College
Biological Sciences Psychology
Psychology

2020 Doctor of Philosophy, University of California, San Diego
Biomedical Sciences

PUBLICATIONS

Xin, T., Yao, H., Gao, H., Zhou, X., Ma, X., Xu, C., Chen, J., Han, J., Pang, X., Xu, R., Song, J., & Chen, S. (2013). Super food Lycium barbarum (Solanaceae) traceability via an internal transcribed spacer 2 barcode. *Food Research International*, *54*(2), 1699–1704. https://doi.org/10.1016/j.foodres.2013.10.007

Zhou, X., St Pierre, C. L., Gonzales, N. M., Zou, J., Cheng, R., Chitre, A. S., Sokoloff, G., & Palmer, A. A. (2020). Genome-Wide Association Study in Two Cohorts from a Multi-generational Mouse Advanced Intercross Line Highlights the Difficulty of Replication Due to Study-Specific Heterogeneity. *G3 (Bethesda, Md.)*, *10*(3), 951–965. https://doi.org/10.1534/g3.119.400763

ABSTRACT OF THE DISSERTATION

Genome-wide Association Discovery, Replication, and Validation in Mouse Models

by

Xinzhu Zhou

Doctor of Philosophy in Biomedical Sciences

University of California, San Diego, 2020

Professor Abraham Palmer, Chair

The “replication crisis” has highlighted the apparent lack of replication across scientific fields. Assessing the true effect size of a genetic variant and designing an adequately powered study remain difficult when replicating genome-wide association studies (**GWAS**) results. Functional replication studies that validate the identified loci have also been scarce. Given the importance of replication, my thesis focuses on the computational and the functional aspects of replication. In chapter 1, I explored replication in a mouse model using an advanced intercross line (**AIL**), which is a multigenerational intercross between two inbred strains. I used genotyped and re-genotyped data from two

cohorts of the LG/J x SM/J AIL mice (F_{34}, n=428; $F_{39\text{-}43}$, n=600). For the subset of traits that were measured in both cohorts (locomotor activity, body weight, and coat color), I attempted to replicate loci identified in either F_{34} or $F_{39\text{-}43}$ in the other cohort. Coat color was robustly replicated; locomotor activity and body weight were only partially replicated, which was inconsistent with our power simulations. We used a random effects model to show that the partial replications could not be explained by Winner's Curse but could be explained by study-specific heterogeneity. Despite this heterogeneity, I performed a mega-analysis by combining F_{34} and $F_{39\text{-}43}$ cohorts (n=1,028). In chapter 2, I followed up on a candidate gene, *Azi2*, which had previously been identified in a GWAS for locomotor response to methamphetamine using an outbred Carworth Farms White (**CFW**) population. To validate *Azi2* as a putative causal gene for methamphetamine sensitivity, I created an *Azi2* knockout mouse line using CRISPR/Cas9 and established the mutant phenotype of locomotor response to methamphetamine. In addition, I investigated a published hypothesis that an independent transcript that matches to the second half of the 3'UTR sequence of *Azi2* is a downregulator of *Slc6a3*, which encodes the dopamine transporter (**DAT**), in the ventral tegmental area (**VTA**) region of midbrain. To test this hypothesis, I examined the possible correlations among *Azi2*, *Azi2* 3'UTR, and *Slc6a3* expression in the *Azi2* KO line and in naïve CFW mice.

INTRODUCTION

The “replication crisis” has been a topic of extensive discussion in many scientific fields. The concern for such a crisis was substantiated in the exemplary field of psychology when the Open Science Collaboration merely yielded 36 successful replication efforts out of the 100 original publications sampled in top-tier journals (Collaboration, 2015; Shrout & Rodgers, 2018). Maxwell, Lau and Howard (2015) (Maxwell et al., 2015) argued that the apparent lack of replication is due to the myriad of complications when designing a replication study. Researchers would have to first determine the appropriate level of power for the replication study. Then they need to adopt an effect size that is generally extrapolated from the original study to estimate the sample size needed to achieve the desired level of power. However, the assumed effect size could be wildly uncertain. Sampling variability that does not represent the population variability in the original study could result in an erroneous estimate of the true effect size. Publication bias for statistically significant results could give the wrong illusion that effect sizes are larger than they are (Shrout & Rodgers, 2018). Even if the estimated effect size in the original study was true, the power analysis for the replication study usually relies on a single value of effect size, instead of a range of values that portray the possible distribution of effect sizes (*e.g.*, a confidence interval for the estimated effect size). The matter is worsened by the fact that the correlation between effect size and sample size is nonlinear; at a given level of power, a small decrease in effect size equates to a much larger increase in sample size. As Maxwell, Lau and Howard (2015) pointed out, although not being able to confirm the original results at the same statistical significance and/or with the same magnitude of effect size do not immediately indicate that the original findings are false positives, designing an adequately powered replication study remains

a challenging feat (Maxwell et al., 2015). In the field of genome-wide association studies (**GWAS**), much of the difficulty in replicating genetic effect is attributed to a phenomenon dubbed “Winner’s Curse” (Poirier et al., 2015; Xiao & Boehnke, 2009; Zöllner & Pritchard, 2007) (also known as “Beavis Effect”; (Beavis et al., 1991, 1994; Keele et al., 2019; King & Long, 2017; Paterson, 2019; Xu, 2003), in which published effect sizes of genetic variants tend to be much larger in magnitude than the subsequent replication studies. Indeed, winning once may seem to exclude the opportunity to win future replication games.

Shrout and Rodgers (2018) elaborated on the distinctions among three definitions of replication: direct replication, systematic replication, and conceptual replication (Shrout & Rodgers, 2018). Direct replication refers to a new study that employs the same set of subjects, materials and protocol as the original study. Systematic replication is similar to direct replication except a few minor changes in the auxiliary measures from the original study. Conceptual replication is an intentional diversion from the original study where the generalizability of the results could be assessed. In addition to the formal definitions of replication, which tend to refer to single replication attempts, meta-analysis is another way of replicating and refining the results. Meta-analysis compares results from multiple studies, yielding not only more accurate estimates of the distribution of effect sizes in a population but also the factors that modulate the effect sizes in subgroups (Shrout & Rodgers, 2018). GWAS as a field has adopted and benefited tremendously from meta-analysis. Complex diseases tend to be highly polygenic where the effect sizes for the risk gene variants, which tend to be rare alleles, are very small. While single GWAS for complex diseases may not be able to identify significant associations due to difficulties in

recruiting, surveying, and testing thousands and thousands of human subjects in one particular place and time, meta-analysis allows for combining multiple large-scale studies, leading to the increase of the power and the identification of the majority of the genetic risk variants for complex diseases (Evangelou & Ioannidis, 2013). Multitudes of methods and software applications that correct for heterogeneity between the studies have been developed and applied in meta-analysis efforts (as reviewed in (Begum et al., 2012)).

Mega-analysis is a special case of meta-analysis. Instead of weighing summary statistics from individual studies, mega-analysis is an approach where samples from different studies are directly combined and analyzed. Mega-analysis is generally used when there is no perceived difference between the study samples; the samples may come from the same population and the samples may have been genotyped together.

Apart from replicating the association between genetic variant and trait, evaluating whether an identified locus is functionally linked to the trait is essential. One could argue that functional validation of genetic variants is another form of replication. Positive validation would enhance the credibility of the variant-trait association markedly, at least in the biological system tested. Negative validation could imply suboptimal validation design, limited evidence for the GWAS association, and/or alternative genetic variants that may contribute to the trait (*e.g.*, variants that reside within the linkage disequilibrium block of the significant genetic variant). Despite the prolific publication of GWAS results across disease areas, functional validation studies have not been routinely performed to follow up on the top candidate genes. A few barriers may discourage the pursuit of validation studies. First, the significance level of the association signal and the magnitude of the estimated effect size from a single study could be overly optimistic, as discussed

in paragraphs above; researchers may not trust the evidence presented and fear for failure. Second, choosing the adequate biological system to perform the functional validation could be tricky. For instance, functional validation of genetic risk loci identified for a complex disease in humans is challenging because 1) no genetic manipulation could be done in humans, 2) designing, executing and interpreting phenotype measurements in the next best system (*e.g.*, other mammalian model organisms such as rats and mice) may not be straightforward, and 3) complex behavioral traits measured in human GWAS may not always be faithfully remodeled in an animal system. The third barrier might simply be the invisible dichotomy between the GWAS community and the experimental scientists who are interested in specific phenotypes, biological pathways, molecular mechanisms or genes: scientists who perform GWAS may not be scientifically invested in every candidate gene they uncover, and experimental scientists may not be aware of every new GWAS report that implicates their favorite gene.

Given the importance of replication in the context of GWAS, my thesis touches on multiple concepts of replication and aims to exemplify both the computational and the functional aspects of replication in GWAS. All of my work is based on the model organism of mice. There are a few advantages of using mice as the system to perform the discovery, replication and validation of GWAS findings. In the discovery stage, using laboratory mice allows for stringent environmental control, thus reducing the non-genetic variance introduced to the phenotypic variance. In the replication stage, multiple cohorts of the same population of mice could be employed for direct replication and mega-analysis. In the validation stage, genetic manipulation could be carried out to create a mutant mouse line; the effect of the mutation on the associated trait and on the

underpinning molecular function could be examined in detail. These advantages are meaningful to consider in light of the fact that using animal models allows for multiple avenues of hypothesis testing, many of which are not attainable using human subjects.

Chapter 1 of my thesis investigates the rate of replication in mouse GWAS. We used an advanced intercross line (**AIL**), which is a multigenerational intercross between two inbred mouse lines, LG/J and SM/J. Genotypes and phenotypes have been collected from multiple generations; the availability of cohorts from the same population serves as a convenient vehicle to conduct replication studies and mega-analysis. In particular, mice from the 34th generation (**F_{34}**; n=428) had been genotyped using a custom Illumina Infinium genotyping microarray to obtain genotypes for 4,593 SNPs (Cheng et al., 2010; Parker et al., 2014) F_{34} mice were later re-genotyped using genotyping-by-sequencing (**GBS**), which is a reduced-representation sequencing method (Davey et al., 2011; Elshire et al., 2011; Fitzpatrick et al., 2013) to obtain a much denser set of SNPs (~60K) in the F_{34} cohort. Using the same cohort of F_{34} mice, we directly compared GWAS results using the array and the GBS SNPs and examined how SNP density influenced loci identification. In later generations, more traits were collected in mice from the 39th – 43rd generations (**$F_{39\text{-}43}$**; n=600). $F_{39\text{-}43}$ mice were also genotyped using GBS in the same batch as the F_{34} animals that were re-genotyped. Between the phenotype data measured in the F_{34} and the $F_{39\text{-}43}$ cohorts, coat color, body weight, and locomotor response to methamphetamine are the three tests that were measured using identical protocol. Therefore, we directly compared GWAS findings for coat color, body weight, and locomotor response to methamphetamine between the F_{34} and the $F_{39\text{-}43}$ cohorts. To set our expectations for replication, we performed simulations to estimate the power for these

replication studies. Because the actual rate of replication was lower than predicted by the power analysis, we used a random effects model to evaluate the role of Winner's Curse and study-specific heterogeneity in the low rate of replication. Finally, because the two cohorts were genotyped together, we combined the two cohorts (n=1028) and performed mega-analysis for the three shared trait coat color, body weight, and locomotor response.

Chapter 2 of my thesis describes a gene validation study following up on a candidate gene, 5-azacytidine–induced gene 2 (***Azi2***), which had previously been identified in a GWAS for locomotor response to methamphetamine. Using 1,200 male Carworth Farms White (**CFW**) mice, a commercially available outbred population, Parker et al (2016) identified a locus on chromosome 9 (chr9.117763640, rs46497021) for locomotor response to methamphetamine (p-value=1.58×10^{-6}; (Parker et al., 2016a)). To further identify the genes responsible for the gene-trait associations, Parker et al (2016) performed an eQTL using CFW brain regions hippocampus (n=79), striatum (n=55), and prefrontal cortex (n=54) and discovered a *cis*-eQTL on chromosome 9 (chr9.118150749, rs234453358) for *Azi2* expression in the striatum (p-value=1.2×10^{-8}; Parker et al 2016). The GWAS locus for locomotor response to methamphetamine injection and the eQTL for *Azi2* expression in the striatum are in the same implicated region (r^2>0.4; Parker et al 2016), indicating that *Azi2* expression may be a causative factor for the locomotor response to methamphetamine injection. To validate *Azi2* as a causal gene for methamphetamine sensitivity, we created an *Azi2* knockout mouse line using the CRISPR/Cas9 system and established the mutant phenotype of locomotor response to methamphetamine in the knockout line.

Despite the plethora of studies on *Azi2* focusing on its functions in NFκB activation, TNF-induced cell death, and immune response (*e.g.*, (Bozóky et al., 2013a; Fujita et al., 2003; Goncalves et al., 2011; Lafont et al., 2018)), no previous study has linked *Azi2* to the dopaminergic processes of methamphetamine sensitivity. Liu et al (2018) identified ^{AZI2}3'UTR, an independent transcript that matches to the second half of the 3'UTR sequence of *Azi2*, as a downregulator of *Slc6a3*, which encodes the dopamine transporter (**DAT**), in the ventral tegmental area (**VTA**) region of midbrain (K. Liu et al., 2018a). If true, these findings could explain the corroborative evidence from both the GWAS and eQTL studies that *Azi2* plays a critical role in the reward sensitization pathway of methamphetamine. To test the hypothesis that *Azi2* or *Azi2* 3'UTR downregulates *Slc6a3* expression in the VTA region of midbrain, we examined the possible correlations among *Azi2*, the *Azi2* 3'UTR, and *Slc6a3* expression in VTA in the *Azi2* KO line and in naïve CFW mice, where the *Azi2* locus was originally identified.

Both chapter 1 and chapter 2 examine the trait methamphetamine sensitivity. Methamphetamine is a psychomotor stimulant of abuse. Affecting the brain and the central nervous system, methamphetamine produces the feeling of euphoria, heightened energy, and enhanced focus. The positive psychoactive sensations induced by methamphetamine have made the drug popular among abusers: in the U.S. alone, approximately 1.9 million people reported using methamphetamine in the past year (2018 National Survey on Drug Use and Health (Lipari, 2018)). Among the large base of users, individual sensitivity to methamphetamine is far from uniform; studies on both humans and animals have shown that individual response to methamphetamine intake is highly varied (Deminiere et al., 1989; Ettenberg, 2009; Glick & Hinds, 1985; Piazza et al., 1989;

Seale, 1991; Smith et al., 2016). In addition, dosage and frequency of methamphetamine use also dictate differential outcome, which would be further categorized into acute response and chronic addiction (Meredith et al., 2005; Russell et al., 2008).

GWAS have facilitated the identification of thousands of loci for complex traits such as substance abuse (Buchwald et al., 2020; Hancock et al., 2018), depression (Forstner et al., 2019; W. Liu et al., 2020), and alcohol addiction (Kinreich et al., 2019; Liu et al., 2019; Sanchez-Roige et al., 2019; Walters et al., 2018; Zhou et al., 2020). Our approach to study methamphetamine sensitivity - discover, replicate and validate significant loci in multiple mouse populations - could be more widely applied to other complex traits. Given the increasing availability of large-scale collaboration datasets and consortiums in humans and model organisms and the relative ease of creating a KO line using the CRISPR/Cas9 system, more replication studies and functional validation studies could be carried out such that the credibility of loci identified in association studies could be assessed and prioritized.

References

Beavis, W. D., Grant, D., Albertsen, M., & Fincher, R. (1991). Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theoretical and Applied Genetics*, *83*(2), 141–145. https://doi.org/10.1007/BF00226242

Beavis, W. D., Smith, O. S., Grant, D., & Fincher, R. (1994). Identification of Quantitative Trait Loci Using a Small Sample of Topcrossed and F4 Progeny from Maize. *Crop Science*, *34*(4), 882–896. https://doi.org/10.2135/cropsci1994.0011183X003400040010x

Begum, F., Ghosh, D., Tseng, G. C., & Feingold, E. (2012). Comprehensive literature review and statistical considerations for GWAS meta-analysis. *Nucleic Acids Research*, *40*(9), 3777–3784. https://doi.org/10.1093/nar/gkr1255

Bozóky, B., Savchenko, A., Csermely, P., Korcsmáros, T., Dúl, Z., Pontén, F., Székely, L., & Klein, G. (2013). Novel signatures of cancer-associated fibroblasts. *International Journal of Cancer*, *133*(2), 286–293. https://doi.org/10.1002/ijc.28035

Buchwald, J., Chenoweth, M. J., Palviainen, T., Zhu, G., Benner, C., Gordon, S., Korhonen, T., Ripatti, S., Madden, P. A. F., Lehtimäki, T., Raitakari, O. T., Salomaa, V., Rose, R. J., George, T. P., Lerman, C., Pirinen, M., Martin, N. G., Kaprio, J., Loukola, A., & Tyndale, R. F. (2020). Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-020-0702-z

Cheng, R., Lim, J. E., Samocha, K. E., Sokoloff, G., Abney, M., Skol, A. D., & Palmer, A. A. (2010). Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics*.

Collaboration, O. S. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/10.1126/science.aac4716

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499.

Deminiere, J. M., Piazza, P. V., Le Moal, M., & Simon, H. (1989). Experimental approach to individual vulnerability to psychostimulant addiction. *Neuroscience and Biobehavioral Reviews*, *13*(2–3), 141–147. https://doi.org/10.1016/s0149-7634(89)80023-5

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Ettenberg, A. (2009). The runway model of drug self-administration. *Pharmacology, Biochemistry, and Behavior*, *91*(3), 271–277. https://doi.org/10.1016/j.pbb.2008.11.003

Evangelou, E., & Ioannidis, J. P. A. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nature Reviews Genetics*, *14*(6), 379–389. https://doi.org/10.1038/nrg3472

Fitzpatrick, C. J., Gopalakrishnan, S., Cogan, E. S., Yager, L. M., Meyer, P. J., Lovic, V., Saunders, B. T., Parker, C. C., Gonzales, N. M., & Aryee, E. (2013). Variation in the form of Pavlovian conditioned approach behavior among outbred male Sprague-Dawley rats from different vendors and colonies: Sign-tracking vs. goal-tracking. *PloS One*, *8*(10), e75042.

Forstner, A. J., Awasthi, S., Wolf, C., Maron, E., Erhardt, A., Czamara, D., Eriksson, E., Lavebratt, C., Allgulander, C., Friedrich, N., Becker, J., Hecker, J., Rambau, S., Conrad, R., Geiser, F., McMahon, F. J., Moebus, S., Hess, T., Buerfent, B. C., Hoffmann, P., Herms, S., Heilmann-Heimbach, S., Kockum, I., Olsson, T., Alfredsson, L., Weber, H., Alpers, G. W., Arolt, V., Fehm, L., Fydrich, T., Gerlach, A. L., Hamm, A., Kircher, T., Pané-Farré, C. A., Pauli, P., Rief, W., Ströhle, A., Plag, J., Lang, T., Wittchen, H.-U., Mattheisen, M., Meier, S., Metspalu, A., Domschke, K., Reif, A., Hovatta, I., Lindefors, N., Andersson, E., Schalling, M., Mbarek, H., Milaneschi, Y., de Geus, E. J. C., Boomsma, D. I., Penninx, B. W. J. H., Thorgeirsson, T. E., Steinberg, S., Stefansson, K., Stefansson, H., Müller-Myhsok, B., Hansen, T. F., Børglum, A. D., Werge, T., Mortensen, P. B., Nordentoft, M., Hougaard, D. M., Hultman, C. M., Sullivan, P. F., Nöthen, M. M., Woldbye, D. P. D., Mors, O., Binder, E. B., Rück, C., Ripke, S., Deckert, J., & Schumacher, J. (2019). Genome-wide association study of panic disorder reveals genetic overlap with neuroticism and depression. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-019-0590-2

Fujita, F., Taniguchi, Y., Kato, T., Narita, Y., Furuya, A., Ogawa, T., Sakurai, H., Joh, T., Itoh, M., Delhase, M., Karin, M., & Nakanishi, M. (2003). Identification of NAP1, a regulatory subunit of IkappaB kinase-related kinases that potentiates NF-kappaB signaling. *Molecular and Cellular Biology*, *23*(21), 7780–7793.

Glick, S. D., & Hinds, P. A. (1985). Differences in amphetamine and morphine sensitivity in lateralized and non-lateralized rats: Locomotor activity and drug self-administration. *European Journal of Pharmacology*, *118*(3), 239–244. https://doi.org/10.1016/0014-2999(85)90134-7

Goncalves, A., Bürckstümmer, T., Dixit, E., Scheicher, R., Górna, M. W., Karayel, E., Sugar, C., Stukalov, A., Berg, T., Kralovics, R., Planyavsky, M., Bennett, K. L., Colinge, J., & Superti-Furga, G. (2011). Functional Dissection of the TBK1 Molecular Network. *PLOS ONE*, *6*(9), e23971. https://doi.org/10.1371/journal.pone.0023971

Hancock, D. B., Markunas, C. A., Bierut, L. J., & Johnson, E. O. (2018). Human Genetics of Addiction: New Insights and Future Directions. *Current Psychiatry Reports*, *20*(2), 8. https://doi.org/10.1007/s11920-018-0873-3

Keele, G. R., Crouse, W. L., Kelada, S. N. P., & Valdar, W. (2019). Determinants of QTL Mapping Power in the Realized Collaborative Cross. *G3: Genes, Genomes, Genetics*, *9*(5), 1707–1727. https://doi.org/10.1534/g3.119.400194

King, E. G., & Long, A. D. (2017). The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster. *G3: Genes, Genomes, Genetics*, *7*(6), 1643–1652. https://doi.org/10.1534/g3.117.041426

Kinreich, S., Meyers, J. L., Maron-Katz, A., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Pandey, G., Subbie-Saenz de Viteri, S., Pitti, D., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., Edenberg, H. J., & Porjesz, B. (2019). Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: A machine learning study. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-019-0534-x

Lafont, E., Draber, P., Rieser, E., Reichert, M., Kupka, S., de Miguel, D., Draberova, H., von Mässenhausen, A., Bhamra, A., Henderson, S., Wojdyla, K., Chalk, A., Surinova, S., Linkermann, A., & Walczak, H. (2018). TBK1 and IKKε prevent TNF-induced cell death by RIPK1 phosphorylation. *Nature Cell Biology*, *20*(12), 1389–1399. https://doi.org/10.1038/s41556-018-0229-6

Lipari, R. N. (2018). *Key Substance Use and Mental Health Indicators in the United States: Results from the 2018 National Survey on Drug Use and Health*. 82.

Liu, K., Yu, J., Zhao, J., Zhou, Y., Xiong, N., Xu, J., Wang, T., Bell, R. L., Qing, H., & Lin, Z. (2018). AZI23'UTR Is a New SLC6A3 Downregulator Associated with an Epistatic Protection Against Substance Use Disorders. *Molecular Neurobiology*, *55*(7), 5611–5622. https://doi.org/10.1007/s12035-017-0781-2

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., 23andMe Research Team, HUNT All-In Psychiatry, Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., Hottenga, J.-J., Huang, H., Jang, S.-K., Jansen, P. R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù, V., Palviainen, T., Pandit, A., Reginsson, G. W., Skogholt, A. H., Smith, J. A., Taylor, A. E., Turman, C., Willemsen, G., Young, H., Young, K. A., Zajac, G.

J. M., Zhao, W., Zhou, W., Bjornsdottir, G., Boardman, J. D., Boehnke, M., Boomsma, D. I., Chen, C., Cucca, F., Davies, G. E., Eaton, C. B., Ehringer, M. A., Esko, T., Fiorillo, E., Gillespie, N. A., Gudbjartsson, D. F., Haller, T., Harris, K. M., Heath, A. C., Hewitt, J. K., Hickie, I. B., Hokanson, J. E., Hopfer, C. J., Hunter, D. J., Iacono, W. G., Johnson, E. O., Kamatani, Y., Kardia, S. L. R., Keller, M. C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K. S., Laakso, M., Lind, P. A., Loukola, A., Lutz, S. M., Madden, P. A. F., Martin, N. G., McGue, M., McQueen, M. B., Medland, S. E., Metspalu, A., Mohlke, K. L., Nielsen, J. B., Okada, Y., Peters, U., Polderman, T. J. C., Posthuma, D., Reiner, A. P., Rice, J. P., Rimm, E., Rose, R. J., Runarsdottir, V., Stallings, M. C., Stančáková, A., Stefansson, H., Thai, K. K., Tindle, H. A., Tyrfingsson, T., Wall, T. L., Weir, D. R., Weisner, C., Whitfield, J. B., Winsvold, B. S., Yin, J., Zuccolo, L., Bierut, L. J., Hveem, K., Lee, J. J., Munafò, M. R., Saccone, N. L., Willer, C. J., Cornelis, M. C., David, S. P., Hinds, D. A., Jorgenson, E., Kaprio, J., Stitzel, J. A., Stefansson, K., Thorgeirsson, T. E., Abecasis, G., Liu, D. J., & Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

Liu, W., Li, W., Cai, X., Yang, Z., Li, H., Su, X., Song, M., Zhou, D.-S., Li, X., Zhang, C., Shao, M., Zhang, L., Yang, Y., Zhang, Y., Zhao, J., Chang, H., Yao, Y.-G., Fang, Y., Lv, L., Li, M., & Xiao, X. (2020). Identification of a functional human-unique 351-bp Alu insertion polymorphism associated with major depressive disorder in the 1p31.1 GWAS risk loci. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*. https://doi.org/10.1038/s41386-020-0659-2

Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *American Psychologist*, *70*(6), 487–498. https://doi.org/10.1037/a0039400

Meredith, C. W., Jaffe, C., Ang-Lee, K., & Saxon, A. J. (2005). Implications of Chronic Methamphetamine Use: A Literature Review. *Harvard Review of Psychiatry*, *13*(3), 141–154. https://doi.org/10.1080/10673220591003605

Parker, C. C., Carbonetto, P., Sokoloff, G., Park, Y. J., Abney, M., & Palmer, A. A. (2014). High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. *Genetics*, *198*(1), 103–116.

Parker, C. C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N. M., Leung, E., Park, Y. J., Aryee, E., Davis, J., Blizard, D. A., Ackert-Bicknell, C. L., Lionikas, A., Pritchard, J. K., & Palmer, A. A. (2016). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, *48*(8), 919–926. https://doi.org/10.1038/ng.3609

Paterson, A. H. (2019). *Molecular Dissection of Complex Traits*. CRC Press.

Piazza, P. V., Deminière, J. M., Le Moal, M., & Simon, H. (1989). Factors that predict individual vulnerability to amphetamine self-administration. *Science (New York, N.Y.)*, *245*(4925), 1511–1513. https://doi.org/10.1126/science.2781295

Poirier, J. G., Faye, L. L., Dimitromanolakis, A., Paterson, A. D., Sun, L., & Bull, S. B. (2015). Resampling to Address the Winner's Curse in Genetic Association Analysis of Time to Event. *Genetic Epidemiology*, *39*(7), 518–528. https://doi.org/10.1002/gepi.21920

Russell, K., Dryden, D. M., Liang, Y., Friesen, C., O'Gorman, K., Durec, T., Wild, T. C., & Klassen, T. P. (2008). Risk factors for methamphetamine use in youth: A systematic review. *BMC Pediatrics*, *8*(1), 48. https://doi.org/10.1186/1471-2431-8-48

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams, M. J., Howard, D. M., Edenberg, H. J., Davies, G., Crist, R. C., Deary, I. J., McIntosh, A. M., & Clarke, T.-K. (2019). Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *The American Journal of Psychiatry*, *176*(2), 107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Seale, T. W. (1991). Genetic Differences in Response to Cocaine and Stimulant Drugs. In J. C. Crabbe & R. A. Harris (Eds.), *The Genetic Basis of Alcohol and Drug Actions* (pp. 279–321). Springer US. https://doi.org/10.1007/978-1-4899-2067-6_8

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, Science, and Knowledge Construction: Broadening Perspectives from the Replication Crisis. *Annual Review of Psychology*, *69*(1), 487–510. https://doi.org/10.1146/annurev-psych-122216-011845

Smith, C. T., Weafer, J., Cowan, R. L., Kessler, R. M., Palmer, A. A., de Wit, H., & Zald, D. H. (2016). Individual differences in timing of peak positive subjective responses to d-amphetamine: Relationship to pharmacokinetics and physiology. *Journal of Psychopharmacology*, *30*(4), 330–343. https://doi.org/10.1177/0269881116631650

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., Foo, J. C., Fox, L., Frank, J., Giegling, I., Gordon, S., Hack, L. M., Hartmann, A. M., Hartz, S. M., Heilmann-Heimbach, S., Herms, S., Hodgkinson, C., Hoffmann, P., Jan Hottenga, J., Kennedy, M. A., Alanne-Kinnunen, M., Konte, B., Lahti, J., Lahti-Pulkkinen, M., Lai, D., Ligthart, L., Loukola, A., Maher, B. S., Mbarek, H., McIntosh, A. M., McQueen, M. B., Meyers, J. L., Milaneschi, Y., Palviainen, T., Pearson, J. F., Peterson, R. E., Ripatti, S., Ryu, E.,

Saccone, N. L., Salvatore, J. E., Sanchez-Roige, S., Schwandt, M., Sherva, R., Streit, F., Strohmaier, J., Thomas, N., Wang, J.-C., Webb, B. T., Wedow, R., Wetherill, L., Wills, A. G., 23andMe Research Team, Boardman, J. D., Chen, D., Choi, D.-S., Copeland, W. E., Culverhouse, R. C., Dahmen, N., Degenhardt, L., Domingue, B. W., Elson, S. L., Frye, M. A., Gäbel, W., Hayward, C., Ising, M., Keyes, M., Kiefer, F., Kramer, J., Kuperman, S., Lucae, S., Lynskey, M. T., Maier, W., Mann, K., Männistö, S., Müller-Myhsok, B., Murray, A. D., Nurnberger, J. I., Palotie, A., Preuss, U., Räikkönen, K., Reynolds, M. D., Ridinger, M., Scherbaum, N., Schuckit, M. A., Soyka, M., Treutlein, J., Witt, S., Wodarz, N., Zill, P., Adkins, D. E., Boden, J. M., Boomsma, D. I., Bierut, L. J., Brown, S. A., Bucholz, K. K., Cichon, S., Costello, E. J., de Wit, H., Diazgranados, N., Dick, D. M., Eriksson, J. G., Farrer, L. A., Foroud, T. M., Gillespie, N. A., Goate, A. M., Goldman, D., Grucza, R. A., Hancock, D. B., Harris, K. M., Heath, A. C., Hesselbrock, V., Hewitt, J. K., Hopfer, C. J., Horwood, J., Iacono, W., Johnson, E. O., Kaprio, J. A., Karpyak, V. M., Kendler, K. S., Kranzler, H. R., Krauter, K., Lichtenstein, P., Lind, P. A., McGue, M., MacKillop, J., Madden, P. A. F., Maes, H. H., Magnusson, P., Martin, N. G., Medland, S. E., Montgomery, G. W., Nelson, E. C., Nöthen, M. M., Palmer, A. A., Pedersen, N. L., Penninx, B. W. J. H., Porjesz, B., Rice, J. P., Rietschel, M., Riley, B. P., Rose, R., Rujescu, D., Shen, P.-H., Silberg, J., Stallings, M. C., Tarter, R. E., Vanyukov, M. M., Vrieze, S., Wall, T. L., Whitfield, J. B., Zhao, H., Neale, B. M., Gelernter, J., Edenberg, H. J., & Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Xiao, R., & Boehnke, M. (2009). Quantifying and correcting for the winner's curse in genetic association studies. *Genetic Epidemiology*, *33*(5), 453–462. https://doi.org/10.1002/gepi.20398

Xu, S. (2003). Theoretical Basis of the Beavis Effect. *Genetics*, *165*(4), 2259–2268.

Zhou, H., Sealock, J. M., Sanchez-Roige, S., Clarke, T.-K., Levey, D. F., Cheng, Z., Li, B., Polimanti, R., Kember, R. L., Smith, R. V., Thygesen, J. H., Morgan, M. Y., Atkinson, S. R., Thursz, M. R., Nyegaard, M., Mattheisen, M., Børglum, A. D., Johnson, E. C., Justice, A. C., Palmer, A. A., McQuillin, A., Davis, L. K., Edenberg, H. J., Agrawal, A., Kranzler, H. R., & Gelernter, J. (2020). Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into biology and relationships with other traits. *Nature Neuroscience*, *23*(7), 809–818. https://doi.org/10.1038/s41593-020-0643-5

Zöllner, S., & Pritchard, J. K. (2007). Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. *The American Journal of Human Genetics*, *80*(4), 605–615. https://doi.org/10.1086/512821

CHAPTER 1: Genome-wide Association Study in Two Cohorts from a Multi-generational Mouse Advanced Intercross Line Highlights the Difficulty of Replication Due to Study-specific Heterogeneity

1.1 Abstract

There has been extensive discussion of the "Replication Crisis" in many fields, including genome-wide association studies (**GWAS**). We explored replication in a mouse model using an advanced intercross line (**AIL**), which is a multigenerational intercross between two inbred strains. We re-genotyped a previously published cohort of LG/J x SM/J AIL mice (F_{34}; n = 428) using a denser marker set and genotyped a new cohort of AIL mice ($F_{39\text{-}43}$; n = 600) for the first time. We identified 36 novel genome-wide significant loci in the F_{34} and 25 novel loci in the $F_{39\text{-}43}$ cohort. The subset of traits that were measured in both cohorts (locomotor activity, body weight, and coat color) showed high genetic correlations, although the SNP heritabilities were slightly lower in the $F_{39\text{-}43}$ cohort. For this subset of traits, we attempted to replicate loci identified in either F_{34} or $F_{39\text{-}43}$ in the other cohort. Coat color was robustly replicated; locomotor activity and body weight were only partially replicated, which was inconsistent with our power simulations. We used a random effects model to show that the partial replications could not be explained by Winner's Curse but could be explained by study-specific heterogeneity. Despite this heterogeneity, we performed a mega-analysis by combining F_{34} and $F_{39\text{-}43}$ cohorts (n = 1,028), which identified four novel loci associated with locomotor activity and body weight. These results illustrate that even with the high degree of genetic and environmental control possible in our experimental system, replication was hindered by study-specific heterogeneity, which has broad implications for ongoing concerns about reproducibility.

1.2 Introduction

Genome-wide association studies (GWAS) in model organisms can use genetically identical cohorts phenotyped under extremely similar conditions, which would be expected to enhance the success of replication. We sought to investigate replication in model organism GWAS using a mouse advanced intercross line (AIL). The use of GWAS in model organisms such as mice (Talbot et al. 1999; Demarest et al. 2001; Yalcin et al. 2004; Valdar et al. 2006; Ghazalpour et al. 2008; Samocha et al. 2010; Churchill et al. 2012; Collaborative Cross Consortium 2012; Parker et al. 2012, 2016; Svenson et al. 2012; Carbonetto et al. 2014; Chesler 2014; Coyner et al. 2014; Gatti et al. 2014; Nicod et al. 2016; Hernandez Cordero et al. 2018, 2019), rats (Baud et al. 2014), chickens (Besnier et al. 2011; Johnsson et al. 2018), fruit flies (King et al. 2012; Mackay et al. 2012; Kislukhin et al. 2013; Marriage et al. 2014; Vonesch et al. 2016), *C. elegans* (Doitsidou et al. 2016) and various plant species (Rishmawi et al. 2017; Cockram and Mackay 2018; Diouf et al. 2018) has become increasingly common over the last decade. These mapping populations can further be categorized as multi-parental crosses, which are created by interbreeding two or more inbred strains, and various outbred populations, in which the founders are of unknown provenance. An F_2 cross between two inbred strains is the prototypical mapping population; however, F_2s provide poor mapping resolution (Parker and Palmer 2011). To improve mapping resolution, Darvasi and Soller (Darvasi and Soller 1995) proposed the creation of advanced intercross lines (AILs), which are produced by intercrossing F_2 mice for additional generations. AILs accumulate additional crossovers with every successive generation, leading to a population with shorter linkage

disequilibrium (LD) blocks, which improves mapping precision, albeit at the expense of power (Parker and Palmer 2011; Gonzales and Palmer 2014).

The longest running mouse AIL was generated by crossing LG/J and SM/J inbred strains, which had been previously selected for large and small body size prior to inbreeding and subsequent intercrossing. We obtained this AIL in 2006 at generation 33 from Dr. James Cheverud (Jmc: LG,SM-G33). Since then, we have collected genotype and phenotype information from multiple generations, including F_{34} (Cheng et al. 2010; Lionikas et al. 2010; Samocha et al. 2010; Parker et al. 2011, 2014; Bartnikas et al. 2012; Carroll et al. 2017; Gonzales et al. 2018) and F_{39}-F_{43}. Our previous publications using the F_{34} generation employed a custom Illumina Infinium genotyping microarray to obtain genotypes for 4,593 SNPs (Cheng et al. 2010; Parker et al. 2014); we refer to this set of SNPs as the 'sparse markers'. Those genotypes were used to identify significant associations for numerous traits, including locomotor activity in response to methamphetamine (Cheng et al. 2010), pre-pulse inhibition (Samocha et al. 2010), muscle weight (Lionikas et al. 2010; Hernandez Cordero et al. 2019), body weight (Parker et al. 2011), open field (Parker et al. 2014), conditioned fear (Parker et al. 2014), red blood cell parameters (Bartnikas et al. 2012), and muscle weights (Carroll et al. 2017). Although not previously published, we also collected phenotype information from the F_{39-43} generations, including body weight, fear conditioning, locomotor activity in response to methamphetamine, and the light dark test for anxiety.

While the prior GWAS using the F_{34} generation detected many significant loci, the sparsity of the markers likely precluded the discovery of some true loci and also made it difficult to clearly define the boundaries of the loci that we did identify. For example, Parker

et al. conducted an integrated analysis of F_2 and F_{34} AILs (Parkeret al. 2011). One of their body weight loci spanned from 87.93–102.70 Mb on chromosome 14. Denser markers might have more clearly defined the implicated region.

In the present study, we used genotyping-by-sequencing (GBS), which is a reduced-representation sequencing method (Davey et al. 2011; Elshire et al. 2011; Fitzpatrick et al. 2013), to obtain a much denser set of SNPs in the F_{34} cohort and, for the first time, genotyped mice from the $F_{39\text{-}43}$ generations. With this denser set of SNPs, we attempted to identify novel loci in the F_{34} cohort that were not detected using the sparse SNPs. We also performed GWAS using the mice from the $F_{39\text{-}43}$ AILs. We explored whether imputation from the array SNPs could have provided the additional coverage we obtained using the denser GBS genotypes. Because $F_{39\text{-}43}$ AILs are the direct descendants of the F_{34}, they are uniquely suited to serve as a replication population for GWAS in the F_{34} generation. For the subset of traits measured in both cohorts, we attempted to replicate the results discovered in one cohort in the other. To set our expectations for replication, we performed simulations to estimate the power for these replication studies. Because the actual rate of replication was lower than predicted by the power analysis, we used a random effects model to evaluate the role of Winner's Curse and study-specific heterogeneity in the low rate of replication. Finally, we also performed a mega-analysis on a subset of traits common to both cohort.

1.3 Materials and methods

Animals

All mice used in this study were members of the LG/J x SM/J AIL that was originally created by Dr. James Cheverud (Loyola University Chicago, Chicago, IL). This AIL has been maintained in the Palmer laboratory since generation F_{33}. Age and exact number of animals tested in each phenotype are described in Table S1. Several previous publications (Samocha et al. 2010; Cheng et al. 2010; Parker et al. 2014; Lionikas et al. 2010; Carroll et al. 2017; Parker et al. 2011; Bartnikas et al. 2012) have reported association analyses of the F_{34} mice (n=428). No prior publications have described the $F_{39\text{-}43}$ generations (n=600). The sample size of F_{34} mice reported in this study (n=428) is smaller than that in previous publications of F_{34} (n=688) because we only genotyped a subset of F_{34} animals using GBS.

F_{34}, $F_{39\text{-}43}$ Phenotypes

All phenotypes are listed in Table S1. We have previously described the phenotyping of F_{34} animals for locomotor activity and locomotor response to methamphetamine (Cheng et al. 2010), fear conditioning (Parker et al. 2014), open field (Parker et al. 2014), coat color, body weight (Parker et al. 2011), complete blood counts (Bartnikas et al. 2012), heart and tibia measurements (Lionikas et al. 2010), muscle weight (Lionikas et al. 2010). Iron content in liver and spleen, which have not been previously reported in these mice, was measured by atomic absorption spectrophotometry, as described in Gardenghi et al. (Gardenghi et al. 2007) and Graziano, Grady and Cerami (Graziano et al. 1974). Although the phenotyping of $F_{39\text{-}43}$

animals has not been previously reported, we followed previously published protocols for locomotor activity and locomotor response to methamphetamine (Cheng et al. 2010), coat color, body weight (Parker et al. 2011), and light/dark test for anxiety (Sittig et al. 2016). We point out here that even though "locomotor activity" was measured in both the F_{34} and $F_{39\text{-}43}$ using the Versamax software (AccuScan Instruments, Columbus, OH), "open field" in the F_{34} cohort was also measured using Versamax, whereas "open field" in the $F_{39\text{-}43}$ cohort was measured using the EthoVision XT software (Noldus system; (Noldus et al. 2001)). Because there are meaningful differences in these experimental procedures, we did not attempt to use "open field" data for replication. In summary, we performed GWAS on all traits collected in individual cohorts. For the replication analysis between the F_{34} and $F_{39\text{-}43}$ cohorts, we only directly compared a number of traits that had been measured in both cohorts: body weight, two Mendelian coat color traits (albino and agouti), and three locomotor activity traits (locomotor activity on day 1 and on day 2, and activity on day 3 following a methamphetamine injection).

F_{34} AIL Array Genotypes

F_{34} animals had been genotyped on a custom SNP array on the Illumina Infinium platform (Cheng et al. 2010; Parker et al. 2014), which yielded a set of 4,593 SNPs on autosomes and X chromosome that we refer to as 'sparse SNPs'.

F_{34} and $F_{39\text{-}43}$ GBS Genotypes

F_{34} and $F_{39\text{-}43}$ animals were genotyped using genotyping-by-sequencing (**GBS**), which is a reduced-representation genome sequencing method (Parker et al. 2016;

Gonzales et al. 2017). We used the same protocol for GBS library preparation that was described in Gonzales et al (Gonzales et al. 2017). We called GBS genotype probabilities using ANGSD (Korneliussen et al. 2014). GBS identified 1,667,920 autosomal and 43,015 X-chromosome SNPs. To fill in missing genotypes at SNPs where some but not all mice had calls, we ran within-sample imputation using Beagle v4.1, which generated hard call genotypes as well as genotype probabilities (Browning & Browning, 2007). After imputation, only SNPs that had dosage $r^2 > 0.9$ were retained. We removed SNPs with minor allele frequency < 0.1 and SNPs with $p < 1.0 \times 10^{-6}$ in the Chi-square test of Hardy–Weinberg Equilibrium (**HWE**) (Table S2). All phenotype and GBS genotype data are deposited in GeneNetwork2 (http://gn2.genenetwork.org/).

QC of individuals

We have found that large genetic studies are often hampered by cross-contamination between samples and sample mix-ups. We used four features of the data to identify problematic samples: heterozygosity distribution, proportion of reads aligned to sex chromosomes, pedigree/kinship, and coat color. We first examined heterozygosity across autosomes and removed animals where the proportion of heterozygosity was more than 3 standard deviations from the mean (Figure S1). Next, we sought to identify animals in which the recorded sex did not agree with the sequencing data. We compared the ratio of reads mapped to the X and Y chromosomes. The 95% CI for this ratio was 196.84 to 214.3 in females and 2.13 to 2.18 in males. Twenty-two F_{34} and $F_{39\text{-}43}$ animals were removed because their sex (as determined by reads ratio) did not agree with their recorded sex; we assumed this discrepancy was due to sample mix-ups. To further

identify mislabeled samples, we calculated kinship coefficients based on the full AIL pedigree using QTLRel. We then calculated a genetic relatedness matrix (**GRM**) using IBDLD (Abney, 2008; L. Han & Abney, 2011), which estimates identity by descent using genotype data. The comparison between pedigree kinship relatedness and genetic kinship relatedness identified 7 pairs of animals that showed obvious disagreement between kinship coefficients and the GRM, these animals were excluded from further analysis. Lastly, we excluded 14 $F_{39\text{-}43}$ animals that showed discordance between their recorded coat color and their genotypes at markers flanking *Tyr*, which causes albinism in mice. The numbers of animals filtered at each step are listed in Table S2. Some animals were detected by more than one QC step, substantiating our evidence that these samples were erroneous.

At the end of SNP and sample filtering, we had 59,561 autosomal and 831 X chromosome SNPs in F_{34}, 58,966 autosomal and 824 X chromosome SNPs in $F_{39\text{-}43}$, and 57,635 autosomal and 826 X chromosome SNPs in the combined F_{34} and $F_{39\text{-}43}$ set (Table S2). GBS genotype quality was estimated by examining concordance between the 66 SNPs that were present in both the array and GBS genotyping results (Figure S3).

LD decay

Average LD (r^2) was calculated using allele frequency matched SNPs (MAF difference < 0.05) within 100,000 bp distance, as described in Parker et al. (Parker et al. 2016).

Imputation to LG/J and SM/J reference panels

F_{34} array genotypes (n=428) and F_{34} GBS genotypes (n=428) were imputed to LG/J and SM/J whole genome sequence data (Nikolskiy et al. 2015) using BEAGLE (Browning & Browning, 2007). For F_{34} array imputation, we used a large window size (100,000 SNPs and 45,000 SNPs overlap). Imputation to reference panels yielded 4.3 million SNPs for F_{34} array and F_{34} GBS imputed sets. Imputed SNPs with $R^2 > 0.9$, MAF > 0.1, HWE p-value > 1.0×10^{-6} were retained, resulting in 4.1M imputed F_{34} GBS SNPs and 4.3M imputed F_{34} array SNPs.

Genome-wide association analysis (GWAS)

We used the linear mixed model, as implemented in GEMMA (X. Zhou & Stephens, 2012), to perform a GWAS that accounted for the complex familial relationships among the AIL mice (Cheng et al. 2010; Gonzales et al. 2017). We used the leave-one-chromosome-out (**LOCO**) approach to calculate the GRM, which effectively circumvented the problem of proximal contamination (Cheng et al. 2013). We used the univariate linear mixed model described in Zhou and Stephens (X. Zhou & Stephens, 2012):

$$y = W\alpha + x\beta + u + \varepsilon;\ u \sim MVN_n\left(0,\ \lambda\tau^{-1}K\right),\ \varepsilon \sim MVN_n\left(0,\ \tau^{-1}I_n\right),$$

where y is a n-vector of traits for n individuals; W is a n×c matrix of covariates (fixed effects); α is a c-vector of the corresponding coefficients; x is an n-vector of genotypes; β is the effect size of the genotype; u is an n-vector of random effects; ε is an n-vector of errors; τ^{-1} is the variance of the residual errors; λ is the ratio between the two variance components; K is a known n × n relatedness matrix and I_n is an n × n

identity matrix. MVN_n stands for the n-dimensional multivariate normal distribution (X. Zhou & Stephens, 2012).

Separate GWAS were performed using the F_{34} array genotypes, the F_{34} GBS genotypes, and the $F_{39\text{-}43}$ GBS genotypes. Apart from coat color (binary trait), raw phenotypes were quantile normalized prior to analysis. Coat color traits were coded as follows: albino: 1 = white, 0 = non-white; agouti: 1 = tan, 0 = black, NA=white. Because F_{34} AIL had already been studied, we used the same covariates as described in Cheng et al. (Cheng et al. 2010) in order to examine whether our array and GBS GWAS would replicate their findings. We included sex and body weight as covariates for locomotor activity traits (see covariates used in (Cheng et al. 2010)) and sex, age, and coat color as covariates for fear conditioning and open field test in F_{34} AILs (see covariates used in (Parker et al. 2014)). We used sex and age as covariates for all other phenotypes. Covariates for each analysis are shown in Table S1. Finally, we performed mega-analysis of F_{34} and $F_{39\text{-}43}$ animals (n=1,028) for body weight, coat color, and locomotor activity, since these traits were measured in the same way in both cohorts. We quantile transformed all continuous phenotypes in each cohort and then combined the transformed phenotypes for the mega-analysis (Coat color traits were not quantile normalized because they are binary).

Identifying dubious SNPs

Some significant SNPs in the F_{34} GWAS were dubious because the flanking SNPs, which would have been expected to be in high LD with the significant SNP (a very strong assumption in an AIL), did not have high -log10(p) values. We only examined SNPs that

obtained significant p-values; close examinations revealed that these SNPs had dubious ratios of heterozygotes to homozygotes calls and had corresponding HWE p-values that were close to our 1.0×10^{-6} threshold (Table S3). We chose the 1.0×10^{-6} as the filter threshold of the HWE p-values based on a gene-dropping exercise. We used the $F_{33\text{-}34}$ family pedigree and the F_{34} genetic map to simulate the genotypes in F_{34} (QTLRel; (Cheng et al. 2011)). The p-value of the chi-square test for Hardy-Weinberg equilibrium in the simulated F_{34} population was 7.24329×10^{-06}, which was close to the HWE threshold used in Gonzales et al. (Gonzales et al. 2018). To avoid counting these as novel loci, we removed those SNPs prior to summarizing our results as they likely reflected genotyping errors.

Selecting independent significant SNPs

To identify independent "lead loci" among significant GWAS SNPs that surpassed the significance threshold, we used the LD-based clumping method in PLINK v1.9. We empirically chose clumping parameters (r^2 = 0.1 and sliding window size = 12,150kb) that gave us a conservative set of independent SNPs (Table S4). For the coat color phenotypes, we found that multiple SNPs remained significant even after LD-based clumping, presumably due to the extremely significant associations at these Mendelian loci. In these cases, we used a stepwise model selection procedure in GCTA (Yang et al. 2011) and performed association analyses conditioning on the most significant SNPs.

Significance thresholds

We used MultiTrans to set significance thresholds for GWAS (B. Han et al. 2009; Joo et al. 2016). MultiTrans is a method that assumes multivariate normal distribution of the phenotypes, which in LMM models, contain a covariance structure due to various degrees of relatedness among individuals. We were curious to see whether MultiTrans produced significance thresholds that were different from the thresholds we obtained from a standard permutation test ('naïve permutation' as per Cheng et al. (Cheng et al. 2013)). We performed 1,000 permutations using the F_{34} GBS genotypes and the phenotypic data from locomotor activity (days 1, 2, and 3). We found that the 95th percentile values for these permutations were 4.65, 4.79, and 4.85, respectively, which were very similar to 4.85, the threshold obtained from MultiTrans using the same data. Thus, the thresholds presented here were obtained from MultiTrans but are similar (if anything slightly more conservative) to the thresholds we would have obtained had we used permutation. Because the effective number of tests depends on the number of SNPs and the specific animals used in GWAS, we obtained a unique adjusted significance threshold for each SNP set in each animal cohort (Table S5).

Credible set analysis

We followed the method described in (The Wellcome Trust Case Control Consortium et al. 2012). Credible set analysis is a Bayesian method of selecting an interval of SNPs that are likely to contain the causal SNPs; we used LD r^2 threshold = 0.8, posterior probability =0.99. The R script could be found on GitHub: https://github.com/hailianghuang/FM-summary/blob/master/getCredible.r

Power analysis

To estimate the power of replication of a SNP from the discovery set in the replication set, we simulated GWAS with 50 varying effect sizes for the discovery SNP using the LMM model. We first fit the trait in a null model (i.e., no genotype effect), and obtained estimates of model parameters including the intercept and the genetic variance component. Using these model parameters, we added the genotype effect to the random numbers generated from the null model to recreate a trait. For each simulated effect size, we scanned every simulated trait 2,500 times and examined the ratio of association tests whose test statistics surpassed the significance thresholds (both the genome-wide significance threshold for the cohort and the nominal p-value of 0.05).

Replication analysis between F_{34} and $F_{39\text{-}43}$ GWAS studies

We modeled the replication between F_{34} and $F_{39\text{-}43}$ GWAS studies using two random effects models (Zou et al. 2019). Both models take as input a set of z-scores for variants computed from an association study ("summary statistics").

The **WC** model accounts only for Winner's Curse. We assume that there is a shared genetic effect (λ) that is responsible for the observed association signal in both studies. To model random noise contributing to Winner's Curse, we model the summary statistics for each variant k from the discovery and replication studies as normally distributed random variables ($s_k^{(1)} \sim N(\lambda,\ 1)$ and $s_k^{(2)} \sim N(\lambda,\ 1)$, respectively). We define the prior probability of the true genetic effect to be $\lambda \sim N(0, \sigma_g^2)$, where the variance in the true genetic effect is learned through a maximum likelihood procedure. We correct for

the effect of Winner's Curse in the discovery study by computing the conditional distribution of the replication summary statistics given the discovery summary statistic.

The **WC+C** model accounts for Winner's Curse and study-specific heterogeneity. In this model, we partition the total effect sizes observed into genetic effects (λ) and study-specific effects ($\delta^{(1)}$ and $\delta^{(2)}$). We model the statistics for each variant k from the initial and discovery studies as normally distributed random variables ($s_k^{(1)} \sim N(\lambda + \delta^{(1)}, 1)$ and $s_k^{(2)} \sim N(\lambda + \delta^{(2)}, 1)$, respectively). In addition to the prior on the genetic effect defined in the **WC** model, we define the prior probabilities of the study-specific effects to be $\delta^{(1)} \sim N(0, \sigma_{c_1}^2)$, and $\delta^{(2)} \sim N(0, \sigma_{c_2}^2)$, where the variance parameters are learned through a maximum likelihood procedure. We correct for the effect of Winner's Curse in the discovery study and study-specific effects by computing the conditional distribution of the replication summary statistic given the discovery summary statistic.

We applied each of these models once using F_{34} as the discovery study and once using $F_{39\text{-}43}$ as the discovery study. We used the genome-wide significance thresholds in Table S5 to identify variants in each discovery study and used the results as input to the random effects models. We then used a Bonferroni corrected threshold (p=0.05/M) for the replication study, where M is the number of genome-wide significant variants in the initial study. We computed the "empirical replication rate" as the proportion of variants passing the genome-wide significant threshold in the discovery study that also passed this Bonferroni corrected threshold in the replication study. Since the estimation of the model parameters requires at least two variants, we only applied this method to phenotypes with at least two genome-wide significant variants in the discovery study.

To assess how well the **WC** and **WC+C** models explained the observed patterns of replication, we computed the predicted replication rates under each model. For each variant that passed the genome-wide significant threshold in the discovery study, we used the conditional distributions previously learned to compute the probability that the variant passed the Bonferroni corrected threshold in the replication study. For each phenotype, we computed the average of these predicted replication rates and compared this average to the empirical replication rates.

Genetic correlation and heritability estimates between F_{34} and $F_{39\text{-}43}$ phenotypes

Locomotor activity, body weight, and coat color traits had been measured in both F_{34} and $F_{39\text{-}43}$ populations. We calculated both SNP heritability and genetic correlations between F_{34} and $F_{39\text{-}43}$ animals using GCTA-GREML analysis and GCTA bivariate GREML analysis (Yang et al. 2011).

LocusZoom Plots

LocusZoom plots were generated using the standalone implementation of LocusZoom (Pruim et al. 2010), using LD scores calculated from PLINK v.1.9 --ld option and mm10 gene annotation file downloaded from UCSC genome browser.

1.4 Results

We used 214 males and 214 females from generation F_{34} (Aap:LG,SM-G34) and 305 males and 295 females from generations $F_{39\text{-}43}$. For the F_{34} AIL 79 traits were available from previously published and unpublished work; for the $F_{39\text{-}43}$ AIL 49 unpublished traits were available (Table S1). F_{34} mice had been previously genotyped on a custom SNP array (Cheng et al. 2010; Parker et al. 2014). The average minor allele frequency (MAF) of those 4,593 array SNPs was 0.388 (Figure 1). To obtain a denser set of SNP markers, we used GBS in F34 and $F_{39\text{-}43}$ AIL mice. Since data on the $F_{39\text{-}43}$ AIL mice had been collected over the span of approximately two years, we carefully considered the possibility of sample contamination and sample mislabeling (Toker et al. 2016) and removed these samples (see Methods; Figure S1 and S2). The final SNP sets included 60,392 GBS-derived SNPs in 428 F34 AIL mice, 59,790 GBS-derived SNPs in 600 $F_{39\text{-}43}$ AIL mice, and 58,461 GBS-derived SNPs that existed in both F_{34} and $F_{39\text{-}43}$ AIL mice (Table S2). The MAF for the GBS SNPs was 0.382 in F_{34}, 0.358 in $F_{39\text{-}43}$, and 0.370 in F_{34} and $F_{39\text{-}43}$ (Figure 1). There were 66 SNPs called from our GBS data that were also present on the genotyping array. The genotype concordance rate for those 66 SNPs, which reflects the sum of errors from both sets of genotypes, was 95.4% (Figure S3). We found that LD decay rates using F_{34} array, F_{34} GBS, $F_{39\text{-}43}$ GBS, and F_{34} and $F_{39\text{-}43}$ GBS genotypes were generally similar to one another, though levels of LD using the GBS genotypes appear to be slightly reduced in the later generations of AILs (Figure S4).

GBS genotypes produced more significant associations than array genotypes in F_{34}

We used a linear mixed model (**LMM**) as implemented in GEMMA (Zhou & Stephens, 2012) to perform GWAS. We used the leave-one-chromosome-out (**LOCO**) approach to address the problem of proximal contamination, as previously described (Listgarten et al. 2012; Cheng et al. 2013; Yang et al. 2014; Gonzales et al. 2017). We performed GWAS using both the sparse array SNPs and the dense GBS SNPs to determine whether additional SNPs would produce more genome-wide significant associations. Autosomal and X chromosome SNPs were included in all GWAS. We obtained a significance threshold for each SNP set using MultiTrans (B. Han et al. 2009; Joo et al. 2016). To select independently associated loci ("lead loci"), we used an LD-based clumping method implemented in PLINK to group SNPs that passed the adjusted genome-wide significance thresholds over a large genomic region flanking the index SNP (Purcell et al. 2007). Applying the most stringent clumping parameters (r^2 = 0.1 and sliding window size = 12,150kb, Table S4), we identified 109 significant lead loci in 49 out of 79 F_{34} phenotypes using the GBS SNPs (Table S7). In contrast, we identified 83 significant lead loci in 45 out of 79 F_{34} phenotypes using the sparse array SNPs (Table S6, Table S7). Among the loci identified in the F_{34}, 36 were uniquely identified using the GBS genotypes, whereas 11 were uniquely identified using the array genotypes. These unique loci could be explained by the disparity of the marker density between the GBS and array genotypes. Some unique loci captured haplotype blocks that were not picked up in the other SNP set. Other unique loci were only slightly above the significance threshold in one SNP set where the corresponding loci in the other SNP set had sub-threshold

significance (*i.e.*, p-value ~ 10^{-5} but below the significance threshold of the cohort; Table S7). Overall, GBS SNPs consistently yielded more significant lead loci compared to array SNPs regardless of the clumping parameter values (Table S4), indicating that a dense marker panel was able to detect more association signals compared to a sparse marker panel.

To determine the boundaries of each locus, we performed a Bayesian-framework credible set analysis, which estimated a posterior probability for association at each SNP (r^2 threshold = 0.8, posterior probability threshold = 0.99; (The Wellcome Trust Case Control Consortium et al. 2012)). The physical positions of the SNPs in the credible set were used to determine the boundaries of each locus. As expected, the greater density of the GBS genotypes allowed us to better define each interval. For instance, the lead locus at chr17:27130383 was associated with distance travelled in periphery in the open field test in F_{34} AILs (Figure 2). However, no SNPs were genotyped between 26.7 and 28.7 Mb in the array SNPs, which makes the size of this LD block ambiguous. In contrast, the LocusZoom plot portraying GBS SNPs in the same region shows that SNPs in high LD with chr17:27130383 are between 27 Mb and 28.3 Mb. The more accurate definition of the implicated intervals allowed us to better refine the list of the coding genes and non-coding variants associated with the phenotype (Table S6).

In our prior studies using the sparse marker set, we did not attempt to increase the number of available markers by using imputation. Therefore, we examined whether the disparity between the numbers of loci identified by the two SNP sets could be resolved by imputation, which should increase the number of markers available for GWAS. We used LG/J and SM/J whole genome sequencing data as reference panels (Nikolskiy et

al. 2015) and performed imputation on array and GBS SNPs using Beagle v4.1 (Browning & Browning, 2007). After QC filtering, we obtained 4.3M SNPs imputed from the array SNPs and 4.1M SNPs imputed from the GBS SNPs. More imputed GBS SNPs were filtered out because GBS SNPs were called from genotype probabilities, thus introducing uncertainty in imputed SNPs. We found that imputed array genotypes and imputed GBS genotypes did not meaningfully increase the number of loci discovered, presumably because the utility of imputation is inherently limited in a two-strain cross.

Under a polygenic model where a large number of additive common variants contribute to a complex trait, heritability estimates could be higher when more SNPs are considered (Yang et al. 2017). Given that there were more GBS SNPs than array SNPs, we used autosomal SNPs to examine whether GBS SNPs would generate higher SNP heritability estimates compared to the sparse array SNPs. Heritability estimates were similar for the two SNP sets, with the exception of agouti coat color, which showed marginally greater heritability for the GBS SNPs (Figure S5; Table S8). Our results show that while the denser GBS SNP set was able to identify more genome-wide significant loci, greater SNP density did not improve the polygenic signal.

Partial replication of loci identified in F_{34} or $F_{39\text{-}43}$ and mega-analysis

We identified 25 genome-wide significant loci for 21 phenotypes in the $F_{39\text{-}43}$ cohort (Table S9). A subset of those traits: coat color, body weight, and locomotor activity, were also phenotyped in the F_{34} AILs. To assess replication, we determined whether the loci that were significant in one cohort (either F_{34} or $F_{39\text{-}43}$) would also be significant in the other. We termed the cohort in which a locus was initially discovered as its “discovery set”

and the cohort we attempted replication in as the “replication set” (Table 1). Coat color phenotypes (both albino and agouti) are Mendelian traits and thus served as positive controls. All coat color and body weight loci were replicated. The three body weight loci identified in the F_{34} were replicated at nominal levels of significance ($p<0.05$) in $F_{39\text{-}43}$; similarly, one body weight locus identified in $F_{39\text{-}43}$ was replicated in F_{34} ($p<0.05$). However, none of the locomotor activity loci were replicated in the reciprocal (replication) cohorts.

We found that using a broader definition of an association region rather than a single SNP did not improve replication between the F_{34} cohort and the $F_{39\text{-}43}$ cohorts. Confidence intervals (e.g., (Baud et al. 2013; Nicod et al. 2016)) and the LOD support interval (Conneally et al. 1985; Lander & Botstein, 1989) have been used to define a QTL. LOD support interval is very sensitive to the density of the SNPs where sparse markers would produce misleadingly large support intervals. In contrast, the credible set interval is an estimate of the posterior probability for association at markers neighboring the discovery SNP, and thus defines the size of the association region. As a result, we extended the replication comparison from the discovery SNP position to the credible set interval. We found that in the replication cohort, the p-value at the discovery SNP and that at the top SNP within the credible set interval (defined by the discovery QTL) were generally similar (Table S10). The replication of the locus chr14.79312393 (discovered in the F_{34} cohort) in the $F_{39\text{-}43}$ cohort was more successful using the discovery QTL region defined by the credible set interval; the p-value at the top SNP within the credible set interval was noticeably more significant (chr14.82586326; p-value = 1.48×10^{-6}) than the p-value at the discovery SNP (chr14.79312393; p-value = 0.0237; Table S10). Our results

suggest that for the most part, the discovery SNP accurately represented the association strength of the loci, presumably because of its strong linkage with the neighboring SNPs. In our case, defining a QTL region by the credible set interval did not increase the count of replicated sites between the two cohorts.

We then considered the more liberal “sign test”, a statistical method to test for consistent differences between pairs of observations, to determine whether the directions of the effect (beta) of the coat color, body weight and activity loci were in the same direction between the discovery and replication cohorts. Specifically, we compared whether the sign (direction) of the beta estimates are consistently above or below zero. We found that 11 of 12 comparisons passed this much less stringent test of replication. The one locus (at chr15.67627183) that did not pass the sign test was the locomotor locus “discovered” in $F_{39\text{-}43}$ (Table 1).

In light of the failure to replicate the locomotor activity findings, we conducted a series of 2,500 simulations per trait to estimate the expected power of our replication cohorts. For each phenotype we used the kinship relatedness matrix and variance components estimated from the replication set. For the coat color traits, we found that we had 100% power to replicate the association at either genome-wide significant levels or the more liberal $p<0.05$ threshold (Figure S6). For body weight and locomotor activity, power to replicate at a genome-wide significance threshold ranged from 20% to 85%, whereas power to replicate at the $p<0.05$ threshold was between 80% and 100% (Figure S6). These power estimates were inconsistent with our empirical observations for the locomotor activity traits, none of which replicated at even the $p<0.05$ threshold, where we should have had almost 100% power (Table 1; Figure S6). However, our power

simulations did not account for Winner's Curse (Zöllner & Pritchard, 2007) or study-specific heterogeneity (Zou et al. 2019).

To determine whether these factors could explain the lower than expected rate of replication, we applied a statistical framework that jointly models Winner's Curse and study-specific heterogeneity in two GWAS studies of the same phenotype (Zou et al. 2019). This framework proposes two random effects models. The first model (**WC**) only accounts for Winner's Curse, while the second model accounts for both Winner's Curse and study-specific heterogeneity due to confounding (**WC+C**). In this context, we define confounding as any biological or technical effect present in one study but not the other. We applied each of these models once using F_{34} as the discovery study and once using $F_{39\text{-}43}$ as the discovery study. The models can be used to assess how well Winner's Curse explains the observed levels of replication. For example, when F_{34} is used as the replication study for the albino coat color phenotype, the expected value of the replication summary statistics after accounting for Winner's Curse is the same as the expected value after accounting for Winner's Curse and confounding (Figure S7). While the 95% confidence intervals for the **WC+C** model are larger than the **WC** model, both models seem to explain the observed data well. However, when F_{34} is used as the discovery study for the locomotor activity on day 1 or body weight, the **WC+C** model explains the data better than the **WC** model.

In order to quantitatively assess how well each of these models explain the observed patterns of replication, we computed the predicted replication rates under each model (Methods) and compared these with the empirical replication rates. In this analysis, we defined the empirical replication rate to be the proportion of variants passing the genome-wide significance threshold in the discovery study that also pass the Bonferroni corrected threshold in the replication study. We used this definition of replication for this analysis instead of replication of lead SNPs to allow for a larger number of variants to be included in the model fitting process. For all phenotypes tested, the **WC** model predicts that all the variants passing the genome-wide significance threshold in the discovery study should pass the Bonferroni corrected threshold in the replication study, which is dramatically different from the observed replication of body weight and locomotor activity on day 1 and 2 phenotypes (Table 2). While the replication in the agouti coat color phenotype is not well predicted by the **WC+C** model, this may be due to the fact that the agouti phenotype is a dominant trait, while our model assumes additive allele effects. These results suggest that the sample sizes are sufficiently large that Winner's Curse cannot account for the lack of replication. However, in these cases, the **WC+C** model has predicted replication rates that are much closer to the true (observed) values, indicating that the lack of replication in these phenotypes is more likely to be due to study-specific heterogeneity that is potentially caused by confounding.

We evaluated whether or not the traits showed genetic correlations across the two cohorts. High genetic correlations would indicate a high degree of additive genetic effect that is shared between the two cohorts, and the low genetic correlations would indicate limited potential for replication. We used all autosomal SNPs to calculate genetic

correlations between the F_{34} and $F_{39\text{-}43}$ generations for body weight, coat color, and locomotor activity phenotypes (Table S11), using GCTA-GREML (Yang et al. 2011). Albino and agouti coat color, body weight and locomotor activity on days 1 and 2 were highly genetically correlated (r_Gs >0.7; Table S11). In contrast, locomotor activity on day 3 showed a significant but weaker genetic correlation (r_G=0.577), perhaps reflecting variability in the quality of the methamphetamine injection, which were only given on day 3. Overall, these results suggest that genetic influences on these traits were largely similar in the two cohorts; however, the genetic correlations were less than 1, suggesting an additional barrier to replication that was not accounted for in our power simulations.

We also calculated the SNP heritability for all traits using GCTA. SNP heritability was consistently lower in the $F_{39\text{-}43}$ cohort compared to the F_{34} cohort, including the Mendelian traits of coat color. The ± 1 × standard error intervals of the F_{34} and $F_{39\text{-}43}$ SNP heritability estimates for the coat color trait albino overlapped. This observation indicates that SNP heritability for albino in the two cohorts is comparable. In contrast, the ± 1 × standard error intervals of the F_{34} and $F_{39\text{-}43}$ SNP heritability estimates for the coat color trait agouti did not overlap. We could not explain the differential SNP heritability for the binary trait agouti in the two cohorts. The lower SNP heritability in $F_{39\text{-}43}$ for the rest of the quantitative traits could be a result of increased experimental variance (Figure 3; Table S12; (Falconer, 1960; Lynch & Walsh, 1996; Mhyre et al. 2005; Zöllner & Pritchard, 2007; Visscher et al. 2008; Zaitlen & Kraft, 2012)).

Due to the relatively high genetic correlations (Table S11), we suspected that a mega-analysis using the combined sample set would allow for the identification of additional loci; indeed, mega-analysis identified four novel genome-wide significant

associations (Figure 4; Table S13). The significance level of five out of six loci identified by the mega-analysis was greater than that in either individual cohort. For instance, the p-values obtained by mega-analysis for chr14:82672838 (p-value = 7.93×10^{-9}) for body weight were lower than the corresponding p-values for the same loci for F_{34} (chr14:79312393, p-value = 7.53×10^{-6}) and $F_{39\text{-}43}$ (chr14.82586326, p-value = 2.63×10^{-6}; Table S13; Table 1).

1.5 Discussion

We used F_{34} and $F_{39\text{-}43}$ generations of a LG/J x SM/J AIL to perform GWAS, SNP heritability estimates, genetic correlations, replication and mega-analysis. We had previously performed several GWAS using a sparse marker set in the F_{34} cohort. In this study we used a denser set of SNPs, obtained using GBS, to reanalyze the F_{34} cohort. We found 109 significant loci, 36 of which had not been identified in our prior studies using the sparse marker set. We used a new, previously unpublished $F_{39\text{-}43}$ cohort for GWAS and showed that genetic correlations were high for the subset of traits that were measured in both cohorts. Despite this, we found that many loci were not replicated between cohorts, even when we used a relatively liberal definition of replication ($p<0.05$). The failure to replicate some of our findings was not predicted by our power simulations. Therefore, we performed an analysis to determine whether Winner's Curse and study-specific heterogeneity could account for the lower than expected replication rate. Winner's Curse alone could not explain the failure to replicate. However, modeling both Winner's Curse and study-specific heterogeneity better explained the observed replication rate. Finally, mega-analysis of the two cohorts allowed us to discover four additional loci. Taken together, our results provide a set of refined regions of association for numerous physiological and behavioral traits in multiple generations of AILs. These loci could serve as benchmarks for future GWAS results in intercross mouse lines. More broadly, this study illustrates the difficulty of replication even when using a highly controlled model system.

Previous publications from our lab used a sparse set of array genotypes for GWAS of various behavioral and physiological traits in 688 F_{34} AILs (Cheng et al. 2010;

Lionikas et al. 2010; Samocha et al. 2010; Parker et al. 2011, 2014; Carroll et al. 2017; Hernandez Cordero et al. 2018; Gonzales et al. 2018). In this study we obtained a much denser marker set for 428 of the initial 688 AIL mice using GBS. The denser genotypes allowed us to identify most of the loci obtained using the sparse set, as well as many additional loci. For instance, using the sparse markers we identified a significant locus on chromosome 8 for locomotor day 2 activity that contained only one gene: Csmd1 (CUB and sushi multiple domains 1). Gonzales et al. (Gonzales et al. 2018) replicated this finding in F_{50-56} AILs and identified a *cis*-eQTL mapped to the same region. Csmd1 mutant mice showed increased locomotor activity compared to wild-type and heterozygous mice, indicating that Csmd1 is likely a causal gene for locomotor and related traits (Gonzales et al. 2018). We replicated this locus in the analysis of the F_{34} cohort that used the denser marker set (Figure S8). We also replicated a locus on chromosome 17 for distance traveled in the periphery in the open field test (Figure 4; (Parker et al. 2014)), three loci on chromosomes 4, 6, and 14 for body weight (Figure S8; (Parker et al. 2011)), one locus on chromosome 7 for mean corpuscular hemoglobin concentrations (MCHC, complete blood count; Figure S8; (Bartnikas et al. 2012)), and numerous loci on chromosome 4, 6, 7, 8, and 11 for muscle weights (Figure S8; (Lionikas et al. 2010)). We noticed that even using original sparse markers, some previously published loci were not replicated in the current GWAS. The most likely explanation is that we had only 428 of the 688 mice used in the previous publications.

QTL mapping studies have traditionally used a 1.0~2.0 LOD support interval to approximate the size of the association region (see (Cervino et al. 2005; Logan et al. 2013)). The LOD support interval, proposed by Conneally et al. (Conneally et al.

1985) and Lander & Botstein (Lander & Botstein, 1989), is a simple confidence interval method involving converting the p-value of the peak locus into a LOD score, subtracting "drop size" from the peak locus LOD score, and finding the two physical positions to the left and to the right of the peak locus location that correspond to the subtracted LOD score. Although Mangin et al. (Mangin et al. 1994) showed via simulation that the boundaries of LOD support intervals depend on effect size, others observed that a 1.0 ~ 2.0 LOD support interval accurately captures ~95% coverage of the true location of the loci when using a dense set of markers (Lander & Botstein, 1989; Dupuis & Siegmund, 1999; Manichaikul et al. 2006). In the present study, we considered using LOD support intervals but found that the sparse array SNPs produced misleadingly large support intervals. Various methods have been proposed for calculating confidence intervals in analogous situations (e.g. (Baud et al. 2013; Nicod et al. 2016)). We performed credible set analysis and compared LocusZoom plots of the same locus region between array SNPs and the GBS SNPs (Figure S8; (Pruim et al. 2010)). For example, the benefit of the denser SNP coverage is easily observed in the locus on chromosome 7 (array lead SNP chr7:44560350; GBS lead SNP chr7:44630890) for the complete blood count trait "retic parameters cell hemoglobin concentration mean, repeat"; denser SNPs delineate the start and the end of an association block much more clearly. Thus, there are advantages of dense SNP sets that go beyond the ability to discover additional loci.

LD in the LG/J x SM/J AIL mice is more extensive than in the Diversity Outbred mice and Carworth Farms White mice (Parker et al. 2016). Some of the loci that we identified are relatively broad, making it difficult to infer which genes are responsible for the association. We focused on loci that contained 5 or fewer genes (Table S6).

We highlight a few genes that are supported by the existing literature for their role in the corresponding traits. The lead SNP at chr1:77255381 is associated with tibia length in F_{34} AILs (Table S6; S8 Fig). One gene at this locus, *EphA4*, codes for a receptor for membrane-bound ephrins. *EphA4* plays an important role in the activation of the tyrosine kinase Jak2 and the signal transducer and transcriptional activator *Stat5B* in muscle, promoting the synthesis of insulin-like growth factor 1 (IGF-1) (Lai et al. 2004; Hyun, 2013; Sawada et al. 2017). Mice with mutated *EphA4* shows significant defect in body growth (Hyun, 2013). Curiously, another gene at this locus, *Pax3*, has been shown as a transcription factor expressed in resident muscle progenitor cells and is essential for the formation of skeletal muscle in mice (Relaix et al. 2006). It is possible that both *EphA4* and *Pax3* are associated with the trait tibia length because they are both involved in organismal growth. Another region of interest is the locus at chr4:66866758, which is associated with body weight (Table S6; Table S13). The lead SNP is immediately upstream of *Tlr4*, Toll-like receptor 4, which recognizes Gram-negative bacteria by its cell wall component, lipopolysaccharide (Hoshino et al. 1999; Takeuchi et al. 1999). TLR4 responds to the high circulating level of fatty acids and induces inflammatory signaling, which leads to insulin resistance (Shi et al. 2006). Kim et al showed TLR4-deficient mice were protected from the increase in proinflammatory cytokine level and gained less weight than wild-type mice when fed on high fat diet (Kim et al. 2012). The association between *Tlr4* and body weight in the AILs corroborates these findings.

We considered both the F_{34} and the $F_{39\text{-}43}$ as both “discovery” and “replication” cohorts. Significant loci for coat color, which are monogenic and served as positive

controls, were replicated, between the two cohorts, as expected. One locus for body weight was replicated ($p<0.05$) between F_{34} and $F_{39\text{-}43}$. However, the loci for locomotor activity were not replicated. Power analyses predicted a much higher rate of replication, which led us to conduct additional analyses to better understand the lower than expected rate of replication.

First, we used a newly introduced method to determine whether Winner's Curse (Zöllner & Pritchard, 2007)) which has also been termed the Beavis Effect (Beavis et al. 1991, 1994; Xu, 2003; King & Long, 2017; Keele et al. 2019; Paterson, 2019) could account for the lower than expected rate of replication. Beavis' original report described a lack of replication of QTL for agronomic traits between small populations of maize (Beavis et al. 1991). Using progeny sizes ranging from 100 to 1000, Beavis simulated interval mapping to evaluate the accuracy of the estimates of phenotypic variance explained at the statistically significant QTL (Beavis et al. 1994; Xu, 2003; Paterson, 2019). Simulations showed that progeny sizes greatly influenced the estimates of phenotypic variance explained; smaller progeny sizes (n=100) generated highly overestimated estimates of phenotypic variances, whereas larger progeny sizes (n=1000) generated estimates of phenotypic variances similar to the actual value (Paterson, 2019; Xu, 2003). King and Long (King & Long, 2017) further examined the Beavis Effect in the next-generation mapping populations in Drosophila melanogaster. The authors found that sample size was the major determinant for the overestimation of phenotypic variance explained at the significant QTL in both the GWAS-based Drosophila Genetic Reference Panel (DGRP) and the multi-parental Drosophila Synthetic Population Resource (DSPR). When sample size remained constant and the true phenotypic variance explained at the

significant QTL was small, the estimation bias was more pronounced. In contrast, estimates for the phenotypic variance explained at all simulated QTL, significant or not, were generally centered at the true values. In an analogous study of power and replication in Collaborative Cross mice, Keele et al. (Keele et al. 2019) found that the Beavis Effect was most striking when the number of strains and true effect size of the QTL were small. This estimation bias indicates that mapping statistically significant QTL across experiments, populations, and panels can be problematic (Macdonald & Long, 2004; Gruber et al. 2007; Najarro et al. 2015). The analyses we performed indicated that Winner's Curse alone could not explain the lack of replication, but a model that also included study-specific heterogeneity could.

Our analysis does cannot explain the source of the study-specific heterogeneity. Possible sources of confounding could include maternal effects, which could differentiate the F_{34} cohort and the $F_{39\text{-}43}$ cohort because F_{33} animals were transported to the University of Chicago from Washington University in St. Louis. In contrast, the breeder animals of the $F_{39\text{-}43}$ cohort have already acclimated to the environment for multiple generations. Another possible source of confounding is that the phenotyping of the $F_{39\text{-}43}$ occurred over five generations (more than a year) during which time numerous environmental factors may have changed (e.g. several technicians performed the data collection). Such factors are known to be an important potential source of confounding; (Falconer, 1960; Lynch & Walsh, 1996; Crabbe et al. 1999; Mhyre et al. 2005; Visscher et al. 2008; Zaitlen & Kraft, 2012; Sorge et al. 2014). Our analyses did not correct for the fact that six phenotypes were examined, thus somewhat increasing the chances that at

least one of our significant associations could have been a false positive that would not be expected to replicate.

Interestingly, we found that the genetic correlations between the discovery and replication samples were relatively high for all traits; however, some traits replicated well and others replicated poorly. Our subsequent analysis showed that study-specific heterogeneity was low for the coat color traits, but higher for the body weight and locomotor traits. This makes an important point, namely that a strong genetic correlation can exists in the presence or absence of study-specific heterogeneity. Finally, it was notable that replication (at $p<0.05$) was relatively successful for body weight, despite the significant evidence of study-specific heterogeneity and low predicted replication (Table 2). Power analyses predicted that the body weight loci should replicate at the genome-wide significance threshold, whereas we only observed replication when at the less stringent $p<0.05$ level (Table 1). The lack of replication at the genome-wide significance threshold for the body weight phenotype was likely due to study-specific heterogeneity due to confounding that was not accounted for in the power analyses. In Table 2, "predicted replication" refers to replication using a Bonferroni significance threshold that accounts for the number of significant SNPs in the discovery study. The low predicted replication rate under the WC+C model for the body weight phenotype is consistent with the low replication (genome-wide) reported in Table 1. Thus, both body weight and locomotor traits were strongly impacted by study specific confounding; however, nominal replication was still possible for body weight but not for the locomotor traits.

Finally, we performed a mega-analysis using F_{34} and $F_{39\text{-}43}$ AIL mice. The combined dataset allowed us to identify four novel genome-wide significant associations

that were not detected in either the F_{34} or the $F_{39\text{-}43}$ cohorts presumably because of the increased sample size in the mega-analysis (Visscher et al. 2017). As is true for all GWAS, the loci identified in the mega-analysis could be false positives.

In addition to performing many GWAS and related analyses that led to the identification of dozens of novel loci for locomotor activity, open field test, fear conditioning, light dark test for anxiety, complete blood count, iron content in liver and spleen, and muscle weight, our study also tested our expectations about replication of GWAS findings. We did not obtain the expected rate of replication. We used a method that can distinguish between Winner's Curse and study-specific heterogeneity to show that the lower than expected rate of replication was due to study-specific heterogeneity. This indicates that study-specific heterogeneity can have a major impact of replication even when in a model system when a genetically identical population is tested under conditions that are designed to be as similar as possible.

1.6 Availability of data and materials

All relevant data are within the paper and its Supporting Information files. Genotypes and phenotypes of F_{34} ("LGSM AI G34 Palmer (Array)": GN655; "LGSM AI G34 Palmer (GBS)": GN656), $F_{39\text{-}43}$ ("LGSM AI G39-43 Palmer (GBS)": GN657), and mega-analysis cohort ("LGSM AI G34 G39-43 Palmer (GBS)": GN654) of AIL are uploaded to GeneNetwork2 (http://gn2.genenetwork.org/). Code used to perform the analyses is included in the supplementary materials on FigShare: https://doi.org/10.25387/g3.11674221.

1.7 Figures

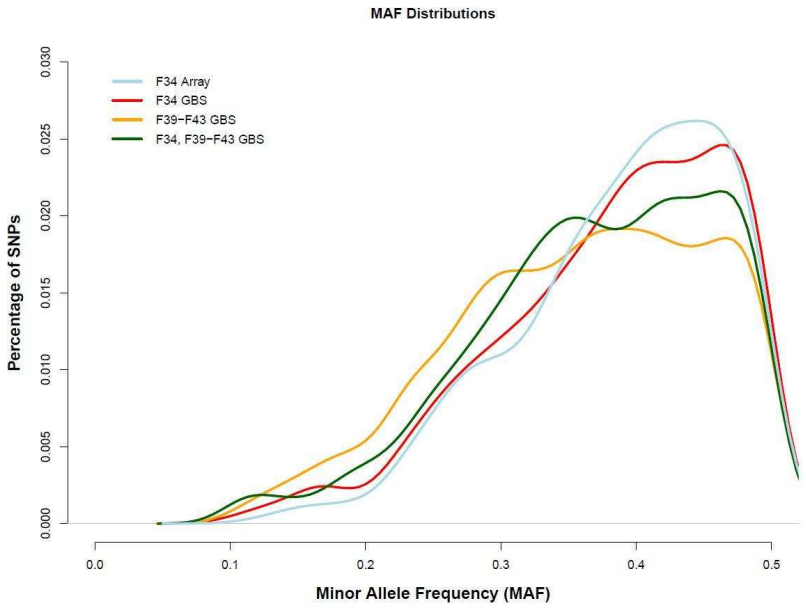


Figure 1.1. Minor allele frequency (MAF) distributions for F_{34} array, F_{34} GBS, F_{39}-F_{43} GBS, and F_{34} and F_{39}-F_{43} GBS SNP sets. The average MAF of those 4,593 array SNPs was 0.388; the average MAF of the 60,392 GBS-derived SNPs in 428 F_{34} AIL mice was 0.382; the average MAF of the 59,790 GBS-derived SNPs in 600 $F_{39\text{-}43}$ AIL mice was 0.358; the average MAF of the 58,461 GBS-derived SNPs that existed in both F_{34} and $F_{39\text{-}43}$ AIL mice was 0.370 (Table S2). MAF distributions are highly comparable between AIL generations.

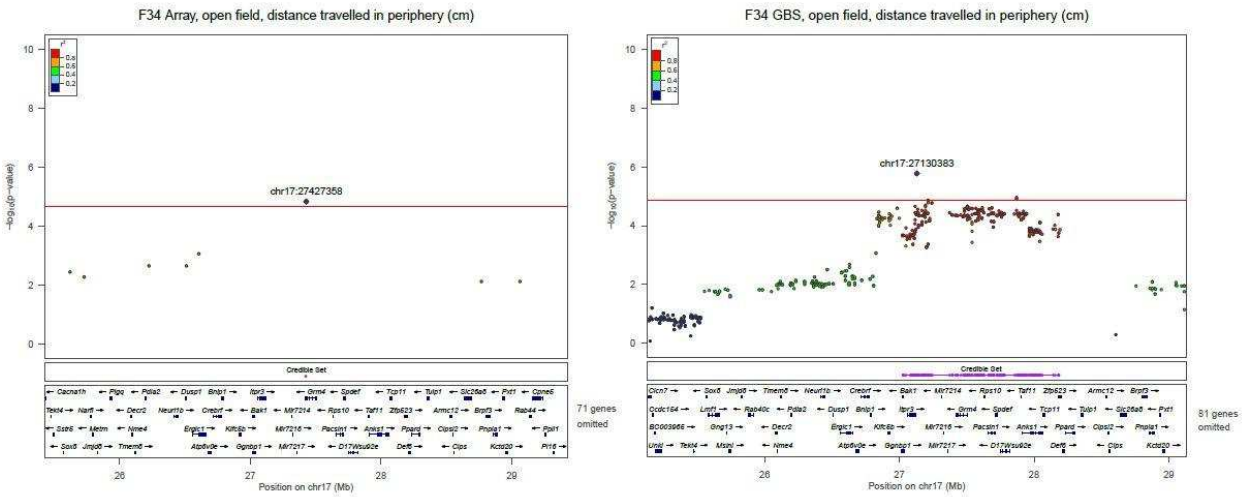


Figure 1.2. Significant loci on chromosome 17 for open field, distance traveled in periphery in F_{34} AIL. As exemplified in this pair of LocusZoom plots, GBS SNPs defined the boundaries of the loci much more precisely than array SNPs. GBS SNPs that are in high LD (r_2 > 0.8, red dots) with lead SNP chr17:27130383 resides between 27 ~ 28.3 Mb. In contrast, too few SNPs are present in the array plot to draw any definitive conclusion about the boundaries or LD pattern in this region. Purple track shows the credible set interval. LocusZoom plots for all loci identified in this paper are in Figure S8.

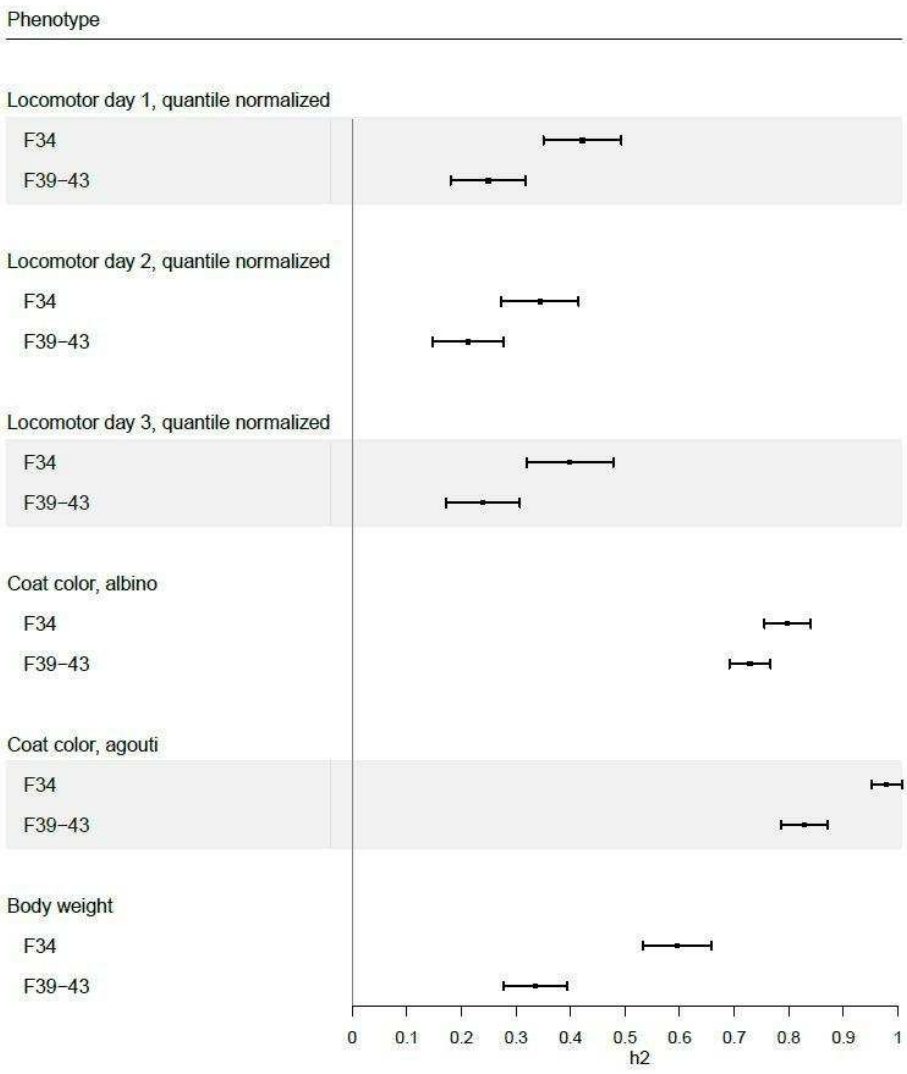


Figure 1.3. SNP-heritability estimates in F_{34} and $F_{39\text{-}43}$ AILs. Square dots represent the SNP heritability estimated by the GCTA-GREML analysis (Yang et al. 2011). The whiskers flanking the square dots show the ± 1 × standard error of the heritability estimate. All heritability estimates are highly significant ($p < 1.0\times10^{-05}$; see Table S12).

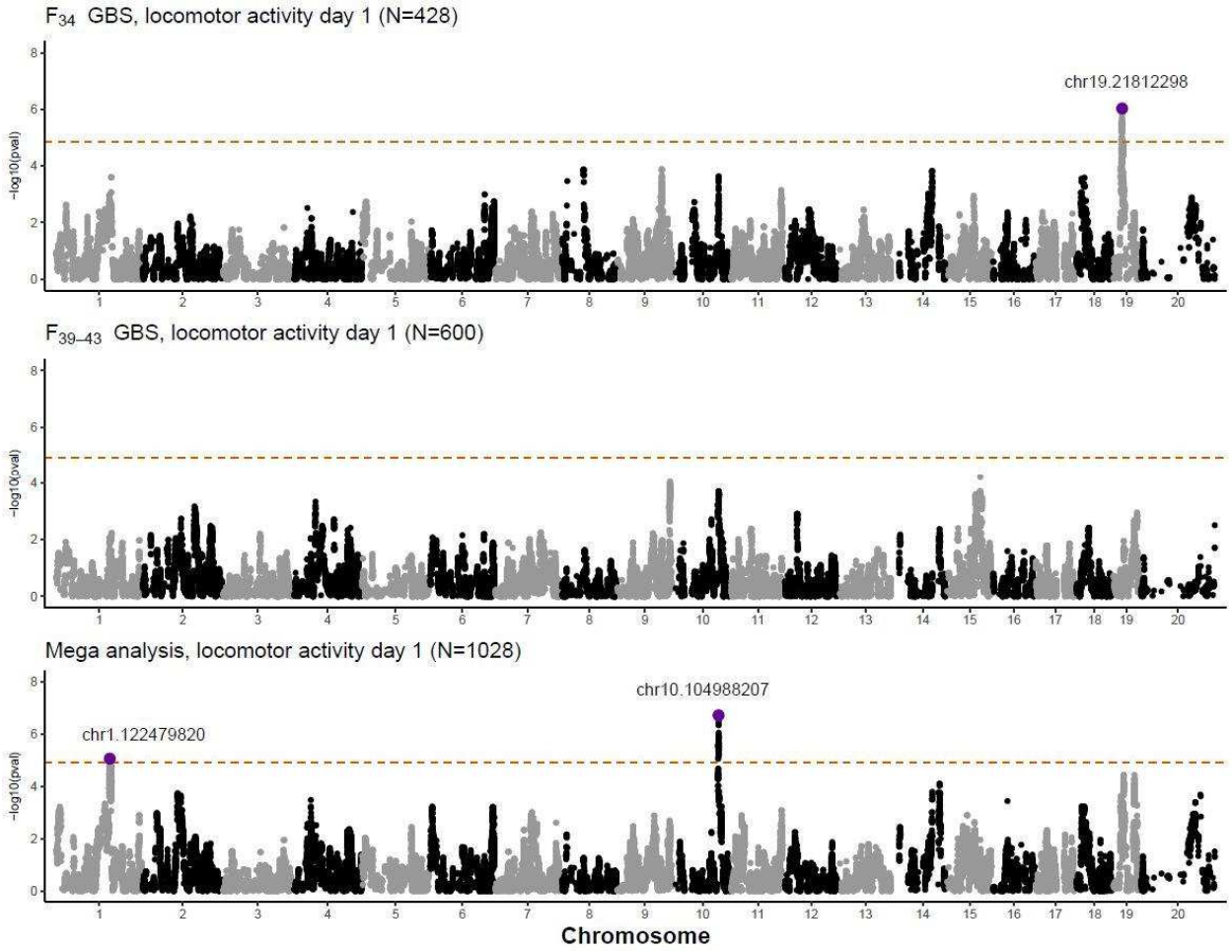


Figure 1.4. Manhattan plots comparing F_{34} GBS, $F_{39\text{-}43}$ GBS, and mega-analysis on locomotor day 1 test using 57,170 shared SNPs in all AIL generations. We performed mega-analysis of F_{34} and $F_{39\text{-}43}$ animals (n=1,028) for body weight, coat color, and locomotor activity, the set of traits that were measured in the same way in both cohorts.

1.8 Tables

Table 1.1. Replication of significant SNPs between F_{34} and $F_{39\text{-}43}$ AIL association analyses.

Table 1 Replication of significant SNPs between F_{34} and $F_{39\text{-}43}$ AIL association analyses. "Discovery set" indicates the AIL generation that significant SNPs were identified. "Replication set" shows the association p-value, β estimates, etc. of the "discovery set" significant SNPs in the replication AIL generation. SNPs that replicated ($p < 0.05$, same sign for the beta) between F_{34} and $F_{39\text{-}43}$ are in bold italics, SNPs that replicated at the genome-wide threshold (see Table S5) are bold, italic and underlined. Genetic correlations (rG) for phenotypes measured in both F_{34} and $F_{39\text{-}43}$ are listed (see also Table S11)

			Discovery set					Replication set				
Phenotype	rG(s.e.)	SNP	P	-log10(p)	af	beta	se	P	-log10(p)	af	beta	se
			F_{34} GBS					$F_{39\text{-}43}$ GBS replicate				
Body weight	0.711(0.25)*	***chr4.66414508***	***8.58×10^{-8}***	***7.07***	0.419	−0.25	0.05	***3.55×10^{-3}***	***2.45***	0.406	−0.13	0.04
		chr6.81405109	***6.22×10^{-6}***	***5.21***	0.497	0.21	0.05	***3.52×10^{-2}***	***1.45***	0.518	0.09	0.04
		chr14.79312393	***7.45×10^{-6}***	***5.13***	0.514	−0.20	0.04	***2.37×10^{-2}***	***1.63***	0.566	−0.10	0.04
Coat color, albino	0.967(0.04)*	***chr7.87642045***	***5.00×10^{-106}***	***105.30***	0.432	−0.58	0.02	***$\underline{1.59\times10^{-162}}$***	***161.80***	0.388	−0.61	0.02
Coat color, agouti	0.971(0.04)*	***chr2.154464466***	***9.43×10^{-191}***	***190.03***	0.129	0.94	0.01	***$\underline{5.7\times10^{-93}}$***	***92.24***	0.207	0.72	0.03
Locomotor test day 1, total distance traveled in 30min	0.968(0.24)*	chr19.21812298	3.98×10^{-7}	6.40	0.461	−0.36	0.07	4.55×10^{-1}	0.342	0.502	−0.05	0.06
Locomotor test day2, total distance traveled in 30min	0.988(0.19)*	chr8.17410225	5.65×10^{-6}	5.248	0.171	0.42	0.09	8.34×10^{-1}	0.079	0.202	0.02	0.08
			$F_{39\text{-}43}$ GBS					F_{34} GBS replicate				
Body weight	0.711(0.25)*	chr1.89192209	6.42×10^{-6}	5.19	0.22	0.22	0.05	5.16×10^{-2}	1.29	0.276	0.10	0.05
		chr14.82586326	***1.48×10^{-6}***	***5.83***	0.658	−0.22	0.04	***3.08×10^{-5}***	***4.51***	0.575	−0.19	0.05
Coat color, albino	0.967(0.04)*	***chr7.87255156***	***3.37×10^{-146}***	***165.47***	0.389	−0.62	0.02	***$\underline{7.80\times10^{-97}}$***	***96.11***	0.444	−0.57	0.02
Coat color, agouti	0.971(0.04)*	***chr2.155091628***	***1.78×10^{-115}***	***114.75***	0.218	0.74	0.02	***$\underline{1.51\times10^{-185}}$***	***184.82***	0.135	0.90	0.01
Locomotor test day 2, total distance traveled in 30min	0.988(0.19)*	chr15.67627183	3.33×10^{-6}	5.478	0.461	0.30	0.06	2.07×10^{-1}	0.683	0.522	−0.08	0.07

Table 1.2. Predicted replication rates.

Table 2 Predicted replication rates. We applied the replication analysis to phenotypes with at least two genome-wide significant variants in the discovery study. These phenotypes include body weight, albino coat color, agouti coat color, locomotor test day 1, and locomotor test day 2. We computed the true replication rate as the fraction of variants that were genome-wide significant in the discovery study that also passed the Bonferroni significance threshold in the replication study ("Empirical replication rate"). The model accounting for Winner's Curse and confounding ("Predicted replication rate WC+C") explains the true replication rate more accurately than the model accounting for only Winner's Curse ("Predicted replication rate WC")

Discovery set	Replication set	Phenotype	Empirical replication rate	Predicted replication rate (WC)	Predicted replication rate (WC+C)
F_{34} GBS	$F_{39\text{-}43}$ GBS	Body weight	0.009	1.000	0.044
		Coat color, albino	1.000	1.000	0.997
		Coat color, agouti	0.932	1.000	0.577
		Locomotor test day 1	0.000	1.000	0.028
		Locomotor test day 2	0.000	1.000	0.140
$F_{39\text{-}43}$ GBS	F_{34} GBS	Body weight	0.297	1.000	0.071
		Coat color, albino	0.911	1.000	0.932
		Coat color, agouti	0.815	1.000	0.925
		Locomotor test day 2	0.000	1.000	0.053

1.9 Author Contributions

AAP and XZ designed the study, oversaw data collection and analysis, and co-wrote the manuscript. XZ imputed genotypes, performed SNP- and subject-level QC, and conducted GWAS in F_{34} and $F_{39\text{-}43}$ AILs under supervision of AAP. CS prepared GBS libraries for sequencing, as well as organizing portions of the $F_{39\text{-}43}$ phenotypes. NMG de-multiplexed GBS sequencing results and performed alignment and variant calling. JZ performed statistical analyses to model replication between F_{34} and $F_{39\text{-}43}$ cohorts. RC helped with kinship relatedness matrix calculated from AIL pedigree and with power analysis. AC provided technical support for running programs and scripts. GS oversaw breeding and phenotyping of the $F_{39\text{-}43}$. We would like to recognize Jackie Lim and Kaitlin Samocha for collecting F_{34} AIL phenotype data and Ryan Walters for collecting $F_{39\text{-}43}$ AIL phenotype data. We wish to acknowledge Alex Gileta for input on a draft of this manuscript. All procedures were approved by the Institutional Animal Care and Use Committee (IACUC protocol: S15226) Euthanasia was accomplished using CO2 asphyxiation followed by cervical dislocation.

1.10 Acknowledgments

Chapter 1, in full, is a reprint of the material as it appears in G3 2020. Xinzhu Zhou, Celine L. St. Pierre, Natalia M. Gonzales, Jennifer Zou, Riyan Cheng, Apurva S. Chitre, Greta Sokoloff, Abraham A. Palmer. Genome-wide Association Study in Two Cohorts from a Multi-generational Mouse Advanced Intercross Line Highlights the Difficulty of Replication Due to Study-specific Heterogeneity. G3: GENES, GENOMES, GENETICS, 2020, 10(3):951-965; https://doi.org/10.1534/g3.119.400763. The dissertation author was the primary investigator and author of this paper.

1.11 References

Abney, M. (2008). Identity-by-descent estimation and mapping of qualitative traits in large, complex pedigrees. *Genetics*, *179*(3), 1577–1590. https://doi.org/10.1534/genetics.108.089912

Bartnikas, T. B., Parker, C. C., Cheng, R., Campagna, D. R., Lim, J. E., Palmer, A. A., & Fleming, M. D. (2012). QTLs for murine red blood cell parameters in LG/J and SM/J F2 and advanced intercross lines. *Mammalian Genome*, *23*(5–6), 356–366. https://doi.org/10.1007/s00335-012-9393-3

Baud, A., Guryev, V., Hummel, O., Johannesson, M., Hermsen, R., Stridh, P., Graham, D., McBride, M. W., Foroud, T., & Calderari, S. (2014). Genomes and phenomes of a population of outbred rats and its progenitors. *Scientific Data*, *1*, 140011.

Baud, A., Hermsen, R., Guryev, V., Stridh, P., Graham, D., McBride, M. W., Foroud, T., Calderari, S., Diez, M., & Sequencing, R. G. (2013). Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nature Genetics*, *45*(7), 767.

Beavis, W. D., Grant, D., Albertsen, M., & Fincher, R. (1991). Quantitative trait loci for plant height in four maize populations and their associations with qualitative genetic loci. *Theoretical and Applied Genetics*, *83*(2), 141–145. https://doi.org/10.1007/BF00226242

Beavis, W. D., Smith, O. S., Grant, D., & Fincher, R. (1994). Identification of Quantitative Trait Loci Using a Small Sample of Topcrossed and F4 Progeny from Maize. *Crop Science*, *34*(4), 882–896. https://doi.org/10.2135/cropsci1994.0011183X003400040010x

Besnier, F., Wahlberg, P., Rönneg\a ard, L., Ek, W., Andersson, L., Siegel, P. B., & Carlborg, O. (2011). Fine mapping and replication of QTL in outbred chicken advanced intercross lines. *Genetics Selection Evolution*, *43*(1), 3.

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097.

Carbonetto, P., Cheng, R., Gyekis, J. P., Parker, C. C., Blizard, D. A., Palmer, A. A., & Lionikas, A. (2014). Discovery and refinement of muscle weight QTLs in B6$\times$ D2 advanced intercross mice. *Physiological Genomics*, *46*(16), 571–582.

Carroll, A. M., Cheng, R., Collie-Duguid, E. S. R., Meharg, C., Scholz, M. E., Fiering, S., Fields, J. L., Palmer, A. A., & Lionikas, A. (2017). Fine-mapping of genes

determining extrafusal fiber properties in murine soleus muscle. *Physiological Genomics*, *49*(3), 141–150.

Cervino, A. C., Li, G., Edwards, S., Zhu, J., Laurie, C., Tokiwa, G., Lum, P. Y., Wang, S., Castellini, L. W., & Lusis, A. J. (2005). Integrating QTL and high-density SNP analyses in mice to identify Insig2 as a susceptibility gene for plasma cholesterol levels. *Genomics*, *86*(5), 505–517.

Cheng, R., Abney, M., Palmer, A. A., & Skol, A. D. (2011). QTLRel: An R package for genome-wide association studies in which relatedness is a concern. *BMC Genetics*, *12*, 66. https://doi.org/10.1186/1471-2156-12-66

Cheng, R., Lim, J. E., Samocha, K. E., Sokoloff, G., Abney, M., Skol, A. D., & Palmer, A. A. (2010). Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics*.

Cheng, R., Parker, C. C., Abney, M., & Palmer, A. A. (2013). Practical considerations regarding the use of genotype and pedigree data to model relatedness in the context of genome-wide association studies. *G3: Genes, Genomes, Genetics*, g3–113.

Chesler, E. J. (2014). Out of the bottleneck: The Diversity Outcross and Collaborative Cross mouse populations in behavioral genetics research. *Mammalian Genome*, *25*(1–2), 3–11.

Churchill, G. A., Gatti, D. M., Munger, S. C., & Svenson, K. L. (2012). The diversity outbred mouse population. *Mammalian Genome*, *23*(9–10), 713–718.

Cockram, J., & Mackay, I. (2018). *Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants*.

Conneally, P. M., Edwards, J. H., Kidd, K. K., Lalouel, J.-M., Morton, N. E., Ott, J., & White, R. (1985). Report of the committee on methods of linkage analysis and reporting. *Cytogenetic and Genome Research*, *40*(1–4), 356–359. https://doi.org/10.1159/000132186

Consortium, C. C. (2012). The genome architecture of the Collaborative Cross mouse genetic reference population. *Genetics*, *190*(2), 389–401.

Coyner, J., McGuire, J. L., Parker, C. C., Ursano, R. J., Palmer, A. A., & Johnson, L. R. (2014). Mice selectively bred for High and Low fear behavior show differences in the number of pMAPK (p44/42 ERK) expressing neurons in lateral amygdala following Pavlovian fear conditioning. *Neurobiology of Learning and Memory*, *112*, 195–203.

Crabbe, J. C., Wahlsten, D., & Dudek, B. C. (1999). Genetics of mouse behavior: Interactions with laboratory environment. *Science (New York, N.Y.)*, *284*(5420), 1670–1672. https://doi.org/10.1126/science.284.5420.1670

Darvasi, A., & Soller, M. (1995). Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics*, *141*(3), 1199–1207.

Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, *12*(7), 499.

Demarest, K., Koyner, J., McCaughran, J., Cipp, L., & Hitzemann, R. (2001). Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. *Behavior Genetics*, *31*(1), 79–91.

Diouf, I. A., Derivot, L., Bitton, F., Pascual, L., & Causse, M. (2018). Water Deficit and Salinity Stress Reveal Many Specific QTL for Plant Growth and Fruit Quality Traits in Tomato. *Frontiers in Plant Science*, *9*, 279.

Doitsidou, M., Jarriault, S., & Poole, R. J. (2016). Next-generation sequencing-based approaches for mutation mapping and identification in Caenorhabditis elegans. *Genetics*, *204*(2), 451–474.

Dupuis, J., & Siegmund, D. (1999). Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics*, *151*(1), 373–386.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Falconer, D. S. (1960). Introduction to quantitative genetics. *Introduction to Quantitative Genetics.* https://www.cabdirect.org/cabdirect/abstract/19601603365

Fitzpatrick, C. J., Gopalakrishnan, S., Cogan, E. S., Yager, L. M., Meyer, P. J., Lovic, V., Saunders, B. T., Parker, C. C., Gonzales, N. M., & Aryee, E. (2013). Variation in the form of Pavlovian conditioned approach behavior among outbred male Sprague-Dawley rats from different vendors and colonies: Sign-tracking vs. goal-tracking. *PloS One*, *8*(10), e75042.

Gardenghi, S., Marongiu, M. F., Ramos, P., Guy, E., Breda, L., Chadburn, A., Liu, Y., Amariglio, N., Rechavi, G., & Rachmilewitz, E. A. (2007). Ineffective erythropoiesis in β-thalassemia is characterized by increased iron absorption mediated by down-regulation of hepcidin and up-regulation of ferroportin. *Blood*, *109*(11), 5027–5035.

Gatti, D. M., Svenson, K. L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., & Palmer, A. (2014). Quantitative trait locus mapping

methods for diversity outbred mice. *G3: Genes, Genomes, Genetics*, *4*(9), 1623–1633.

Ghazalpour, A., Doss, S., Kang, H., Farber, C., Wen, P.-Z., Brozell, A., Castellanos, R., Eskin, E., Smith, D. J., & Drake, T. A. (2008). High-resolution mapping of gene expression using association in an outbred mouse stock. *PLoS Genetics*, *4*(8), e1000149.

Gonzales, N. M., & Palmer, A. A. (2014). Fine-mapping QTLs in advanced intercross lines and other outbred populations. *Mammalian Genome*, *25*(7–8), 271–292.

Gonzales, N. M., Seo, J., Cordero, A. I. H., Pierre, C. L. S., Gregory, J. S., Distler, M. G., Abney, M., Canzar, S., Lionikas, A., & Palmer, A. A. (2018). Genome wide association analysis in a mouse advanced intercross line. *Nature Communications*, *9*(1), 5162. https://doi.org/10.1038/s41467-018-07642-8

Gonzales, N. M., Seo, J., Hernandez-Cordero, A. I., Pierre, C. L. S., Gregory, J. S., Distler, M. G., Abney, M., Canzar, S., Lionikas, A., & Palmer, A. A. (2017). Genome wide association study of behavioral, physiological and gene expression traits in a multigenerational mouse intercross. *BioRxiv*, 230920. https://doi.org/10.1101/230920

Graziano, J. H., Grady, R. W., & Cerami, A. (1974). The identification of 2, 3-dihydroxybenzoic acid as a potentially useful iron-chelating drug. *Journal of Pharmacology and Experimental Therapeutics*, *190*(3), 570–575.

Gruber, J. D., Genissel, A., Macdonald, S. J., & Long, A. D. (2007). How Repeatable Are Associations Between Polymorphisms in achaete–scute and Bristle Number Variation in Drosophila? *Genetics*, *175*(4), 1987–1997. https://doi.org/10.1534/genetics.106.067108

Han, B., Kang, H. M., & Eskin, E. (2009). Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS Genetics*, *5*(4), e1000456.

Han, L., & Abney, M. (2011). Identity by descent estimation with dense genome-wide genotype data. *Genetic Epidemiology*, *35*(6), 557–567. https://doi.org/10.1002/gepi.20606

Hernandez Cordero, A. I., Carbonetto, P., Riboni Verri, G., Gregory, J. S., Vandenbergh, D. J., P Gyekis, J., Blizard, D. A., & Lionikas, A. (2018). Replication and discovery of musculoskeletal QTLs in LG/J and SM/J advanced intercross lines. *Physiological Reports*, *6*(4).

Hernandez Cordero, A. I., Gonzales, N. M., Parker, C. C., Sokolof, G., Vandenbergh, D. J., Cheng, R., Abney, M., Sko, A., Douglas, A., Palmer, A. A., Gregory, J. S., & Lionikas, A. (2019). Genome-wide Associations Reveal Human-Mouse Genetic

Convergence and Modifiers of Myogenesis, CPNE1 and STC2. *American Journal of Human Genetics*, *105*(6), 1222–1236. https://doi.org/10.1016/j.ajhg.2019.10.014

Hoshino, K., Takeuchi, O., Kawai, T., Sanjo, H., Ogawa, T., Takeda, Y., Takeda, K., & Akira, S. (1999). Cutting Edge: Toll-Like Receptor 4 (TLR4)-Deficient Mice Are Hyporesponsive to Lipopolysaccharide: Evidence for TLR4 as the Lps Gene Product. *The Journal of Immunology*, *162*(7), 3749–3752.

Hyun, S. (2013). Body size regulation and insulin-like growth factor signaling. *Cellular and Molecular Life Sciences*, *70*(13), 2351–2365. https://doi.org/10.1007/s00018-013-1313-5

Johnsson, M., Henriksen, R., Höglund, A., Fogelholm, J., Jensen, P., & Wright, D. (2018). Genetical genomics of growth in a chicken model. *BMC Genomics*, *19*(1), 72.

Joo, J. W. J., Hormozdiari, F., Han, B., & Eskin, E. (2016). Multiple testing correction in linear mixed models. *Genome Biology*, *17*(1), 62.

Keele, G. R., Crouse, W. L., Kelada, S. N. P., & Valdar, W. (2019). Determinants of QTL Mapping Power in the Realized Collaborative Cross. *G3: Genes, Genomes, Genetics*, *9*(5), 1707–1727. https://doi.org/10.1534/g3.119.400194

Kim, K.-A., Gu, W., Lee, I.-A., Joh, E.-H., & Kim, D.-H. (2012). High Fat Diet-Induced Gut Microbiota Exacerbates Inflammation and Obesity in Mice via the TLR4 Signaling Pathway. *PLOS ONE*, *7*(10), e47713. https://doi.org/10.1371/journal.pone.0047713

King, E. G., & Long, A. D. (2017). The Beavis Effect in Next-Generation Mapping Panels in Drosophila melanogaster. *G3: Genes, Genomes, Genetics*, *7*(6), 1643–1652. https://doi.org/10.1534/g3.117.041426

King, E. G., Macdonald, S. J., & Long, A. D. (2012). Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics*, genetics–112.

Kislukhin, G., King, E. G., Walters, K. N., Macdonald, S. J., & Long, A. D. (2013). The genetic architecture of methotrexate toxicity is similar in Drosophila melanogaster and humans. *G3: Genes, Genomes, Genetics*, g3–113.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356.

Lai, K.-O., Chen, Y., Po, H.-M., Lok, K.-C., Gong, K., & Ip, N. Y. (2004). Identification of the Jak/Stat Proteins as Novel Downstream Targets of EphA4 Signaling in Muscle IMPLICATIONS IN THE REGULATION OF ACETYLCHOLINESTERASE

EXPRESSION. *Journal of Biological Chemistry*, *279*(14), 13383–13392. https://doi.org/10.1074/jbc.M313356200

Lander, E. S., & Botstein, D. (1989). Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*, *121*(1), 185–199.

Lionikas, A., Cheng, R., Lim, J. E., Palmer, A. A., & Blizard, D. A. (2010). Fine-mapping of muscle weight QTL in LG/J and SM/J intercrosses. *Physiological Genomics*, *42*(1), 33–38.

Listgarten, J., Lippert, C., Kadie, C. M., Davidson, R. I., Eskin, E., & Heckerman, D. (2012). Improved linear mixed models for genome-wide association studies. *Nature Methods*, *9*(6), 525.

Logan, R. W., Robledo, R. F., Recla, J. M., Philip, V. M., Bubier, J. A., Jay, J. J., Harwood, C., Wilcox, T., Gatti, D. M., & Bult, C. J. (2013). High-precision genetic mapping of behavioral traits in the diversity outbred mouse population. *Genes, Brain and Behavior*, *12*(4), 424–437.

Lynch, M., & Walsh, B. (1996). *Genetics and Analysis of Quantitative Traits*.

Macdonald, S. J., & Long, A. D. (2004). A Potential Regulatory Polymorphism Upstream of hairy Is Not Associated With Bristle Number Variation in Wild-Caught Drosophila. *Genetics*, *167*(4), 2127–2131. https://doi.org/10.1534/genetics.104.026732

Mackay, T. F., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., & Cridland, J. M. (2012). The Drosophila melanogaster genetic reference panel. *Nature*, *482*(7384), 173.

Mangin, B., Goffinet, B., & Rebai, A. (1994). Constructing confidence intervals for QTL location. *Genetics*, *138*(4), 1301–1308.

Manichaikul, A., Dupuis, J., Sen, S., & Broman, K. W. (2006). Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. *Genetics*.

Marriage, T. N., King, E. G., Long, A. D., & Macdonald, S. J. (2014). Fine-mapping nicotine resistance loci in Drosophila using a multiparent advanced generation inter-cross population. *Genetics*, *198*(1), 45–57.

Mhyre, T. R., Chesler, E. J., Thiruchelvam, M., Lungu, C., Cory-Slechta, D. A., Fry, J. D., & Richfield, E. K. (2005). Heritability, correlations and in silico mapping of locomotor behavior and neurochemistry in inbred strains of mice. *Genes, Brain and Behavior*, *4*(4), 209–228. https://doi.org/10.1111/j.1601-183X.2004.00102.x

Najarro, M. A., Hackett, J. L., Smith, B. R., Highfill, C. A., King, E. G., Long, A. D., & Macdonald, S. J. (2015). Identifying Loci Contributing to Natural Variation in Xenobiotic Resistance in Drosophila. *PLoS Genetics*, *11*(11), e1005663. https://doi.org/10.1371/journal.pgen.1005663

Nicod, J., Davies, R. W., Cai, N., Hassett, C., Goodstadt, L., Cosgrove, C., Yee, B. K., Lionikaite, V., McIntyre, R. E., & Remme, C. A. (2016). Genome-wide association of multiple complex traits in outbred mice by ultra-low-coverage sequencing. *Nature Genetics*, *48*(8), 912.

Nikolskiy, I., Conrad, D. F., Chun, S., Fay, J. C., Cheverud, J. M., & Lawson, H. A. (2015). Using whole-genome sequences of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides. *BMC Genomics*, *16*(1), 415.

Noldus, L. P. J. J., Spink, A. J., & Tegelenbosch, R. A. J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers*, *33*(3), 398–414. https://doi.org/10.3758/BF03195394

Parker, C. C., Carbonetto, P., Sokoloff, G., Park, Y. J., Abney, M., & Palmer, A. A. (2014). High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. *Genetics*, *198*(1), 103–116.

Parker, C. C., Cheng, R., Sokoloff, G., Lim, J. E., Skol, A. D., Abney, M., & Palmer, A. A. (2011). Fine-mapping alleles for body weight in LG/J × SM/J F<Subscript>2</Subscript> and F<Subscript>34</Subscript> advanced intercross lines. *Mammalian Genome*, *22*(9–10), 563. https://doi.org/10.1007/s00335-011-9349-z

Parker, C. C., Cheng, R., Sokoloff, G., & Palmer, A. A. (2012). Genome-wide association for methamphetamine sensitivity in an advanced intercross mouse line. *Genes, Brain and Behavior*, *11*(1), 52–61.

Parker, C. C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N. M., Leung, E., Park, Y. J., Aryee, E., Davis, J., Blizard, D. A., & Ackert-Bicknell, C. L. (2016). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, *48*(8), 919.

Parker, C. C., & Palmer, A. A. (2011). Dark matter: Are mice the solution to missing heritability? *Frontiers in Genetics*, *2*, 32.

Paterson, A. H. (2019). *Molecular Dissection of Complex Traits*. CRC Press.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., Boehnke, M., Abecasis, G. R., & Willer, C. J. (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, *26*(18), 2336–2337.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., & Daly, M. J. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559–575.

Relaix, F., Montarras, D., Zaffran, S., Gayraud-Morel, B., Rocancourt, D., Tajbakhsh, S., Mansouri, A., Cumano, A., & Buckingham, M. (2006). Pax3 and Pax7 have distinct and overlapping functions in adult muscle progenitor cells. *J Cell Biol*, *172*(1), 91–102. https://doi.org/10.1083/jcb.200508044

Rishmawi, L., Bühler, J., Jaegle, B., Hülskamp, M., & Koornneef, M. (2017). Quantitative trait loci controlling leaf venation in Arabidopsis. *Plant, Cell & Environment*, *40*(8), 1429–1441.

Samocha, K. E., Lim, J. E., Cheng, R., Sokoloff, G., & Palmer, A. A. (2010). Fine mapping of QTL for prepulse inhibition in LG/J and SM/J mice using F2 and advanced intercross lines. *Genes, Brain and Behavior*, *9*(7), 759–767.

Sawada, T., Arai, D., Jing, X., Miyajima, M., Frank, S. J., & Sakaguchi, K. (2017). Molecular interactions of EphA4, growth hormone receptor, Janus kinase 2, and signal transducer and activator of transcription 5B. *PLOS ONE*, *12*(7), e0180785. https://doi.org/10.1371/journal.pone.0180785

Shi, H., Kokoeva, M. V., Inouye, K., Tzameli, I., Yin, H., & Flier, J. S. (2006). TLR4 links innate immunity and fatty acid–induced insulin resistance. *The Journal of Clinical Investigation*, *116*(11), 3015–3025. https://doi.org/10.1172/JCI28898

Sittig, L. J., Carbonetto, P., Engel, K. A., Krauss, K. S., Barrios-Camacho, C. M., & Palmer, A. A. (2016). Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron*, *91*(6), 1253–1259. https://doi.org/10.1016/j.neuron.2016.08.013

Sorge, R. E., Martin, L. J., Isbester, K. A., Sotocinal, S. G., Rosen, S., Tuttle, A. H., Wieskopf, J. S., Acland, E. L., Dokova, A., Kadoura, B., Leger, P., Mapplebeck, J. C. S., McPhail, M., Delaney, A., Wigerblad, G., Schumann, A. P., Quinn, T., Frasnelli, J., Svensson, C. I., Sternberg, W. F., & Mogil, J. S. (2014). Olfactory exposure to males, including men, causes stress and related analgesia in rodents. *Nature Methods*, *11*(6), 629–632. https://doi.org/10.1038/nmeth.2935

Svenson, K. L., Gatti, D. M., Valdar, W., Welsh, C. E., Cheng, R., Chesler, E. J., Palmer, A. A., McMillan, L., & Churchill, G. A. (2012). High-resolution genetic mapping using the Mouse Diversity outbred population. *Genetics*, *190*(2), 437–447.

Takeuchi, O., Hoshino, K., Kawai, T., Sanjo, H., Takada, H., Ogawa, T., Takeda, K., & Akira, S. (1999). Differential Roles of TLR2 and TLR4 in Recognition of Gram-

Negative and Gram-Positive Bacterial Cell Wall Components. *Immunity*, *11*(4), 443–451. https://doi.org/10.1016/S1074-7613(00)80119-3

Talbot, C. J., Nicod, A., Cherny, S. S., Fulker, D. W., Collins, A. C., & Flint, J. (1999). High-resolution mapping of quantitative trait loci in outbred mice. *Nature Genetics*, *21*(3), 305.

The Wellcome Trust Case Control Consortium, Maller, J. B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J. M. M., Auton, A., Myers, S., Morris, A., Pirinen, M., Brown, M. A., Burton, P. R., Caulfield, M. J., Compston, A., Farrall, M., Hall, A. S., Hattersley, A. T., Hill, A. V. S., Mathew, C. G., Pembrey, M., Satsangi, J., Stratton, M. R., Worthington, J., Craddock, N., Hurles, M., Ouwehand, W., Parkes, M., Rahman, N., Duncanson, A., Todd, J. A., Kwiatkowski, D. P., Samani, N. J., Gough, S. C. L., McCarthy, M. I., Deloukas, P., & Donnelly, P. (2012). Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics*, *44*(12), 1294–1301. https://doi.org/10.1038/ng.2435

Toker, L., Feng, M., & Pavlidis, P. (2016). Whose sample is it anyway? Widespread misannotation of samples in transcriptomics studies. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.9471.2

Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M. S., Rawlins, J. N. P., Mott, R., & Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics*, *38*(8), 879.

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era—Concepts and misconceptions. *Nature Reviews Genetics*, *9*(4), 255–266. https://doi.org/10.1038/nrg2322

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation. *The American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Vonesch, S. C., Lamparter, D., Mackay, T. F., Bergmann, S., & Hafen, E. (2016). Genome-wide analysis reveals novel regulators of growth in Drosophila melanogaster. *PLoS Genetics*, *12*(1), e1005616.

Xu, S. (2003). Theoretical Basis of the Beavis Effect. *Genetics*, *165*(4), 2259–2268.

Yalcin, B., Willis-Owen, S. A., Fullerton, J., Meesaq, A., Deacon, R. M., Rawlins, J. N. P., Copley, R. R., Morris, A. P., Flint, J., & Mott, R. (2004). Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nature Genetics*, *36*(11), 1197.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, *88*(1), 76–82. https://doi.org/10.1016/j.ajhg.2010.11.011

Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., & Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100.

Yang, J., Zeng, J., Goddard, M. E., Wray, N. R., & Visscher, P. M. (2017). Concepts, estimation and interpretation of SNP-based heritability. *Nature Genetics*, *49*(9), 1304–1310. https://doi.org/10.1038/ng.3941

Zaitlen, N., & Kraft, P. (2012). Heritability in the genome-wide association era. *Human Genetics*, *131*(10), 1655–1664. https://doi.org/10.1007/s00439-012-1199-6

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821.

Zöllner, S., & Pritchard, J. K. (2007). Overcoming the Winner's Curse: Estimating Penetrance Parameters from Case-Control Data. *The American Journal of Human Genetics*, *80*(4), 605–615. https://doi.org/10.1086/512821

Zou, J., Zhou, J., Faller, S., Brown, R., & Eskin, E. (2019). Accurate modeling of replication rates in genome-wide association studies by accounting for winner's curse and study-specific heterogeneity. *BioRxiv*, 21. https://doi.org/doi: https://doi.org/10.1101/856898

CHAPTER 2: A Null Allele of *Azi2* Enhances Sensitivity to Methamphetamine, Replicating a Finding from a Mouse Genome-wide Association Study

2.1 Abstract

Methamphetamine is a widely abused psychostimulant. In a previous genome-wide association study (**GWAS**), we identified a locus that influenced the locomotor sensitivity to methamphetamine and found that this locus was also an eQTL for the gene *Azi2*. In this study, we proposed that heritable differences in the expression of *Azi2* were causally related to the differential response to methamphetamine. We created an *Azi2* KO line. We found that homozygous *Azi2* mutant mice with lower *Azi2* expression in the striatum showed an enhanced locomotor response to methamphetamine; this direction of effect is opposite to the findings from the previously published GWAS. A recent report suggested that in the ventral tegmental area (**VTA**) the 3'UTR of *Azi2* mRNA downregulates the expression of *Slc6a3*, a gene that encodes the dopamine transporter (**DAT**), which is a direct target of methamphetamine. We evaluated the relationship between *Azi2*/*Azi2* 3'UTR and *Slc6a3* expression in the VTA in the *Azi2* KO line and in a new cohort of CFW mice. We did not observe any correlation between *Azi2* and *Slc6a3* in the VTA in either cohort. However, gene expression analysis confirmed that the mutation in *Azi2* mutant mice altered *Azi2* expression and also revealed a number of potentially important genes and pathways that were regulated by *Azi2*, including the metabotropic glutamate receptor group III pathway and nicotinic acetylcholine receptor signaling pathway. Our results support a role for *Azi2* in methamphetamine sensitivity;

however, the exact mechanism does not appear to involve regulation of *Slc6a3* and thus remains unknown.

2.2 Introduction

Methamphetamine is a widely abused psychomotor stimulant. In the United States, approximately 1.6 million people reported using methamphetamine in the past year (NIDA, 2020). Methamphetamine can produce feelings of euphoria, heightened energy, and enhanced focus. Although most users experience stimulants such as methamphetamine, amphetamine and cocaine as pleasurable, studies in both humans and animals have found marked individual differences (Deminiere et al., 1989; Ettenberg, 2009; Glick & Hinds, 1985; Piazza et al., 1989; Seale, 1991; Smith et al., 2016). Significant evidence suggests that these differences are at least partially genetic (Hart et al., 2012; Phillips & Shabani, 2015). These differences are believed to alter the behavioral and subjective response to methamphetamine and may therefore alter risk for methamphetamine abuse. Such individual differences are often modeled by studying the acute locomotor response to drug use in rodents, which may influence risk of progressing from initial use to problematic use and addiction (de Wit & Phillips, 2012; Steketee & Kalivas, 2011)

Over the last decade, large scale genome-wide association studies (**GWAS**) have facilitated the identification of thousands of loci that influence complex traits including abuse of alcohol and other substances (Buchwald et al., 2020; Kinreich et al., 2019; Liu et al., 2019; Sanchez-Roige et al., 2019; Walters et al., 2018; Zhou et al., 2020). GWAS in model organisms provide a complementary approach to human GWAS and have also identified loci for numerous traits, including several that are related to drug abuse (*e.g.* (Gonzales et al., 2018; Parker et al., 2016)). An advantage of GWAS in model organisms is that putatively identified genes can be directly manipulated to assess causality and to

better understand the underlying molecular mechanisms; however, such manipulations are not frequently performed.

We previously reported a GWAS that examined a number of behavioral and physiological traits, including the acute locomotor stimulant response to methamphetamine using 1,200 outbred Carworth Farms White (**CFW**) male mice (Parker et al., 2016). CFW mice are a commercially available outbred population that have relatively small linkage disequilibrium blocks, which allow the identification of small loci and thus narrow the number of genes that might be causally related to the trait being measured (Parker et al., 2016). One of the many genome-wide significant findings from that study was an association between a locus on chromosome 9 (rs46497021) and the locomotor response to methamphetamine injection. That locus also contained a *cis*-expression quantitative trait locus (***cis*-eQTL**) for the gene 5-azacytidine–induced gene 2 (***Azi2***) in the striatum (the peak eQTL SNP was rs234453358, which is in strong LD with rs46497021). Based on these data, we suggested that heritable differences in *Azi2* expression may be the molecular mechanism by which that locus influenced the acute locomotor response to methamphetamine. However, in our prior publication we did not directly test that hypothesis by manipulating *Azi2* expression.

Azi2 is known to activate NFκB (Fujita et al., 2003), to be involved in TNF-induced cell death (Fukasaka et al., 2013; Goncalves et al., 2011; Lafont et al., 2018; Maruyama et al., 2015), and to influence immune response (Bozóky et al., 2013). However, its role in methamphetamine sensitivity remains largely unknown. One possible mechanism by which *Azi2* could influence the response to methamphetamine was proposed by another group after the publication of Parker et al (2016) (Parker et al., 2016). Liu et al (2018)

identified the second half of 3' UTR of *Azi2* mRNA (**AZI2 3'UTR**) as a regulator of *Slc6a3* (Liu et al., 2018), which encodes the dopamine transporter (**DAT**), a critical membrane protein that regulates the level of neurotransmitter dopamine in the synaptic cleft and is directly bound by methamphetamine (Kish, 2008). In addition to a number of cellular assays, Liu et al (2018) also showed that rats that had been bi-directionally selected for alcohol preference showed differential AZI2 3'UTR expression and differential expression of *Slc6a3* expression in the ventral tegmental area (**VTA**) (Liu et al., 2018). Because methamphetamine causes dopamine release by acting at the dopamine transporter, we considered it highly probable that the effect of *Azi2* on methamphetamine sensitivity could be mediated through *Azi2* induced changes in *Slc6a3* expression.

To test the hypothesis that *Azi2* was the gene responsible for the association detected in our GWAS, and that its action was mediated via *Slc6a3*, we created an *Azi2* KO line using CRISPR/Cas9 to generate a frameshifting mutation in exon 3 of *Azi2*. In this study, we will refer to our version of the *Azi2* mutation as the '*Azi2* mutant allele', all progeny of the *Azi2* KO founders as the '*Azi2* KO mice', and the homozygous *Azi2* mutant mice as the '*Azi2* mutant mice'. Using the *Azi2* KO mice, we examined the acute locomotor response to methamphetamine. We evaluated gene expression in the striatum, the brain tissue in which the eQTL for *Azi2* expression was identified in Parker et al (2016) (Parker et al., 2016), to validate the elimination of *Azi2*. In an effort to determine whether our mutant allele altered the expression of *Azi2* 3'UTR, and whether such changes might alter *Slc6a3*, as predicted by Liu et al (2018), we measured gene expression in the VTA of wildtype, heterozygous and mutant *Azi2* KO mice. We also performed parallel studies in a new cohort of CFW mice to confirm that the allele identified in Parker et al (2016) was

indeed associated with changes in *Azi2* and to determine whether it also associated with differential *Slc6a3* expression in the VTA, a brain region that was not examined in Parker et al (2016).

2.3 Results

Creation and characterization of *Azi2* mutant allele using CRISPR/Cas9

To investigate whether *Azi2* might be the gene underlying the association observed in CFW mice (Parker et al., 2016) between the locus on chromosome 9 and the locomotor response to methamphetamine, we created a mutant allele of *Azi2* using CRISPR/Cas9. Because of technical difficulties generating embryos from CFW, and because of the more complicated breeding programs necessary for maintaining a mutant allele on an outbred background, we generated the mutant alleles on the C57BL/6J background. We designed a sgRNA targeting exon 3 of *Azi2* (Figure 1a; Table S1), because exon 3 harbors the start codon of *Azi2* and is included in all four annotated transcripts of *Azi2*. We selected a mutant mouse line that carried a 7bp frameshifting deletion in exon 3 of *Azi2* (Table S2).

Because Parker et al (2016) had identified a coincident eQTL for *Azi2* in the striatum, we sought to confirm that the mutant allele would reduce *Azi2* expression in that same brain region. In addition, because the Liu et al study (Liu et al., 2018) study focused on the role of *Azi2* in the VTA, we also examined *Azi2* expression in that brain region. Using qRT-PCR in wildtype, heterozygote and mutant *Azi2* KO mice, we confirmed that the 7bp deletion led to highly significantly decreased abundance of *Azi2* mRNA in the striatum (Figure 1c, Figure S1). In a separate cohort of mice, we used RNA-Seq in the *Azi2* KO mice to further demonstrate that that the 7bp deletion led to highly significantly decreased abundance of *Azi2* mRNA in the striatum (Figure 1b&d, Figure S2a). In addition to the decreased mRNA abundance associated with the mutant allele, many of

the remaining transcripts are not expected to encode functional protein because they will be frameshifted.

Modulated by time bin, locomotor response to methamphetamine was greater in *Azi2* mutant mice

Having created a *Azi2* mutant allele, we next sought to examine whether *Azi2* mutant mice showed an altered locomotor response to methamphetamine. In particular, we sought to precisely replicate the protocol used in Parker et al (2016), in which saline was given on days 1 and 2 and 1.5 mg/kg methamphetamine was given on day 3. Parker et al. (2016) found that mice with more 'A' alleles at rs46497021 exhibited greater sensitivity to methamphetamine; those results are plotted in Figure 2a-c. We reanalyzed data from Parker et al. (2016) to demonstrate that *Azi2* expression was higher in individuals that had the 'A' allele at rs46497021 (Figure S3a). The same was also true at rs234453358, which was in LD with rs46497021 (Figure S3b). In our *Azi2* KO line, we found that locomotor responses to saline on days 1 and 2 did not differ among the three genotype groups (Figure 2d, e), but the response to methamphetamine was different among the three genotype groups where the mutant mice had lower *Azi2* expression in the striatum and higher methamphetamine sensitivity (Figure 2f; Figure S4). This direction of effect is opposite to that of the CFW mice from Parker et al (2016) where mice with more 'A' alleles at rs46497021 had higher *Azi2* expression and higher methamphetamine sensitivity (Figure 2c).

Azi2 mRNA and *Azi2* 3'UTR mRNA did not downregulate *Slc6a3* expression in *Azi2* KO mice

Liu et al (2018) reported that the 3'UTR of *Azi2* mRNA negatively regulates *Slc6a3* in the midbrain of rats, and that there was an increase in expression of the 3'UTR of *Azi2* in the non-alcohol preferring rats compared to alcohol preferring rats (Liu et al., 2018). Based on those data, Liu et al (2018) argued that regulation of *Slc6a3* expression by the 3'UTR of *Azi2* is important for substance use related traits. Based on these data, we hypothesized that *Azi2* expression in CFW mice might have led to altered response to methamphetamine because of its ability to regulate *Slc6a3*.

We tested this hypothesis using both qRT-PCR and RNA-Seq. Using qRT-PCR, we measured the level of *Azi2*, *Azi2* 3'UTR and *Slc6a3* mRNA in the VTA (Figure 1a). The *Azi2* 3'UTR on exon 8 that we amplified using qRT-PCR is homologous to the 3'UTR amplified in the rat alcohol model (Liu et al., 2018) and is only present in two of the four full-length *Azi2* transcripts, ENSMUST0000044454.11 and ENSMUST00000133580.7 (Figure 1b). We found that the expression of both *Azi2* and *Azi2* 3'UTR amplicons were decreased in a genotype-dependent manner in the VTA in the mutant mice (Figure S1a; Figure S5a). However, there was no significant effect of genotype on *Slc6a3* expression in the VTA (Figure S1b; Figure S5b). Furthermore, we did not observe any correlation between the expression of *Azi2*, *Azi2* 3'UTR or *Slc6a3* in the VTA (Figure 3a&b).

We also used RNA-Seq to examine the hypothesis that *Azi2*/*Azi2* 3'UTR could downregulate *Slc6a3* in the VTA in the *Azi2* KO line. Expression of *Azi2* and *Azi2* 3'UTR were lower in the heterozygote and mutant mice; however, there was no effect of genotype on *Slc6a3* expression (Figure 3c&d). Taken together, our results do not support

the negative correlation between *Azi2*/*Azi2* 3'UTR and *Slc6a3* in the VTA in our *Azi2* KO mice.

Azi2 mRNA did not downregulate *Slc6a3* regulation in naïve CFW mice in VTA

We also examined the relationship between *Azi2*, *Azi2* 3'UTR and *Slc6a3* in a behaviorally and drug naïve CFW mice to address the possibility that the strain difference between C57BL/6J and CFW may have contributed to the previously reported negative correlation between *Azi2*/*Azi2* 3'UTR and *Slc6a3* expression. Using 31 male CFW mice, we found that there was no correlation between *Azi2* nor *Azi2* 3'UTR and *Slc6a3* in the VTA (Figure 3e&f; Figure S6; Figure S7).

Analysis of gene expression differences using RNA-Seq in *Azi2* KO and in CFW mice

Next, we sought to identity genome-wide changes in gene expression observed in the *Azi2* KO line using the RNA-Seq data. When comparing wildtype to mutant mice, we identified 23 differentially expressed genes in the striatum at FDR < 0.1 (Figure 4a; Table S6). In the VTA, the same comparison identified four differentially expressed genes (Figure 4b; Table S6). For both tissues, *Azi2* was by far the most significantly differentially expressed gene. When comparing wildtype to heterozygous mice, we found that *Azi2* was the only differentially expressed gene in both the striatum and the VTA (Figure S8a, S8b).

We sought to identify genome-wide changes in gene expression observed in the CFW mice after stratifying them by rs234453358, which was the peak eQTL SNP for *Azi2*. When comparing 'AA' to the 'GG' homozygotes, we identified five differentially expressed genomic features in the striatum at FDR < 0.1 (Figure 4c; Table S6). In the VTA, the same comparison identified only *Azi2* as being differentially expressed (Figure 4d; Table S6). When comparing the 'AG' to the 'GG' mice, we found five differentially expressed genomic features in the striatum (Figure S8c; Table S6) but none in the VTA (Figure S8d; Table S6). As discussed in the prior section, *Slc6a3* was not differentially expressed in any of these comparisons.

Finally, we performed PANTHER gene list analysis to reveal pathways implicated by these results (Thomas et al., 2003). The mutant vs wildtype comparison in the striatum in the *Azi2* KO mice and the 'GG' vs 'AA' comparison in the striatum in the CFW mice both identified the Wnt signaling pathway (Table S6). Additional pathways that were identified include angiogenesis, Alzheimer disease-presenilin pathway, TGF-beta signaling pathway, metabotropic glutamate receptor group III pathway, cadherin signaling pathway, Notch signaling pathway, Huntington disease, nicotinic acetylcholine receptor signaling pathway, inflammation mediated by chemokine and cytokine signaling pathway, and cytoskeletal regulation by Rho GTPase (Table S6). A few genes had larger than five $\log_2$ fold changes but did not pass the FDR < 0.1 threshold, and thus were not considered as differentially expressed genes. Nevertheless, PANTHER gene list analysis showed that these genes are involved in pathways similar to those of differentially expressed genes (Table S7).

2.4 Discussion

This purpose of the current study was to follow up on findings from the mouse GWAS by Parker et al (2016), which identified a locus that influenced the locomotor simulant response to methamphetamine and identified a co-localized *cis*-eQTL for the gene *Azi2*, which was hypothesized to be the casual variant. To experimentally test this hypothesis, we created a mutant allele of *Azi2.* The mutant allele reduced *Azi2* expression in the striatum (Figure 1), which was the tissue that showed an eQTL in Parker et al (2016) (Parker et al., 2016). Furthermore, most of all of the remaining transcripts were frameshifted and thus non-functional. Importantly, we observed significantly greater methamphetamine sensitivity in *Azi2* mutant mice (Figure 2F), supporting a role for this gene in the responses to methamphetamine. However, based on data from Parker et al (2016), we had predicted a positive relationship between *Azi2* expression and the locomotor response to methamphetamine. Instead, the mutant mice showed that elimination of *Azi2* increased sensitivity. After the publication of Parker et al (2016), Liu et al (Liu et al., 2018) reported that the 3'UTR of *Azi2* regulated the expression of the dopamine transporter, which is the target of methamphetamine, and suggested that findings from Parker et al might be mediated by this mechanism. However, we did not observe any correlation between *Azi2* and *Sla6a3* in either the *Azi2* mutant mice or the CFW mice that harbor an eQTL for *Azi2* (Figure 3). Those observations do not support the hypothesis that *Azi2's* effects are mediated by regulation of *Slc6a3* expression. While we did observe effects of the *Azi2* mutant allele on the expression of other genes (Figure 4), future studies will be needed to define the molecular pathway by which *Azi2* regulates stativity to methamphetamine.

A major conclusion from our work is that *Azi2* alters the locomotor response to methamphetamine. Although we did not exhaustively characterize these mice, they did not present any overt physical or behavioral abnormalities, and they did not show locomotor differences in the absence of methamphetamine administration (Figure 2d&e), suggesting that the effect of the mutation is at least somewhat specific to methamphetamine sensitivity. However, Parker et al (2016) found that the eQTL allele associated with greater *Azi2* expression was also associated with greater locomotor response to methamphetamine, whereas the current study found the opposite, namely that loss of *Azi2* was associated with greater response to methamphetamine. One possible explanation for this finding is that the effect of *Azi2* is modified by genetic background – the eQTL was observed in CFW mice, whereas the mutant allele was characterized using a C57BL/6J background. Consistent with this, we have previously reported that genetic background can induce directionally opposite effects of other mutant alleles (Sittig et al., 2016). Another possibility is that the total loss of *Azi2* in the mutant line could have different consequences than the differential expression observed in the CFW mice. While both explanations are plausible, the difference in direction complicates our interpretation of the behavioral results and calls into question whether they should be considered to “replicate” or “recapitulate” the findings from Parker et al (2016).

Along the same lines, Parker et al (2016) examined total activity after 30 minutes in males only, whereas in the current study we used males and females and used ANOVA to examine the factors sex, genotype and time, in which time was modeled as six 5-minute bins making up the full 30-minute test. We observed a significant interaction between genotype and time bin; however, post-hoc tests did not identify any specific time bins that

were different from one another. Had we instead decided *a priori* to only examine total activity over all 30 minutes, the effect of *Azi2* on behavior would not be significant.

One of the rationales for using an inbred C57BL/6J background was that expressing a mutant allele on an isogenic background would enhance our ability to detect an effect of the mutation, since it would remove other genetic differences that could be confounding. It is notable that despite this advantage, a relatively large sample size was required to obtain significant results. We observed a similar result in a prior study in which we examined a mutant allele of the gene *Csmd1* (Gonzales et al., 2018), which had been implicated by a separate mouse GWAS (Cheng et al., 2010). While these two examples do not represent a large enough sample to draw general conclusions, it may be that genes identified using mouse GWAS have relatively subtle effects that require sample sizes that are larger than those often employed when examining mutant mice. Because studies like these do not use the alleles identified in the GWAS, power analyses are difficult because the expected effect size is unknown. Our observations imply that future studies following up on mouse GWAS should consider using relatively large samples (in this case more than 100 total subjects) before drawing conclusions.

The goal of our study was to determine whether *Azi2* is the gene responsible for the association detected by Parker et al (2016). That locus contained a second candidate gene, COX assembly mitochondrial protein 1 (***Cmc1***), which could also have contained regulatory variants for other nearby genes. it is possible that the locus harbored multiple causal variants. Studies such as ours are not able to refute other possible causal alleles, which is a limitation.

In an effort to define the causal pathway by which *Azi2* might alter sensitivity to the locomotor effects of methamphetamine, we tested the hypothesis that the 3'UTR of *Azi2* regulates *Slc6a3*, as described by Liu et al (Liu et al., 2018). Using a luciferase reporter assay in the human neuroblastoma SK-N-AS cells, Liu et al (2018) identified *Azi2* 3'UTR as a putative downregulator of the promoter activity in only one allele of a dinucleotide polymorphism in Intron 1 of *SLC6A3*, but not the other (Liu et al., 2018). It is possible that this allele specificity is the reason for the lack of correlation we observed between the 3'UTR of *Azi2* and the expression of *Slc6a3*. However, Liu et al (2018) did not detect any *SLC6A3* allele-dependence in the downregulation of endogenous *SLC6A3* mRNA level by *Azi2* 3'UTR in the human neuroblastoma BE(2)-M17 cells. Furthermore, Northern blot and qRT-PCR results on *Azi2*/ *Azi2* 3'UTR and *Slc6a3* expression in the VTA of alcohol preferring and non-preferring rats did not distinguish the two alleles of *Slc6a3* (Liu et al., 2018). Thus, we investigated the correlation between *Azi2*/ *Azi2* 3'UTR and *Slc6a3* expression in both our *Azi2* KO mice and naïve CFW mice, intending to replicate the experiments performed on alcohol preferring and non-preferring rats (Liu et al., 2018). Our results do not contradict the findings of Liu et al (2018) but they strongly suggest that *Azi2*'s actions observed in our studies are not mediated by *Slc6a3*.

Although we did not find evidence to support a role for *Slc6a3* in the effects of *Azi2*, we did identify a number of other genes that were differentially expressed in both the *Azi2* mutant line and in the CFW mice (Figure 4). One differentially expressed gene, *Slc16a6*, is mapped to the metabotropic glutamate receptor group III pathway. Another gene, *Myh1*, is mapped to the nicotinic acetylcholine receptor signaling pathway. Future studies should investigate whether these genes and their involvement in synaptic signaling

pathways might hold clues about the relationship between *Azi2* and sensitivity to methamphetamine.

Our study is not without limitations. For example, the creation of the mutant allele using CRISPR/Cas9 could have induced unintended off-target mutations. We backcrossed the mutant allele for several generations, and unless putative off-target mutations were nearby and thus linked to the mutant allele, they should have segregated independently, since all experiments uses wildtype littermate controls. Nevertheless, it is technically possible that a linked, off-target mutation may have interfered with our results. We also did not characterize the effect of the *Azi2* mutation on other doses of methamphetamine nor did we examine other behavioral traits of these mice, which might have provided clues about the possible role of *Azi2* in substance abuse-related traits. Finally, we conducted all of the studies intended to examine the role of *Slc6a3* in naïve C57BL/6J and CFW mice, whereas Liu et al (2018) examined human cell lines, human postmortem dopamine neurons, and alcohol preferring and alcohol non-preferring rats (Liu et al., 2018). Thus, our conclusions about the lack of correlation between *Azi2* and *Slc6a3* only apply to the mouse systems that we examined.

The present study is notable because it remains rare to experimentally test specific genes identified using model organism or human GWAS. Our results highlight the potential for such studies, including their ability to contribute to a molecular understanding of how a specific gene influences a specific trait, which is essential for deriving new biological insights from GWAS results. However, our results also illustrate challenges, including the choice of background, and the criteria needed to claim replication.

2.5 Methods and Materials

Establishment of an *Azi2* knockout mouse line using CRISPR/Cas9

We followed the JAX protocol of microinjection of CRISPR mix using sgRNA and *Cas9* mRNA (https://www.jax.org/news-and-insights/1998/july/superovulation-technique). We designed a sgRNA targeting exon 3 of *Azi2* (Vector Name: pRP[CRISPR]-hCas9-U6>{20nt_GGGCCGAGAACAAGTGAATA}; Table S1).

All animal procedures were approved by the local Institutional Animal Care and Use Committee and were conducted in accordance with the NIH Guide for the Care and Use of Laboratory Animals. The CRISPR microinjection procedures were performed at the University of California San Digeo, Moores Cancer Center, Transgenic Mouse Core. We ordered five C57BL/6J stud males (7-8 weeks old) and five C57BL/6J females (3-4 weeks old) from the Jackson Laboratory (Bar Harbor, ME). Upon arrival at the vivarium, the stud males were singly-housed and the females were housed in groups of four. On Day 1 of the microinjection week, all five females were super-ovulated via 0.1ml pregnant mare serum (PMS) intraperitoneal injection per animal. On Day 3, all females were super-ovulated via 0.1ml human chorionic gonadotropin (HCG) intraperitoneal injection per animal. After hormonal priming, each female was placed into the home cage of one stud male for mating. On Day 4, ovulation was expected to occur, and females were separated from the stud males. Fallopian tubes were dissected out from the mated females and were collected in M2 medium. Zygotes were harvested and injected with the CRISPR mix (625ng = 3.1ul×200ng/ul of *Azi2* sgRNA + 1250ng = 5ul×250ng/ul of *Cas9* mRNA + 16.9ul ph7.5 IDTE; total volume 25ul). Injected zygotes were surgically transplanted to pseudopregnant female C57BL/6JOlaHsd (Harlan) mice. Pregnant surrogate dams were

singly caged one week before the expected birth date of the pups. C-sections were carried out if necessary.

Azi2 knockout line breeding and genotyping scheme

We obtained 14 *Azi2* CRISPR founders (five males, nine females). The founders were genotyped via Sanger sequencing to verify the presence of deletions. The male founders were then backcrossed to wildtype C57BL/6J mice to minimize the effect of off-targeting. F_1/F_2 *Azi2* KO mice were genotyped via Sanger/NGS again to ensure the transmission of mutation. Heterozygous F_1s were paired to produce F_2s, which were genotyped via next-generation sequencing. Among others, we identified one 7bp deletion in F_2s. This deletion was predicted to cause mRNA degradation of the four full-length RefSeq supported *Azi2* transcripts, ENSMUST00000044454.11, ENSMUST00000133580.7, ENSMUST00000134433.7, and ENSMUST00000154583.7, and three shorter, predicted *Azi2* transcripts, ENSMUST000000143024.1, ENSMUST000000130735.7, and ENSMUST000000127189.1. We genotyped this deletion via restriction fragment length polymorphism (RFLP); this mutation harbors the StuI restriction enzyme target site, which allows us to easily genotype the mice. From the 5^{th} generation (F_5) and onward, we decided to only keep the line with the 7bp deletion. We genotyped the following generations of *Azi2* KO animals via PCR and RFLP. The sgRNA for the CRISPR/Cas9 procedure and the genotyping primers for *Azi2* are included in Table S1.

Locomotor response to methamphetamine

We followed the locomotor response to methamphetamine protocol as described in Parker et al 2016 (Parker et al., 2016). Adult male and female mice were tested over a three-day period between 0800 and 1700 h. During the experiment mice were group housed 2-5 per cage on a 12h/12h light-dark cycle with lights on at 0600 h. Mice were transported to the procedure room at least 30 min before testing, which allowed them to habituate to the new environment in their home cages. On each day of testing, each animal was placed in an individual clean cage. Animals were weighed to determine the volume of injection (0.01 ml/g body weight). On day one and day two, mice received an *i.p.* injection of 0.9% saline solution; on day three, mice received an *i.p.* injection of methamphetamine solution (1.5 mg/kg of (+)-Methamphetamine hydrochloride; Sigma Life Science, St. Louis, MO). Immediately following injection, each mouse was placed in the test chamber for activity recording. All animals were measured using the Versamax software (AccuScan Instruments, Columbus, OH). At the end of the 30 min test, mice were returned to their home cages. Test chambers were sprayed with 10% isopropanol between tests. At the end of each test day, animals were returned to the vivarium.

We performed the locomotor test in the F_9 of *Azi2* KO mice. We removed one heterozygote and one mutant mice whose locomotor activity on day 3 were more than three standard deviations away from the mean. We analyzed a total of 135 mice and the ratio of genotypes was as expected: 33 wildtype littermates (18 females, 15 males), 67 heterozygotes (31 females, 36 males), and 35 mutants (18 females, 17 males).

Genotyping CFW naïve mice

CFW mice were genotyped via Sanger sequencing. Genotyping primers for CFW mice are included in Table S5. We genotyped the GWAS top SNP for the trait "Distance traveled, 0–30 min, on day 3" at rs46497021 (rs46497021) and the eQTL top SNP for *Azi2* expression in the striatum at rs234453358 (rs234453358), as described in Parker et al (2016) (Parker et al., 2016).

Brain tissue collection

Mouse brain tissue was harvested via an adult mouse brain slicer matrix with 1.0 mm coronal section slice intervals (ZIVIC instruments, Pittsburgh, PA, USA). Striatum was collected from slice Bregma 0 to 2 and VTA was collected from slice Bregma -4 to -2. Four tissue punches, two on the left and two on the right hemisphere, were collected for each animal. After dissection, brain tissue was placed in an Eppendorf tube that was fully submerged in dry ice.

Analysis of CFW data from Parker et al (2016)

Because genotypes in Parker et al (2016) were represented as genotype probabilities, we first converted probabilities to dosages and then coded dosages < 0.2 as homozygous reference, dosages > 0.8 and <1.2 as heterozygous, dosages > 1.8 homozygous alternative. A hundred and seven mice with intermediate dosage values are excluded from the plots. We used likelihood ratio test of nested models (*lmtest* R package; (Zeileis & Hothorn, 2002)) to examine the genotype effect.

qRT-PCR

Primers and probes selected for *Azi2*, 3'UTR of *Azi2*, *Slc6a3*, and *Gapdh* gene expression assays are shown in Table S2. We used pre-designed TaqMan gene expression assays for *Azi2*, *Slc6a3*, and *Gapdh*. We custom-designed the gene expression assay for the 3'UTR of *Azi2*. We custom designed the TaqMan primers and the FAM-MGB probe for 3'UTR of *Azi2* according to the Custom TaqMan Assay Design Tool (https://tools.thermofisher.com/content/sfs/manuals/cms_042307.pdf). We only used two replicates per sample instead of the recommended three replicates because of the low RNA/cDNA content in our tissue samples. We ran all qRT-PCR experiments on the StepOnePlus™ Real-Time PCR System (Applied Biosystems, Waltham, MA, USA).

To demonstrate that the 7bp deletion led to the degradation of full-length *Azi2* mRNA transcripts, we performed qRT-PCR that amplified the exon 6-7 junction in *Azi2* mRNA transcripts ENSMUST00000044454.11, ENSMUST00000133580.7, and ENSMUST00000134433.7; in ENSMUST00000154583.7 this same sequence corresponds to exon 5-6. This amplicon would detect the four RefSeq *Azi2* transcripts and three predicted transcripts ENSMUST00000135251.1, ENSMUST00000130735.7, and ENSMUST00000133814.1. Our CRISPR/Cas9 deletion scheme ensured that all four full-length, RefSeq supported *Azi2* transcripts and a few short *Azi2* predicted transcripts would be degraded via nonsense-mediated mRNA decay due to the deletion. Given the mRNA degradation, our qRT-PCR design for whole-gene *Azi2* expression would only detect the two short, predicted *Azi2* transcripts, ENSMUST0000013525.1 and ENSMUST00000133814,1, in heterozygous and mutant mice.

The *Azi2* 3'UTR on exon 8 we amplified using qRT-PCR is homologous to the 3'UTR amplified in the rat alcohol model (Liu et al., 2018) and is only present in two of the four full-length *Azi2* transcripts, ENSMUST0000044454.11 and ENSMUST00000133580.7.

We measured *Azi2* expression in 44 mice from the *Azi2* KO line, which included 15 homozygous mutants. We measured the *Azi2* 3'UTR expression in an additional 33 mice from the *Azi2* KO line. Each batch of the *Azi2* KO mice used for gene expression assays were of similar age (199-201 days at sacrifice for the mice used for measuring *Azi2* mRNA and 226-232 days at sacrifice for the mice used for measuring *Azi2* 3'UTR mRNA).

We also performed qRT-PCR in CFW mice. We removed three animals whose *Azi2* or *Slc6a3* gene expression was more than three standard deviations away from the mean. We analyzed a final set of 31 CFW mice for the *Azi2* and *Slc6a3* expression in the striatum and in the VTA (rs46497021: 'GG' n=5, 'GA' n=20, 'AA' n=6; rs234453358: 'AA' n=12, 'AG' n=13, 'GG' n=6).

RNA-Sequencing

We extracted RNA from the striatum and VTA of the mouse brain and prepared cDNA libraries from 68 samples with RNA integrity scores ≥7.0 (32 *Azi2* KO line, 36 CFW mice) as measured on a TapeStation (Agilent, Santa Clara, CA, USA). The cDNA libraries were prepared with the NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina

(NEW ENGLAND BioLabs, Ipswich, MA, USA) and sequenced on two lanes (two chips, one lane on each chip) of an Illumina NovaSeq S4 using 100 bp, paired-end reads.

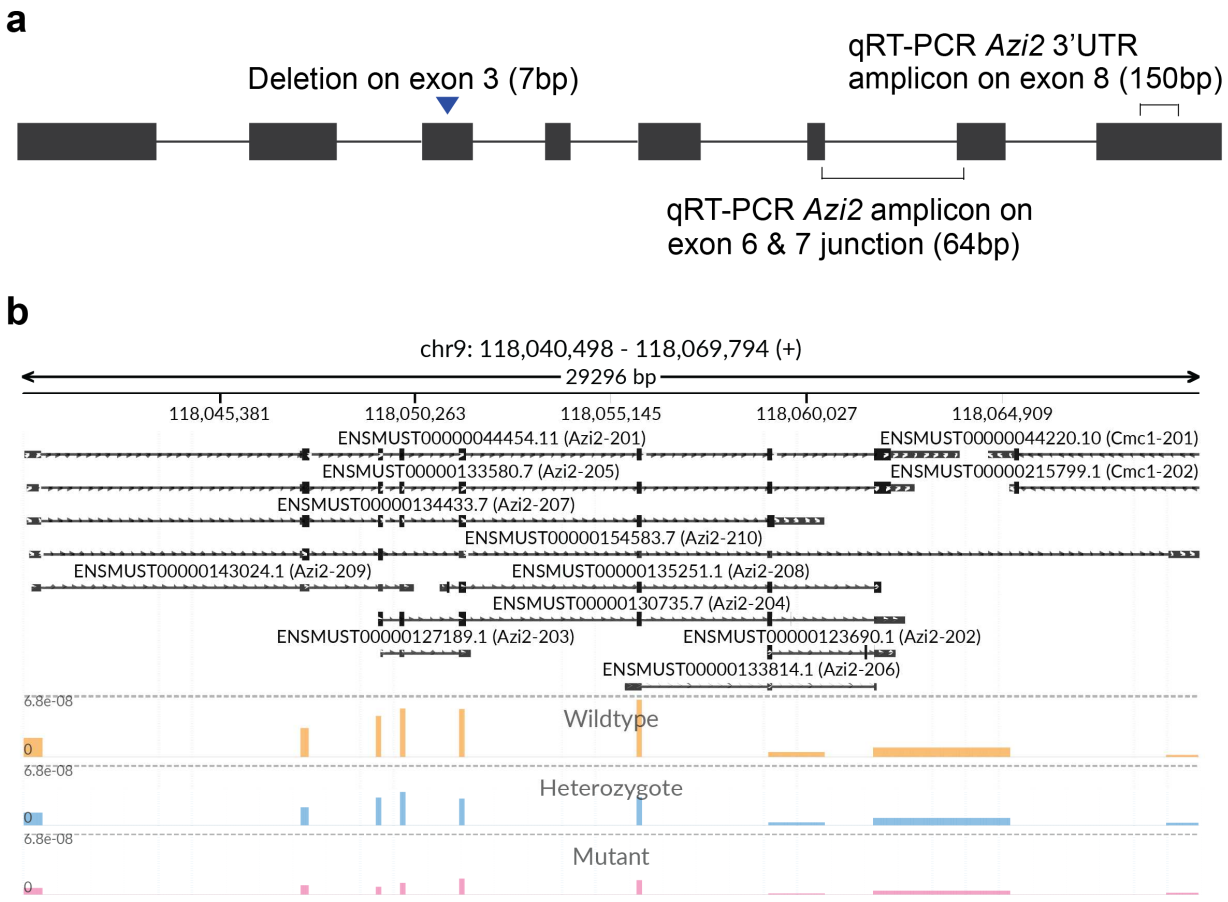
All sequencing reads passed the Illumina sequencing quality score of 20. We used HISAT2 (Kim et al., 2015) to align the adapter-trimmed paired-end reads simultaneously to mouse reference genome mm10. We used HTSeq to assign reads to gene features, in which the union of all the sets of all features overlapping each position *i* in the read was counted (Anders et al., 2015). We then examined potential expression outliers due to technical variance in PCA plots and removed one sample from the F_9 of *Azi2* KO cohort and two samples from the CFW cohort. We used the final set of 31 F_9 of Azi2 KO samples (16 striatum samples: 5 wildtype, 6 heterozygous, 5 mutant; 15 VTA samples: 3 wildtype, 7 heterozygous, 5 mutant) and 34 CFW samples (17 striatum samples: 5 'AA', 6 'AG', 6 'GG'; 17 VTA samples: 6 'AA', 6 'AG', 5 'GG').

We calculated the read counts aligned to each exon feature of *Azi2* by providing Samtools the genomic coordinates of the exon features (Li et al., 2009). Then, we normalized the read counts by dividing the raw reads by the length of the exon feature and the total number of reads in the sample. Normalized read counts for *Azi2* whole gene was calculated by summing normalized read counts aligned to exons 1-8 for transcripts ENSMUST00000044454.11, ENSMUST00000133580.7, and ENSMUST00000134433.7 and exons 1-7 for transcript ENSMUST00000154583.7 (chr9.118040522-chr9.118069794). Normalized read counts between chr9.118063214-chr9.118063336, the mouse genomic region homologous to the second half of *Azi2* 3'UTR in rats, as described in Liu et al (2018) (Liu et al., 2018), were assigned as *Azi2* 3'UTR normalized read counts.

We used DESeq2 to perform differential expression analysis (Love et al., 2014). Prior to count normalization and differential expression analysis, we calculated the average count per million (CPM) within each cohort and tissue combination across genes and samples. We only retained genes with CPM larger than 1. Then, we disabled independent filtering, which identifies the maximum number of adjusted *p* values lower than a significance level alpha based on the mean of normalized counts. We kept the "Cook's distance" parameter in DESeq2, which removes genes with extreme count outliers that do not fit well to the negative binomial distribution. These procedures ensured that genes with extremely low and large raw counts are removed and the same set of genes are used in differential expression analysis in between-genotype comparisons. At the end of filtering steps, we have 14,680 genes in *Azi2* KO striatum, 15,003 genes in *Azi2* KO VTA, 14,594 genes in CFW striatum, 14,936 genes in CFW VTA. All differential expression analyses performed on the *Azi2* KO mice have the design of ~ genotype + sex because we observed a separation by sex effect in PCA plots; analyses on the CFW mice only included the genotype factor because all mice were male.

2.6 Figures

Figure 2.1. A 7bp deletion on exon 3 of *Azi2* was generated by CRISPR/Cas9. The genomic position of the CRISPR/cas9 deletion and the qRT-PCR amplicons for *Azi2* mRNA and *Azi2* 3'UTR mRNA are indicated in **a**. **b** RNA-Seq reads aligned to the exons of *Azi2* in a wildtype, a heterozygous, and a mutant mouse in the *Azi2* KO line show the effect of the CRISPR/Cas9 deletion on mRNA abundance in the striatum. The mutant mouse had fewer reads across all the exons than the heterozygote, which in turn had fewer reads than the wildtype. The RNA-Seq reads aligned to each exon feature were normalized to the length of exon and the total read counts of the sample. To choose the most representative sample for each genotype for *Azi2* expression in striatum, we calculated the average normalized read counts for each genotype and identified the sample closest to the average. Note that the tracks include all *Azi2* transcripts annotated in the comprehensive gene annotation file of GRCm38.p6; only the top four transcripts are supported by RefSeq. The genome tracks were plotted using the Python visualization tool svist4get (Egorov et al., 2019). **c** Using qRT-PCR, we showed that *Azi2* expression in the striatum in mutant *Azi2* KO mice is significantly lower than the heterozygous and the wildtype mice ($F(2,41) = 319.41$, $p < 2.2\times10^{-16}$). Delta CT was calculated as the mean CT of target gene (*Azi2*) – the mean CT of the control gene (*Gapdh*); larger values of -delta CT indicate higher gene expression level. *Azi2* expression was measured using real time PCR in 15 wildtype, 14 heterozygous, and 15 mutant *Azi2* KO mice. We used Welch two sample t-test to make between group comparisons, which show that all groups were different (wildtype vs heterozygote $t(26.301) = -9.365$, $p = 7.312\times10^{-10}$; heterozygote vs mutant $t(25.906) = -16.297$, $p = 3.919\times10^{-15}$; wildtype vs mutant $t(27.944) = -23.3$, $p < 2.2\times10^{-16}$). **d** Using RNA-Seq we showed that normalized read counts mapped to *Azi2* in the striatum of the *Azi2* KO line also show a significant effect of the CRISPR/Cas9 deletion ($F(2,13) = 146.04$, $p = 1.236\times10^{-9}$). RNA-Seq was performed in 5 wildtype, 6 heterozygous, and 5 mutant *Azi2* KO striatum samples. Between group comparisons show that all groups are different (wildtype vs heterozygote $t(7.8331) = 8.7235$, $p = 2.654\times10^{-5}$; heterozygote vs mutant $t(8.537) = 9.3156$, $p = 9.272\times10^{-6}$; wildtype vs mutant $t(7.8068) = 16.399$, $p = 2.501\times10^{-7}$).



Azi2 mRNA in striatum (qRT-PCR)

Azi2 reads in striatum (RNA-Seq)

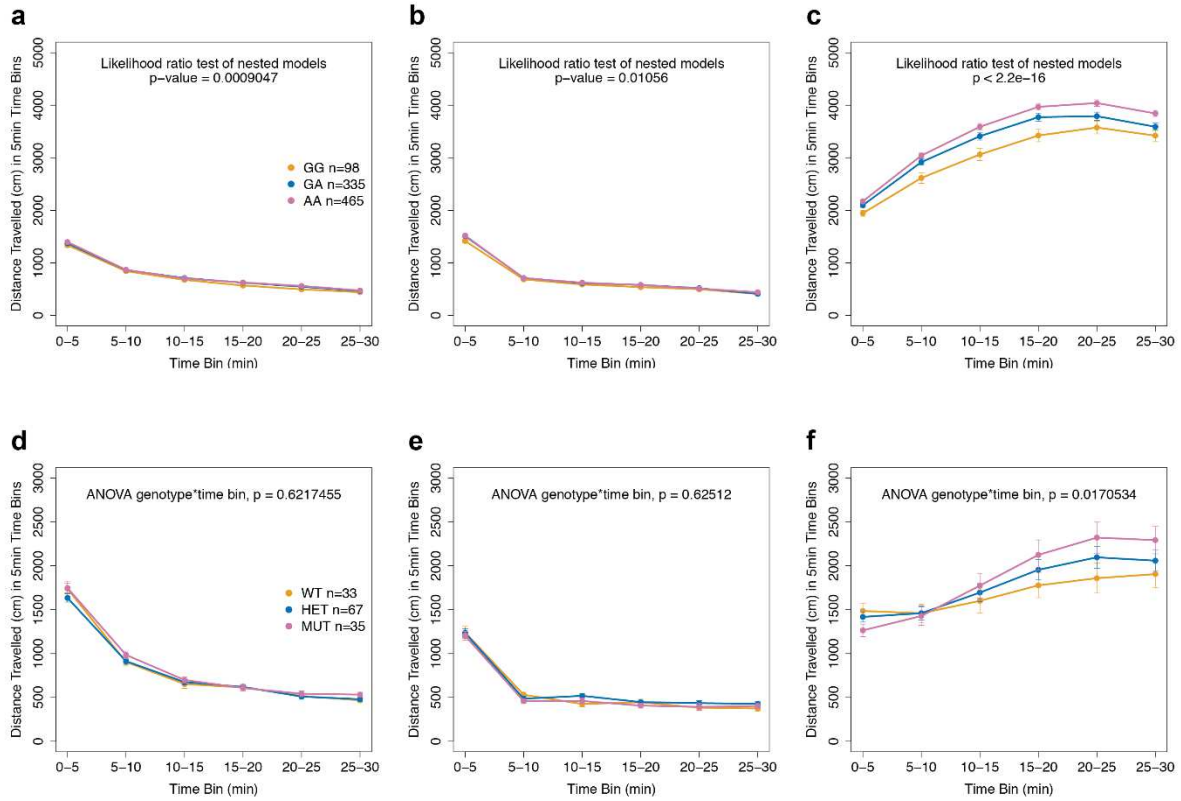
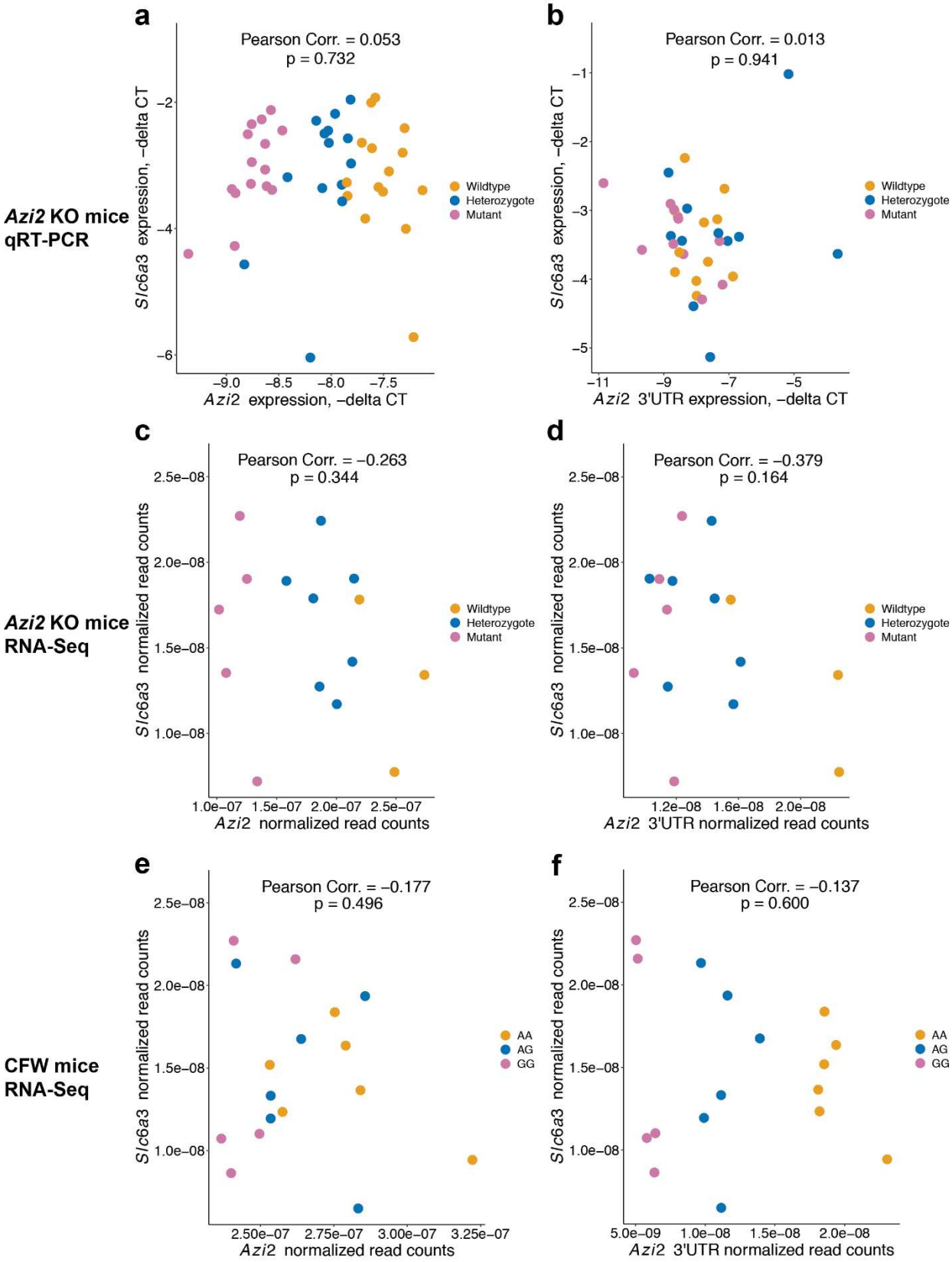


Figure 2.2. Locomotor response to methamphetamine was moderately heightened in mutant *Azi2* KO mice. a - **c** Locomotor activity of CFW (n=898 as previously reported in Parker et al 2016) following administration of saline (days 1 and 2) and methamphetamine (day 3), plotted in 5 min time bins. The genotype effect was noticeably more significant on day 3 of methamphetamine injection ($X^2(1,2) = 70.598$, $p < 2.2\times10^{-16}$) than that on day 1 ($X^2(1,2) = 11.013$, $p = 0.0009047$) and day 2 of saline injection ($X^2(1,2) = 6.5381$, $p = 0.01056$). **d** - **f** A total of 135 mice from the *Azi2* KO line were tested in the locomotor response to methamphetamine experiment. To evaluate the effect of the *Azi2* mutant allele, we used an ANOVA to analyze the effects of genotype, time bin and sex on the locomotor response to methamphetamine. The 3-way interaction was not significant (see Table S3); however, there was a significant interaction between genotype and time bin ($F(2,798)= 4.09230$, $p = 0.0170534$). Post-hoc tests did not identify any particular time bin that was different, though there were some trends towards differences between the wildtype and mutant mice (Tables S3 and S4).

Figure 2.3. *Azi2* and *Azi2* 3'UTR did not down-regulate *Slc6a3* in the VTA in the *Azi2* KO mice; *Azi2* did not down-regulate *Slc6a3* in the VTA of naïve CFW mice. We used **a** 44 *Azi2* KO mice (wildtype = 15, heterozygote= 14, mutant= 15) for examining the correlation between *Azi2* and *Slc6a3* expression measured by qRT-PCR, **b** 33 *Azi2* KO mice (wildtype = 11, heterozygote= 11, mutant= 11) for examining the correlation between *Azi2* 3'UTR and *Slc6a3* expression measured by qRT-PCR, **c** & **d** 15 *Azi2* KO mice (wildtype = 3, heterozygote= 7, mutant= 5) for examining the correlation between *Azi2*/*Azi2* 3'UTR and *Slc6a3* normalized read counts measured by RNA-Seq, and **e** & **f** 17 CFW mice ('AA' = 6, 'AG' = 6, 'GG' = 5) for examining the correlation between *Azi2*/*Azi2* 3'UTR and *Slc6a3* normalized read counts measured by RNA-Seq. Neither *Azi2* nor *Azi2* 3'UTR expression was negatively correlated to *Slc6a3* expression in the VTA in either cohort of mice. **a & b** Neither the level of *Azi2* ($r(42) = 0.05321076$, $p = 0.7316$) nor *Azi2* 3'UTR expression ($r(31) = 0.01339348$, $p = 0.941$) was negatively correlated to the expression of *Slc6a3* in the VTA. **c & d** No significant correlation between *Azi2* and *Slc6a3* ($r(13) = -0.2627568$, $p = 0.3441$) or *Azi2* 3'UTR and *Slc6a3* ($r(13) = -0.3786666$, $p = 0.164$) was observed. **e & f** Neither *Azi2* ($r(15) = -0.1771307$, $p = 0.4964$) nor *Azi2* 3'UTR ($r(15) = -0.1370425$, $p = 0.5999$) was negatively correlated with *Slc6a3* expression in the VTA at the eQTL top SNP.


a
Pearson Corr. = 0.053
p = 0.732
Slc6a3 expression, −delta CT
Azi2 expression, −delta CT
Wildtype
Heterozygote
Mutant
b
Pearson Corr. = 0.013
p = 0.941
Slc6a3 expression, −delta CT
Azi2 3'UTR expression, −delta CT
Wildtype
Heterozygote
Mutant
Azi2 KO mice
qRT-PCR
c
Pearson Corr. = −0.263
p = 0.344
Slc6a3 normalized read counts
Azi2 normalized read counts
Wildtype
Heterozygote
Mutant
d
Pearson Corr. = −0.379
p = 0.164
Slc6a3 normalized read counts
Azi2 3'UTR normalized read counts
Wildtype
Heterozygote
Mutant
Azi2 KO mice
RNA-Seq
e
Pearson Corr. = −0.177
p = 0.496
Slc6a3 normalized read counts
Azi2 normalized read counts
AA
AG
GG
f
Pearson Corr. = −0.137
p = 0.600
Slc6a3 normalized read counts
Azi2 3'UTR normalized read counts
AA
AG
GG
CFW mice
RNA-Seq

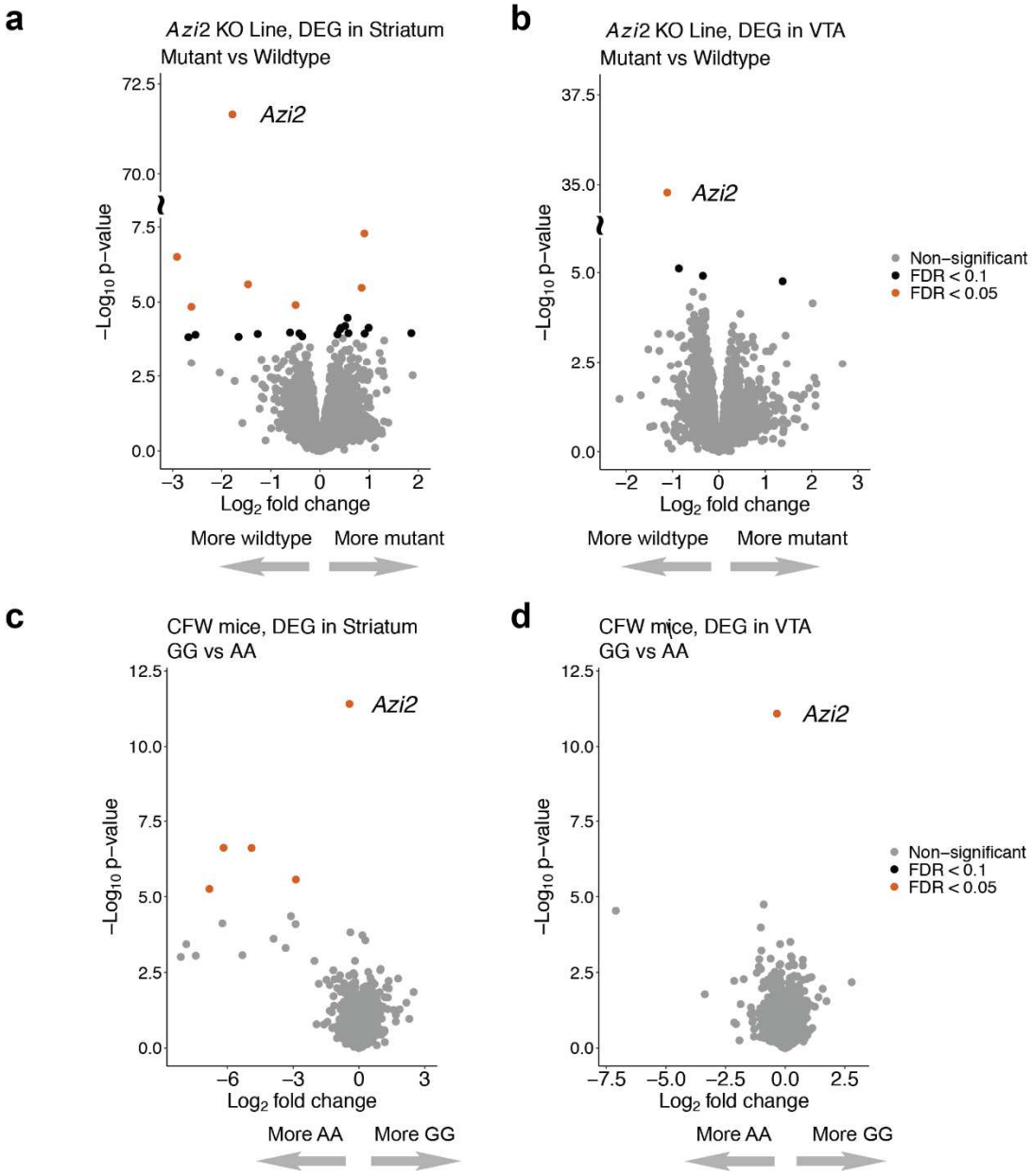


Figure 2.4. *Azi2* (ENSMUSG00000039285.12) was consistently differentially expressed between mutant vs wildtype mice in the *Azi2* KO line and between homozygous alternative ('GG') vs homozygous reference ('AA') mice at the top eQTL SNP for *Azi2* expression (Parker et al 2016; rs234453358) in the naïve CFW mice. a & **b** In the *Azi2* KO line, differential expression is performed on 16 striatum samples (wildtype = 5, heterozygote = 6, mutant = 5) and 15 VTA samples (wildtype = 3, heterozygote = 7, mutant = 5). **c** & **d** In the CFW mice, differential expression is performed on 17 striatum samples ('AA' = 5, 'AG' = 6, 'GG' = 6) and 17 VTA samples ('AA' = 6, 'AG' = 6, 'GG' = 5). Genes with FDR p-value < 0.05 are in shown in orange, and genes with FDR p-value < 0.1 are shown in black.

2.7 Author Contributions

AAP and XZ designed the study, oversaw data collection and analysis, and co-wrote the manuscript. XZ designed the CRISPR/Cas9 knockout scheme, maintained mouse colonies, conducted behavioral experiments, collected brain tissues, extracted RNA and DNA from mice, genotyped the mice, performed qRT-PCR, and analyzed the behavioral and gene expression data under supervision of AAP. ABL provided council for mouse handling, mouse behavioral tests and brain dissection. PM performed quality control of RNA and prepared cDNA libraries for RNA-Seq under supervision of FT.

2.8 Acknowledgements

We gratefully acknowledge helpful discussion and guidance on RNA-Seq design from Francesca Telese and Patricia Montilla-Perez.

Chapter 2, in full, is currently being prepared for submission for publication of the material. Xinzhu Zhou, Amanda Barkley-Levenson, Patricia Montilla-Perez, Francesca Telese, Abraham A. Palmer. The dissertation author was the primary investigator and author of this material.

2.9 References

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. https://doi.org/10.1093/bioinformatics/btu638

Bozóky, B., Savchenko, A., Csermely, P., Korcsmáros, T., Dúl, Z., Pontén, F., Székely, L., & Klein, G. (2013). Novel signatures of cancer-associated fibroblasts. *International Journal of Cancer*, *133*(2), 286–293. https://doi.org/10.1002/ijc.28035

Buchwald, J., Chenoweth, M. J., Palviainen, T., Zhu, G., Benner, C., Gordon, S., Korhonen, T., Ripatti, S., Madden, P. A. F., Lehtimäki, T., Raitakari, O. T., Salomaa, V., Rose, R. J., George, T. P., Lerman, C., Pirinen, M., Martin, N. G., Kaprio, J., Loukola, A., & Tyndale, R. F. (2020). Genome-wide association meta-analysis of nicotine metabolism and cigarette consumption measures in smokers of European descent. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-020-0702-z

Cheng, R., Lim, J. E., Samocha, K. E., Sokoloff, G., Abney, M., Skol, A. D., & Palmer, A. A. (2010). Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. *Genetics*.

de Wit, H., & Phillips, T. J. (2012). Do initial responses to drugs predict future use or abuse? *Neuroscience & Biobehavioral Reviews*, *36*(6), 1565–1576. https://doi.org/10.1016/j.neubiorev.2012.04.005

Deminiere, J. M., Piazza, P. V., Le Moal, M., & Simon, H. (1989). Experimental approach to individual vulnerability to psychostimulant addiction. *Neuroscience and Biobehavioral Reviews*, *13*(2–3), 141–147. https://doi.org/10.1016/s0149-7634(89)80023-5

Ettenberg, A. (2009). The runway model of drug self-administration. *Pharmacology, Biochemistry, and Behavior*, *91*(3), 271–277. https://doi.org/10.1016/j.pbb.2008.11.003

Fujita, F., Taniguchi, Y., Kato, T., Narita, Y., Furuya, A., Ogawa, T., Sakurai, H., Joh, T., Itoh, M., Delhase, M., Karin, M., & Nakanishi, M. (2003). Identification of NAP1, a regulatory subunit of IkappaB kinase-related kinases that potentiates NF-kappaB signaling. *Molecular and Cellular Biology*, *23*(21), 7780–7793.

Fukasaka, M., Ori, D., Kawagoe, T., Uematsu, S., Maruyama, K., Okazaki, T., Kozaki, T., Imamura, T., Tartey, S., Mino, T., Satoh, T., Akira, S., & Takeuchi, O. (2013). Critical Role of AZI2 in GM-CSF–Induced Dendritic Cell Differentiation. *The Journal of Immunology*, *190*(11), 5702–5711. https://doi.org/10.4049/jimmunol.1203155

Glick, S. D., & Hinds, P. A. (1985). Differences in amphetamine and morphine sensitivity in lateralized and non-lateralized rats: Locomotor activity and drug self-administration. *European Journal of Pharmacology*, *118*(3), 239–244. https://doi.org/10.1016/0014-2999(85)90134-7

Goncalves, A., Bürckstümmer, T., Dixit, E., Scheicher, R., Górna, M. W., Karayel, E., Sugar, C., Stukalov, A., Berg, T., Kralovics, R., Planyavsky, M., Bennett, K. L., Colinge, J., & Superti-Furga, G. (2011). Functional Dissection of the TBK1 Molecular Network. *PLOS ONE*, *6*(9), e23971. https://doi.org/10.1371/journal.pone.0023971

Gonzales, N. M., Seo, J., Cordero, A. I. H., Pierre, C. L. S., Gregory, J. S., Distler, M. G., Abney, M., Canzar, S., Lionikas, A., & Palmer, A. A. (2018). Genome wide association analysis in a mouse advanced intercross line. *Nature Communications*, *9*(1), 5162. https://doi.org/10.1038/s41467-018-07642-8

Hancock, D. B., Markunas, C. A., Bierut, L. J., & Johnson, E. O. (2018). Human Genetics of Addiction: New Insights and Future Directions. *Current Psychiatry Reports*, *20*(2), 8. https://doi.org/10.1007/s11920-018-0873-3

Hart, A. B., de Wit, H., & Palmer, A. A. (2012). Genetic factors modulating the response to stimulant drugs in humans. *Current Topics in Behavioral Neurosciences*, *12*, 537–577. https://doi.org/10.1007/7854_2011_187

Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, *12*(4), 357–360. https://doi.org/10.1038/nmeth.3317

Kinreich, S., Meyers, J. L., Maron-Katz, A., Kamarajan, C., Pandey, A. K., Chorlian, D. B., Zhang, J., Pandey, G., Subbie-Saenz de Viteri, S., Pitti, D., Anokhin, A. P., Bauer, L., Hesselbrock, V., Schuckit, M. A., Edenberg, H. J., & Porjesz, B. (2019). Predicting risk for Alcohol Use Disorder using longitudinal data with multimodal biomarkers and family history: A machine learning study. *Molecular Psychiatry*. https://doi.org/10.1038/s41380-019-0534-x

Kish, S. J. (2008). Pharmacologic mechanisms of crystal meth. *CMAJ*, *178*(13), 1679–1682. https://doi.org/10.1503/cmaj.071675

Lafont, E., Draber, P., Rieser, E., Reichert, M., Kupka, S., de Miguel, D., Draberova, H., von Mässenhausen, A., Bhamra, A., Henderson, S., Wojdyla, K., Chalk, A., Surinova, S., Linkermann, A., & Walczak, H. (2018). TBK1 and IKKε prevent TNF-induced cell death by RIPK1 phosphorylation. *Nature Cell Biology*, *20*(12), 1389–1399. https://doi.org/10.1038/s41556-018-0229-6

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 (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Liu, K., Yu, J., Zhao, J., Zhou, Y., Xiong, N., Xu, J., Wang, T., Bell, R. L., Qing, H., & Lin, Z. (2018). AZI23'UTR Is a New SLC6A3 Downregulator Associated with an Epistatic Protection Against Substance Use Disorders. *Molecular Neurobiology*, *55*(7), 5611–5622. https://doi.org/10.1007/s12035-017-0781-2

Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., Datta, G., Davila-Velderrain, J., McGuire, D., Tian, C., Zhan, X., 23andMe Research Team, HUNT All-In Psychiatry, Choquet, H., Docherty, A. R., Faul, J. D., Foerster, J. R., Fritsche, L. G., Gabrielsen, M. E., Gordon, S. D., Haessler, J., Hottenga, J.-J., Huang, H., Jang, S.-K., Jansen, P. R., Ling, Y., Mägi, R., Matoba, N., McMahon, G., Mulas, A., Orrù, V., Palviainen, T., Pandit, A., Reginsson, G. W., Skogholt, A. H., Smith, J. A., Taylor, A. E., Turman, C., Willemsen, G., Young, H., Young, K. A., Zajac, G. J. M., Zhao, W., Zhou, W., Bjornsdottir, G., Boardman, J. D., Boehnke, M., Boomsma, D. I., Chen, C., Cucca, F., Davies, G. E., Eaton, C. B., Ehringer, M. A., Esko, T., Fiorillo, E., Gillespie, N. A., Gudbjartsson, D. F., Haller, T., Harris, K. M., Heath, A. C., Hewitt, J. K., Hickie, I. B., Hokanson, J. E., Hopfer, C. J., Hunter, D. J., Iacono, W. G., Johnson, E. O., Kamatani, Y., Kardia, S. L. R., Keller, M. C., Kellis, M., Kooperberg, C., Kraft, P., Krauter, K. S., Laakso, M., Lind, P. A., Loukola, A., Lutz, S. M., Madden, P. A. F., Martin, N. G., McGue, M., McQueen, M. B., Medland, S. E., Metspalu, A., Mohlke, K. L., Nielsen, J. B., Okada, Y., Peters, U., Polderman, T. J. C., Posthuma, D., Reiner, A. P., Rice, J. P., Rimm, E., Rose, R. J., Runarsdottir, V., Stallings, M. C., Stančáková, A., Stefansson, H., Thai, K. K., Tindle, H. A., Tyrfingsson, T., Wall, T. L., Weir, D. R., Weisner, C., Whitfield, J. B., Winsvold, B. S., Yin, J., Zuccolo, L., Bierut, L. J., Hveem, K., Lee, J. J., Munafò, M. R., Saccone, N. L., Willer, C. J., Cornelis, M. C., David, S. P., Hinds, D. A., Jorgenson, E., Kaprio, J., Stitzel, J. A., Stefansson, K., Thorgeirsson, T. E., Abecasis, G., Liu, D. J., & Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics*, *51*(2), 237–244. https://doi.org/10.1038/s41588-018-0307-5

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. https://doi.org/10.1186/s13059-014-0550-8

Maruyama, K., Fukasaka, M., Uematsu, S., Takeuchi, O., Kondo, T., Saitoh, T., Martino, M. M., & Akira, S. (2015). 5-Azacytidine-induced Protein 2 (AZI2) Regulates Bone Mass by Fine-tuning Osteoclast Survival. *Journal of Biological Chemistry*, *290*(15), 9377–9386. https://doi.org/10.1074/jbc.M114.631374

NIDA. (2020, April). *What is the scope of methamphetamine misuse in the United States?* National Institute on Drug Abuse. https://www.drugabuse.gov/publications/research-

reports/methamphetamine/what-scope-methamphetamine-misuse-in-united-states

Parker, C. C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N. M., Leung, E., Park, Y. J., Aryee, E., Davis, J., Blizard, D. A., Ackert-Bicknell, C. L., Lionikas, A., Pritchard, J. K., & Palmer, A. A. (2016). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, *48*(8), 919–926. https://doi.org/10.1038/ng.3609

Phillips, T., & Shabani, S. (2015). An animal model of differential genetic risk for methamphetamine intake. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00327

Piazza, P. V., Deminière, J. M., Le Moal, M., & Simon, H. (1989). Factors that predict individual vulnerability to amphetamine self-administration. *Science (New York, N.Y.)*, *245*(4925), 1511–1513. https://doi.org/10.1126/science.2781295

Sanchez-Roige, S., Palmer, A. A., Fontanillas, P., Elson, S. L., 23andMe Research Team, the Substance Use Disorder Working Group of the Psychiatric Genomics Consortium, Adams, M. J., Howard, D. M., Edenberg, H. J., Davies, G., Crist, R. C., Deary, I. J., McIntosh, A. M., & Clarke, T.-K. (2019). Genome-Wide Association Study Meta-Analysis of the Alcohol Use Disorders Identification Test (AUDIT) in Two Population-Based Cohorts. *The American Journal of Psychiatry*, *176*(2), 107–118. https://doi.org/10.1176/appi.ajp.2018.18040369

Seale, T. W. (1991). Genetic Differences in Response to Cocaine and Stimulant Drugs. In J. C. Crabbe & R. A. Harris (Eds.), *The Genetic Basis of Alcohol and Drug Actions* (pp. 279–321). Springer US. https://doi.org/10.1007/978-1-4899-2067-6_8

Sittig, L. J., Carbonetto, P., Engel, K. A., Krauss, K. S., Barrios-Camacho, C. M., & Palmer, A. A. (2016). Genetic Background Limits Generalizability of Genotype-Phenotype Relationships. *Neuron*, *91*(6), 1253–1259. https://doi.org/10.1016/j.neuron.2016.08.013

Smith, C. T., Weafer, J., Cowan, R. L., Kessler, R. M., Palmer, A. A., de Wit, H., & Zald, D. H. (2016). Individual differences in timing of peak positive subjective responses to d-amphetamine: Relationship to pharmacokinetics and physiology. *Journal of Psychopharmacology*, *30*(4), 330–343. https://doi.org/10.1177/0269881116631650

Steketee, J. D., & Kalivas, P. W. (2011). Drug wanting: Behavioral sensitization and relapse to drug-seeking behavior. *Pharmacological Reviews*, *63*(2), 348–365. https://doi.org/10.1124/pr.109.001933

Thomas, P. D., Campbell, M. J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., & Narechania, A. (2003). PANTHER: A Library of Protein

Families and Subfamilies Indexed by Function. *Genome Research*, *13*, 2129–2141. https://doi.org/doi: 10.1101/gr.772403

Walters, R. K., Polimanti, R., Johnson, E. C., McClintick, J. N., Adams, M. J., Adkins, A. E., Aliev, F., Bacanu, S.-A., Batzler, A., Bertelsen, S., Biernacka, J. M., Bigdeli, T. B., Chen, L.-S., Clarke, T.-K., Chou, Y.-L., Degenhardt, F., Docherty, A. R., Edwards, A. C., Fontanillas, P., Foo, J. C., Fox, L., Frank, J., Giegling, I., Gordon, S., Hack, L. M., Hartmann, A. M., Hartz, S. M., Heilmann-Heimbach, S., Herms, S., Hodgkinson, C., Hoffmann, P., Jan Hottenga, J., Kennedy, M. A., Alanne-Kinnunen, M., Konte, B., Lahti, J., Lahti-Pulkkinen, M., Lai, D., Ligthart, L., Loukola, A., Maher, B. S., Mbarek, H., McIntosh, A. M., McQueen, M. B., Meyers, J. L., Milaneschi, Y., Palviainen, T., Pearson, J. F., Peterson, R. E., Ripatti, S., Ryu, E., Saccone, N. L., Salvatore, J. E., Sanchez-Roige, S., Schwandt, M., Sherva, R., Streit, F., Strohmaier, J., Thomas, N., Wang, J.-C., Webb, B. T., Wedow, R., Wetherill, L., Wills, A. G., 23andMe Research Team, Boardman, J. D., Chen, D., Choi, D.-S., Copeland, W. E., Culverhouse, R. C., Dahmen, N., Degenhardt, L., Domingue, B. W., Elson, S. L., Frye, M. A., Gäbel, W., Hayward, C., Ising, M., Keyes, M., Kiefer, F., Kramer, J., Kuperman, S., Lucae, S., Lynskey, M. T., Maier, W., Mann, K., Männistö, S., Müller-Myhsok, B., Murray, A. D., Nurnberger, J. I., Palotie, A., Preuss, U., Räikkönen, K., Reynolds, M. D., Ridinger, M., Scherbaum, N., Schuckit, M. A., Soyka, M., Treutlein, J., Witt, S., Wodarz, N., Zill, P., Adkins, D. E., Boden, J. M., Boomsma, D. I., Bierut, L. J., Brown, S. A., Bucholz, K. K., Cichon, S., Costello, E. J., de Wit, H., Diazgranados, N., Dick, D. M., Eriksson, J. G., Farrer, L. A., Foroud, T. M., Gillespie, N. A., Goate, A. M., Goldman, D., Grucza, R. A., Hancock, D. B., Harris, K. M., Heath, A. C., Hesselbrock, V., Hewitt, J. K., Hopfer, C. J., Horwood, J., Iacono, W., Johnson, E. O., Kaprio, J. A., Karpyak, V. M., Kendler, K. S., Kranzler, H. R., Krauter, K., Lichtenstein, P., Lind, P. A., McGue, M., MacKillop, J., Madden, P. A. F., Maes, H. H., Magnusson, P., Martin, N. G., Medland, S. E., Montgomery, G. W., Nelson, E. C., Nöthen, M. M., Palmer, A. A., Pedersen, N. L., Penninx, B. W. J. H., Porjesz, B., Rice, J. P., Rietschel, M., Riley, B. P., Rose, R., Rujescu, D., Shen, P.-H., Silberg, J., Stallings, M. C., Tarter, R. E., Vanyukov, M. M., Vrieze, S., Wall, T. L., Whitfield, J. B., Zhao, H., Neale, B. M., Gelernter, J., Edenberg, H. J., & Agrawal, A. (2018). Transancestral GWAS of alcohol dependence reveals common genetic underpinnings with psychiatric disorders. *Nature Neuroscience*, *21*(12), 1656–1669. https://doi.org/10.1038/s41593-018-0275-1

Zeileis, A., & Hothorn, T. (2002). Diagnostic Checking in Regression Relationships. *R News*, *2*(3), 7–10.

Zhou, H., Sealock, J. M., Sanchez-Roige, S., Clarke, T.-K., Levey, D. F., Cheng, Z., Li, B., Polimanti, R., Kember, R. L., Smith, R. V., Thygesen, J. H., Morgan, M. Y., Atkinson, S. R., Thursz, M. R., Nyegaard, M., Mattheisen, M., Børglum, A. D., Johnson, E. C., Justice, A. C., Palmer, A. A., McQuillin, A., Davis, L. K., Edenberg, H. J., Agrawal, A., Kranzler, H. R., & Gelernter, J. (2020). Genome-wide meta-analysis of problematic alcohol use in 435,563 individuals yields insights into

biology and relationships with other traits. *Nature Neuroscience*, *23*(7), 809–818. https://doi.org/10.1038/s41593-020-0643-5

CONCLUSION

The present thesis has investigated the discovery, replication, and validation of GWAS results using the mouse as the model organism. In chapter 1, we used multiple generations of a LG/J x SM/J AIL to perform GWAS, SNP heritability estimates, genetic correlations, replication and mega-analysis. Despite the high genetic correlations in a subset of traits measured in two cohorts of the AIL population, we found that many loci were not replicated between cohorts. We found that the lack of replication was not a result of inadequate power in the replication study but a combined result of Winner's Curse and study-specific heterogeneity. Finally, mega-analysis of the two cohorts allowed us to discover four additional loci. This work highlights the difficulty of replication due to study-specific heterogeneity even when tight genetic and environmental control is allowed for. In chapter 2, we created an *Azi2* KO line, established the mutant *Azi2* phenotype of locomotor response to methamphetamine, and investigated the hypothesis that an independent *Azi2* 3'UTR transcript could downregulate *Slc6a3* expression in the VTA of the midbrain. We found a significant gene and 5 min time bin interaction on day three of the locomotor test when mice received a methamphetamine injection. We showed that locomotor response to methamphetamine was moderately heightened in mutant *Azi2* mice, a trend that is consistent with the CFW mice. We did not observe a negative correlation between *Azi2*/ *Azi2* 3'UTR expression and *Slc6a3* expression in the VTA in either the *Azi2* KO mice or in naïve CFW mice. Our results support the findings from Parker et al (2016) that *Azi2*, which is localized within the linkage disequilibrium (**LD**) block of the top GWAS and top eQTL loci for methamphetamine sensitivity, may be the causal gene for this trait (Parker et al., 2016). Our results do not support findings from Liu

et al (2018) that *Azi2* 3'UTR downregulates *Slc6a3* expression in the VTA (Liu et al., 2018).

The CRSIPR/Cas9 method for creating a knockout mouse line yields cleaner mutations than the traditional method of gene targeting using embryonic stem (**ES**) cells. Targeting vectors that include the homology arms to the targeted region, positive selection markers (*e.g.,* neo^r) and negative selection markers (*e.g.*, HSV-tk) are taken up into the ES cells by electroporation (Hall et al., 2009; Limaye et al., 2009). While the selection markers facilitate the enrichment of the incorporation of the vectors and the exclusion of random integration, the selection markers are left in the modified ES cells, which could bring collateral damage to the expression of genes flanking the targeted region (Conlon, 2011). As a solution, loxP and Frt sites could be designed to the drug selection gene cassette; after drug selection, those markers could be deleted by expressing Cre or Flp recombinase in a temporal and spatial-specific manner. However, this procedure involves a more complex design of the selection cassette and requires additional efforts to express the recombinase, either by inserting a Cre/Flp expression vector in ES cells or by mating the KO mouse with a mouse constitutively expressing the recombinase (Conlon, 2011). In contrast, CRISPR/Cas9 does not introduce extra selection constructs into the targeted region; the resulting transgenic founder mice simply have the mutation at the genomic position homologous to the sgRNA. Given the ease and precision of creating a KO allele using CRISPR/Cas9, more functional validation studies could be carried out to assess genetic variants identified in GWAS using the mouse model.

We created a null *Azi2* allele and investigated the impact of the *Azi2* mutation on methamphetamine sensitivity and *Slc6a3* expression in the VTA. While we replicated the pattern of methamphetamine sensitivity as shown in Parker et al (2016) (Parker et al., 2016) in the *Azi2* KO line, we did not observe any influence of *Azi2* mutation on *Slc6a3*. This lack of consequence could be explained by the important concept of degeneracy. Degeneracy refers to the wide-spread phenomenon in biological systems where multitudes of structurally different elements can yield similar outcome given similar constraints on the environment (Edelman & Gally, 2001). Single gene mutations have not always resulted in phenotypic changes, which suggests that compensatory mechanisms are at play in manipulated organisms. The lack of *Azi2*/*Azi2* 3'UTR regulation of *Slc6a3* in mouse could be a case of degeneracy, in which many pathways, networks and systems other than *Azi2*-related could make up the loss of function of *Azi2*. Thus, understanding the network effect of *Azi2* on methamphetamine sensitivity and *Slc6a3* expression could be an alternative approach to replicate the findings of Parker et al (2016) (Parker et al., 2016). Neurobiological processes, such as response to dopaminergic drugs, are highly affected by the interactions among behavior, environment, and neural mechanisms (Seth, 2005). The study of complex networks has been adopted in the field of neuroscience to illuminate the structural and functional connectivity of brain modules, either in isolation or in response to external stimuli (Sporns et al., 2004; G. Tononi et al., 1994; G Tononi et al., 1996). Furthermore, multiple methods that infer causality in network interactions have been developed, including Granger causality (*i.e.*, (Larvie et al., 2016; Seth, 2005; Tam et al., 2012)) and selective perturbations (*i.e.*, (Keinan et al., 2004; Pearl, 2000; Giulio

Tononi & Sporns, 2003). These methods could be employed to investigate the potential causal relation between *Azi2* and *Slc6a3* in a network context.

Evaluating genetic signals in GWAS, either via replication studies or functional validation studies, is not straightforward. Much of the challenge stems from the inherent properties of GWAS. GWAS results are not causative; LD regions instead of particular genetic variants are implicated in the association. Furthermore, only a small fraction of the identified genetic variants reside within coding regions (MacArthur et al., 2017). In chapter 2 of this thesis, we considered *Azi2* as a putative causal gene for methamphetamine sensitivity because both the GWAS and the *cis*-eQTL loci for methamphetamine sensitivity include this gene (Parker et al., 2016). Other methods, such as transcriptome-wide association studies (**TWAS**) (Gusev et al., 2016; Mancuso et al., 2017; Nica et al., 2010) and two-sample, univariable and multivariable Mendelian randomization studies (Burgess & Thompson, 2015; Inoue & Solon, 2010; Porcu et al., 2019; Zhu et al., 2016) can provide additional evidence for the directional causal effect from SNP to gene to trait.

Complex traits are highly polygenic in which the common alleles tend to have relatively small effect sizes. This poses additional hurdles to assess the credibility of the collection of genetic signals in GWAS. Replicating the genetic basis of a polygenic trait requires an exceedingly large sample size. Functional validation of multiple risk variants is also difficult. In order to fully capture the composite effect of multiple risk variants on the phenotype, one might wish to create multiple null alleles in one organism. However, this strategy may seriously compromise the fit, fertility and survival of the organism, which hinders subsequent experiments. Nevertheless, recent development in polygenic risk

scores (**PRS**) methods proves to be useful in profiling individuals that are probabilistically at risk for complex diseases (Torkamani et al., 2018).

Finally, other -omics could be combined with GWAS to demonstrate additional support for the identified loci. Genetic and structural variants, gene expression, transcriptome modifications, and chromosome conformation patterns associated with a trait provide a synergistic view of the processes that contribute to phenotypic variation (Gorkin et al., 2019).

There are many limitations to this thesis. In particular, the lack of replication presented in both chapters underlines the limitations of designing studies that do not have identical genetic background of the subjects and experimental conditions. For example, the two AIL cohorts included in the study presented in chapter 1 are a few generations apart. As a result, multiple technicians collected, processed, and recorded genotype and phenotype data spanning a couple of years. This may contribute to the study-specific heterogeneity that we observed between the two cohorts. A more ideal study design would include two consecutive generations of AIL mice handled by the same technician, at least for each phenotype. Another example is demonstrated in the study presented in chapter 2. We created the *Azi2* KO on the C57BL/6J background, which is different from the outbred population of CFW mice where the *Azi2* locus was initially identified. While we intended to generate the *Azi2* KO using the most popular laboratory strain so that the results may be more generalizable to behavioral studies performed on the C57BL/6J background, the strain difference between the C57BL/6J and the CFW mice may lead to differential behavioral performance in methamphetamine sensitivity. In addition, we conducted all of the *Azi2* gene validation experiments in the *Azi2* KO line and in naïve

CFW mice, whereas Liu et al (2018) examined *Azi2*/ *Azi2* 3'UTR in human cell lines, human postmortem nigral dopamine neurons, and in rats that had been selected for the binary trait alcohol preferring and alcohol non-preferring (Liu et al., 2018). As a result, despite the fact that we did not corroborate the findings from Liu et al (2018) that *Azi2* 3'UTR is a regulator of *Slc6a3*, we could not definitively argue that *Azi2* 3'UTR does not downregulate *Slc6a3* expression in the VTA in any biological system. The limitations of my thesis reveal the dilemma of the replication studies: replicating a result exactly requires conforming to the design of the original study, a practice that confines the generalizability of scientific discoveries.

Improvements in GWAS design and analysis facilitate better definition of the genetic architecture of disease. This is the first step of enhancing human health; ultimately, insights from human and model organism GWAS can be translated to efforts that improve diagnosis, fasten drug discovery, and invent new therapeutics (Boerwinkle & Heckbert, 2014). Replication and validation of GWAS findings, together with the multitude of methods that provide auxiliary evidence for the identified genetic variants, are essential in this undertaking.

References

Boerwinkle, E., & Heckbert, S. R. (2014). Following-up GWAS Signals: Lessons learned from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Targeted Sequencing Study. *Circulation. Cardiovascular Genetics*, *7*(3), 332–334. https://doi.org/10.1161/CIRCGENETICS.113.000078

Burgess, S., & Thompson, S. G. (2015). Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *American Journal of Epidemiology*, *181*(4), 251–260. https://doi.org/10.1093/aje/kwu283

Conlon, R. A. (2011). Transgenic and Gene Targeted Models of Dementia. In P. P. De Deyn & D. Van Dam (Eds.), *Animal Models of Dementia* (pp. 77–90). Humana Press. https://doi.org/10.1007/978-1-60761-898-0_5

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences*, *98*(24), 13763–13768. https://doi.org/10.1073/pnas.231499798

Gorkin, D. U., Qiu, Y., Hu, M., Fletez-Brant, K., Liu, T., Schmitt, A. D., Noor, A., Chiou, J., Gaulton, K. J., Sebat, J., Li, Y., Hansen, K. D., & Ren, B. (2019). Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biology*, *20*(1), 255. https://doi.org/10.1186/s13059-019-1855-4

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B. W. J. H., Jansen, R., de Geus, E. J. C., Boomsma, D. I., Wright, F. A., Sullivan, P. F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A. J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O. T., Kuusisto, J., Laakso, M., Price, A. L., Pajukanta, P., & Pasaniuc, B. (2016). Integrative approaches for large-scale transcriptome-wide association studies. *Nature Genetics*, *48*(3), 245–252. https://doi.org/10.1038/ng.3506

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: Generation of Gene Knockout Mice. *Current Protocols in Cell Biology / Editorial Board, Juan S. Bonifacino ... [et Al.]*, *CHAPTER*, Unit-19.1217. https://doi.org/10.1002/0471143030.cb1912s44

Inoue, A., & Solon, G. (2010). Two-Sample Instrumental Variables Estimators. *The Review of Economics and Statistics*, *92*(3), 557–561.

Keinan, A., Sandbank, B., Hilgetag, C. C., Meilijson, I., & Ruppin, E. (2004). Fair attribution of functional contribution in artificial and biological networks. *Neural Computation*, *16*(9), 1887–1915. https://doi.org/10.1162/0899766041336387

Larvie, J. E., Sefidmazgi, M. G., Homaifar, A., Harrison, S. H., Karimoddini, A., & Guiseppi-Elie, A. (2016). Stable Gene Regulatory Network Modeling From Steady-State Data. *Bioengineering*, *3*(2). https://doi.org/10.3390/bioengineering3020012

Limaye, A., Hall, B., & Kulkarni, A. B. (2009). Manipulation of Mouse Embryonic Stem Cells for Knockout Mouse Production. *Current Protocols in Cell Biology / Editorial Board, Juan S. Bonifacino ... [et Al.]*, *CHAPTER*, Unit-19.1324. https://doi.org/10.1002/0471143030.cb1913s44

Liu, K., Yu, J., Zhao, J., Zhou, Y., Xiong, N., Xu, J., Wang, T., Bell, R. L., Qing, H., & Lin, Z. (2018). AZI23'UTR Is a New SLC6A3 Downregulator Associated with an Epistatic Protection Against Substance Use Disorders. *Molecular Neurobiology*, *55*(7), 5611–5622. https://doi.org/10.1007/s12035-017-0781-2

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., Junkins, H., McMahon, A., Milano, A., Morales, J., Pendlington, Z. M., Welter, D., Burdett, T., Hindorff, L., Flicek, P., Cunningham, F., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(D1), D896–D901. https://doi.org/10.1093/nar/gkw1133

Mancuso, N., Shi, H., Goddard, P., Kichaev, G., Gusev, A., & Pasaniuc, B. (2017). Integrating Gene Expression with Summary Association Statistics to Identify Genes Associated with 30 Complex Traits. *American Journal of Human Genetics*, *100*(3), 473–487. https://doi.org/10.1016/j.ajhg.2017.01.031

Nica, A. C., Montgomery, S. B., Dimas, A. S., Stranger, B. E., Beazley, C., Barroso, I., & Dermitzakis, E. T. (2010). Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. *PLOS Genetics*, *6*(4), e1000895. https://doi.org/10.1371/journal.pgen.1000895

Parker, C. C., Gopalakrishnan, S., Carbonetto, P., Gonzales, N. M., Leung, E., Park, Y. J., Aryee, E., Davis, J., Blizard, D. A., Ackert-Bicknell, C. L., Lionikas, A., Pritchard, J. K., & Palmer, A. A. (2016). Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, *48*(8), 919–926. https://doi.org/10.1038/ng.3609

Pearl, J. (2000). *CAUSALITY: MODELS, REASONING, AND INFERENCE*. https://search.proquest.com/openview/28a611423e61457e5c1379d1100a3418/1/advanced

Porcu, E., Rüeger, S., Lepik, K., Santoni, F. A., Reymond, A., & Kutalik, Z. (2019). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-10936-0

Seth, A. K. (2005). Causal connectivity of evolved neural networks during behavior. *Network (Bristol, England)*, *16*(1), 35–54. https://doi.org/10.1080/09548980500238756

Sporns, O., Chialvo, D. R., Kaiser, M., & Hilgetag, C. C. (2004). Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, *8*(9), 418–425. https://doi.org/10.1016/j.tics.2004.07.008

Tam, G. H. F., Chang, C., & Hung, Y. S. (2012). Application of Granger causality to gene regulatory network discovery. *2012 IEEE 6th International Conference on Systems Biology (ISB)*, 232–239. https://doi.org/10.1109/ISB.2012.6314142

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proceedings of the National Academy of Sciences*, *91*(11), 5033–5037. https://doi.org/10.1073/pnas.91.11.5033

Tononi, G, Sporns, O., & Edelman, G. M. (1996). A complexity measure for selective matching of signals by the brain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(8), 3422–3427.

Tononi, Giulio, & Sporns, O. (2003). Measuring information integration. *BMC Neuroscience*, *4*, 31. https://doi.org/10.1186/1471-2202-4-31

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, *19*(9), 581–590. https://doi.org/10.1038/s41576-018-0018-x

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M. R., Powell, J. E., Montgomery, G. W., Goddard, M. E., Wray, N. R., Visscher, P. M., & Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature Genetics*, *48*(5), 481–487. https://doi.org/10.1038/ng.3538