THE UNIVERSITY OF CHICAGO

INTEGRATIVE GENETIC ANALYSIS OF BEHAVIORAL AND METABOLIC TRAITS IN AN ADVANCED INTERCROSS LINE OF MICE

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY
NATALIA M. GONZALES

CHICAGO, ILLINOIS
DECEMBER 2017

"Man had always assumed that he was more intelligent than dolphins because he had achieved so much - the wheel, New York, wars and so on... In fact there was only one species on the planet more intelligent than dolphins, and they spent a lot of their time in behavioural research laboratories running round inside wheels and conducting frighteningly elegant and subtle experiments on man. The fact that once again man completely misinterpreted this relationship was entirely according to these creatures' plans."

Douglas Adams, *The Hitchhiker's Guide to the Galaxy*, 1979.

Table of Contents

List of Figures

List of Tables [1]

1. Note: Due to the large size of some tables, the tables have been provided in a supplementary file accompanying the dissertation. In such cases, the page number provided above directs the reader to a table's caption.

ACKNOWLEDGMENTS

I've incurred a lot of debt in graduate school. It's not the financial kind, it's the good kind - the kind of debt that's impossible to repay.

I doubt I'd have been able to complete this journey without the lessons I learned from Elaine and Roger Gonzales, my parents. Both of them grew up in rural New Mexico and come from large families. They managed to make their way through college (fueled primarily by potted meat and saltines, as my mom tells it), and my dad also served in the army and earned a master's degree. They've always put a high value on education; I remember them emphasizing the importance of having a college degree when I was quite young. To some extent, I think this is because they probably struggled a lot along the way and didn't want me to take knowledge for granted. I can't thank them enough for the incredible foresight, selflessness, and hard work that provided me with the best gift I could ever ask for: an education. Mom and Dad, you raised me to be a self-sufficient and curious young woman who loves learning more than anything else, and I am so grateful to have you as my parents. You were proud when I got A's in school, but my GPA was never as important as things like effort, fortitude, and compassion. Grades may have helped me get into graduate school, but it's those other qualities that helped me make it through.

Mom, thank you for teaching me how to read, how to spell, and for taking me to the library every week during the summer. Thank you for reversing your circadian rhythm for 30 years to support Dad and to pick Noah and I up from school every day. Thank you for somehow managing to put a delicious, nutritious, homemade dinner on the table 7 days a week (and for teaching me how to cook and eat like an adult). Dad, thank you for reading to me and drawing with me as a kid (these are still my favorite hobbies). Thank you for providing critical defense during my epic battle with Sister Margeurite. Thank you for supporting my decision to go to school thousands of miles away even though I could have stayed closer to home. Guys, I'm also really thankful that I have a brother (even if he is

part Neanderthal). I try to play it cool but I love him more than anyone else on the planet.

I have a large family, but Cathy Tulley and Lino Gonzales are the aunt and uncle that know me the best. They've visited me every fall since I moved to Chicago, and on top of the love and support they've given me over the years, they also spoil me terribly. Thank you for feeding me steak and bacon, for restoring broken crayons, and for teaching me that "life is too short" to waste on things that don't matter in the long run.

I owe a lot to Mike Brucato, Elliot Heilman, Dave Samuels, Natalie Diaz, Frank Naughton, Steph Reist and Yawen Lu for their friendship and support. I'm grateful for mentors like Ed Smith, Clarissa Parker, and Shyam Gopalakrishnan for providing crucial life and career advice at various points during my time in grad school. I'm also grateful to my thesis committee, which included Carole Ober, Yoav Gilad, Mark Abney, and Xiaoxi Zhuang, for helping me to grow as a scientist and communicator. Mark, Carole, and John Novembre gave me a space to work, provided advice, and made me feel welcome after the Palmer Lab moved to UCSD - thank you all for your generosity.

Speaking of the Palmer Lab, the last person I want to acknowledge is the man himself, Abraham Palmer. I'm very fond of Abe and I'm happy to have him as an advisor. It usually takes me a long time to feel comfortable around new people, but every once in a while I meet someone who disarms me right away, and Abe is one of them. Even if we hadn't shared an interest in behavior, it was immediately clear to me that we would get along, and I was right. The main reason is because in addition to being a good mentor and friend, Abe is an unusually genuine person (I'm not sure it's possible to wear mutton chops in earnest without possessing this quality). This is obvious from the way he expresses his thoughts directly, informally, and without filter - and since he also has a sense of humor, communicating with him has been pretty easy. All of this explains how in 2012, I let a man with mutton chops convince me that if I took 1000 mice, some meth, and a computer, I could turn it into a Ph.D.

In years that followed, Abe trusted me with an ambitious project, improved my written and oral communication skills, taught me how to write grants, and he let me draw a portrait of him as a T-rex in permanent marker on the wall outside of his office. Abe also taught me the art of leveraging, which I combined with my newly-acquired grant writing skills to convince him to send me to meetings in Boulder, D.C., Madison, Galveston, L.A., Bar Harbor, Phoenix, Blacksburg, Bar Harbor again, Portland, Germany, Italy, Japan, and New Buffalo, MI (all between 2012-2015). This amount of freedom is somewhat unusual, and I'm not sure if I owe it to Abe's generosity or his forgetfulness, but I'm grateful either way. Thank you, Abe.

To the rest of the people who helped me with this project, thank you. I've acknowledged your work under *Contributions* in Chapters 2 and 3.

Finally, because I love animals, I have to say this - I am so sorry. I'm sorry that I personally killed hundreds of mice with my own hands. I put a high value on knowledge, but it was never an easy thing to do and I still think about it often. If you use mice to study behaviors like fear, anxiety and depression, you *have* to believe that they can experience these things (or something like them) - otherwise, what are you studying them for? That's part of why it's difficult not to see them as individuals sometimes. So, thank you, little mammals, for your lives that I took, thank you a thousand times.

ABSTRACT

Mice are one of the primary organisms used to study mammalian behavior and physiology. The goal of this study was to identify genes in the mouse that would allow us to develop specific, testable hypotheses about the mechanisms underlying complex human traits, including psychiatric and metabolic diseases. We used a simple mouse outbred population, the LG/J × SM/J advanced intercross line (**AIL**; Aap:LG,SM-G50-56), to identify 126 quantitative trait loci (**QTLs**) for the locomotor response to methamphetamine, prepulse inhibition of the acoustic startle reflex, body weight, muscle weight, and various other behavioral and physiological traits.

A key advantage of studying model organisms is the ability to procure primary tissue for functional genomic analysis. Integrating phenotype, genotype, and gene expression data collected from the same individuals can accelerate the process of gene identification and provide insight into the biological mechanisms that influence complex traits. Therefore, we used RNA-sequencing to measure gene expression QTLs (**eQTLs**) in the hippocampus (**HIP**), prefrontal cortex (**PFC**), and striatum (**STR**) of ~200 AIL mice that were phenotyped and genotyped. We identified over 2,000 *cis*-eQTLs and over 400 *trans*-eQTLs in each brain region and integrated these data with QTLs to identify candidate genes. Our work demonstrates that an integrative analysis of genotype, phenotype and gene expression data is a powerful approach that can accelerate the process of gene identification and provide insight into the biology of complex traits.

CHAPTER 1

INTRODUCTION

A major objective of human genetics is to understand how variation in the sequence and structure of DNA leads to phenotypic diversity. Phenotypic diversity may be caused by environmental factors, genetic variability, or more commonly, by a combination of the two. By identifying the molecular factors that generate phenotypic diversity and studying their effects, genetic research has the potential to broaden our understanding of biology, possibly leading to better diagnosis andtreatment of disease.

This chapter describes two common strategies used to study the relationship between genotypes and phenotypes, with special emphasis on genetic mapping. I highlight three major insights that emerged from early genome-wide association studies (**GWAS**) in humans and consider their implications for psychiatric disease. I argue that orthogonal research in mice can address some of the challenges presented by human genetic complexity. In particular, simple outbred models, including the advanced intercross line (**AIL**) used in this study, have several appealing properties that simplify the interpretation of association studies, and in many cases, provide support for human findings. I conclude by describing some important statistical issues that arise in GWAS ofstructured populations.

1.1 Genetic approaches for studying phenotypic variation

There are two main ways to investigate the genetic causes of phenotypic variation. Gene-based approaches typically begin with a hypothesis about the role of a specific gene or sequence. In humans, genotypes at a candidate gene for disease might be compared among cases and controls. Often, a gene or its regulatory machinery is perturbed so that the consequences of partial or complete loss of its function can be studied in living cells or animal models. The relative simplicity of such experiments makes them valuable for exploring the

role of specific genes. However, gene-based approaches are limited by our lack of knowledge about the relationship between genes and phenotypes; that is, at this point in time, *we simply do not know what most genes are doing*. What we do know is that genes can have different effects in different tissues, environments, and genetic backgrounds. We also know that most segregating polymorphisms exist outside of protein-coding genes. Although the molecular consequences of noncoding mutations remain poorly understood, evidence from functional genomic studies alludes to a critical role for noncoding sequences in transcriptional and epigenetic regulation. Clearly, an exhaustive survey of molecular function cannot be accomplished by focusing solely on coding sequences.

Much of our current knowledge about human disease was discovered using a complementary set of hypothesis-free genetic mapping approaches, including linkage mapping and GWAS. Genetic mapping studies are designed to detect correlations between traits and chromosomal regions, which can then be targeted for functional analysis. Early linkage studies successfully used families to locate large-effect mutations responsible for rare Mendelian diseases (which were often attributed to a single gene). However, they were less effective for studying common diseases, which are characterized by low penetrance and high polygenicity.

GWAS, in contrast, rely on the many recombinations present in samples of unrelated individuals to detect associations, and are more appropriate for studying common traits and diseases. Recombination disrupts linkage disequilibrium (**LD**) between nearby markers by separating them onto smaller, more independent haplotypes. This is desirable because it tends to reduce the number of single nucleotide polymorphisms (**SNPs**) tagged by each genotyped marker, potentially making it easier to determine which one is causal. However, because higher genotyping density is necessary for GWAS in highly recombinant populations, a greater number of independent hypothesis tests must be performed, which results in a stringent threshold for genome-wide significance. Therefore, high recombination improves mapping precision (the average width of the confidence interval for associated loci), but

reduces power (the likelihood of detecting an association with a given effect size at a given significance threshold).

GWAS were designed under the assumptions that the majority of variants implicated in common diseases would be common (that is, with minor allele frequencies, or **MAFs**, of 0.05 or greater) and have intermediate effect sizes (i.e. an odds ratio of 1.5-3.0) [1]. Early GWAS samples of <1,000 individuals were largely unsuccessful, aside from a few exceptions that identified SNPs with what we now know to be unusually large effects [2–4]. Discoveries from GWAS using samples of a few thousand individuals were far greater than what had been achieved by linkage mapping, but they were only able to explain a fraction of the heritability for common diseases, which family and twin studies had suggested was much higher [5]. More associations were identified as samples increased to the tens of thousands [6], and it became clear that the first GWAS were dramatically underpowered. Rapid developments in sequencing technology restored optimism to the search for disease genes and renewed an interest in family studies, which used targeted sequencing to identify a role for rare and *de novo* mutations in diseases like autism and schizophrenia [7]. Subsequently, the field of human genetics developed a greater appreciation for polygenic methods [8] which helped to dispel the fear that GWAS were inadequate to explain the missing heritability of common diseases [9, 10]. As Boyle *et al.* [11] and others [12, 13] have suggested, the prevalence of polygenic architectures may entail a shift in focus from disease genes to disease networks, which may eventually require a new framework for designing and interpreting future GWAS. Even so, GWAS have led to a number of invaluable insights that have transformed our understanding of complex traits.

1.2 Major insights from human GWAS and implications for psychiatric traits

GWAS unexpectedly revealed that most trait-associated variants are in noncoding regions, an observation that has sparked interest in using functional genomic data to explore their underlying mechanisms [14–18]. Functional genomics is the study of dynamic genomic processes (as opposed to the static effects of mutations) which include transcription, translation, and gene regulation. Functional genomics studies offer a way to test the assumption that the gene closest to a trait-associated variant is most likely to be causal, which is not always true [19–22]. Studies of gene expression have been particularly useful for identifying functional candidates within trait-associated regions [22–28]. However, it is often challenging to obtain large samples of human tissue for functional studies, particularly from the brain, which can only be collected opportunistically or *post mortem*.

Another significant discovery from GWAS is that polygenicity is ubiquitous. Nearly all common diseases appear to be caused by a very large number of common and rare alleles whose effects are individually very small [11]. This pattern holds true for psychiatric disease [29] and most quantitative traits, such as height and weight [30, 31]. Although widespread polygenicity may seem daunting from an analytic perspective, it ultimately inspired a variety of clever experimental designs [32–36] new statistical models [10,25,37–39], and an abundance of community resources [15,16,40–44] that improve our ability to describe genetic phenomena and importantly, focus our efforts on tractable questions. As an example, one consequence of polygenicity is that extremely large samples are required to detect genome-wide significant associations. Accordingly, building larger, more diverse genomic resources will allow us to identify novel polymorphisms and associations [42, 45]. However, genotyping more SNPs in larger samples can be costly, and depending on the trait, may not be worth the effort. For instance, environmental influences, which introduce noise into phenotype measurements and diminish heritability, are integral to psychiatric conditions like post-traumatic stress

disorder, which requires exposure to trauma, and drug abuse, which requires exposure to a drug. A better strategy in these cases might be to study a specific symptom or aspect of the disease that can be quantitatively measured in large samples [46].

Findings from GWAS also suggest that pleiotropy, which describes a single genetic factor with many phenotypic effects, is a common phenomenon. Pleiotropy has had a profound impact on the study of psychiatric disease, for which clinical diagnosis is based heavily on observation, phenomenology, and categorical criteria [47]. Evidence that bipolar disorder, schizophrenia, and autism share many of the same genetic associations has forced us to reconsider clinical definitions of psychiatric disease [48]. Our definitions of normal vs. disordered states may also need revision: comparing psychiatric patients to healthy relatives has demonstrated that that certain symptoms are also present at sub-clinical levels in unaffected individuals [29, 46]. *COMT*, a gene implicated in schizophrenia, also affects working memory in healthy individuals [49] and variants involved in schizophrenia and bipolar disorder have been associated with creativity in healthy subjects [50]. This indicates that quantitative traits that can be measured in healthy *and* diseased individuals could be useful for understanding certain aspects of psychiatric disease. It also suggests that current psychiatric nosology may one day be replaced withquantitative assessments informed by genetic and neurobiological discoveries [51].

GWAS have enriched our understanding of complex traits, but they have also introduced challenges. Some of these are familiar: phenotypic and environmental heterogeneity and the difficulty of procuring primary human tissue are limitations that were acknowledged long before the first GWAS. Likewise, the idea of redefining psychiatric phenotypes is hardly new, but evidence for a shared etiology among clinically distinct conditions adds to the urgency of developing more rigorous criteria for classifying psychiatric and behavioral conditions. Other challenges are less familiar, and a variety of methods may be required to address them. More associations will be discovered by genotyping new alleles and studying diverse

populations, and improvements in annotation, genotyping and imputation accuracy will become increasingly important as GWAS expand their scope. For traits like depression and bipolar disorder, which been fairly resistant to genetic dissection, the hope is that these efforts will lead to a lucrative set of candidate genes and networks; however, more work may be needed to determine if this expectation is realistic [11]. In the meantime, what else can be done? I argue that GWAS in model organisms provide a powerful, complementary set of approaches that can generate valuable biological insights and in many cases, inform interpretation of human GWAS.

1.3 Model organisms can be used to circumvent some limitations of human GWAS

Here I describe key benefits and limitations of using rodents to study aspects of human disease. Many of these qualities also generalize to other model organisms. However, as fellow mammals, rodents are frequently used as models for human disease, and their ability to learn and perform complex tasks makes them the preferred models for studying behavior. Although it is possible to study a more complex range behaviors in rats than in mice [52], there are practical advantages that have made large genetic studies in mice more feasible until relatively recently [53, 54]. For example, the state of mouse genome and transcriptome annotation is highly advanced [55], and a wealth of mouse genetic resources exist due to a long history of genetic studies in the mouse [55–57]. For these reasons, and because they are central to my research, I will focus specifically on mice for the remainder of the discussion.

Certain limitations of human GWAS can be circumvented by using the mouse. Unlike some human data, phenotypes are relatively fast and easy to collect in mice due to their small size, short generation time, and impressive breeding capacity. Environmental variables can be controlled to minimize confounding, or deliberately manipulated to study traits like drug abuse, which depend on environmental exposures. Mice can often be used to obtain

measurements that would be considered invasive or unfeasible in humans. Obvious physical differences prevent the mouse from being a precise model for human biology; however, mouse measurements are typically quantitative and may be more informative for specific aspects of human disease than the clinical categories often used in human GWAS [58–60]. Importantly, phenotypes, genotypes, and primary tissue for functional genomic assays can be collected from the same mice under optimal conditions. This is important for psychiatric traits, not only because human brain tissue is difficult to obtain, but because medication use and other external variables can confound analyses of gene expression in humans [46].

1.4 Genetic analysis in advanced intercross lines

Crosses among well-characterized inbred strains have been used to identify hundreds of loci associated with quantitative traits (**QTLs**), but due to low recombination density, populations such as F2 intercrosses, backcrosses, and congenic lines have rarely led to the identification of individual genes [54,61]. In recent years, low genotyping costs have motivated GWAS in more recombinant populations, which provide higher resolution for mapping QTLs [62–64]. Given the ubiquity of polygenic architecture and the complexity of human populations, the ability to manipulate genetic diversity, allele frequency, and recombination density in mice is an enormous advantage. Advanced intercross lines (**AILs**) provide one example of how the balance between power and precision can be tailored to fit the goals of GWAS.

An AIL is generated by continuing to cross two inbred strains beyond the F2 generation. Each successive generation of outbreeding leads to additional recombination, which improves mapping precision [65]. Because only two inbred strains are used, the MAF for all variants starts at 0.5, which maximizes power to detect the effects of each allele. Although it is possible for rare alleles to arise from inadvertent selection (e.g. for fecundity) or *de novo* mutations, these situations are uncommon. Genetic drift can cause variants to shift in frequency, but if the population size is sufficiently large, allele frequencies are expected to

remain high [65]. Thus, the problematic relationship between allele frequency and effect size observed in unrelated humans is minimized in an AIL. Furthermore, having only two parental strains means that alleles and haplotypes are interchangeable, which reduces number of markers that must be genotyped (potentially allowing for a larger sample without a proportional increase in cost). Any two alleles at a polymorphic site that are identical by state (**IBS**) are necessarily identical by descent (**IBD**), which makes them fully informative for haplotype phasing and genotype imputation. If pedigree information or genome sequences from the parental strains are available, phasing and imputation can be simplified even further.

Of course, producing a highly recombinant AIL requires considerable time and resources. An AIL breeding population should be as large as possible to minimize genetic drift, inbreeding, and relatedness among individuals; this allows recombinations to accumulate in each generation, improving mapping precision [65]. However, the expense of maintaining mice typically limits populations to 50-100 breeder pairs. A random breeding scheme has the benefit of simplicity, but a better strategy is to avoid pairing full siblings. The *optimal* strategy is to use pedigree or genotype information to minimize the level of relatedness among breeder pairs. Another recommended strategy is to use only one male and one female from each litter as breeders. Limiting familial contributions to the next generation preserves genetic diversity by doubling the effective population size relative to its actual size, thus helping to prevent the fixation of parental alleles [66].

A second limitation of AILs is that they have low genetic diversity. Thus, at some loci, there will be no polymorphism between the parental strains, and therefore, some genes that have the potential to affect a trait may go undetected. Even so, the relative simplicity of analysis in an AIL and the ability to detect associations using a moderately-sized sample are compelling advantages for GWAS.

1.5 GWAS in structured populations

Early GWAS in humans clearly demonstrated that a failure to account for the confounding effects of population structure could lead to false positive associations. Population structure refers to the presence of systematic differences in allele frequencies among subsets of a population. Structured samples require careful analysis because correlations exist among phenotypes (more closely related individuals are more phenotypically similar) *and* among genotypes (the probability of having the same genotype at a given marker is higher among more closely related individuals). This non-independence violates an assumption of the linear regression model commonly used to test for association in GWAS, and failure to properly control for population structure can diminish power [67–69].

The earliest methods to account for population structure involved structured association analysis or principle components analysis (**PCA**). Structured association analysis assumes that a population is comprised of a number of subpopulations or clusters, each tracing back to a distinct ancestral group [70]. Genetic markers are used to determine cluster membership, which provides a summary of population structure. PCA may also be used to cluster individuals by genetic background [71].

Although these methods are useful for identifying population structure due to ancestry, they are not designed to account for recent familial relationships [72]. Therefore, genomic control is often applied with these methods to prevent the inflation of test statistics caused by cryptic relatedness [73]. A more recent alternative approach is to use a linear mixed model (**LMM**), which is very effective for controlling the false positive rate in GWAS of structured populations, including multigenerational pedigrees [68, 74, 75].

Modeling the covariance between phenotypes is important for GWAS, which usually assume an additive model of inheritance. In outbred populations, pairs of individuals covary randomly with respect to genetic background. If phenotypic variation is caused by the sum of multiple genetic factors with individually small effects, then the more alleles that individuals

share, the more similar they are likely to be. The key point is that covariance is higher for AILs since individuals are related to one another within and between families. Simple linear models do not account for this covariance and can lead to spurious associations when applied inappropriately. However, LMMs simultaneously account for the fixed effect of genotype at a locus and for latent effects that influence the phenotype at random, making it a very powerful approach for GWAS.

1.6 Overview

This chapter is intended to provide a framework for understanding the broad questions that motivated my research and the general approach applied to address them: how do genetic mutations produce phenotypic diversity? What are the genetic mechanisms that influence human behavior and disease? How can research in model organisms enhance our understanding of human biology?

In summary, I adopted methods from human and mouse genetics to study complex traits in an AIL derived from a cross of the LG/J (**LG**) and SM/J (**SM**) inbred strains. I used 1,063 male and female mice from generations 50-56 (**G50-56**) of the LG × SM AIL to map QTLs for 123 traits related to the development of human psychiatric and metabolic disorders such as drug abuse, schizophrenia, and obesity. For a subset of the mice, I used RNA sequencing to measure mRNA abundance (which I refer to as gene expression) in three brain regions and used these data to map expression QTLs (**eQTLs**). In Chapter 2, I will introduce the concept of an intermediate phenotype to explain how mice can be used to model aspects of human disorders. I will describe the genetic architecture of behavior and physiology in the LG × SM AIL and summarize QTL mapping results. In Chapter 3, I will show that integrating QTLs and eQTLs identified in the mouse is a useful way to learn about the genetic basis of complex traits. Here, I will present *cis*- and *trans*-eQTLs identified in hippocampus (**HIP**), prefrontal cortex (**PFC**), and striatum (**STR**) of approximately 200

mice. I will also discuss specific QTLs in greater detail, using information about eQTLs to identify promising candidate genes for further investigation. In Chapter 4, I will discuss some broad implications of this work for future studies of complex traits in humans and model organisms.

CHAPTER 2

MOUSE GWAS OF INTERMEDIATE PHENOTYPES FOR HUMAN PSYCHIATRIC AND METABOLIC DISEASE

2.1 Introduction

As outlined in Chapter 1, my intention in using an outbred mouse population was to circumvent some of the challenges that are frequently encountered in human GWAS. Previous work from the Palmer Lab established the LG × SM AIL as an effective tool for fine-mapping QTLs identified in F2 crosses. Equipped with methods adapted from human and plant genetics, this study is the first to demonstrate that AILs are a powerful mapping population in their own right.

This chapter describes how I used a mouse AIL to identify genetic factors related to psychiatric and metabolic disease. It is difficult, if not impossible, to recapitulate human disease in a model system. Model organisms are nonetheless useful tools for developing and testing specific hypotheses about mechanisms that affect disease [76–80]. In an attempt to identify genes and pathways with potential relevance to a broad spectrum of disorders, I focused my efforts on intermediate phenotypes.

2.1.1 What is an intermediate phenotype?

The concept of an intermediate phenotype is central to my work, but it has a complex etymology that warrants further discussion. The term *intermediate phenotype* has often been used interchangeably with terms such as *endophenotype*, *biomarker*, or *sub-clinical trait*, each of which can be interpreted differently among geneticists and clinicians [81]. Most definitions agree that an intermediate phenotype lies somewhere along the pathway connecting genetic variation to a phenotypic outcome. However, they make various assumptions about causality, heritability and genetic architecture that are not necessarily accurate, and this can

complicate the interpretation of GWAS [82]. I use the term *intermediate phenotype* to denote a mechanism-related manifestation of a broader, more complex trait, usually a disease [83]. This assumes that diseases can be meaningfully deconstructed into distinct phenotypic units, and implies that these units can provide insight about disease-related mechanisms.

Blood lipid levels, which have been studied for their relationship to coronary artery disease (**CAD**), provide a straightforward example of the intermediate phenotype approach and demonstrate how it can be used to develop hypotheses about the mechanisms that influence disease. This example allows me to clarify some important assumptions of the approach before discussing advantages and challenges of using intermediate phenotypes to study psychiatric traits. I then introduce the intermediate phenotypes measured in this study, which include blood glucose levels, behavioral responses to methamphetamine, locomotor activity, and others.

GWAS of lipid levels have been used to elucidate mechanisms that influence the development of CAD. In 2010, a meta-analysis of 46 GWAS for lipid traits replicated an association between HDL cholesterol levels and *GALNT2* [84]. The authors used a mouse model to confirm that *GALNT2* regulates HDL cholesterol levels in blood, leading to specific hypotheses about its role in the development of CAD and in mammalian metabolism more generally [84, 85]. The same meta-analysis replicated an association between LDL cholesterol and *SORT1* [84, 86], which had previously been identified in GWAS of CAD [87]. In a follow-up study, the authors demonstrated that the associated SNP disrupted a transcription factor binding site within a liver-specific enhancer, resulting in decreased expression of *SORT1* [88,89]. Subsequent knockdown of *SORT1* in mouse liver implicated it as a regulator of hepatic cholesterol secretion, providing a mechanistic explanation for its association with CAD [88, 90].

The relationship between intermediate phenotype and disease is very clear in these examples: lipid accumulation in blood vessels promotes inflammation, which can lead to

atherosclerosis of the coronary arteries, a primary symptom of CAD. However, my definition does not assume that intermediate phenotypes are causal. This is a critical point for psychiatric genetics because the biological basis of mental illness is not well understood. Due to the difficulty of distinguishing cause from consequence in psychiatric disease, intermediate phenotypes are typically based on statistical correlation or clinical symptoms [91]. Thus, an important outcome of studying intermediate phenotypes will be to determine which ones are actually on the causal pathway to disease [82].

The use of intermediate phenotypes also has practical advantages. Since they are not diseases in themselves, they can often be measured systematically and quantitatively in healthy subjects or model organisms. I argue that for psychiatric genetics, traits that can be translated to and from model organisms provide the greatest advantage, as they offer a system in which hypotheses about the brain and behavior can be tested by manipulating genetic or environmental variables [58–60, 77, 80, 92]. In the following sections, I will describe how the intermediate phenotypes measured in this study relate to human drug abuse, schizophrenia, and metabolic syndrome.

2.1.2 Intermediate phenotypes for human substance use disorders

The etiology of substance use disorders is multifactorial

Substance use disorders (addictions) are defined as a set of chronic, relapsing medical and psychiatric diseases characterized by persistent and compulsive drug use despite negative consequences [47]. The enormous medical and financial burden imposed by illicit drug use [93] underscores a need to understand what variables influence risk for drug abuse so that addiction can be prevented and treated.

However, isolating genetic factors that influence risk for drug abuse in humans is challenging for a variety of reasons. Most notably, substance use disorders are heterogeneous and progress in stages that include initial use, drug abuse, relapse, and withdrawal [94, 95].

Each stage is influenced by an unknown combination of genetic and environmental risk factors that interact in complex ways [94–96]. Risk for drug exposure might depend on personality traits such as impulsivity or sensation seeking [97–99], preexisting psychiatric disorders [100, 101], or social factors [94, 95, 98, 102], to name a few. Risk for progressing to drug abuse after initial use may depend on these factors as well as other variables, such as the subjective effects of the drug (the experience of reward or aversion) [103–105] or external stressors [100, 101]. However, drug use does not always lead to abuse, and drug abuse does not necessarily result in addiction, suggesting that the risk factors influencing the early stages of drug use may differ from those important in later stages [96, 98, 103, 106]. Finally, the neurobiological changes (e.g. altered neurotransmission or synaptic plasticity) that occur after repeated drug use make it difficult to disentangle cause from consequence in the study of addiction [94, 96, 107, 108].

The subjective experience of reward in early stages of drug use as an intermediate phenotype for substance abuse disorders

Rodents and other model organisms are useful for studying the neurobiological, pharmacological, and behavioral effects of addictive drugs [77, 109]. In particular, a variety of rodent models are available for exploring intermediate phenotypes related to addiction, including drug reward, drug sensitivity, impulsivity, and others [110–112]. I chose to study the acute subjective response to drug administration, which is an appealing intermediate phenotype for several reasons. The subjectively positive effects of drugs are thought to contribute to the early stages of drug abuse in humans [103,113]. Humans and rodents display inter-individual variability in response to and preference for drugs of abuse [103, 114] and this variability is known to be heritable [115–118]; moreover, the positive subjective response to drugs, or drug liking, is correlated with the development of drug abuse [103, 104, 109]. Accordingly, I and others have proposed that individuals who are genetically predisposed to respond positively

to drugs are more likely to continue using them, putting them at greater risk to develop addiction [106, 119]. Thus, a broad goal of this approach was to identify genes that predispose an individual to like or dislike drugs of abuse so that future experiments might determine how such genes increase or decrease risk for drug abuse.

Another reason I chose to study drug liking is that it can be measured in both rodents and humans using the conditioned place preference (**CPP**) paradigm. In rodents, CPP is measured as the amount of time spent in an environment paired with a reward, usually a drug [112,120,121]. Drug and vehicle are alternately paired with separate environments over a conditioning period of several days. On the last day, the animal is allowed to explore both settings, and preference for the drug is measured as the amount of time spent in the drug-paired environment. CPP has been widely used to study drug-seeking behavior and reward in rodents [112] and has also been applied to zebrafish [122–124] and invertebrates [125,126]. Results from studies of CPP for methamphetamine [114,127–129], d-amphetamine [130,131], and alcohol [128] in healthy human subjects demonstrates that self-reported preference for a drug-paired room is correlated with the self-reported pleasant effects of psychostimulants [127,128,130,131], suggesting that CPP is a reasonable way to measure the hedonic properties of rewarding drugs. However, I note that the relationship between CPP and drug liking is ambiguous in model organisms because I rely on behavioral observation rather than self-report to infer their internal states. Therefore, other aspects of reward, including associative learning, craving, or drug aversion [112, 132] should also be considered when interpreting CPP in animals.

Multiple studies of transgenic and inbred strains of mice have demonstrated a genetic basis for CPP [133–145]. However, little is known about its genetic architecture. Preliminary data from the Palmer Lab had demonstrated that LG × SM AIL mice exhibit high variability in CPP for methamphetamine [146]. I had also identified multiple QTLs for the locomotor response to methamphetamine in G34 of the LG × SM AIL [75, 147], suggesting that the

population would be a useful tool for identifying genes related to multiple features of drug abuse.

Locomotor activity as an intermediate phenotype for sensitivity to psychostimulant drugs

Locomotor activity forms the foundation of several widely used behavioral paradigms for studying aspects of human psychiatric traits in rodents. Locomotion is relevant to psychostimulant addiction because psychostimulant drugs share the ability to potently increase synaptic levels of dopamine in the brain. In addition to modulating the experience of pleasure and reward, dopamine influences locomotor behavior and coordination, explaining why drugs like methamphetamine are associated with locomotor activation. The acute effects of psychostimulants are nearly identical in mice and humans due to the highly conserved brain structures and circuits that regulate dopamine release and reuptake. Nonetheless, individuals of both species exhibit high variability in response to psychostimulant drugs, and differences in drug sensitivity are thought to influence continued drug use. In mice, heightened locomotor sensitivity to acute methamphetamine is protective against future self-administration of the drug [109, 119]. When administered intermittently, some animals exhibit heightened locomotor sensitivity to psychostimulants (despite receiving an equal or lesser dose), an effect that may be observed after a single administration of the drug [109, 148]. This is known as locomotor sensitization, a phenomenon that has been proposed as a way to study reward sensitivity [107, 148].

The standard equipment used to measure CPP can also measure locomotor behavior, enabling a comparison of baseline activity (following administration of vehicle, usually saline) to drug-induced activity in the same animal over the course of the experiment. A common unit of locomotor activity in mice is the total distance traveled within the testing environment over a period of time. In previous experiments, I used the open field testing environment to

measure the locomotor response to saline and methamphetamine in the LG x SM AIL and identified QTLs for both traits [75,147]. The open field testing chamber is very similar to that of CPP, except the latter is divided into two distinct sides separated by a divider that restricts the mouse to one side during conditioning [146]. On days when preference is measured, the mouse is permitted to explore both sides of the chamber through an open passage in the divider, and the number of side changes are measured as counts. Side changes made by drug-naïve mice in a novel environment may be used to model exploratory behavior, and measuring side changes after place conditioning with a rewarding drug has been proposed as a way to model drug-seeking [149]. However, interpretation of side changes as an intermediate phenotype for drug abuse requires caution, as its behavioral and neurological properties have not been studied as extensively as other forms of rodent activity. Furthermore, it is unclear whether or not they capture aspects of behavior that are independent from locomotor activity expressed as distance.

Sensorimotor gating as a readout of dysfunctional mesolimbic circuitry in schizophrenia and other psychiatric diseases

Sensorimotor gating is an information filtering mechanism that permits an organism to process and respond to environmental stimuli without being overwhelmed by irrelevant or redundant cues; this response is impaired in individuals with schizophrenia [150–153], obsessive-compulsive disorder [154, 155], Huntingtons disease [156] and other psychiatric illnesses [46, 157, 158]. The neurobiological underpinnings of sensorimotor gating have been studied extensively in humans and animal models [46,157,159], and genes associated with psychiatric illness have been shown to affect sensorimotor gating in transgenic mice [160–163]. However, relatively few studies have mapped QTLs for sensorimotor gating in humans [152, 164].

Sensorimotor gating can be measured by prepulse inhibition of the acoustic startle reflex (**PPI**) in vertebrate and invertebrate systems [159, 165]. The startle reflex is elicited by an

intense stimulus, such as a loud tone, and refers to a fast-twitch of the skeletal muscle; this directs attention to the environment so that the startled animal can respond to a potential threat [165]. Although the startle reflex is automatic, it can be inhibited to some degree by a preceding stimulus of lower intensity; this is called a prepulse [166]. In humans, PPI is measured from the startle-blink reflex, whereas in rodents, muscle contractions are measured from the whole body [46, 152, 167–169]. However, in both experiments, the startle reflex is elicited by a loud tone, and a transducer is used to output a voltage proportional to the magnitude of the animals response.

PPI may be related to aspects of inhibitory control, attention, working memory, and other traits that are correlated with sensorimotor gating in humans [170] and animal models [159]. Its specific relationship to psychiatric dysfunction nonetheless remains ambiguous; for example, there is no normal level of PPI, it is sensitive to environmental variables, and unlike the startle response itself, it does not appear to be an adaptive trait [46]. However, PPI remains stable within individuals over time [152, 171], it can be predictably manipulated by administration of anti-psychotic drugs, which are used to treat schizophrenia [152, 172], and interestingly, it does not appear to be correlated with other automatic inhibitory mechanisms [46, 173]. Furthermore, it is not clear whether PPI deficits in schizophrenia are caused by the same neurological mechanisms as those responsible for PPI deficits in other diseases [46]. For these reasons and others, Swerdlow *et al.* [46] have suggested that PPI might be more accurately described as a readout of mesolimbic circuitry than as a psychiatric symptom.

The broad relevance of PPI to a range of psychiatric disorders make it an appealing intermediate phenotype to explore in humans and animal models. Similarly, the startle response varies among individuals and may be important for psychiatric disorders whose symptoms include fear and anxiety [165, 174]. The value of identifying genetic factors that affect PPI and startle in mice is that it may elucidate how disruptions in neural circuits and genetic pathways lead to dysfunctional behavior in psychiatric disease.

2.1.3 Modeling metabolic syndrome in the LG × SM AIL

Metabolic syndrome encompasses several related conditions that put individuals at increased risk for developing type 2 diabetes and CAD (e.g. obesity, dyslipidemia, high blood pressure, and insulin resistance) [175]. Mice have been useful for understanding the genetic contribution to metabolic syndrome in humans in spite of obvious physiological differences between the two species [175, 176].

The LG and SM strains were derived by independently selecting for high and low body weight [177], making them popular models for studying traits related to body size and growth. LG and SM are also useful for studying diabetes, lipid levels, and other metabolic phenotypes: SM mice are more susceptible to atherosclerosis, have higher cholesterol, triglyceride and basal glucose levels, gain more adipose weight in adulthood, and have stronger responses to a high fat diet [178]. Years before the Palmer Lab established its own LG × SM AIL using his G33 AIL mice as progenitors [75], Dr. James Cheverud had explored these phenotypes in the inbred strains and developed the AIL for the purpose of studying metabolism [175, 178, 179].

In addition to body weight [147], the Palmer Lab has mapped QTLs for hind limb muscle weight [180] and fiber composition [181] in the LG × SM AIL. Skeletal muscle is the most abundant tissue in the body, and low muscle strength is correlated with lower life expectancy and higher rates of obesity, non-alcoholic fatty liver disease, and CAD [182–184]. The muscle weights included in this study differ in size and fiber constitution; they include two dorsiflexors (tibialis anterior, TA; and extensor digitorum longus; EDL) and three plantar flexors (gastrocnemius, plantaris and soleus). Soleus is the only slow-twitch muscle and is comprised mainly of type 1 fibers; the rest are fast-twitch muscles comprised primarily of type 2 muscle fibers [185]. Different morphological and functional properties are associated with each fiber type [186], and I reasoned that muscles composed of different types might be regulated by distinct genetic mechanisms. A secondary goal was to replicate and refine muscle weight QTLs identified in G34 of the LG × SM AIL [180].

Because body weight, tail length, and fasting blood glucose levels are fast, simple and inexpensive to measure, the Palmer Lab has routinely studied these traits even though my primary interest is in behavior [62, 147]. I made it a priority to establish collaborations with scientists interested in the genetic basis of metabolism, growth, and physiology so that the LG × SM AIL can be used to its full potential.

Mapping QTLs for these phenotypes they allowed me to survey multiple domains of behavior and metabolism that are broadly relevant to human biology and disease. The Palmer lab had the infrastructure and prior expertise required to measure CPP and PPI in large samples of mice. Importantly, the LG × SM AIL had previously been used to identify QTLs for locomotor activity in response to methamphetamine and saline [75], PPI, startle [167], muscle weight [180], body weight [147], and glucose traits [175], which provided me with an opportunity to replicate previous findings.

2.2 Methods

2.2.1 Animals

In 2006, the Palmer Lab established a LG × SM AIL colony using 140 G33 mice obtained from Dr. James Cheverud (Jmc: LG,SM-G33), who initiated the AIL at Washington University in St. Louis in the early 2000s [187]. Since 2009, pairwise kinship coefficients estimated from the AIL pedigree have been used to select the most unrelated breeder pairs for each new generation (additional details about animal breeding and care are provided in the Supplementary Methods). Approximately 100 breeder pairs were maintained in G49-55 to generate the mice for this study. One female and one male from each nuclear family were selected for phenotyping and housed with up to three other same-sex mice that were weaned on the same day. Up to three additional mice from each family were reserved as potential breeders.

I subjected 1,123 AIL mice (562 female, 561 male) (Aap:LG,SM-G50-56) to a four-week

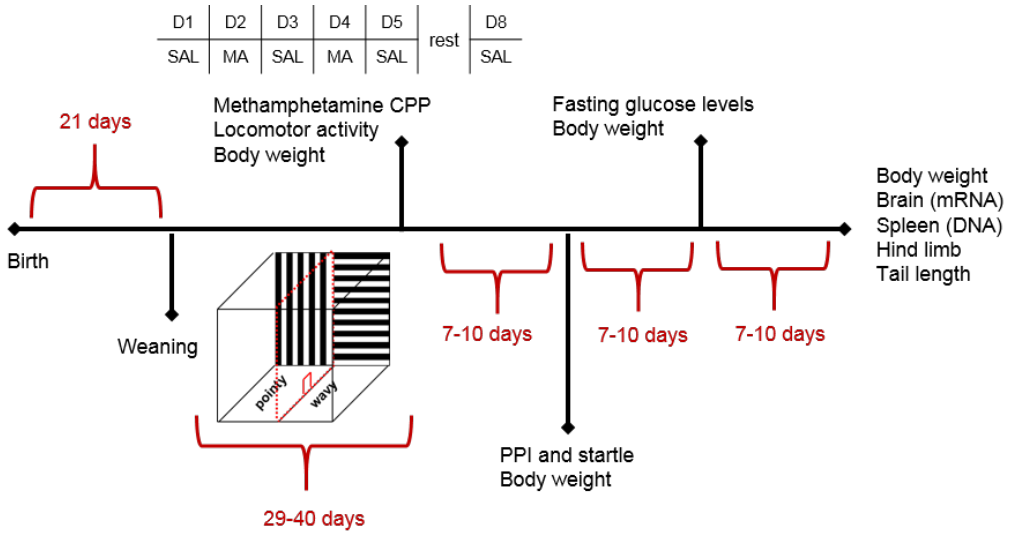

Figure 2.1: **AIL phenotypes**. I measured a variety of behavioral and physiological traits in the LG × SM AIL, including CPP, locomotor activity, body weight, PPI, startle, blood glucose levels after a four-hour fast, and tail length. A schematic of CPP D1-D8 is shown; SAL indicates that the mouse received a saline injection and MA indicates that the mouse received an injection of 1 mgkg methamphetamine. AIL mice were weaned at approximately 21 days after birth and housed in groups of 2-4 same-sex mice per cage. Mice were phenotyped in 22 batches of 24-71 individuals. Median age was 54 days (mean=55.09, range=35-101) on the first day of CPP and 83 days (mean=84.4, range=64-129) at death. Mice were maintained on a 12:12h light-dark cycle and all measurements took place in the light phase. Standard lab chow and water were available *ad libitum*, except during testing. Muscle and tibia traits were measured *post mortem*.

battery of tests. Mice were tested over a period of two years (Figure2.1). All procedures were approved by the Institutional Animal Care and Use Committee at the University of Chicago in accordance with NIH guidelines for the care and use of laboratory animals. I briefly describe each phenotype in the following sections to provide context for interpreting the results. A detailed description of how I measured each phenotype and prepared the data for QTL mapping is provided in the Supplementary Methods.

2.2.2 CPP and locomotor activity

CPP is measured as the amount of time spent in a drug-associated environment relative to a neutral environment over the course of 30 minutes. The full procedure takes eight days.

Initial preference is measured on day 1 (**D1**) after a saline injection. On D2-5, mice are trained to associate one environment with 1 mg/kg methamphetamine and the other with vehicle (0.9% saline). After two days of rest, CPP is measured on D8 after an injection of saline. CPP, locomotor activity (total distance traveled in cm) and side changes (number of times a mouse moves between chambers on D1 and D8) are measured automatically by a computer. I define locomotor sensitization to methamphetamine as the difference in activity between D4 and D2. I also used the difference between final and initial preference for the drug-paired environment as a second measure of CPP. All phenotypes were measured in five-minute bins for a total of 30 minutes.

2.2.3 PPI and startle

PPI is measured across multiple trials, which are consolidated into two blocks for each prepulse intensity (3, 6, and 12 dB above 70 dB background noise). I refer to these as **PPI3**, **PPI6**, and **PPI12**. There are also four blocks of startle-alone trials; the primary startle phenotype is the mean of the four blocks. Habituation to startle is defined as the difference between startle block 1 and block 4.

2.2.4 Fasting blood glucose levels

I measured blood glucose levels after a four-hour fast 4-14 days (mean=7.3 days, median=7) after PPI testing. Median age of mice at the time of testing was 75 days (mean=76.4, range=56-122). Mice were brought into the testing room between 09:00 and 09:30 and transferred to new cages that did not contain food. After four hours, I weighed each mouse and used a razor blade to make a small incision at the tip of the tail, which allowed me to obtain a small drop of blood that I analyzed with a glucometer. Glucose levels are expressed in mg/dL units. Once all of the mice in a cage were tested, I replenished their food and returned them to the colony.

2.2.5 Body weight

I measured body weight (g) before each behavioral test and at time of death. Six measurements were taken within the same week during CPP. After that, weight measurements were taken approximately one week apart until death. I also generated an average phenotype using body weight measurements taken approximately one week apart (i.e. measurements from D2-D5 of CPP were excluded from the average).

2.2.6 Muscle weight and bone length

I measured tail length (cm) at the time of death. I transferred frozen hind limb tissue to Dr. Arimantas Lionikas, who brought them to his laboratory at the University of Aberdeen for dissection. Five hind limb muscles, gastrocnemius, extensor digitorum longus (**EDL**), soleus, plantaris, and tibialis anterior (**TA**), were separated under a dissection microscope and weighed to 0.1 mg precision.

Tibia length was measured to 0.01 mm precision with a digital caliper. Mineral content of the tibia was measured with the Faxitron MX-20 scanner (Faxitron Bioptics LLC, AZ, USA) and quantified with ImageJ (V1.48p, National Institutes of Health, USA). Apparent bone mineral content was characterized by the median of the optical density of the bone image. A variable for tibia thickness was generated by combining thickness measurements with the central axis of the bone image. Both bone mineral content and thickness were measured in units of pixel intensity.

2.2.7 Genotyping by sequencing

I used GBS to obtain genotypes for 1,110 AIL mice. GBS is a reduced-representation genotyping method designed for the Illumina sequencing platform [188, 189]. DNA libraries were prepared by digesting DNA with a restriction enzyme and annealing oligonucleotide adapters to the resulting overhangs.

I extracted DNA from spleen using a standard salting-out protocol and prepared GBS libraries with PstI, as previously described [62,190]. I sequenced 24 uniquely indexed samples per lane of an Illumina HiSeq 2500 using single-end, 100 bp reads.

Twenty-four mice were selected for sequencing on the Giga Mouse Universal Genotyping Array (**GigaMUGA**), which contains over 140,000 SNPs that segregate in laboratory mice [191]; however, only a subset of these markers were expected to be polymorphic between LG and SM. Obtaining GigaMUGA genotypes allowed me to validate GBS genotype calls and compare the performance of the two methods (details are provided in the Supplementary Methods 2.5).

2.2.8 DNA alignment

1,110 GBS libraries were aligned to the mm10 reference genome using BWA [192]. The LG and SM genomes are fully sequenced [193]; this enabled me to realign reads around known indels and recalibrate base quality scores with GATK [194]. On average, 3.2 million reads were obtained per sample. I discarded 32 samples with less than one million reads aligned to the main chromosome contigs (1-19, X, Y) or with a primary alignment rate less than three standard deviations under the mean (<77%).

2.2.9 Variant calling and imputation

I used ANGSD [195] to obtain genotype likelihoods for 1,078 mice at 49,373 GBS SNPs. After identifying and correcting sample mix-ups, I discarded 15 samples whose identities could not be resolved (Supplementary Methods 2.5). I used Beagle [196, 197] to impute 4.4 million known SNPs in LG and SM [193] into the AIL sample. I discarded uncommon variants (MAF<0.1), Hardy-Iinberg Equilibrium violations ($p \leq 7.62 \times 10^{-6}$), and SNPs with low imputation quality (dosage $r^2 < 0.9$) before pruning SNPs in high LD ($r^2 < 0.95$) (Supplementary Methods 2.5). 523,028 SNPs were retained for mapping.

2.2.10 QTL mapping model

I used GEMMA [198] to map QTLs with an LMM. Covariates for each trait are listed in Supplementary Table 2.2. To avoid the loss of power that results when the SNP being tested is also included in the random, polygenic effect, I used a leave-one-chromosome-out LMM (**LOCO-LMM**) to test SNPs on each autosome. Unlike the standard LMM, the LOCO-LMM calculates a different polygenic effect for each chromosome. That is, the polygenic effect that does *not* include markers on a given chromosome is defined as the random effect in the model used to test SNPs on that chromosome.

2.2.11 Heritability estimates

I estimated the proportion of phenotypic variance explained by 523,028 GWAS SNPs (SNP heritability) using a restricted maximum likelihood algorithm in GEMMA [198, 199]. I ran another GWAS for each trait, this time dropping the fixed effect of dosage and including the full polygenic effect estimated from genotypes on the 19 autosomes. I then repeated the procedure using the top SNP at each QTL as a covariate and used the difference between the two values as an estimate of the effect size for the top SNP at each QTL.

2.2.12 Genome-wide significance

MultiTrans [200] and SLIDE [201] were used to determine a threshold for genome-wide significance. MultiTrans estimates the phenotypic covariance under an LMM and uses it to transform the genotype data such that the correlation among transformed genotypes is equivalent to the correlation among test statistics sampled from an multivariate normal distribution [200]. This obviates the need to generate null phenotypes and calculate their p-values. Instead, p-values are sampled directly from the multivariate normal distribution using SLIDE, which adjusts the threshold for LD between nearby markers. I specified a large sliding window of 5,000 SNPs to ensure that local LD was adequately accounted for

in regions where SNP density is low in LG and SM. I drew 2.5 million samples to obtain a threshold of $p = 8.06 \times 10^{-6}$ at a genome-wide significance level of 0.05. Because the data were quantile normalized, I was able to use the same threshold for all traits.

2.2.13 QTL intervals

I converted p-values to LOD scores and used a 1.5-LOD support interval to approximate a critical region around each QTL. Though simplistic, the LOD drop approach provides a quick, straightforward way to gauge mapping precision and systematically identify overlap between QTLs and other features. However, proximity to a QTL is neither necessary nor sufficient to determine whether a gene or SNP is functionally relevant; this is one reason for integrating QTLs with gene expression data in Chapter 3.

2.3 Results

In the following sections, I will summarize the genetic architecture and heritability of complex traits measured in the LG × SM AIL and show that a LOCO-LMM can improve power to detect QTLs in populations with complex relationships. I will also describe general results for each set of related traits while highlighting some advantages and disadvantages of genetic analysis in the LG × SM AIL. I will discuss candidate genes identified for a subset of traits alongside eQTL results in Chapter 3, which provide additional context for interpreting QTLs.

2.3.1 Genetic architecture in the LG × SM AIL

SNP density and distribution

I used GBS to genotype 1,063 of the 1,123 mice that were phenotyped (60 were not successfully genotyped for technical reasons described in the Supplementary Methods 2.5). GBS yielded 49,373 empirically derived autosomal SNPs (Figure 2.2a,b). In contrast, in the 24

AIL mice that were also genotyped on the GigaMUGA [191], only 24,934 markers were polymorphic in LG and SM (Supplementary Figure: 2.8), demonstrating that for this population, GBS is a more powerful and cost-efficient approach for genotyping.

MAF distribution

One unique property of an AIL is that all SNPs are expected to be common because the population is created by crossing two inbred strains. Consistent with this, Figure 2.2c shows that the MAF distribution consists primarily of common variants. Notably, we observed little to no relationship between MAF and QTL effect size in the LG $\times$ SM AIL (Figure 2.3). Of the 4.4 million SNPs imputed from LG and SM, only 2,798 had MAF <0.1; these SNPs were flagged as genotyping errors and excluded from further analyses. Following imputation and LD pruning, I obtained a final set of 523,028 autosomal SNPs.

LD decay

I calculated the decay of LD because it is critical to mapping resolution. As shown in Figure 2.2d, there is an increase in the rate of LD decay relative to a previous mapping study in G34 [62, 75]. The mean width of QTLs mapped in G50-56 was 2.59 Mb (median=2.28 Mb, range=0.07-8.27). On average, each QTL contained 40.7 genes (median=24.5, range=0-254).

Heritability

To estimate heritability, I calculated the proportion of trait variance explained (**pve**) by the additive effects of 523,028 SNPs for each trait (Supplementary Table 2.2). In general, heritability estimates were larger for metabolic and physiological traits2.5 than for behavioral traits**??**, which is consistent with findings in other rodent GWAS [53, 62, 63]. Mean heritability for physiological traits was 0.327 (se=0.043). Mean heritability for behaviors estimated to have a genetic component was 0.168 (se=0.038). Notably, there was no clear evidence of her-

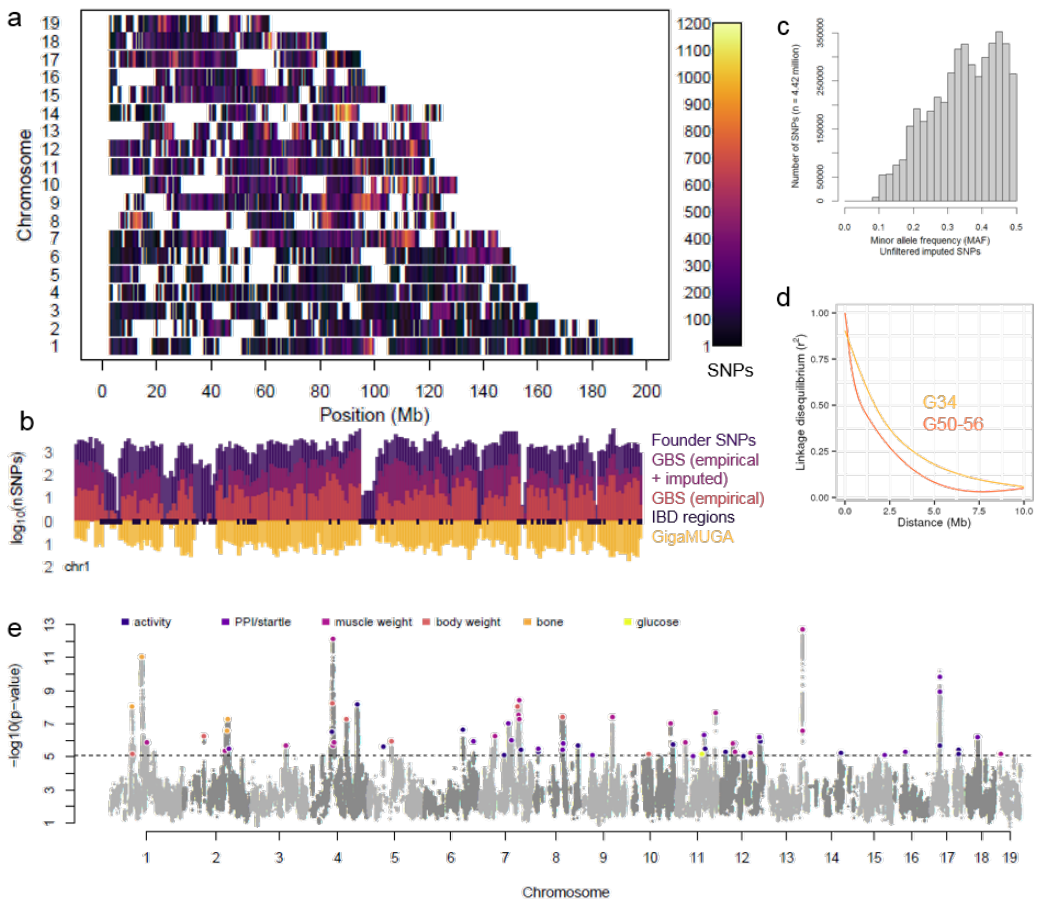

Figure 2.2: **Genetic architecture in the LG × SM AIL. (a) Genome-wide distribution of GWAS variants.** Genome-wide distribution and density of 523,028 GWAS SNPs are plotted for each chromosome in 500 Kb windows. **(b) SNP density on chromosome 1.** A detailed view of SNPs on chromosome 1 shows the log_{10}-scaled density of known SNPs in LG and SM from the reference panel (dark purple, top background), imputed and quality-filtered GWAS SNPs (medium violet, top middle), GBS SNPs before reference panel imputation (orange, top foreground), and GigaMUGA SNPs that were polymorphic in LG and SM (gold, bottom). At the x-axis, predicted IBD regions in LG and SM [193] are marked in navy. **(c) MAF distribution of high-quality imputed variants.** The majority of AIL alleles are common, as shown by the MAF distribution of 4.42 million variants imputed from LG and SM. **(d) LD decay in G50-56 of the LG × SM AIL.** Comparison of LD decay in G34 (gold) and G50-56 (orange) shows that 95% of G50-56 SNP pairs are spaced ~2.5 Mb apart have an $r^2 \leq 0.25$, compared to an $r^2 \approx 0.37$ in G34. **(e) Manhattan plot of QTLs.** The dashed line indicates a permutation-derived significance threshold of $p = 8.06 \times 10^{-6}$ corresponding to a genome-wide threshold of 0.05. For clarity, QTLs mapping to the same locus were only highlighted once.

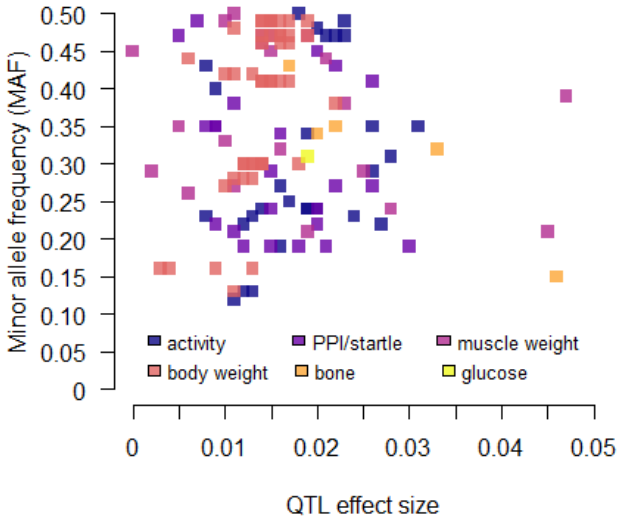

Figure 2.3: **MAF by QTL effect size in the LG × SM AIL.** There is little to no relationship between MAF and QTL effect size (proportion of phenotypic variance explained by the top SNP at each QTL) in the AIL.

itability for CPP, locomotor sensitization to methamphetamine, or habituation to acoustic startle (Supplementary Table 2.2).

2.3.2 Validation of the LOCO-LMM

I mapped QTLs using a leave-one-chromosome-out LMM (**LOCO-LMM**) to control for confounding due to proximal contamination and relatedness [67,69]. As a proof of principle, I mapped QTLs for albinism using a simple linear model, a standard LMM and a LOCO-LMM (Figure 2.6). Albino coat is an autosomal recessive trait caused by loss of function mutations in the *albino* locus (*Tyr*) on chromosome 7 [202] (Figure 2.6a); LG mice have the null allele and SM mice do not.

As expected, p-values from a genome-wide scan using a linear model, which does not account for relatedness, appeared highly inflated (Figure 2.6b). Inflation is reduced by fitting a standard LMM, which includes SNPs from chromosome 7 in both the fixed and random

Figure 2.4: **Heritability of metabolic and physiological traits in the LG × SM AIL.** Heritability is expressed as the percentage (rather than the proportion) of phenotypic variance explained by GWAS SNPs. Rounded values are displayed for each trait for comparison. Estimates for body weight from D2-D5 of CPP, which are similar to other body weight traits, are omitted for clarity.

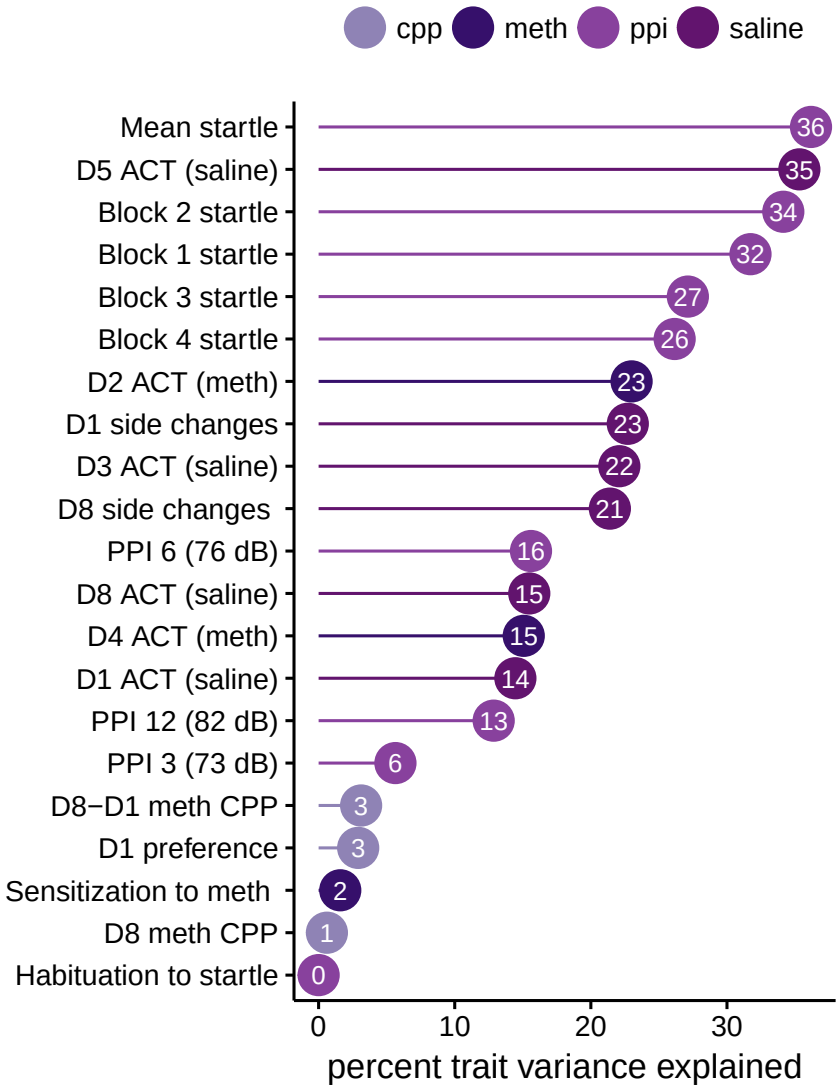

Figure 2.5: **Heritability of behavioral traits in the LG × SM AIL.** Heritability is expressed as the percentage (rather than the proportion) of phenotypic variance explained by GWAS SNPs. Rounded values are displayed for each trait for comparison. Estimates for individual time bins on each day are omitted; only heritability for 0-30 min (total) is displayed for activity, CPP, and sensitization.

effects (Figure 2.6c). The LOCO-LMM, which excludes markers from the chromosome being tested in the random term, showed an intermediate level of inflation (Figure 2.6d). To determine how well each model controls the false positive rate under the null hypothesis, I repeated the analyses after excluding chromosome 7 SNPs from the fixed effect. Even in the absence of the causal locus, I observed substantial inflation under the linear model, which can only be explained by population structure (Figure 2.6e). The standard LMM appeared overly conservative, which is likely a result of proximal contamination (Figure 2.6f). In contrast, the LOCO-LMM showed no inflation, consistent with the absence of *Tyr* and linked markers in the fixed effect(Figure 2.6g). These results support the use of a LOCO-LMM.

2.3.3 Identifying potentially spurious QTLs

The significance threshold obtained after correcting for multiple hypothesis testing was $p = 8.06 \times 10^{-6}$ (genome-wide significance threshold = 0.05); on average, this is the proportion of null SNPs in the data that are called significant. However, the true number of null SNPs is likely to vary across different traits, and some unknown proportion of significant QTLs could be false discoveries.

In Manhattan plots of human data, it is common to see the signal from one SNP appear much stronger than that of adjacent markers, but low LD makes it challenging to determine if these associations are false positives. However, the LD structure in the LG $\times$ SM AIL makes it unlikely for most SNPs to be independent from other SNPs nearby. QTL peaks in an AIL typically extend across multiple SNPs, and I rarely observed a QTL at a single SNP that was not supported by signals at nearby markers.

With this in mind, I visually inspected each peak for singleton SNPs and re-mapped suspicious QTLs using the full set of unpruned variants on the QTL chromosome. I considered p-values, allele frequencies, and proximity to adjacent, sub-threshold SNPs to determine whether a QTL was caused by imputation error.

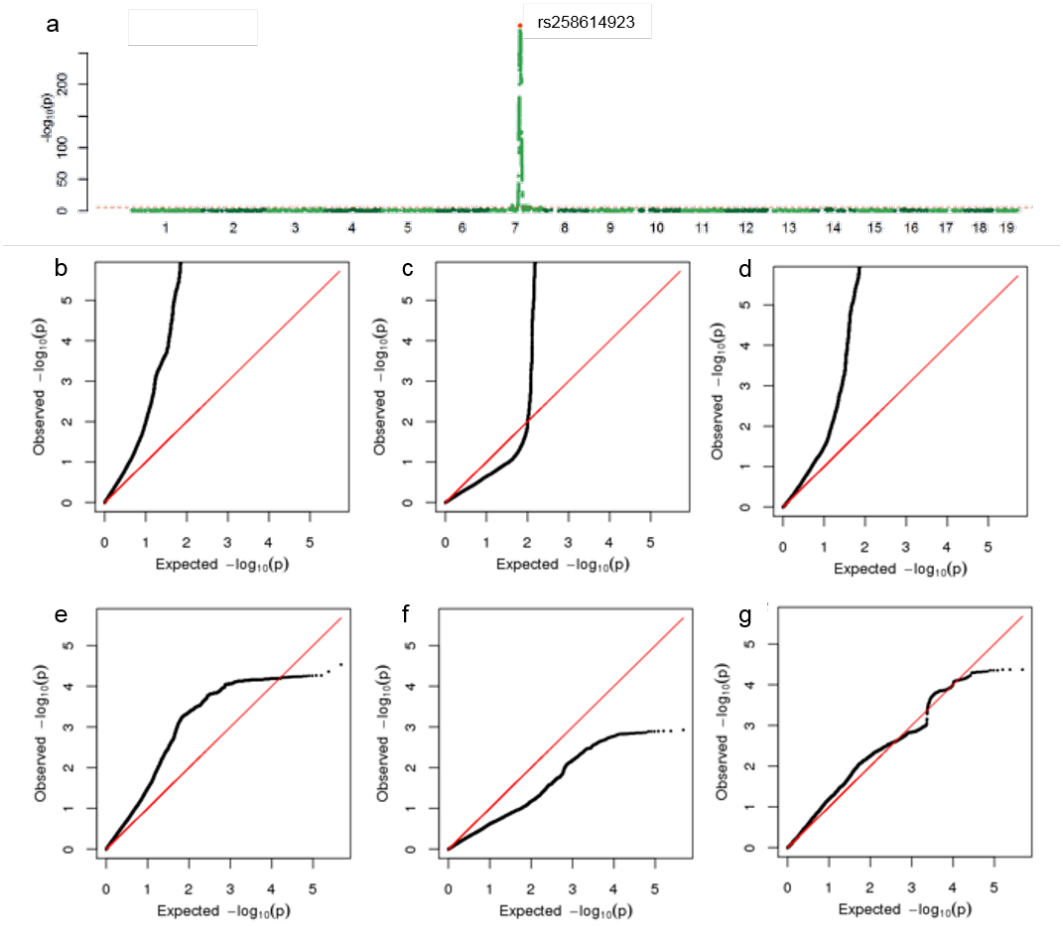

Figure 2.6: **Coat color GWAS validates the use of a LOCO-LMM**. I used three methods to conduct a GWAS for albino coat, a Mendelian trait caused by the *Tyr* locus on mouse chromosome 7. A Manhattan plot of GWAS results for albinism using the LOCO-LMM is shown in **(a)**. I repeated each scan after excluding SNPs on chromosome 7 to check for residual inflation of p-values. Quantile-quantile plots of expected vs. observed p-values from albino GWAS are shown for three different methods: **(b)** simple linear regression with no relatedness term; **(c)** a standard LMM that accounts for relatedness by including all GWAS SNPs in the GRM; **(d)** the LOCO-LMM whose GRM excludes SNPs located on the chromosome being tested, and **(e)** results from simple linear regression **(f)** the standard LMM, and **(g)** the LOCO-LMM after excluding chromosome 7.

I identified four QTLs in which the associated SNPs had unusual allele frequencies relative to non-significant SNPs nearby (D5 saline activity on chromosome 5, startle block 4 on chromosome 15, D8 side changes on chromosome 14, and tibia length on chromosome 19). Two other QTLs (glucose on chromosome 11 and gastrocnemius weight on chromosome 4) were ambiguous. For glucose and gastrocnemius, the top SNPs showed evidence of being errors, but unlike the others, they were surrounded by a wide span of elevated, but sub-threshold p-values. Although QTLs for both traits were mapped in previous generations of the LG $\times$ SM AIL [175, 180], I found no evidence for replication at these loci. However, I noted that two other muscle weights (for TA and plantaris) had sub-threshold QTLs at the gastrocnemius region on chromosome 4, and that they coincided with a set of strong QTLs for saline activity (Supplementary Table 2.3). I suggest interpreting these QTLs with caution (association statistics and QTL coordinates are provided in Supplementary Table 2.3).

I also determined that the top SNP within a set of startle QTLs on chromosome 17 was likely to be a genotyping error. Although the signals were very clear, using the most significant SNP to estimate the critical QTL region would have produced a 1.5-LOD interval that was overly narrow. Therefore, I calculated the critical region using another significant SNP in the region that was in high LD with its neighbors.

2.3.4 QTLs for behavior

Locomotor activity and methamphetamine reward

Multiple convergent QTLs for locomotor activity. I identified 28 activity QTLs on nine chromosomes, including 19 for saline activity (four of which were for side changes) and six for the locomotor response to methamphetamine (Figure 2.2e; Supplementary Table 2.3; Supplementary Figure 2.14). In nearly all cases, the SM allele was associated with increased activity, consistent with behavioral data from the inbred strains [75, 146]. The exceptions

were QTLs for locomotor sensitization to methamphetamine on chromosome 7, a QTL for D3 saline activity on chromosome 8, and QTLs for D8 activity on chromosome 12, where LG ancestry was associated with greater activity levels (Supplementary Table 2.3).

Activity is correlated on days when the testing environment and injection are the same (Supplementary Figure **??**), but because behavior is a dynamic response to the environment, I found it useful to treat each day's measurements as a different set of traits. For example, a mouse's preference for an environment may change after associating it with a rewarding (or aversive) drug experience, and a drug-naïve mouse may have a different response to methamphetamine than a mouse that has already experienced its effects [148]. Multiple activity traits mapped to five loci on chromosomes 4, 7, 12 and 17 (Supplementary Table 2.3). QTLs associated with multiple correlated traits measured on the same day may be redundant; in other cases, overlap could be due to pleiotropy.

It is unclear whether or not activity measurements from different five-minute time bins measured on the same day capture the same underlying phenotype (the bins are arbitrarily defined; I could have just as easily mapped QTLs for 10 or 15 minute intervals). Consistent with a previous study demonstrating that QTLs for total activity are driven primarily by activity in one or two five-minute bins [203], I observed that the locomotor response tends to be stronger in the first half of each day's test. One explanation for this is that exploratory locomotor activity in mice tends to be higher in a novel environment than in a familiar one [204, 205], and a decrease in activity over time may reflect habituation to the testing environment [146].

Still, activity QTLs associated with a single trait are less convincing than QTLs with multiple related traits that converge on an overlapping interval. Therefore, I visually inspected regional association plots for all traits that had at least one SNP with a p-value <0.0001 for evidence of trending QTLs that did not pass the threshold for genome-wide significance. I found evidence for non-significant associations with other, related activity traits

at the same loci for all QTLs except one (Supplementary Figure 2.14). The singleton QTL was associated with saline activity (D5, 10-15 min) and mapped to a region on chromosome 8 that I had previously identified as being associated with saline activity in G34 of the LG × SM AIL [75]. As before, the top SNP was located in an intron of *Csmd1*, a gene that has been associated with schizophrenia [206, 207]. However, the QTL identified here was more subtle, possibly due to differences in study design and the test used to measure locomotor activity.

Low heritability for methamphetamine CPP and locomotor sensitization. I measured three main phenotypes related to CPP: initial preference for the left side of the testing chamber on D1, final preference on D8, and the difference between D8 and D1 preference (as for activity, CPP traits were measured in five-minute bins). In disagreement with a wealth of data from rodents that supports a genetic basis for CPP [133–145], heritability estimates in this study were uniformly low (pve<=0.039, se<=0.025), suggesting that the additive genetic component for CPP is negligible in the LG × SM AIL (Supplementary Table 2.2).

A prior study using a small sample of SM, LG, and G45-46 AIL mice also found CPP heritability to be low [146], although that study was underpowered. I conclude that in this AIL, this CPP protocol, in conjunction with 1 mg/kg methamphetamine has low or possibly no heritability. Indeed, when I accounted for covariates like sex and testing chamber, I saw very little preference for the drug-paired side on D8, suggesting that little positive (or negative) conditioning occurred. It is possible that CPP would have been more robust (and perhaps more heritable) if I had used a different dose. The use of 2 mg/kg methamphetamine did not result in higher heritability in G45-46 despite more robust CPP; however, Bryant *et al.* [146] did not account for the effect of testing chamber, which I observed to have a large effect on both CPP and locomotor activity. The robust CPP observed by Bryant *et al.* [146] could therefore be an artifact of uncontrolled environmental factors. It remains possible that methamphetamine reward may be heritable under conditions where it can be measured more

precisely.

Heritability was also low for locomotor sensitization to methamphetamine. Nevertheless, I identified a QTL for two sensitization traits (10-15 and 15-20 minute intervals) on chromosome 7 (Supplementary Figure 2.14; Supplementary Table 2.3). A lack of signal for the other time points makes this QTL difficult to interpret, since I would expect sensitization to be associated with an *overall* difference in activity levels. However, sensitization is highly correlated with D4 activity (Supplementary Figure 2.11), and it is possible that the QTL is capturing some of the locomotor variation on D4.

Sensorimotor gating and the acoustic startle response

Evidence for a strong genetic component to startle. I identified 20 QTLs for five startle response traits on nine chromosomes (Figure 2.2e; Supplementary Table 2.3). Most QTLs were associated with multiple startle traits, with the exception of QTLs for startle block 4 on chromosomes 2, 15, and 16. I consider block 4 to be the noisiest as it is less correlated with the rest of the startle traits (Supplementary Figure 2.12) and has the lowest heritability (pve = 0.261, sd = 0.043). Heritability for block 3 was similar (pve = 0.271, sd = 0.043), suggesting habituation to stimuli from previous blocks as a likely cause of diminished heritability later in the assay. Notably, the mean startle response had the highest heritability out of all the behaviors I measured (pve = 0.362, sd = 0.045; Supplementary Table 2.2).

Robustness of PPI varies with prepulse intensity. I identified one QTL for PPI12 on chromosome 6 and three QTLs for PPI6 and PPI12 that mapped to two separate regions on chromosome 8 (Figure 2.2e; Supplementary Table 2.3). Interestingly, one of the PPI6 QTLs was located near *Csmd1*, which has been associated with schizophrenia in humans [206, 207]. This was potentially interesting, since PPI is deficient in schizophrenia patients (as well as in other psychiatric disorders) [46]. As for startle, heritability was slightly higher for traits

measured in early blocks (Supplementary Table 2.2), implying habituation. PPI6 had the highest heritability (pve = 0.156, sd = 0.037) and PPI3 had the lowest heritability (pve = 0.056, sd = 0.028; Figure **??**). I did not observe any QTLs for PPI3, suggesting that a 73 dB prepulse is not loud enough to inhibit the startle response in the LG × SM AIL. In support of this, I observed a trending QTL for PPI3 at the startle locus on chromosome 17 (Supplementary Figure 2.14).

Agreement with startle and PPI QTLs identified in G34. Previously, the Palmer Lab had established that LG startles more than SM (males only) and exhibits greater PPI (both sexes) [167]. In the current study, the LG allele was associated with a weaker response at half of the startle QTLs, but was consistently associated with increased PPI (Supplementary Table 2.3). Samocha *et al.* [167] used 472 LG x SM F2 and 135 G34 AIL mice to identify startle QTLs on chromosomes 7 and 17, PPI QTLs on chromosomes 11 and 12, and one QTL for habituation on chromosome 4. Only one QTL was replicated in this study (startle, chromosome 17). Given that Samocha *et al.* [167] had a much smaller sample size and used a different model to map QTLs, this may not be surprising. In this study, sex, weight, age and testing chamber were used as additive covariates for both PPI and startle (Supplementary Table 2.3), whereas Samocha *et al.* [167] mapped startle QTLs using a strain by sex interaction term and did not specify any covariates for PPI or habituation.

This study also differs in that I mapped QTLs for individual startle and PPI blocks, which provides information about how these behaviors change over time. I observed that QTLs for the mean startle response tended to co-localize with early startle blocks, although I did find QTLs exclusive to startle block 4 on chromosomes 2 and 16 (Supplementary Figure 2.14). Chromosome 17 was an exception; all startle traits (along with several for locomotor activity) mapped to the same region at approximately 28 Mb (Supplementary Table 2.3). Notably, habituation was not found to be heritable, implying that the QTL identified in G34 [167] could be a false positive. It is also possible that in the current study, QTLs

exclusive to startle block 4 are capturing some aspect of habituation, but this is difficult to determine from the current data. I note that on chromosome 17, block 4 is the only startle QTL where the SM allele was associated with an increased startle response (Supplementary Table 2.3). I observed the same pattern on chromosome 9, where a greater startle response in block 3, but not block 4, was associated with SM. In contrast, shared startle QTLs on chromosomes 7, 11, 12, and 18 exhibited the same effect of ancestry across traits.

A QTL on chromosome 17 is associated with many different behaviors

I observed a QTL on chromosome 17 at ~27 Mb that was associated with both saline- and methamphetamine-induced locomotor activity. The same region was strongly associated with startle (Supplementary Table 2.3), and previous studies in G34 of the LG × SM AIL identified QTLs for anxiety, conditioned fear [208], and the startle response [167] in the same region. The chromosome 17 QTLs identified in G50-56 range from 1.14-2.84 Mb and contain several genes that appear important for brain function and behavior (details are provided in Chapter 3).

2.3.5 QTLs for metabolism and physiology

Fasting blood glucose levels

The single QTL for fasting blood glucose levels that I identified was on chromosome 11 at 68.06-76.34 Mb, an ~8 Mb region encompassing over 250 genes ($p = 7.38 \times 10^{-6}$; MAF=0.31; Supplementary Figure 2.14). I was somewhat surprised that I did not identify more glucose QTLs, since LG and SM exhibit striking differences in glucose metabolism [175, 209] and QTLs for glucose traits were identified in earlier generations of the LG × SM AIL [210]. Heritability for fasting blood glucose levels was 0.187 (sd = 0.039), but this does not account for epistasis, imprinting, or maternal effects, which have been shown to influence metabolic traits

in this population [175,210–213]. Furthermore, there is evidence that context-dependent genetic effects (e.g. alleles that only affect the phenotype of animals on a specific diet) also contribute to differences in LG × SM AIL metabolism [214].

Tibia and tail length

I identified QTLs for tibia length on chromosome 1 at 59.38-61.42 Mb ($p = 7.38 \times 10^{-6}$; MAF=0.32) and 83.76-87.27 Mb ($p = 8.64 \times 10^{-12}$; MAF=0.15). The two regions contained 21 and 53 genes each. On chromosome 2, I found overlap between another tibia QTL at 118.99-123.74 Mb ($p = 5.25 \times 10^{-8}$; MAF=0.35) and a QTL for tail length at Mb ($p = 2.84 \times 10^{-7}$; MAF=0.34); both QTLs contained over 100 genes. As expected, the LG allele was associated with increased length at all bone QTLs (Supplementary Table 2.3). I did not identify QTLs for tibia bone mineral content or thickness, which was consistent with their low heritabilities (pve<=0.10) relative to the other muscle and bone phenotypes (Supplementary Table 2.2). Manhattan plots for each trait are presented in Supplementary Figure 2.14.

Body weight

I identified 46 QTLs for body weight measured at 10 time points (Supplementary Table 2.2), most of which converge at the same loci on chromosomes 2, 4, 5, 7, and 8 (Figure 2.2e; Supplementary Table 2.3). Heritability for body weight ranged from 0.373-0.434 (sd=0.043-0.044). If I assume that most overlapping QTLs are capturing the same effect, then I effectively identified eight body weight QTLs.

Four QTLs on chromosomes 2, 4, and 7 were associated with at least eight body weight measurements (chromosome 4 contained two body weight loci). QTLs on chromosomes 5 and 8 were associated with four traits each, and the remaining QTLs on chromosome 1 and 10 were singleton associations. All QTLs with less than 10 body weight associations were

supported by at least one sub-threshold body weight QTL (Supplementary Figure 2.14). The chromosome 1 QTL for weight on CPP D4 also overlapped with a QTL for tibia length (Supplementary Table 2.3), implying that the association with body weight might be driven by a factor regulating bone growth (or muscle mass, which is related to bone size) as opposed to adiposity or organ weight.

Muscle and bone

I identified 21 QTLs for muscle weight on 10 chromosomes despite a smaller sample size (N=878-883) (Figure 2.2e; Supplementary Table 2.3). Heritability for muscle weight ranged from 0.202 for soleus to 0.429 for EDL (Supplementary Table 2.2). Each muscle had at least five QTLs, except for soleus, which had none. If overlapping muscle QTLs are related to the same underlying factor, then I effectively identified 16 associations with muscle weight.

LG and SM exhibit a twofold difference in soleus muscle size [180] that is largely due to differences in the cross-sectional area of soleus muscle fibers [215]. Soleus is a slow-twitch muscle whose fiber composition is distinct from the other four; EDL, TA, gastrocnemius, and plantaris are fast-twitch muscles [185, 186]. If muscles composed of different fiber types are regulated by distinct genetic mechanisms, it is possible that they also have distinct genetic architectures (addressing this hypothesis was one motivation for mapping muscle QTLs). Given that I identified QTLs for soleus weight [180] and muscle fiber composition [181] in LG $\times$ SM G34, it was somewhat surprising that I failed to find any here. I did observe trending, non-significant QTLs for soleus on chromosomes 12 and 15; the former was located near QTLs for TA and plantaris, and the latter overlapped another non-significant QTL for EDL (Supplementary Figure 2.14). I suggest that a meta-analysis of G34 and G50-56 data could be used to resolve the genetic architecture of different muscle weights in more detail.

Overlap with muscle and body weight QTLs identified in G34

QTLs for body weight, muscle weight, and muscle fiber properties were previously identified using F2 and G34 mice. Despite differences in sample size, genotyping density, and the models used for mapping, many of the QTLs identified in G50-56 converged on the same loci. In some cases, recombination appears to have facilitated LD decay within the larger regions identified in G34, allowing multiple, adjacent QTLs to emerge in G50-56.

Parker *et al.* [147] identified 11 body weight QTLs in a population of F2 and G34 mice that were weighed at approximately 62 days of age [147] (this corresponds to weight on CPP D8 in this study; details are provided in the Supplementary Methods 2.5). I identified QTLs on many of the same chromosomes as Parker *et al.* [147], but most mapped to different intervals. The exception was a region on chromosome 7 that was associated with all 10 body weight measurements. In G34, the QTL was located at 107.65-113.46 Mb [147]; in G50-56, the ten overlapping QTLs ranged from 105.67-110.81 Mb. Lionikas *et al.* [180] also identified five muscle weight QTLs at 110.87-114.93 Mb, which overlaps the G34 body weight QTL but stops just short of the G50-56 region. I speculate that additional recombination in G50-56 precipitated LD decay within the chromosome 7 region identified by Parker *et al.* [147], allowing me to distinguish between the two adjacent loci. The identification of G50-56 QTLs for TA, plantaris and gastrocnemius weight located adjacent to the body weight peak supports this hypothesis (Supplementary Table 2.3). I note that there is still some overlap between the muscle and body weight QTLs in G50-56, but this is expected, since muscle mass contributes to body mass. The size and shape of the body weight QTLs suggest that additional variants regulating body size are located upstream of muscle loci and that their effects may be less visible in muscle than in other tissues.

A G50-56 QTL for body weight at death on chromosome 4 (57.76-65.35 Mb) overlapped an association with the percentage of type 1 fibers in G34 soleus muscle at 57.70-62.70 Mb (variation in muscle fiber size accounts for a large portion of the difference in muscle mass

between LG and SM) [181]. Nine other body weight traits also mapped to the region, but their intervals were considerably narrower, ranging from 62.1-66.4 Mb (Supplementary Table 2.3). A closer inspection of the QTLs suggested that the larger region associated with weight at death actually contains two QTLs, one of which was under the threshold for significance. I remapped QTLs by alternately conditioning on the top SNP in each apparent peak to determine if the signals were independent; however, I found no evidence that this was the case. The large QTL for weight at death had the weakest signal ($p = 1.62 \times 10^{-6}$), which might explain why its 1.5-LOD drop interval captured the sub-threshold signal at 57-60 Mb. I think it is likely that there are multiple loci within the region that affect body weight, perhaps through different mechanisms. For example, the distal QTL may be related to muscle weight, since G50-56 QTLs for EDL, TA, and plantaris mapped to different intervals between 63.87-69.90 Mb (Supplementary Table 2.3).

A locus on chromosome 11 (28.0-31.5 Mb) associated with type 1 and type 2A fiber area in G34 soleus [181] overlapped a G50-56 QTL for EDL weight at 29.79-33.05 Mb. EDL is a small fast-twitch muscle comprised mostly of type 2 fibers, whereas fibers in soleus, a slow-twitch muscle, are primarily of type 1 [185]. Overlap of the two QTLs could imply that the G50-56 EDL locus specifically affects muscle fibers. Interestingly, a nearby region (28.34-29.57 Mb) was associated with various other muscle weights, not including EDL, in a separate study of G34 mice [180], suggesting that multiple genes on chromosome 11 may be involved in muscle regulation. Even so, all of the muscle QTLs agree that the LG allele confers an increase in muscle mass.

Finally, a G50-56 body weight QTL at 80.80-85.10 Mb on chromosome 8 was identified as a muscle weight QTL in G34 [180]. The associated traits were TA, EDL, gastrocnemius, and the first principal component of muscle weight, which mapped between 83.37-85.45 Mb [180]. Unexpectedly, the body weight locus was slightly larger in G50-56, which could suggest the presence of multiple causal factors within the region. Another G50-56 body weight QTL on

chromosome 5 (67.07-67.69 Mb) was found 1 Mb upstream of a large G34 QTL for EDL weight (68.78-82.13 Mb) [180].

The overlap of body weight and muscle QTLs may imply that the QTLs for body weight are caused by variation in muscle mass, or simply that the loci affect multiple tissues in a similar direction. I think it is likely that the traits are affected by multiple different genes within each region; however, more data is needed to elucidate the underlying mechanisms. Differences in marker density, mapping resolution, sample size, and QTL models might explain why some G34 QTLs did not replicate in G50-56. A meta-analysis of the current data along with densely genotyped G34 mice could help to disentangle the genetic architecture of body and muscle weight.

2.3.6 Discussion

I identified 52 QTLs for behavioral traits and 74 QTLs for metabolic and physiological traits in G50-56 of the LG × SM AIL, most of which were novel associations. I demonstrated that there is little to no relationship between MAF and QTL effect size is weak in the AIL, which is not the case for humans (this is likely due to differences in population history, genetic background, and environmental variables; I discuss this at length in Chapter 4). I showed that a LOCO-LMM is the optimal approach for controlling for confounding due to relatedness, which is an important concern for GWAS in samples of individuals that are related to each other in complex ways.

I also replicated several associations that were identified in combined analyses of F2 and G34 AIL mice, including QTLs for activity, startle, body weight, and muscle weight. I found considerable overlap between body and muscle weight QTLs identified in G50-56 and QTLs for the same traits that were identified in G34. In some cases, G34 body weight QTLs overlapped G50-56 muscle QTLs (but not G50-56 body weight QTLs), which was somewhat unexpected. This could suggest that certain body weight QTLs identified in the AIL are

driven by differences in muscle mass. More generally, it demonstrates that fine-mapping QTLs for distinct, but related traits in multiple AIL generations is a useful way to begin disentangling multiple effects within larger QTLs. I would be interested to map QTLs for muscle weight, organ weight and fat mass in future generations of the AIL; this data could be used to identify general and tissue-specific factors that influence body weight.

The biggest surprise to come out of this study was the observation of low heritability for CPP, which was a primary focus of this study. Although this was somewhat disappointing, I knew at the onset that GWAS of CPP could be risky. Because CPP could be influenced by genetic variants that influence learning, spatial cognition, memory, incentive salience, or other aspects of behavior that are distinct from reward, I had assumed that the risk would be in not knowing how to interpret CPP QTLs. As it turned out, there were no CPP QTLs to interpret. *How could this be?* There are a number of studies in knockout and transgenic animals that imply a genetic basis for CPP, and furthermore, AIL mice had exhibited high phenotypic variability in preference for methamphetamine. A likely explanation is that there is not enough genetic variation in the LG $\times$ SM AIL to produce CPP, and the differences we observed are due to environmental factors. It is also possible that I may have detected CPP QTLs if I had used a different drug, dose, or a longer conditioning period. However, I would suggest a different approach for future studies that aim to identify genetic factors important for reward. Instead of adjusting the CPP procedure or using different mouse strains, I would recommend using the rat to study phenotypes related to drug self-administration, which provides a more direct way to measure variation in reward [112] (it is also easier to measure in rats than in mice for a variety of reasons [52]). At the time this study was developed, large GWAS were impractical in rats, but this is no longer the case. As the rat genetic toolkit continues to expand, I expect that GWAS in rats will be very useful for exploring the genetic basis of complex behaviors such as drug reward.

Discussion of QTL results will continue in Chapter 3, where I will describe how I used

functional annotations, eQTLs identified in three brain regions, and differentially expressed genes in LG and SM muscle to interpret QTL results and identify promising candidate genes.

2.4 Contributions

Dr. Abraham A. Palmer designed the experiment and provided financial support for this project (R01 DA021336). Additional funding came from F31 DA03635803 and T32 GM007197.
I wrote my dissertation in the first-person perspective, but this work was a team effort. Dr. Clarissa C. Parker, Dr. Margaret G. Distler, and Michael A. Guido provided me with training in mouse husbandry and behavior. I performed all mouse behavioral, glucose and body weight phenotyping at the University of Chicago Animal Research and Care Facility, except on a few instances when Celine St. Pierre or Emily Leung assisted with PPI, or when Matthew Cheung assisted with glucose testing. Celine, Emily, Matthew, Yeonhee (Jenny) Park, Emmanuel Aryee and I dissected tissue. Celine and Emily extracted DNA and prepared samples for GBS, which were sequenced in Dr. Yoav Gilad's lab in the Department of Human Genetics at the University of Chicago. Dr. Arimantas Lionikas measured muscle and tibia traits in his laboratory at the University of Aberdeen. Dr. Shyam Gopalakrishnan initiated the analysis of GBS data and provided Natalia the computational training required to finish it. The Center for Research Informatics at the University Chicago computational resources to process the data; Mike Jarsulic oversaw the computational cluster and assisted with technical issues. Dr. Mark Abney advised me on pedigree error checking and permutation. I processed mouse phenotypes, analyzed GBS data, and mapped QTLs.

2.5 Supplementary Materials

2.5.1 Supplementary Methods

Phenotypes

I processed phenotypes for 1,063 mice (530 female, 533 male) that had high-quality GBS data. Factor covariates (sex, batch, generation, coat color, and testing chamber) were represented as indicator variables. Sex was included as a covariate in all trait models. Other covariates were selected in three stages: (1) I used the adjusted r-squared statistic from a univariate linear regression model to estimate the proportion of phenotypic variance explained by each variable separately. (2) Variables that explained 1% or more of the trait variance were used in the model selection R package, leaps, to identify an optimal set of predictors for each trait. (3) I then reviewed each list of selected covariates and made revisions if necessary (details below). Residuals from the final trait models were plotted to identify outliers. An individual was considered an outlier if its residual value was more than three standard deviations from the mean and fell outside the 99% confidence interval of the normal distribution. Traits were quantile normalized after removing outliers. The final sample size for each trait is shown in Supplementary Table 2.2 along with a list of covariates used for QTL mapping.

Conditioned place preference and locomotor activity

CPP paradigm and testing environment. CPP is an associative learning paradigm that measures the motivational properties of a drug and the ability to associate its effects with a particular environment. Mice learn to distinguish between two environments that are paired with either administration of a drug or administration of saline. After repeated pairings, mice are given a choice between the two environments. An increased amount of time spent in the drug-paired environment indicates drug reward and is interpreted as 'preference' for

the drug.

I created two visually and tactilely distinct environments by dividing a transparent acrylic testing chamber (37.5 × 37.5 × 30 cm; AccuScan Instruments, Columbus, OH, USA) into equally sized arenas using an opaque divider with a passage (5 × 5 cm) in the bottom center [146]. The other three walls of each partition are distinguished by visual cues (stripes on the walls) and tactile cues (floor textures). The chamber is placed inside a frame that transmits an evenly spaced grid of infrared photo beams through the chamber walls. Beam breaks used to monitor the mouse's activity and location are converted to time and distance units by the AccuScan VersaMax Software. Each testing chamber is encased within a sound attenuating PVC/lexan environmental chamber. Overhead lighting provides low illumination (∼80 lux), and a fan provides both ventilation and masking of background noise.

I reversed the divider during conditioning trials to restrict the mouse to one arena within the testing chamber. 1 mg/kg methamphetamine was always paired with the left arena (white horizontal stripes, smooth floor) and physiological saline was always paired with the right arena (black vertical stripes, textured floor). I had previously established that mice did not prefer one arena over the other [146]. The 1 mg/kg dose of methamphetamine was intended to generate preference and locomotor stimulation without inducing stereotyped behaviors.

Measurement of CPP and locomotor activity. CPP and locomotor activity was measured simultaneously between the hours of 08:30 to 17:30. Up to a dozen mice were tested simultaneously in 12 separate CPP chambers. Median age at the start of CPP was 54 days (mean=55.09, range=35-101). On each day of the assay, mice were placed on a portable shelf and were transported from the colony to an adjacent testing room. Mice were given 30-45 minutes to acclimate to the room before being removed from their home cages. Before each 30 minute test, mice were weighed and momentarily separated into clean holding cages. After injection and placement into test chambers, mice were free to travel between arenas

on days 1 and 8 after receiving an intraperitoneal injection of physiological saline. On days 2-5 (conditioning trials) mice were intraperitoneally administered either methamphetamine (1 mg/kg, days 2 and 4) or vehicle (saline, days 3 and 5) in a volume of 0.01 ml/g body weight. I also measured locomotor activity (total distance traveled in cm) on each day and recorded the number of times the mouse switched between sides of the chamber on days 1 and 8. After each 30 min testing session, I returned mice to their home cages. Test chambers were cleaned with 10% isopropanol between runs. Home cages were returned to the colony at the end of each experiment.

Analysis of CPP and locomotor activity I define CPP as the increase in time spent in the methamphetamine-paired arena on day 8 compared to day 1. To account for the possibility of initial preference for one arena, I also considered preference on the final test day (without regard to preference on day 1) as a second outcome measure for CPP. I refer to locomotor activity on day 1 as the locomotor response to a novel environment. I also interpret side changes on day 1 as a response to novelty. The locomotor response to saline is measured on days 3 and 5. I also measure activity on day 8, but the environment is different because the mouse has access to the entire CPP chamber. Side changes measured on days 1 and 8 provide additional measures of locomotor activity in response to novelty and saline, respectively. The locomotor response to 1 mg/kg methamphetamine is measured on days 2 and 4. I also calculated the increase in methamphetamine-induced activity on day 4 relative to day 2 as a crude measure of locomotor sensitization. Locomotor sensitization is an increase in the magnitude of drug-induced activity after repeated administration of the same (or sub-threshold) dose of the drug.

All CPP and related phenotypes were measured in 5-minute bins over the course of 30 minutes. I summed measurements across bins to obtain a value for total time in seconds, distance in cm or number of side changes. Therefore, I have seven individual measurements for each CPP and locomotor phenotype. Binned measurements are highly correlated; there-

fore, I used the same set of covariates for all binned phenotypes within a given day and phenotype class.

Occasional software malfunctions that occurred at the time of testing (in which the Accuscan software was unable to record movement in certain chambers for 6.9-13.8 seconds during the test) were automatically detected and included in the output for each 5-minute time interval in which the error occurred. Data for these specific intervals and total activity on the affected day were marked as missing in 61 mice. All data were quantile normalized prior to QTL mapping.

Prepulse inhibition of the acoustic startle response (PPI)

PPI paradigm and testing environment When a mouse is startled by a loud noise, its facial and skeletal muscles contract rapidly. PPI is the reduction of the acoustic startle response when the startle stimulus is preceded by a low decibel tone [166] and is considered to be an intermediate phenotype for schizophrenia [46]. To measure PPI, each mouse is placed inside a 5 cm Plexiglas cylinder within a lit and ventilated testing chamber (San Diego Instruments, San Diego, CA, USA). The mouses movements are detected by a piezoelectric sensor. Once inside the chamber, mice have 5 minutes to acclimate to 70 dB white noise, which remains in the background for the duration of the 18 minute test. After acclimation, mice are repeatedly exposed to acoustic startle stimuli (120 dB pulse; 40 ms) which is sometimes preceded by a 20 msec prepulse (3-12 dB above background noise) at variable intervals.

The acclimation period is followed by 62 trials that are a mixture of the following five types: a pulse alone trial, which consists of a 40-millisecond 120 dB burst (startle stimulus), a no stimulus trial where no stimulus is presented, and three prepulse trials containing a 20 millisecond prepulse that is either 3, 6 or 12 dB above the 70 dB background noise level followed 100 milliseconds later by a 40 millisecond 120 dB pulse. Trials are split into four

consecutive blocks. Blocks 1 and 4 each contain 6 pulse-alone trials. Blocks 2 and 3 are a mixture of 25 trials (6 pulse-alone, 4 no stimulus and 15 prepulse trials). The variable inter-trial interval is 920 seconds (mean=15 seconds) throughout all 62 trials.

Measurement of PPI and startle I measured PPI 4-9 days after the last day of CPP (mean=7.13 days; median=7 days). Median age of mice at the time of testing was 68 days (mean=69.2, range=49-115). The PPI system was calibrated at the start of each testing day according to the manufacturers instructions. Mice were transferred to the testing room one cage at a time, weighed, and then placed into the testing chamber. Mice were returned to their home cage after testing. PPI cylinders were cleaned between sessions. After returning mice to the colony, the procedure was repeated for consecutive cages.

Analysis of PPI and startle The startle response is calculated as the mean startle response across the startle alone trials. I also mapped QTLs for startle amplitude in blocks 1-4 separately.

I define PPI as the normalized difference between two values: the mean startle response during pulse-alone trials and the mean startle response during prepulse trials. The first value is the raw startle phenotype. The second value was calculated for each level of prepulse intensity (73, 76 and 82 dB) and divided by the raw startle value to obtain a proportion, which I transformed with the logit function. To avoid extreme values caused by logit transformation of negative PPI values, I projected the transformed data onto an interval between 0.01 and 0.99, as described in Parker *et al.* [62]. Startle, which is positive, was transformed with the log10 function. I also measured habituation to startle, which is the difference between the mean startle response during the first and last block of pulse-alone trials.

After transformation, I examined the distribution of startle responses during the no-stimulus trials to check for technical errors. I identified 44 mice that seemed to startle in the absence of a pulse (Supplementary Figure 2.7), which I interpreted as a technical effect

since all 44 mice were tested in PPI box 3. I retained these mice in the analysis and included box 3 as a covariate for all PPI and startle phenotypes. I identified a group of mice with unusually low startle responses (Supplementary Figure 2.7). It is likely that these individuals are hearing-impaired because their startle responses overlapped the trait distribution for the no-stimulus trials (excluding the 44 mice from box 3). PPI-related phenotypes for 13 mice with a mean startle response of 1.1 units or lower were marked as missing. Data were quantile normalized prior to mapping QTLs.

Fasting blood glucose levels I measured blood glucose levels after a 4 hour fast 4-14 days (mean=7.3 days, median=7) after PPI testing. Median age of mice at the time of testing was 75 days (mean=76.4, range=56-122). Mice were brought into the testing room between 09:00 and 09:30 and transferred to new cages that did not contain food. After 4 hours I weighed each mouse and used a razor blade to make a small incision at the tip of the tail, which allowed me to obtain a small drop of blood that I analyzed with glucose strips (Bayer Contour TS Blood Glucose Test Strips) and a glucometer (Bayer Contour TS Blood Glucose Monitoring System). Glucose levels are expressed in mg/dL units. Once all of the mice in a cage were tested, I gave them food and returned them to the colony. Glucose measurements were quantile normalized before mapping QTLs. Covariates are shown in Supplementary Table 2.2.

Coat color The LG × SM AIL segregates three coat color phenotypes. LG has a white (albino) coat. The SM strain is fully inbred except at the *agouti* locus, where attempts to maintain a homozygous state have been unsuccessful; thus, LG × SM AIL mice can be either albino, black or agouti [216]. I transformed coat color into three indicator variables and treated them as quantitative traits in GEMMA. I found this approach acceptable for testing this method because the genetic basis of coat color is well-known and because I did not expect to identify novel QTLs. Although GEMMA's LMM was intended for quantitative

analysis, its robustness to model misspecification makes it acceptable for mapping factors expressed as binary case-control variables [37].

Tissue collection I collected tissues for DNA and RNA extraction and further phenotyping by collaborators 4-15 days (mean=7.46, median=7) after measuring glucose. Median age at sacrifice was 83 days (mean=84.4, range=64-129). Mice were removed from the colony immediately prior to dissection, weighed and sacrificed using cervical dislocation followed by rapid decapitation. Tail length (cm from base to tip of the tail) was also measured at this time. The left hind limb of each mouse was cut off just below the pelvis, placed in a tube and stored on dry ice. Spleens and tails were stored in a 1.5 mL solution of 0.9% saline for DNA extraction and placed on dry ice immediately after dissection.

HIP, STR and PFC were dissected from intact brains for RNA sequencing and gene expression analysis. I chose these tissues because they can be isolated from whole brain in a consistent and straightforward manner, and because they have been repeatedly linked to drug abuse, learning, and motor phenotypes [217, 218]. Intact brains were extracted and submerged in chilled RNALater (Ambion) for one minute before dissecting each tissue. Individual tissues were stored separately in 0.5 mL tubes of RNALater, which were placed on ice immediately after dissection. All brain tissue in this study was dissected by the same experimenter and subsequently stored at -80C.

A concern when measuring gene expression is that the events immediately before sacrifice might alter gene expression. To alleviate this problem, I only collected brain tissue from the first mouse taken from each cage, which I selected in advance (to avoid sampling brain regions from mice that were easier to catch). I avoided choosing full siblings so that I would have better power to detect eQTL, but otherwise mice were selected randomly. Sampling only one mouse per cage also reduces the amount of time between the mouse's removal from the colony (resting state) and storage of the tissue in RNALater to under 5 minutes.

Hind limb muscle and bone I transferred frozen hind limb tissue to Dr. Arimantas Lionikas, who brought them to his laboratory at the University of Aberdeen for dissection. Each limb was kept frozen until the day of dissection. Muscles were separated under a dissection microscope and weighed to 0.1 mg precision on a balance (Pioneer, Ohaus). The soft tissue was then removed from the tibia and measured to 0.01 mm precision with a digital caliper (Z22855, OWIM GmbH & Co). Mineral content of the tibia was measured with the Faxitron MX-20 scanner (Faxitron Bioptics LLC, AZ, USA) using methods adapted from Nicod et al., 2016. Three types of materials; 0.8 mm of aluminium, 1.0 mm of polystyrene and 0.8 mm of steel, were scanned together with the bones for calibration. ImageJ (V1.48p, National Institutes of Health, USA) was used to quantify the apparent bone mineral content, which was characterized by the median of the optical density of bone image which was measured in pixel intensity. I also generated a variable for bone thickness by combining thickness measurements with the central axis of the bone image, the measurement unit for this variables was also pixel intensity.

Phenotyped muscles include two dorsiflexors, TA and EDL, and three plantar flexors, gastrocnemius, plantaris and soleus. Muscles were selected because they differ in size and constitution of fiber types. Of the fast-twitch muscles, TA and gastroc are largest and express the entire range of type 2 fibers and some type 1 fibers. EDL and plantaris are smaller fast-twitch muscles comprised mainly of type 2B, 2X and 2A fibers. Soleus, a slow-twitch muscle, is comprised mostly type 1, 2A and 2X fibers [185]. Different morphological and functional properties are associated with each fiber type [186], and I reasoned that muscles composed of different types might be regulated by distinct genetic mechanisms. Tibia length is indicative of skeleton size of the skeleton, whereas mineral content and thickness of bone are predictors of bone strength,

Elongation of bones is associated with longer, larger muscles. Therefore, in order to isolate muscle-specific QTLs as opposed to QTLs that regulate growth across multiple tissues

I included tibia length as a covariate in muscle mass GWAS. Skeletal muscle is abundant tissue contributing substantially to body weight in mammals. To avoid circular correction which would reduce power to detect muscle weight QTLs I did not use body weight as a covariate for muscle traits. All muscle weights and bone measurements were quantile normalized before QTL mapping. A final list of covariates used for leg muscle and tibia phenotypes is provided in Supplementary Table 2.2.

Genotype quality control

Identification and correction of sample mix-ups Mislabeling and sample mix-ups are common in large genetic studies and can reduce power to detect QTLs. I was concerned about the possibility of samples being mistakenly swapped or mislabeled. Therefore, I called variants in two stages. First-pass variant calls were used to identify and resolve sample mix-ups. In stage two, after correcting or omitting errors in the data, I repeated variant calling from scratch.

I compared relationships estimated from genetic data against kinship estimated from the AIL pedigree to identify and resolve mixed-up samples. Sample information for apparent mix-ups was checked against breeding records, mouse cage cards, and other experiment logs to identify the source of each error. I examined the number of reads mapping to the sex chromosomes, verified homozygous LG genotypes at the Tyr locus for albino mice, cross-checked genotypes with RNA sequencing data and compared genotypes called from replicate GBS samples sequenced on separate dates. I also measured genotype concordance for 24 individuals regenotyped on the Giga Mouse Universal Genotyping Array (GigaMUGA) [191].

Sex swaps I counted the number of reads that mapped to the X and Y chromosomes to validate the sex of each mouse. LG, SM, F1 and GigaMUGA samples were used as benchmarks. For true females, I consistently observed that the number of X chromosome

reads was an order of magnitude greater than the number of Y chromosome reads. A difference greater than one order of magnitude was never observed for true males that had more reads on the X chromosome than the Y. Nineteen samples with an X:Y read ratio that violated this heuristic were flagged as potential errors before proceeding to call variants.

Genetic relatedness Most mice in the sample have an opposite-sex sibling, which allowed me to identify errors by comparing pedigree kinship to the realized relationships estimated from genetic data. To estimate the genetic relatedness matrix (**GRM**), I called variants using ANGSD as described for the final genotypes, except I required that only 15% of the samples have reads in order for a call to be made. I removed first-pass variants with MAF ¡ 0.01 before using Beagle v.4.1 to fill in missing genotypes at 106,180 loci [196]. I did not input a reference panel at this stage; instead, I inferred missing data from LD within the sample. This ensured that all mice had a genotype at each empirically typed GBS locus while avoiding perpetuating widespread errors by imputing from a reference panel or pedigree.

Approximately 37,000 Beagle variants with a dosage r2 value (**DR2**, a measure of imputation quality) above 0.7 and MAF above 0.1 were used to calculate realized kinship coefficients with IBDLD [219, 220]. Pedigree kinship was calculated using a custom R script (`https://github.com/pcarbo/breedail`). I identified 21 non-sibling pairs with unusually high kinship and 22 sibling pairs with unusually low kinship, 8 of which overlapped with mice flagged as sex swaps.

Genotype concordance with GigaMUGA Concordance between GBS and array genotypes was evaluated at multiple stages to guide filtering and gauge the efficacy of my variant calling pipeline. Twenty-four mice with moderate to high GBS coverage were selected for genotyping on the GigaMUGA. The GigaMUGA contains probes for over 143,000 SNPs [191]. After removing chip SNPs with an Illumina quality score ¡ 0.7, I was left with 115,478 autosomal SNPs, 24,934 of which are known to be polymorphic in LG and SM [193]. Genotype

concordance for 18,278 overlapping SNPs after correcting sample mix-ups, imputation, and filtering was 97.4% (Supplementary Table 2.1). This is similar to concordance rates observed for other animal populations genotyped with GBS [62, 221] and falls within the range of imputation concordance rates reported in human studies [222, 223].

Error correction After making changes, I re-estimated kinship coefficients with IBDLD and compared error-corrected GBS genotypes to error-corrected RNA sequencing data. To create genotypes from RNA sequencing data, I retrieved allele counts at biallelic sites that had at least 25 high-quality reads and used the ratio as a proxy for genotype dosage. I converted dosages to hard calls and measured concordance with GBS genotypes at overlapping sites for each mouse and tissue separately.

Ultimately, there were 15 out of 1,078 samples whose identities could not be resolved. These mice were included in variant calling for the final sample because they provide additional information for obtaining genotype likelihoods. However, they were discarded before imputation and were not used for mapping.

Variant calling

Genotype likelihoods I used an implementation of the Samtools variant calling algorithm in ANGSD to obtain genotype likelihoods at 899,436 sites for which at least 20% of samples had reads [195,224]. GBS produces variable coverage across individuals, which leads to highly heterogeneous call rates and a bias toward homozygous calls [188]. Accordingly, I expected ANGSD's allele frequency estimates to be biased and used a lenient MAF threshold of 0.005 to filter raw genotype likelihoods.

Imputation I used Beagle to call genotypes from ANGSD likelihoods at 221,091 autosomal sites that passed quality control filters. When a reference panel is provided, Beagle requires hard genotype calls as input; however, ANGSD only outputs likelihoods. Therefore,

I imputed missing genotypes in three steps. First, I used Beagle to phase and fill in missing calls using within-sample LD (no reference panel or pedigree was provided). This produces a file with hard genotype calls, genotype probabilities and dosages that can be used to impute additional genotypes from an external reference panel. Next, I excluded SNPs with a MAF under 0.1, leaving 49,373 variants for step three. I used a MAF threshold of 0.1 because both alleles are expected to be common in an AIL; this was confirmed at generation 34 [75]. I then used Beagle to impute untyped SNPs from LG and SM reference haplotypes, which I obtained from whole-genome sequencing data [193]. The JAX Mouse Map Converter was used to create a genetic map from mm10 base pair coordinates [225]. I retained 3.4 million variants with high imputation quality (DR2 $\geq$ 0.9) and MAF >0.1 for further analysis.

Hardy-Iinberg equilibrium (HWE) An AIL is not a randomly mating population, its effective population size is not infinitely large, and LD is extensive. Selecting an appropriate threshold for excluding HWE deviations is further complicated by the tendency of GBS to falsely call heterozygous genotypes as homozygous when coverage is low. I ran 1,000 gene dropping simulations in the R package QTLRel v.0.2-15 to simulate null genotypes consistent with the AIL pedigree (but not impacted by the overrepresentation of homozygotes that is observed when using GBS) [226]. I used the R package HardyIinberg v.1.5.6 to test simulated genotypes for deviation from HWE [227] . To reduce the computational burden of gene dropping, I restricted my analysis to 372,995 SNPs with unique cM positions. Chi-squared p-values $\leq 7.62 \times 10^{-6}$ were only detected for 1% of simulated loci. I used this value to identify loci where the observed genotype proportions constituted a significant deviation from HWE given that the data are from an AIL. 52,466 loci that were found to deviate from HWE at a significance level $\leq 7.62 \times 10^{-6}$ were excluded. Finally, I used Plink v.1.90 [228] to remove variants in high LD (r^2>0.95), leaving 523,028 SNPs for mapping QTLs and eQTLs.

2.5.2 Supplementary Figures

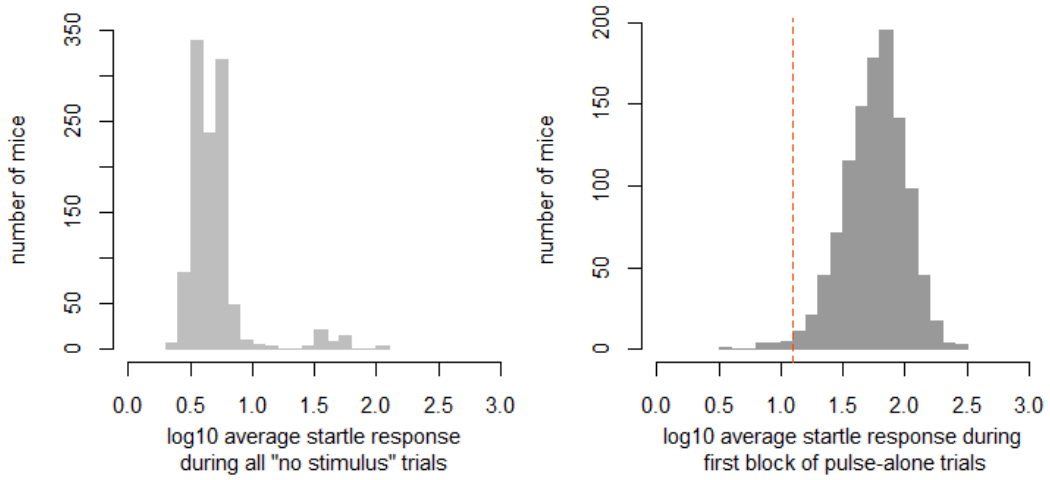

Figure 2.7: **Identification of outliers for startle and PPI. (a) Distribution of the mean response measured across eight no stimulus trials, when no startle pulse was sounded**. The upper tail shows 44 mice, all of whom were tested in box 3, that appear to startle in the absence of a pulse. I interpreted this as a technical effect and included box 3 as a covariate for all PPI and startle traits. **(b) Distribution of the mean startle response during the first block of pulse-alone trials**. Mice falling within the tail of the startle response distribution are not responding to the startle cues, possibly due to a hearing impairment. I excluded PPI and startle measures for 13 mice whose mean response overlapped the distribution of no-stimulus trials (not including mice from box 3). The cutoff point of 1.1 is marked with a dashed line. Each panel includes data for 1,123 phenotyped mice.

2.5.3 Supplementary Tables

Table 2.1: **GBS genotype concordance.** (The supplementary file associated with this dissertation contains concordance rates between the GigaMUGA genotyping array and GBS genotype calls (unfiltered and filtered for MAF, DR2, HWE, and LD.

Table 2.2: **Trait heritabilities, sample sizes, and covariates used for GWAS.** (The supplementary file associated with this dissertation contains trait descriptions, mean and standard deviations for the raw phenotype data, the sample size and covariates used for GWAS, and heritability estimates with standard error from GEMMA.

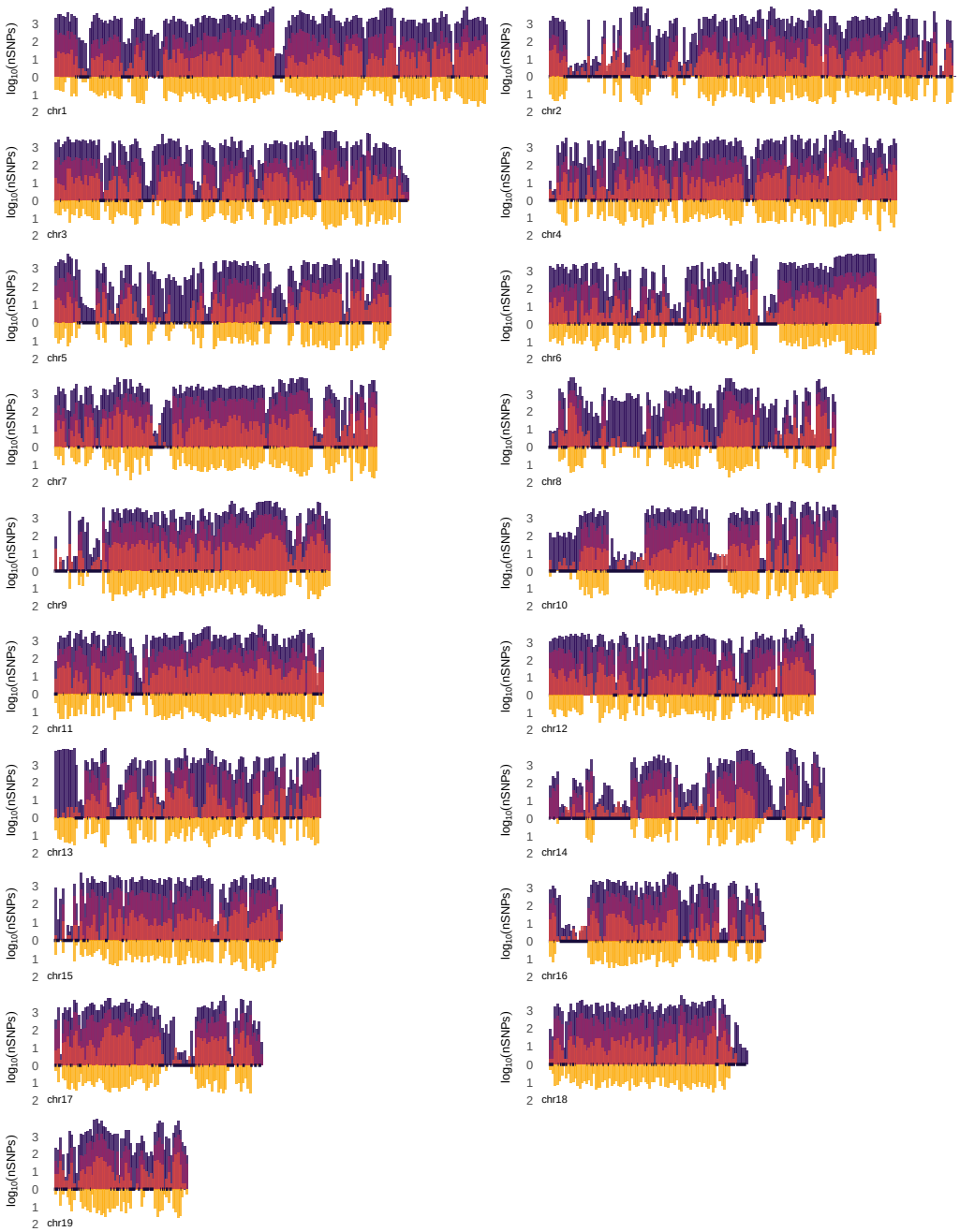

Figure 2.8: **SNP density.**. A detailed view of SNPs on chromosome 1 shows the log_{10}-scaled density of known SNPs in LG and SM from the reference panel (dark purple, top background), imputed and quality-filtered GWAS SNPs (medium violet, top middle), GBS SNPs before reference panel imputation (orange, top foreground), and GigaMUGA SNPs that were polymorphic in LG and SM (gold, bottom). At the x-axis, predicted IBD regions in LG and SM [193] are marked in navy.

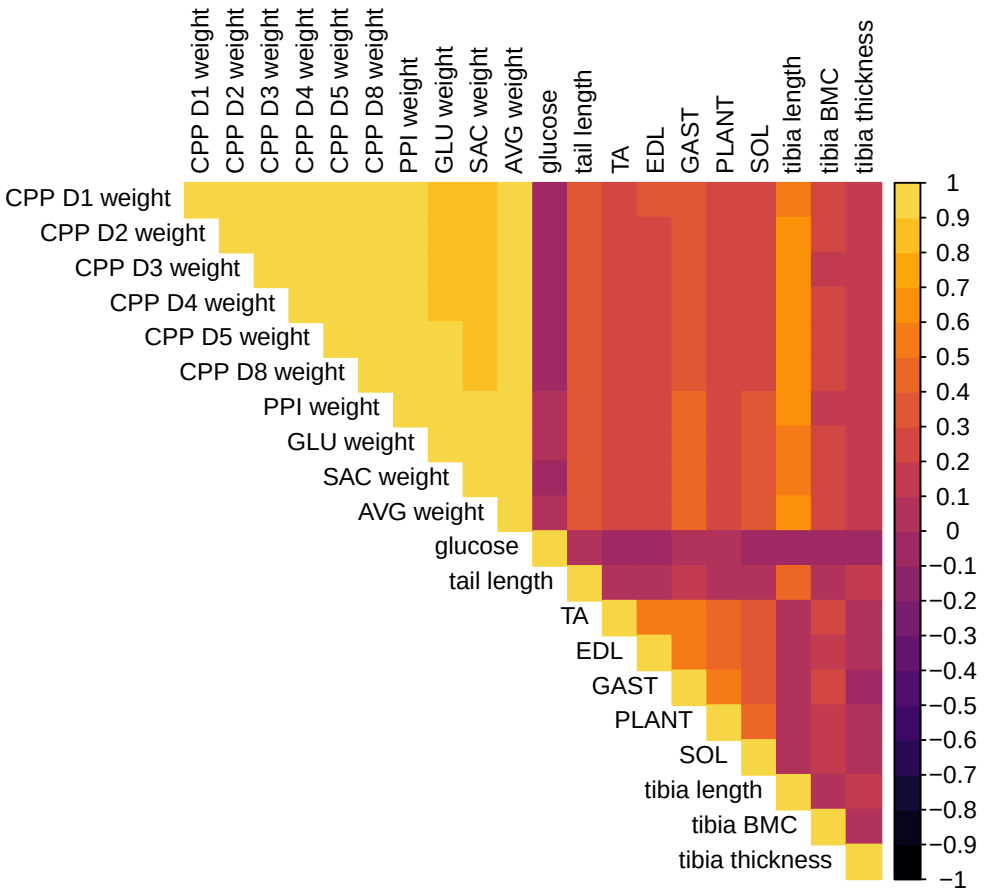

Figure 2.9: **Heat map of correlations among body weight, muscle, bone and glucose traits.**. Pearson's r^2 for the quantile-normalized residuals (with covariate effects removed) of physiological measurements taken in G50-56 AIL mice are depicted in a heat map. AVG weight is average weight. CPP, PPI, GLU weights are weights for days when conditioned place preference, prepulse inhibition, or glucose levels were tested. AVG is an average of CPP D1, CPP D8, and the other body weight traits (i.e. weights measured one week apart). BMC is bone mineral content. TA is tibialis anterior, EDL is extensor digitorum longus, GAST is gastrocnemius, PLANT is plantaris, and SOL is soleus muscle.

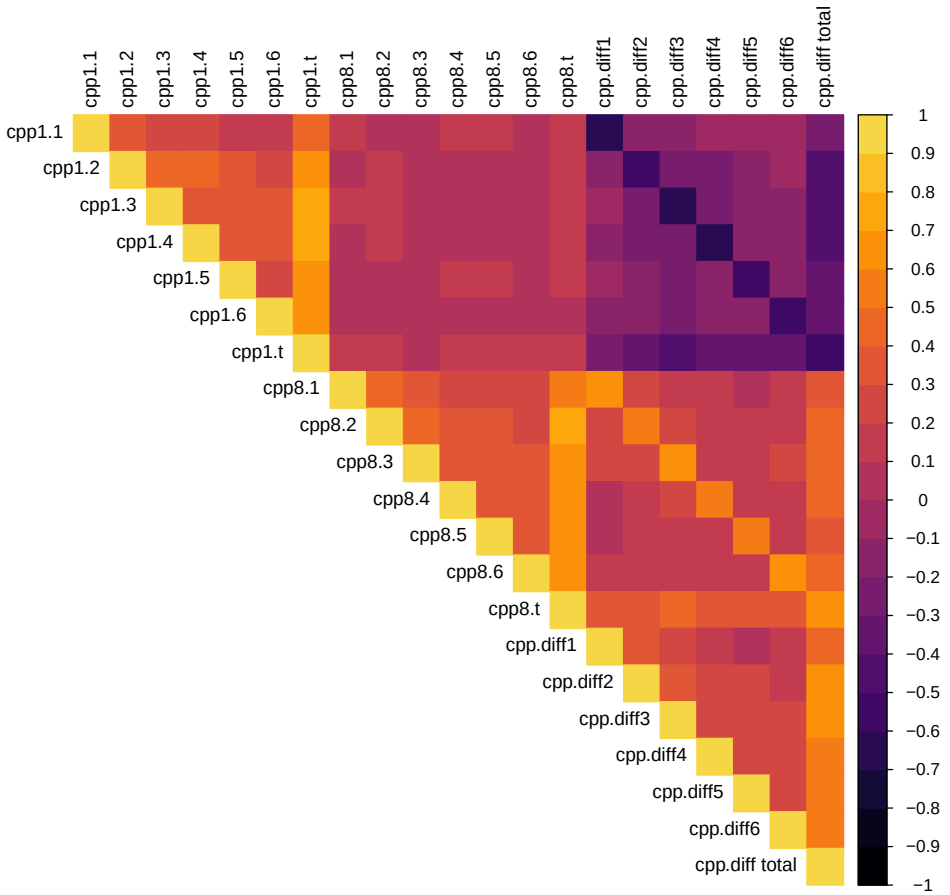

Figure 2.10: **Heat map of correlations among CPP traits.**. Pearson's r^2 for the quantile-normalized residuals (with covariate effects removed) of CPP measurements taken in G50-56 AIL mice are depicted in a heat map. cpp.diff is the difference between CPP on D8 and D1. Labels include the testing day and the time bin. For example, cpp1.1 refers to CPP on D1, bin 1 (0-5 min). Full trait names are provided in Supplementary Table 2.2.

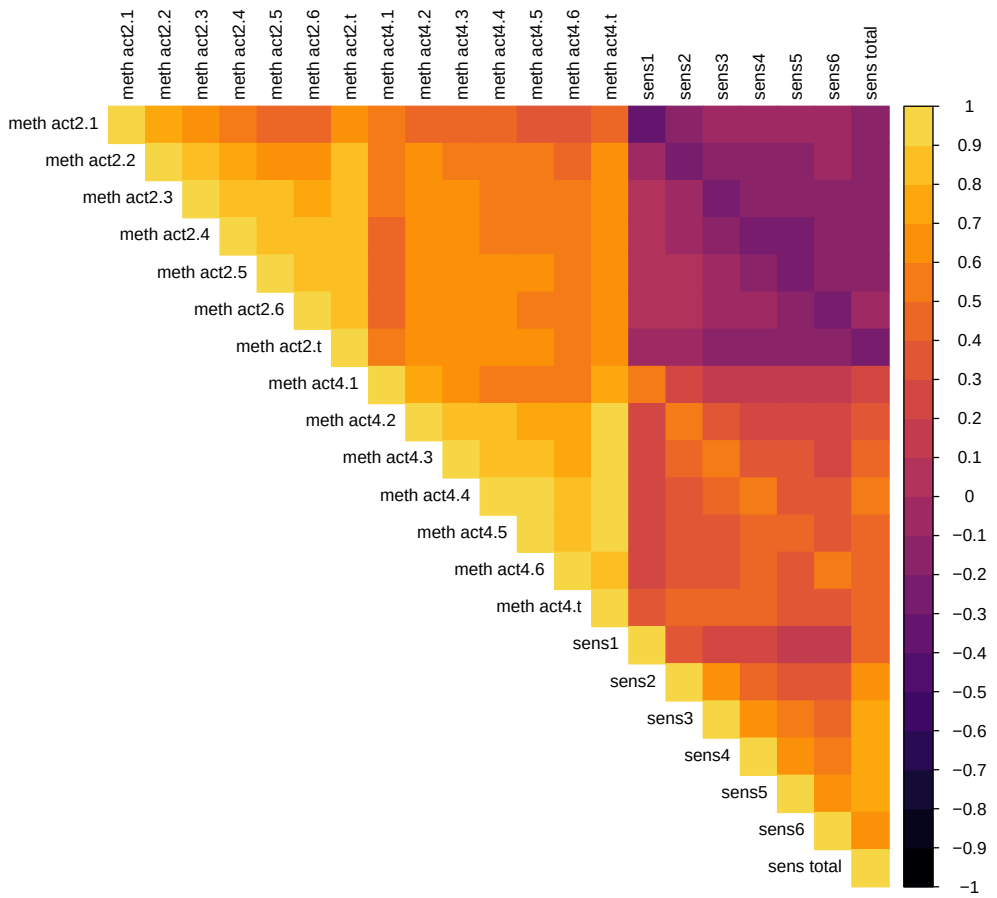

Figure 2.11: **Heat map of correlations among methamphetamine activity traits.**. Pearson's r^2 for the quantile-normalized residuals (with covariate effects removed) of methamphetamine locomotor response measurements taken in G50-56 AIL mice are depicted in a heat map. sens refers to sensitization, which is the difference in activity on D4 and D2. Full trait names are provided in Supplementary Table 2.2.

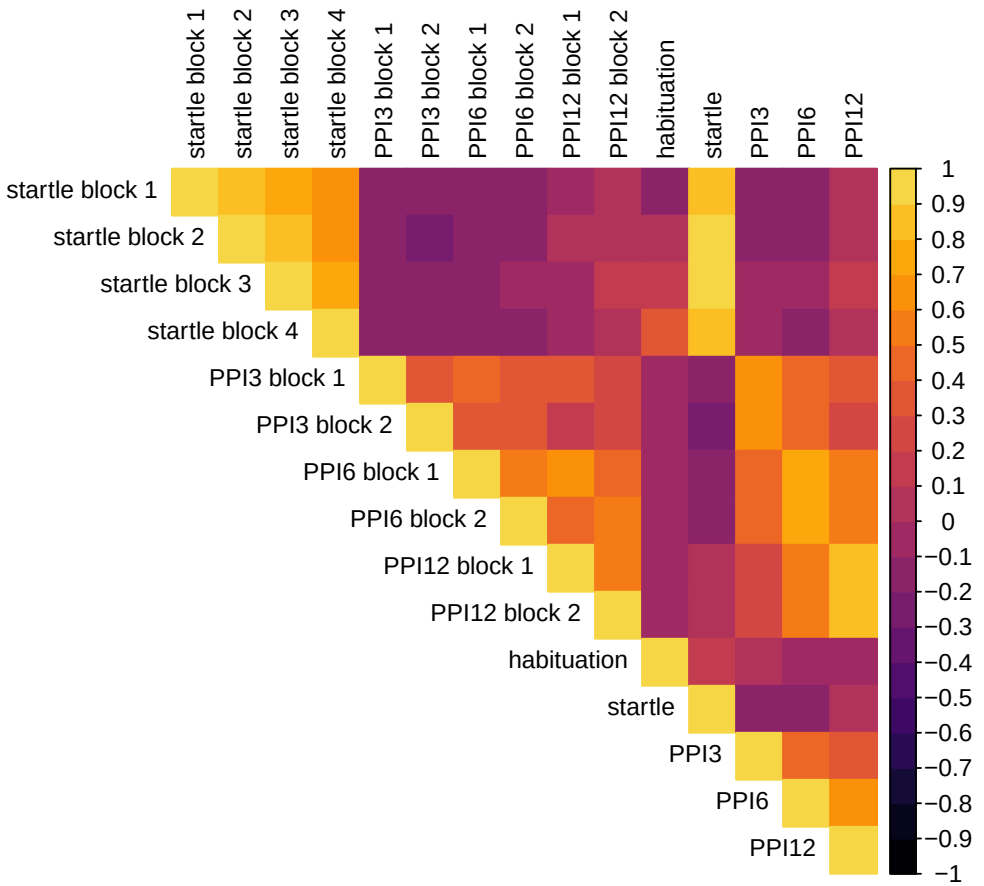

Figure 2.12: **Heat map of correlations among PPI and startle traits.**. Pearson's r^2 for the quantile-normalized residuals (with covariate effects removed) of PPI and startle response measurements, including habituation, taken in G50-56 AIL mice are depicted in a heat map. PPI3, PPI6, and PPI12 refer to prepulse inhibition after a 73, 76, or 82 dB prepulse. Full trait names are provided in Supplementary Table 2.2.

Table 2.3: **Summary of QTLs.** (The supplementary file associated with this dissertation contains information on all QTLs identified in this study, including: p-values, 1.5-LOD intervals, QTL width, chromosome, position, rsid, proportion of variance explained (effect size), MAF, LG and SM alleles, reference and alternative alleles, variant annotations for the top SNP, proportion and number of LG, SM, and heterozygous genotypes, raw and quantile normalized phenotype means (and their standard deviations) stratified by genotype at the top SNP, number of QTL genes, number of eQTL genes, and genes with deleterious mutations (missense, stop-gain, stop-loss) as annotated by dbSNP v.142.

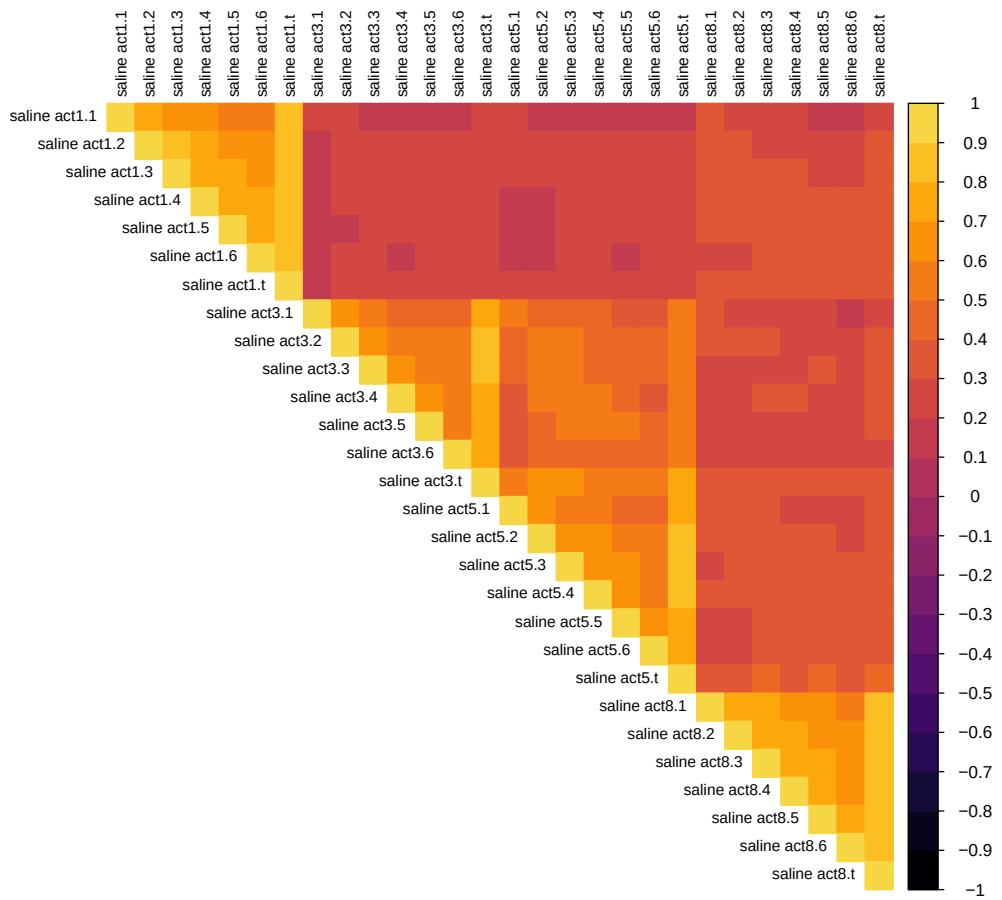

Figure 2.13: **Heat map of correlations among saline activity traits.**. Pearson's r^2 for the quantile-normalized residuals (with covariate effects removed) of saline activity measurements taken in G50-56 AIL mice are depicted in a heat map. Full trait names are provided in Supplementary Table 2.2.

Figure 2.14: **Manhattan plots.**.(The supplementary file associated with this dissertation contains Manhattan plots for each of the 123 traits mapped in this study. Trait heritability and standard deviation output from GEMMA is listed at the top of the plot. SNP rsid numbers are depicted below each genome-wide significant QTL. The dotted line represents a genome-wide significance level of $p = 8.06 \times 10^{-6}$, based on permutation with SLIDE.

CHAPTER 3

INTEGRATIVE ANALYSIS OF GENOTYPES, PHENOTYPES AND GENE EXPRESSION DATA

3.1 Introduction

In Chapter 2, I described properties of the genetic landscape in the LG × SM AIL and summarized QTLs identified for each set of phenotypes. I discussed overlap of QTLs for related measurements as well as replication of QTLs identified in LG × SM G34. However, I avoided discussing specific genes. Here, I examine a subset of QTLs in greater detail, integrating the results with convergent eQTLs and functional information about their target genes, which I assembled from online databases. This approach provided context for interpreting QTLs and guided my search for promising candidate genes. I also summarize *cis*- and *trans*-eQTL results for each tissue. My work demonstrates that an integrative analysis of genotypes and gene expression data in the mouse is a suitable approach for studying complex human traits, particularly when the underlying mechanisms are unknown.

Gene expression is considered to be an intermediate phenotype on the pathway connecting genetic variation to behavior. Therefore, I sequenced brain-derived mRNA obtained from a subset of mice that were phenotyped and genotyped and used their gene expression data to map eQTLs. I then integrated QTLs with eQTLs identified in hippocampus (**HIP**), prefrontal cortex (**PFC**) and striatum (**STR**), using the transcriptional context within each region to guide selection of candidate genes. Here, I describe my rationale for integrating QTLs with gene expression data and explain why I chose to study HIP, PFC and STR. I will also discuss findings from comparative genomic studies in humans and mice, which suggest that in general, gene expression patterns (and to some extent, regulatory mechanisms) in the two organisms are highly conserved.

3.1.1 Rationale for an integrative analysis of QTLs and eQTLs

Integrating QTLs with eQTLs mapped in disease-relevant tissues has become a standard approach for prioritizing candidate genes [229]. The main benefit of integrating QTLs with gene expression is that it allows for interpretation of QTLs within a biological context, which can generate specific, mechanistic hypotheses about the trait of interest [88, 230, 231]. For example, overlap between a QTL and an eQTL might suggest that the QTL affects the associated trait by modifying expression of the eQTL gene; in this sense, gene expression can be considered an intermediate phenotype on the pathway from polymorphism to organismal phenotype. Another benefit of an integrative approach is that it improves power by providing additional support for associations, which can be useful for controlling the false negative error rate [26].

eQTLs are also interesting in their own right because they can provide insight into global properties of gene regulation [229,232]. *cis*-eQTLs, which are usually located near the target gene, are allele-specific, whereas *trans*-eQTLs, which usually affect the expression of distant genes, are not. There is evidence that many *cis*-eQTLs are widely shared across different tissues [233–236], suggesting that in some cases, it will be possible to detect gene expression differences even if their functional consequences occur in other structures. In contrast, *trans*-eQTLs have been difficult to identify because they are more subtle in effect than *cis*-eQTLs, they are often confounded with technical variables, they are more likely to be tissue-specific, and they incur a greater penalty for multiple testing [229,232]. Consequently, there have been fewer attempts to map them, and most results have failed to replicate across studies [232]. A recent study in humans has suggested that *trans*-eQTL hotspots, which regulate the expression of many genes, may be more robust than *trans*-eQTLs with a single target [237], but it is not yet clear if this will generalize to other organisms and tissues. Unlike in humans, *trans*-eQTL hotspots identified in model organisms have frequently been associated with the expression of hundreds of different genes [238–241], which begs the question of whether

the genetic architecture of *trans*-regulation in humans is truly distinct from that of other organisms, or merely an artifact of low power [242].

The questions that motivated this study have less to do with interspecies gene regulation than with a desire to understand behavior. However, an additional outcome of this work will be to build upon existing eQTL maps [76] in the mammalian brain, which can provide support for QTLs and eQTLs identified in other studies and build on our current understanding of the genetic architecture of gene expression.

3.1.2 Relevance of HIP, PFC and STR to behavior and disease

I measured various metabolic traits as part of this study, but my primary interest is in behavior. Therefore, I mapped eQTLs in three brain regions selected for their relevance to the behaviors of interest (Chapter 2). HIP has an established role in regulating attention and memory [243–246]; there is also evidence for its involvement in spatial learning [218], cognitive flexibility, and mood [97, 244]. PFC is known for its role in behavioral inhibition, decision-making and motivational behavior [247, 248]; it also interacts with HIP to regulate memory [245, 246] and cognition [249]. As a critical component of reward and reinforcement learning, striatal functions have been studied extensively in the context of drug abuse [94, 97, 105, 250, 251]. STR also plays a central role in motor control and goal-directed behavior [94, 250, 252]. Clearly, each region regulates a number of important behavioral processes, and much of their activity is coordinated by a common circuitry. Accordingly, PFC, HIP, and STR are considered important for a variety of psychiatric and behavioral phenotypes, including Parkinsons disease [252], impulsivity [97], sensorimotor gating [46, 253, 254], and drug abuse [94].

By selecting three specific regions, I was unable to identify eQTLs specific to other tissues or developmental time points, but this limitation applies to all eQTL studies. Furthermore, a survey of gene expression throughout development was not my goal; if I had selected an

earlier time point to collect brain tissue, it would have been impossible to measure all of the phenotypes that I did (because the mice would be dead). Since it was feasible for a single experimenter to dissect all of the brains for this study within minutes after killing each mouse, I was able to collect nearly 300 brains, which made it reasonable to exclude lower-quality samples at various stages of processing (an unfortunate reality of high-throughput experiments) and yet still have a large sample for analysis. Measuring gene expression from brains of the same mice that were phenotyped and genotyped is a major advantage of this design; I argue that this outweighs the benefits of a more comprehensive survey, which would require substantially more effort for a reward whose magnitude is difficult to foresee. Moreover, this approach would have been impossible to do in humans for obvious ethical reasons, and also because in most cases, human brain tissue must be collected opportunistically or *post-mortem*.

Several large efforts are underway that aim to provide the human genetics community with a catalog of gene expression, methylation, regulatory annotations and other functional genomic data across multiple tissues and cell types [235, 255, 256]. However, the ability to measure gene expression from individuals phenotyped and genotyped in GWAS will likely remain impossible for tissues like the brain. Studies in rodents and other model systems are therefore essential for exploring the genomic environments that lead to disease vulnerability, particularly in tissues that are difficult to collect from humans.

3.1.3 Global patterns of gene expression appear highly conserved in humans and mice

An assumption of using rodents to model human disease is that even if specific findings do not generalize from model to human, the pathways that contribute to the phenotype will be shared [229]. Genetic studies have found substantial overlap in the factors that contribute to complex traits in rodents and humans [80,257,258], and rodents have frequently

been helpful for interpreting findings from human GWAS [85, 88, 259, 260], suggesting that this assumption is reasonable. Due to its advanced genetic toolkit, the mouse is a logical starting point for exploring how gene regulatory networks and neural circuits function to regulate behavior [59, 259–262] (although this may change as resources for the rat continue to develop [54, 263]).

This raises an important question: are the mouse and human transcriptomes similar enough to make meaningful comparisons? A tremendous amount of work has been devoted to addressing this concern [55]. For now, the answer appears to be yes, but with some important caveats. Essentially, the ability to identify orthologous transcriptional patterns, cell types, and organismal phenotypes across species is confounded by extreme biological complexity [11, 55]. Straightforward comparisons across species, tissues and cells are complicated by differences in the technology used to generate the data and various aspects of experimental design and statistical analysis [264–268]. Biological variables, such as RNA integrity [269], menstrual/estrus cycle or circadian rhythm [270], may also exaggerate molecular differences between species.

Nonetheless, considerable similarity between humans and mice is apparent at the molecular level. For example, human and mouse genomes are comparable in terms of size, nucleotide composition and structure [271–273]. Over 70% of mouse genes have one-to-one orthologs in humans, and 80% of human genes have one-to-one orthologs in the mouse [274]. Mean levels of gene expression in humans and mice tend to be positively correlated, even when comparing cell lines to primary tissue [275]. This seems to extend to the brain as well, despite significant expansion of the human neocortex. A single-cell transcriptome analysis of the developing midbrain revealed differences in developmental timing, neuron development and cell proliferation between human and mouse, but observed high conservation of gene expression and cell types overall [276]. The post-synaptic proteomes of mice and humans have a similar composition, although some protein families vary in abundance between the

two organisms [277]. Comparison of DNA methylation signatures in the mouse and human cortex recently revealed neuron cell-type specific regulatory differences between species, but global patterns suggest high conservation of differentially methylated sequences in neuronal subtypes of human and mouse [278]. Moreover, there is a general consensus that variation in gene expression differs more between tissues than between species [55, 264, 279] (encouragingly, one study observed that even disease-associated genes are more likely to have variable expression across tissues than across species [280]). Finally, chromatin state [281, 282] and gene regulation [18, 283, 284] appear to be conserved despite some interspecies differences in regulatory sequences [17, 285].

Taken together, this evidence suggests that with careful consideration of functional and structural conservation between the two organisms (e.g. among tissues, transcripts and pathways relevant to disease), studying gene expression in the mouse is a reasonable strategy for generating hypotheses about human biology. Trait-associated eQTLs identified in the mouse may benefit human GWAS by providing strong prior hypotheses for association with an analogous trait; quite possibly, the additional evidence would make it reasonable to accept associations that are under the threshold for genome-wide significance. Similarly, candidate genes identified in mice might implicate their human orthologs as promising candidates for targeted sequencing, which could be used to identify rare variants with functional relevance to the trait of interest.

3.2 Methods

3.2.1 RNA sequencing

I extracted mRNA from 741 tissue samples (280 HIP, 254 STR and 255 PFC) as described in Parker *et al.* (ref. [62]) and prepared cDNA libraries from 265 HIP, 236 STR and 240 PFC samples that had RNA integrity scores ≥ 8.0. Barcoded sequencing libraries were prepared

with the Illumina TruSeq RNA Kit, pooled in sets of 24, and sequenced on two lanes of an Illumina HiSeq 2500 machine using 100 bp, single-end reads.

Because mapping quality tends to be higher for reads that closely match the reference genome [286], read mapping in an AIL may be biased if one founder is more similar to the reference strain [287]. I addressed this concern by aligning RNA sequencing reads to custom genomes created from LG and SM genomes [193]. I used HISAT for alignment [288], assigning each read to a gene as defined by Ensembl [289]. Mapping position and CIGAR strings for each read were then reassigned to match mm10 reference genome coordinates. I combined the two alignment files for each sample by choosing the best mapping for each. Only uniquely mapped reads were included in the final alignment file for each sample.

I used the Bioconductor package GenomicAlignments for assembly [290] followed by DESeq to obtain normalized read counts for each gene [291]. I excluded genes detected in less than 95% of samples within each tissue. I also excluded 10 HIP, 16 STR and 4 PFC samples that had less than five million mapped reads or an alignment rate under 91.48%, retaining a total of 562 samples (208 HIP, 169 STR, and 185 PFC) for further analysis. Additional details describing RNA sequencing quality control are provided in the Supplementary Methods 3.5.

RNA sequencing alignment to custom genomes

I built custom genomes by incorporating known SNPs and indels from the LG and SM strains into the mm10 reference sequence [193]. I aligned RNA sequencing reads to custom LG and SM genomes using the default parameters in HISAT [288].

I then used the Bioconductor package GenomicAlignments for genome assembly (Lawrence et al. 2013). Each RNA sequencing read was assigned to a gene as defined by Ensembl (GRCm38 v.85) [289]. I required that each read overlap one unique disjoint region of the gene. If a read contained a region overlapping multiple genes, the genes were split into dis-

joint intervals, and any shared regions between them were hidden. If the read overlapped one of the remaining intervals, it was assigned to the gene that the interval originated from; otherwise, it was discarded.

Next, I quantile normalized gene expression data and used PCA to remove the effects of unknown confounding variables [14]. For each tissue, I calculated the first 100 principal components (**PCs**) of the $k \times n$ matrix of expression values, where k is the number of genes and n is the number of samples. I quantile normalized PCs and tested them for association with AIL SNPs using the same approach as described for QTL mapping. I specified two covariates, sex and batch, for each expression trait. If a PC showed evidence of association with a SNP, I retained it to avoid removing *trans*-eQTL effects. I evaluated significance with the same permutation-based threshold used for QTLs ($p \leq 8.06 \times 10^{-6}$). I then used linear regression to remove the effects of the remaining PCs (71 in HIP, 81 in STR and 93 in PFC) and quantile normalized the residuals prior to mapping eQTLs.

3.2.2 cis-eQTLs

I used a LOCO-LMM to map eQTLs. I conservatively included sex and batch as covariates even though PC regression may account for them. SNPs within 1 Mb upstream or downstream of the gene were considered as potential *cis*-eQTLs. The mean number of SNPs per *cis* region was 541.8 (median=357, range=1-4,002). I excluded 2,143 genes because there were no SNPs within their *cis*-regions.

I used eigenMT to correct for multiple hypothesis testing for *cis*-eQTLs [292]. A Bonferronni test that corrects for all SNPs in each *cis* region is too conservative for an AIL because many SNPs are in high LD. eigenMT estimates the number of independent tests for each gene, accounting for LD among SNPs in *cis*. The effective number of SNPs is then used in a modified Bonferronni test to obtain an adjusted p-value for each gene. The mean number of effective tests per gene was 22.5 for HIP (median=7, range=1-837), 37.4 for STR

(median=8, range=1-1627) and 30.8 for PFC (median=7, range=1-1185).

3.2.3 trans-eQTLs

Only SNPs outside of the chromosome containing the gene being tested were considered as potential *trans*-eQTLs. This decision was motivated by recent studies in humans and mice which found that the majority of *trans*-eQTLs located on the same chromosome as the gene being tested were redundant with *cis*-effects [237, 241]. I reasoned that it I would not miss many *trans*-eQTLs by using this approach.

I determined a *trans*-eQTL significance threshold for each tissue by permuting data for a randomly chosen gene 1,000 times. Since expression data was quantile normalized, I only permuted one randomly chosen gene per tissue. I did not remove *cis* SNPs before permuting the data, so I expect the thresholds to be slightly conservative for some genes. At a genome-wide significance level of 0.05, the permutation-based threshold for *trans*-eQTLs was $p = 8.68 \times 10^{-6}$ in STR, $p = 9.01 \times 10^{-6}$ in HIP, and $p = 1.04 \times 10^{-5}$ in PFC.

I define a *trans*-eQTL hotspot as a *trans*-eQTL that regulates the expression of 10 or more genes. To identify master regulators, I divided chromosomes into 5 Mb bins and assigned each *trans*-eQTL target gene to the bin containing its top eQTL SNP. I used bins because I expected SNPs in high LD to capture the same signal. I also repeated the procedure using larger windows to account for eQTL regions that spanned two adjacent bins. I then counted the number of genes regulated by SNPs in each bin.

3.2.4 Integration of QTLs and eQTLs

I integrated QTLs with brain eQTLs, coding annotations from dbSNP (v.142), differentially expressed genes identified in LG and SM muscle [181, 293], and online databases to identify candidate genes and pathways for further investigation. Throughout the text, I refer to *cis*-eQTL and *trans*-eQTL target genes as *cis*-eGenes and *trans*-eGenes. For a *cis*-eGene to

overlap a QTL, I required at least one SNP within its *cis* region to fall inside the 1.5-LOD interval for the QTL. In contrast, QTL genes are defined by their transcription start and end points. Since the *cis* region extends 1 Mb in either direction of the gene, it is possible for a region to contain more eGenes than QTL genes.

3.2.5 URLs

I used the Mouse Genome Informatics Database (**MGI**; `http://www.informatics.jax.org/`) to query genes for association with mouse phenotypes, OMIM disease associations, and transgenic organisms [40]. I queried the GWAS Catalog (`https://www.ebi.ac.uk/gwas/`) [294] and Gene2Function database (`http://www.gene2function.org/search/`) to find overlap between genes implicated by GWAS (e.g. for obesity) and QTL genes identified in this study. Finally, I used Enrichr (`http://amp.pharm.mssm.edu/Enrichr`) to explore annotations for genes targeted by the same *trans*-eQTL regions [295, 296].

3.3 Results

3.3.1 cis and trans-eQTLs

Thousands of eQTLs identified in the brain

For a subset of phenotyped and genotyped mice, I used RNA sequencing to measure gene expression in HIP (n=208), PFC (n=185) and STR (n=169) (Supplementary Figure 3.15). I identified *cis*-eQTLs for 2,902 genes in HIP, 2,125 genes in PFC, and 2,054 genes in STR (FDR=0.05); 1,087 genes had *cis*-eQTLs in all three tissues (Supplementary Figure 3.15). I also mapped 515 *trans*-eQTLs in HIP, 413 *trans*-eQTLs in PFC, and 444 *trans*-eQTLs in STR (FDR=0.05); see Supplementary Figure 3.15 and Figure 3.1). 18.5% of genes with significant *trans*-eQTLs also had a significant *cis*-eQTL.

In general, *trans*-eGenes shared between tissues were regulated by eQTLs on different chromosomes. Notably, 389 out of all 1,476 *trans*-eQTLs in HIP, PFC and STR had only one SNP passing the threshold for genome-wide significance (26.4%); the remaining *trans*-eGenes were associated with at least 14 significant SNPs. This could be due to tissue-specific regulation or a lack of power to detect *trans*-eQTLs.

trans-eQTL hotspots

I defined *trans*-eQTL hotspots as 5 Mb regions containing the top *trans*-eQTL SNP for 10 or more genes. I identified four *trans*-eQTL hotspots in HIP, two in PFC, and seven in STR (Figure 3.1; Supplementary Figures 3.16, 3.17, 3.18). Extending *trans* bins from 5 Mb to 10 Mb produced two additional regions in PFC and one in STR (Supplementary Table 3.1). Despite considerable overlap among genes measured in HIP, PFC and STR, I identified only four *trans*-eGenes in all three tissues (*Sfi1, Eif3j2, Col6a5*, and *Aars*), none of which had *cis*-eQTLs. PFC and STR shared the most *trans*-eGenes (32), 19 of which were regulated by the same region.

Previous studies of *trans*-eQTLs in model organisms have identified *trans*-eQTL hotspots that regulate the expression of hundreds of genes [229, 238–241, 297]. I found one *trans* hotspot on chromosome 12 between ~70.75-73.75 Mb that was associated with 85 target genes in HIP (Supplementary Table 3.1). The hotspot contains 44 genes and transcripts, 19 of which have *cis*-eQTLs in HIP.

The presence of a regulatory hotspot is intriguing, but its underlying mechanisms are unclear. One or more *cis*-eQTLs within the hotspot could control the expression of genes encoding transcription factors or other downstream effectors that regulate the distal targets. However, LD can cause nearby *cis*-eQTLs to have similar expression patterns, which makes it difficult to tell whether the *trans*-effect is mediated by *cis*-eQTLs within the hotspot or some other mechanism. Similarly, a hotspot might appear to regulate a large number of

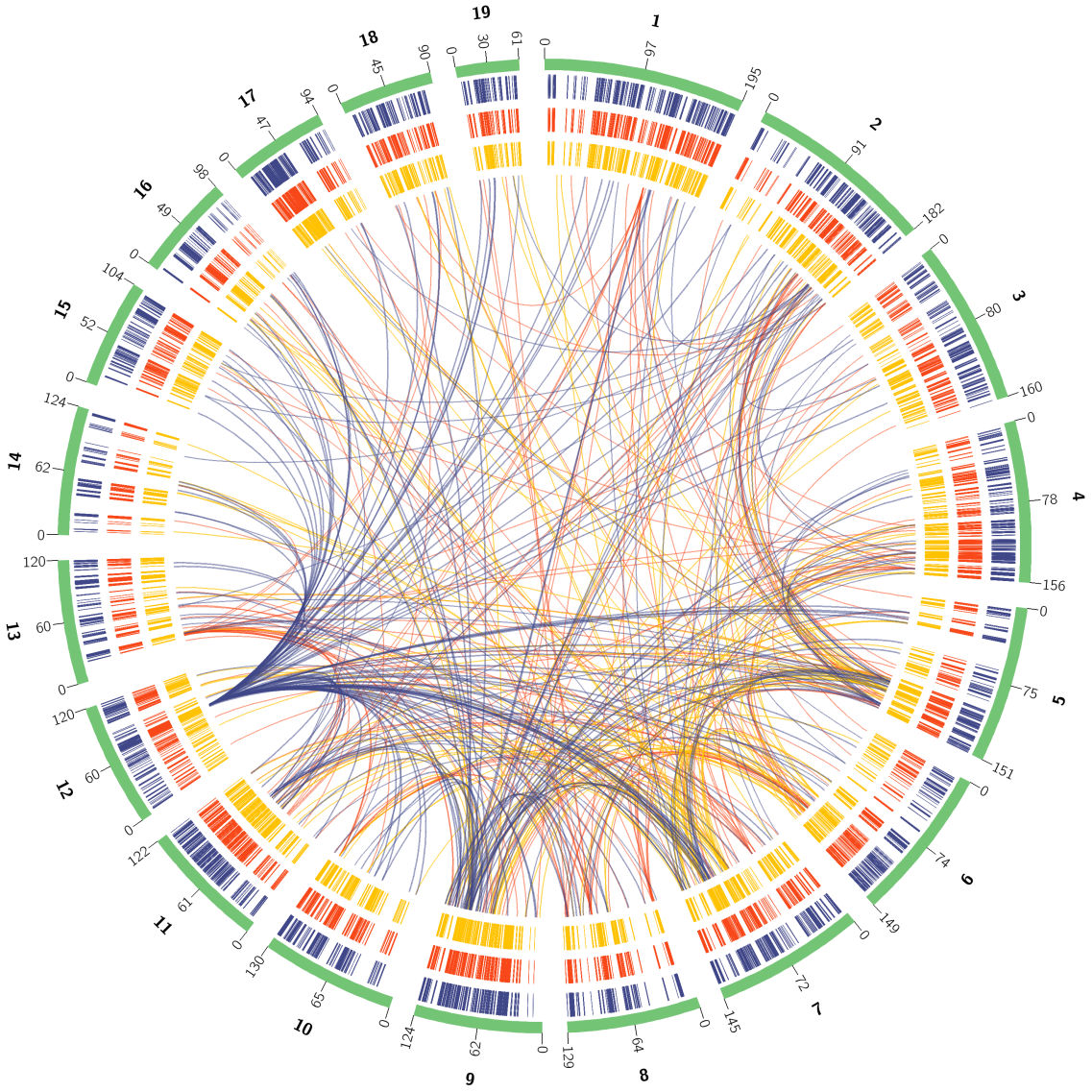

Figure 3.1: ***cis*-eQTLs and *trans*-eQTL hotspots in HIP, PFC and STR**. I mapped over 2,000 *cis*-eQTLs and over 400 *trans*-eQTLs in each brain region (FDR=0.05). The 19 autosomes are delineated by green bars on the outer edge of the plot. Each ring inside the circle shows locations of *cis*-eQTLs for HIP (blue, outer ring), PFC (gold, inner ring), and STR (orange, middle ring). For clarity, only *trans*-eQTLs that regulate 10 or more genes are shown in the inner circle. Individual Circos plots showing the full set of *trans*-eQTLs identified in each tissue are provided in the Supplementary Figures 3.16, 3.17, 3.18.

genes if its targets are in high LD (e.g. if they share a *cis*-eQTL). If this were true, then it could explain why *trans* hotspots that regulate large numbers of genes have frequently been found in model organisms but only occasionally identified in humans [229], since model organisms tend to have longer haplotypes. If *trans* associations tend to be mediated by shared *cis*-eQTLs at target loci, this could also explain why most *trans* hotspots have failed to replicate across populations [229].

The target genes of this particular hotspot were widely distributed across 17 chromosomes, and only seven of the 85 genes also had a *cis*-eQTL (Supplementary Table 3.1). There were six chromosomes that each contained a pair of target genes located within 1-2 Mb of one another, but most targets were widely distributed throughout individual chromosomes. Therefore, it seemed unlikely that multiple *cis*-eQTLs in the target regions were responsible for the *trans* hotspot. I reasoned that the target loci might share a common regulatory domain, such as a transcription factor binding site, and used Enrichr [295, 296] to scan multiple databases for shared microRNA and transcription factor binding motifs among the target genes. I did not find evidence of enrichment, but noted that several target genes have roles in transcriptional activation (*Yap1, Notch2, Sfrp1, Hfe, Igf2, Tcf20, Mitf, Clu* and *Ywah*). There is a chance that they might also contribute to the *trans* effect, but I did not explore this possibility.

Instead, I tested for over-representation of shared phenotypes and tissue expression patterns within the target gene set. I used Enrichr to query the mouse and human phenotype ontology databases and gene expression atlases for the two organisms. Two genes related to ataxia (neurologically-mediated abnormalities in muscle coordination) in humans, *Cacna1a* and *Sptbn2* were also related to ataxia in mutant mice, along with six other genes (*Gnaq, Plxna2, Tpp1, Bsn, Slc4a2*, and *Slc12a2*). Several genes were expressed in human retina and mouse retinal pigment epithelium, and four were implicated in human retinal disease (*Mfrp, Abca4, Slc7a14* and *Myo7a*). I considered that the implicated genes could be part

of a common pathway or process affected by the trans mechanism, and queried the mouse phenotype ontology database to identify genes within the chromosome 12 hotspot that have been associated with ataxia or ocular phenotypes in mutant mice. I also scanned for hotspot genes with known regulatory roles.

No hotspot genes were associated specifically with ataxia, but *Six1*, a transcription factor, was associated with several muscle abnormalities, and a related gene, *Six6* was associated with abnormal retinal morphology and retinal hypoplasia in knockout mice [298]. *Six1, Six6* and *Six4*, a third transcription factor in the hotspot region, are known to interact with the transcriptional coactivator *Eya3* [299]. *Arid4a* encodes a DNA-binding protein that interacts with *pRB* (retinoblastoma protein). *ARID4A* and *pRB* repress transcription by recruiting histone deacetylases [300]. *ARID4A* also regulates imprinted regions by modifying DNA and histone methylation marks [301]. *Hif1a*, a master regulator of transcription, encodes the alpha subunit of the transcription factor *HIF-1*. Conditional knockout of *Hif1a* in the mouse brain resulted in loss of hippocampal neurons, impaired spatial memory, and abnormal brain development [302]. *Mnat1* regulates RNA transcription by activating RNA polymerase II [303], and *Ppm1a* is a phosphatase that can suppress the activity of *SMAD2/3*, which regulate transcription [304]. This evidence implies that the chromosome 12 region is truly a regulatory hotspot and likely has multiple downstream effects. However, more work is needed to identify the specific effects of these transcription factors (and potentially, other hotspot genes) on their targets, and importantly, to understand what significance these genes have in HIP.

3.3.2 Integration of eQTLs and behavioral QTLs

A locomotor response QTL on chromosome 4 confirms a role for *Oprd1* in regulating baseline activity.

Multiple saline activity traits from D5 and D8 mapped to a region on chromosome 4 between 130.96-133.74 Mb (Supplementary Table 2.3). In each case, the SM allele was associated with increased activity. The strongest QTL was for total activity on D5 of CPP and contained 45 genes, 25 of which had *cis*-eQTLs ($p = 6.75 \times 10^{-9}$; MAF=0.24; Figure 3.2). Because it completely encompassed the narrowest QTL (D5 activity, 25-30 min; $p = 4.66 \times 10^{-8}$; MAF=0.24), which contained only one gene and 11 *cis*-eGenes (including *A930031H19Rik*, a lncRNA), I focused on the smaller region to prioritize candidates.

One gene, *Oprd1* (opioid receptor delta 1), had a *cis*-eQTL in the three brain regions that I analyzed. The SNP most strongly associated with locomotor activity was also associated with decreased expression of *Oprd1*. Consistent with the hypothesis that differential expression of *Oprd1* causes the behavioral QTLs, this gene has been shown to increase locomotor activity in knockout mice [305] by modulating neurotransmission in HIP and STR [217].

Despite this compelling evidence, I can not exclude the possibility that multiple genes are involved in this QTL. For example, another *cis*-eGene in the interval, *Trnau1ap*, contains a missense mutation in the SM background and is important for selenocystine production. Selenoprotein deficiencies have been linked to motor deficits and neuronal loss in STR [306], and conditional knockout of *Trnau1ap* in mouse neurons results in impaired motor coordination [307], suggesting that *Trnau1ap* might also modulate locomotor activity.

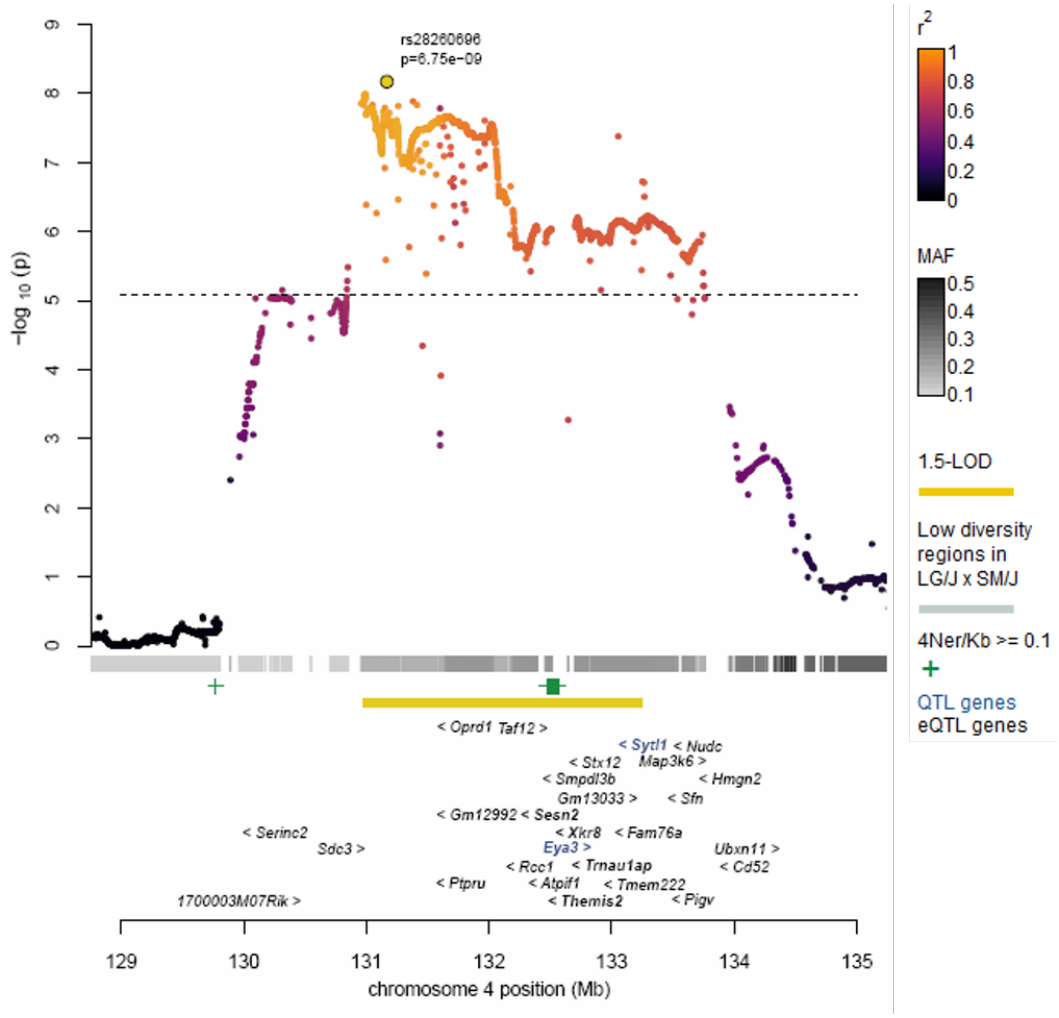

Figure 3.2: **QTL for locomotor activity (D5 saline, 0-30 min) on chromosome 4**. Points are colored by LD (r^2) with the top QTL SNP, which is highlighted in gold and labeled with its p-value and rsid. MAFs in the region are shown in the band below the plot. Regions predicted to have low genetic diversity among LG and SM are shown as grey bands. The 1.5-LOD interval for the QTL is shown as a gold band. Areas that have an elevated rate of recombination as estimated by Brunschwig *et al.* [308] are shown as green plus symbols. Finally, genes within the 1.5-LOD region are shown below the x-axis. For clarity, I only listed genes that have a *cis*-eQTL within the interval (black text) or genes that have missense mutations (bold). eGenes with a missense mutation that are *not* eGenes, if present, are listed in blue.

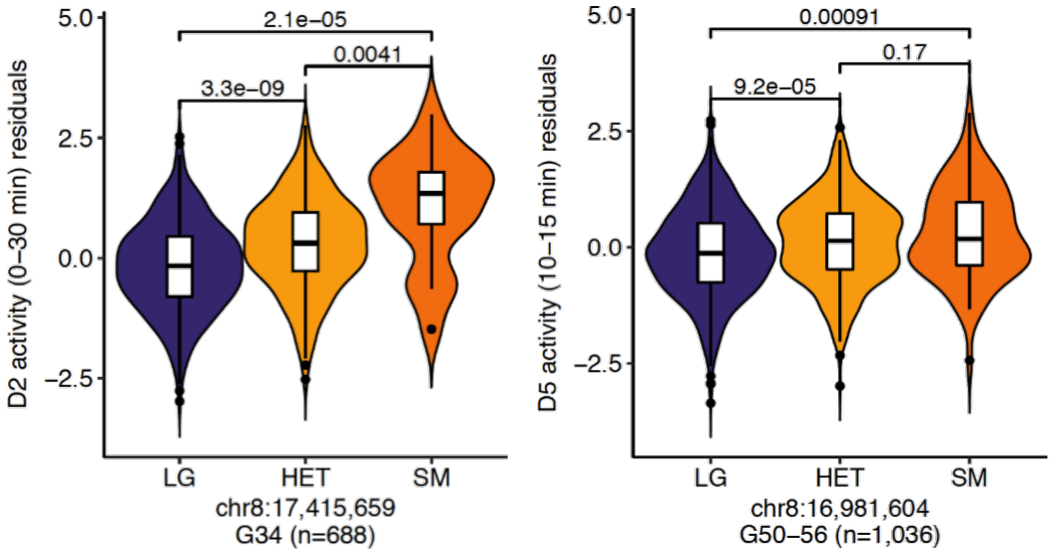

Figure 3.3: ***Csmd1* and in G34 and G50-56 of the LG × SM AIL**. The SM allele at the chromosome 8 activity QTLs containing *Csmd1* in AIL G34 and G50-56 confers an increase in saline activity. In G34 (left), there are 455 mice with the LG genotype (T/T), 215 heterozygotes (T/C), and 18 SM homozygotes (C/C). In G50-56, 570 mice are homozygous LG (T/T), 412 are heterozygous (T/G), and 81 are homozygous SM (G/G).

Replication of an association between *Csmd1* and locomotor activity identified in LG × SM G34

I replicated a saline activity QTL on chromosome 8 at 16.50-16.99 Mb ($p = 5.12 \times 10^{-6}$; MAF=0.24) that I had previously identified in a combined analysis of LG × SM AIL F2 and G34 mice [75]. In both studies, the region spanned ~0.5 Mb, the SM allele was associated with increased activity3.3, and the QTL contained only one gene, *Csmd1* (CUB and sushi multiple domains 1). Here, I also identified *Csmd1* as a *cis*-eGene in HIP3.5.

Human GWAS have implicated *Csmd1* in schizophrenia [309, 310], behavioral disinhibition [311], and cognition [312]. The Palmer Lab and others have used knockout mice to explore the effect of *Csmd1* on a variety of intermediate phenotypes for psychiatric disease [313, 314]. *Csmd1* is positioned at 15.89-17.53 Mb and contains 70 coding exons; however, it produces only four transcripts. As described by Distler *et al.* [313], knockout mice were created by disrupting the first exon of the gene, which results in a 70% decrease of *Csmd1*

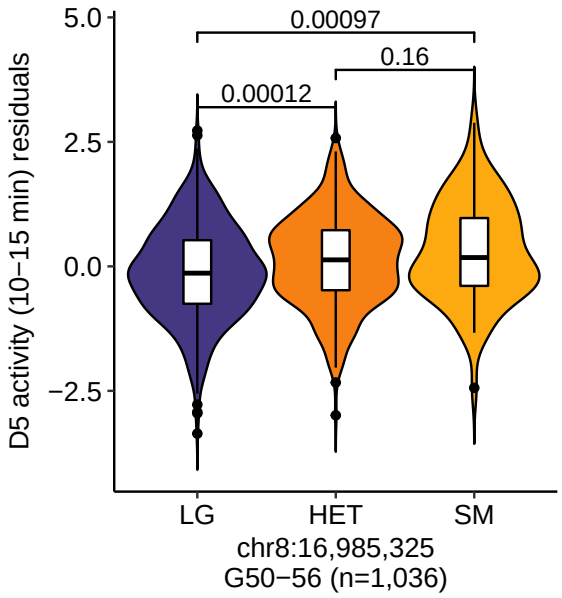

Figure 3.4: ***Csmd1* and in G34 and G50-56 of the LG × SM AIL**. The SM allele at the chromosome 8 HIP eQTL for *Csmd1* confers an increase in *Csmd1* expression. Of the 208 mice with expression data from HIP, 102 are homozygous LG (T/T), 85 are heterozygous (T/G), and 21 are homozygous SM (G/G).

expression in the brain. Steen *et al.* [314] observed that *Csmd1* mutants displayed abnormal anxiety- and depression-like behaviors, but did not observe differences in locomotor activity. Strikingly, the only behavioral difference observed among wild-type and *Csmd1* knockout mice in the Palmer Lab was in locomotor activity levels (unpublished data). Due to the different genetic backgrounds and behavioral tests used in the two studies, the specific role of *Csmd1* remains inconclusive. However, data from knockout experiments suggest a mechanism by which *Csmd1* could regulate locomotor activity in the LG × SM AIL. Homozygous knockout mice exhibited increased activity compared to heterozygous and wild-type mice**??**. This suggests that the SM allele, which produces the same effect in the AILs, could be a loss of function variant. However, given the size of *Csmd1*, I suggest this with caution, as there could be other variants causing the effect.

I identified another QTL for PPI6 at 15.34-16.99 Mb on chromosome 8, which contained *Csmd1* and *Mir3106*, a microRNA ($p = 3.55 \times 10^{-6}$; MAF=0.24); however, *Mir3106* did

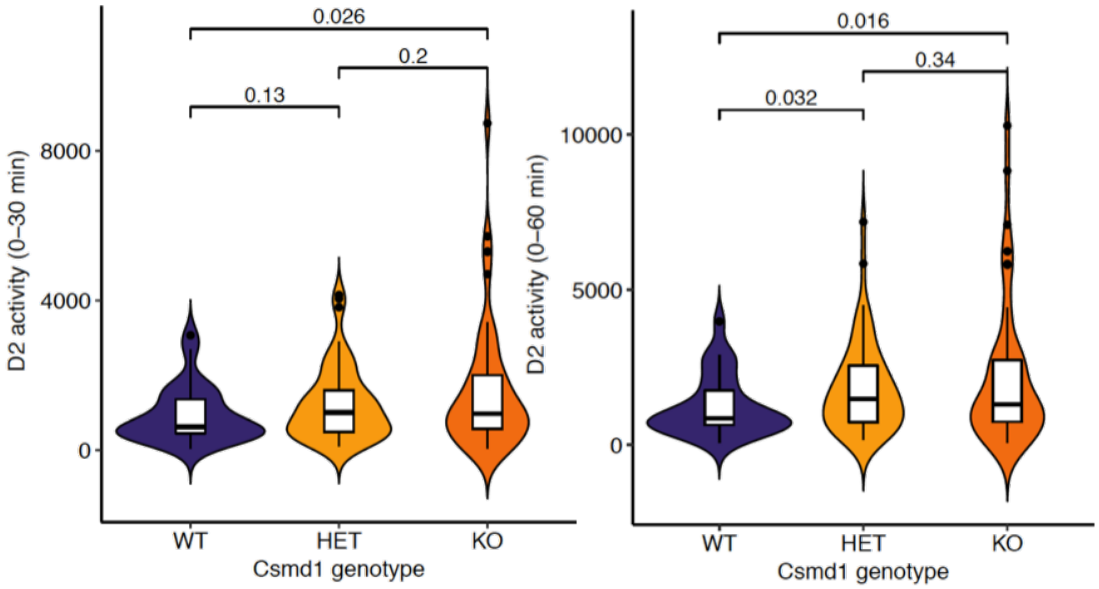

Figure 3.5: **Locomotor activity in *Csmd1* knockout mice**. Locomotor activity between 0-30 and 0-60 minutes of activity (cm distance traveled) in the open field test is shown for *Csmd1* knockout mice. Knockout mice (n=45) show increased activity relative to wild-type (WT; n=26) or HET (n=54) genotypes. Details about the strain background are described in Distler *et al.* [313].

not have an eQTL. Here, the SM allele was associated with decreased PPI. Neither Distler *et al.* [313] nor Steen *et al.* [314] observed abnormal PPI in *Csmd1* knockout mice. This discrepancy may be explained by differences in genetic background of the mice used in the two studies.

QTLs for novelty and methamphetamine-induced locomotor activity map to a region implicated in motor abnormalities and schizophrenia.

One methamphetamine activity QTL (D2, 15-20 min) and three saline activity QTLs from D1 (side changes at 15-20 and 0-30 min; distance traveled 0-5 min) mapped to chromosome 17 between ~75.7-78.9 Mb (Supplementary Table 2.3). I also observed suggestive QTLs for D2 and D5 activity at this locus. The strongest genome-wide significant QTL was for side changes on D1 (15-20 min, $p = 3.60 \times 10^{-6}$, MAF=0.50; 3.6). In each case, the SM allele

was associated with increased activity.

Three genes with *cis*-eQTLs in HIP appeared at all loci: *Crim1, Qpct*, and *Vit*. The SM allele was associated with decreased expression of the genes3.7. *Crim1* contains a missense mutation in the SM background and is important for development and motor neuron survival [315]. *Qpct* has been implicated in the accumulation of β-amyloid proteins in Alzheimers disease and has been associated with schizophrenia in human GWAS [316,317]. *Vit* has been associated with motor neuron deficits in transgenic mice and is thought to play a role in cell adhesion [318]. However, none of the three genes have been implicated in locomotor activity or drug response phenotypes. I suggest that these genes would make good candidates for a functional follow-up study.

QTLs on chromosomes 7, 17 and 18 suggest a role for glutamatergic signaling, cell adhesion, and neuronal development in regulating the startle response.

I chose to focus on the three startle QTLs that were associated with the mean startle response on chromosomes 7, 17 and 18. QTLs on chromosomes 7 and 17 were also identified in a smaller GWAS of startle in LG $\times$ SM F2 and G34 [167]. Dissection of startle QTLs on chromosomes 7, 17 and 18 suggests a complex network of genes contributing to the startle response.

A large startle QTL on chromosome 7 ranging from 81.4-89.6 Mb ($p = 4.18 \times 10^{-7}$; MAF=0.35) contained 78 QTL genes and 30 *cis*-eGenes (Supplementary Table 2.3). Two narrower QTLs for block 1 ($p = 3.13 \times 10^{-6}$; 3.50 Mb) and block 2 ($p = 9.39 \times 10^{-8}$; 0.87 Mb) mapped inside the region and shared the same top SNP as the main startle trait, located in an intron of *Ap3b2* (adaptor-related protein complex 3, beta 2 subunit). The critical region for a second block 1 QTL overlapped the larger region by 260 Kb at the distal end, but its top SNP did not overlap the other startle QTLs ($p = 9.56 \times 10^{-7}$; MAF=0.49). However, I focused on the block 2 QTL because it was the strongest association and the narrowest

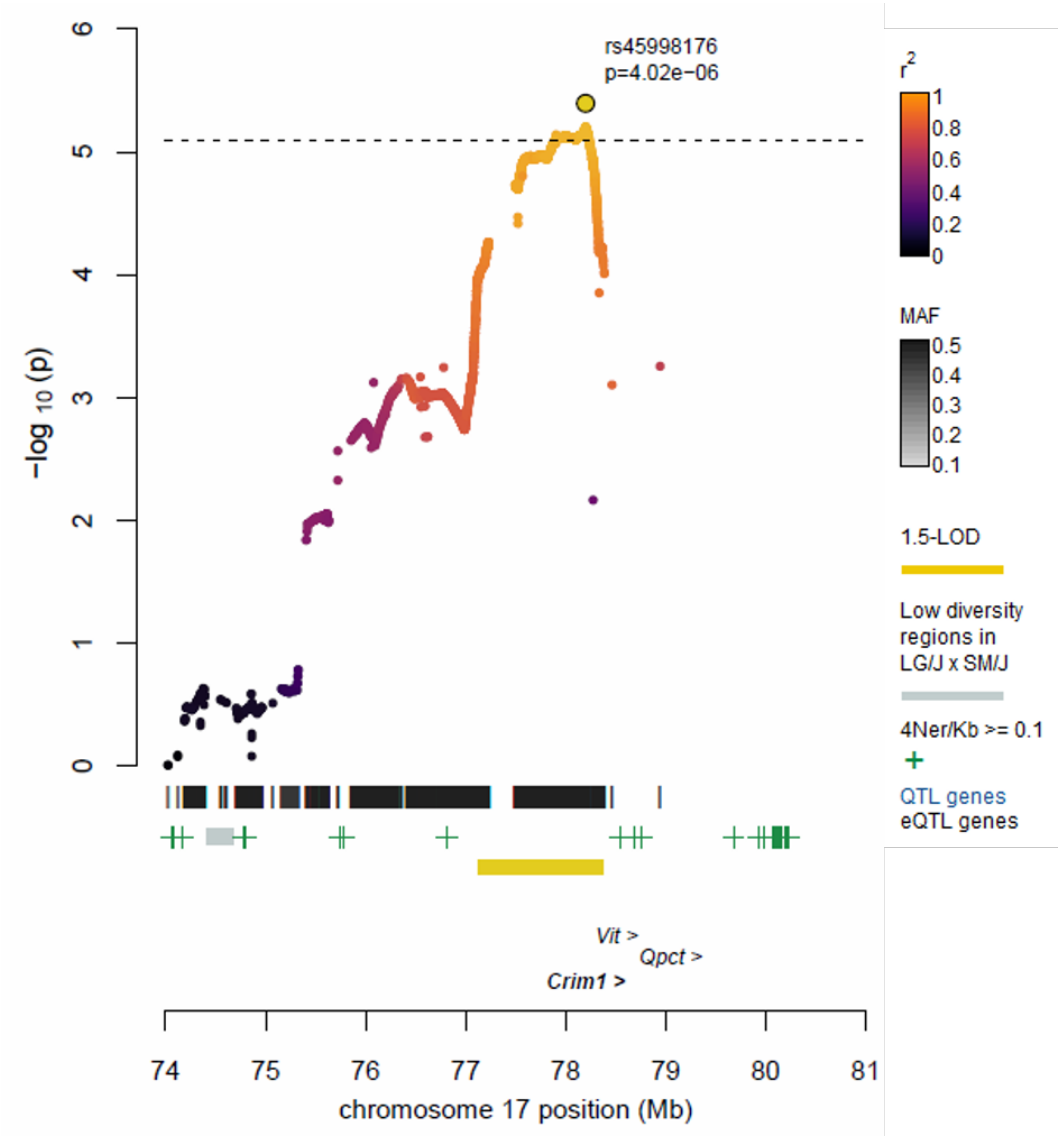

Figure 3.6: **QTL for locomotor activity (D1 side changes, 0-30 min) on chromosome 17**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

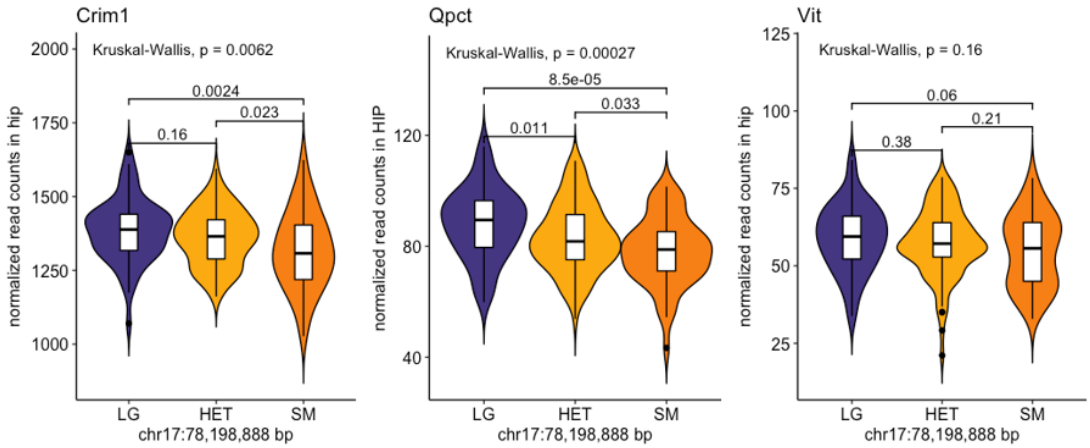

Figure 3.7: **Expression of *Crim1, Qpct,* and *Vit* in HIP**. The SM allele was associated with lower expression (expressed as normalized read counts) of all *cis*-eGenes within the chromosome 17 QTL for activity. 57 mice are homozygous LG, 55 are homozygous SM, and 96 are heterozygous.

interval, containing 14 QTL genes and 12 *cis*-eGenes.

Five *cis*-eGenes overlapping the block 2 QTL have missense mutations (*Ap3b2, Iqgap1, Crtc3, Homer2*, and *Saxo2*). *Homer2* has *cis*-eQTLs in HIP ($p = 1.31 \times 10^{-5}$) and PFC ($p = 3.41 \times 10^{-15}$); it is also a *trans*-eGene in PFC, regulated by a region on chromosome 5 (FDR=0.05). *Ap3b2* is a neuron-specific isoform of the adaptor protein 3 complex [319] that has been associated with hyperactivity, involuntary movement, and seizures in mutant mice [320]. *Homer2* (homer scaffolding protein 2) has been linked to associative learning and behavioral responses to cocaine [144]. *Iqgap1* (IQ motif containing GTPase activating protein 1) is thought to affect cognition by decreasing dendritic spine density [321]. Other genes in the region that are known to regulate behavior include *Grm5* (glutamate receptor, metabotropic 5) and *Folh1* (folate hydrolase 1). The presence of several interesting genes makes it difficult to identify a single candidate, but it raises the possibility that the startle QTL could be caused by genes that regulate glutamatergic signaling and neural development.

All five startle traits mapped to a region on chromosome 17 between 27.14-29.71 Mb (Supplementary Table 2.3). The QTL for mean startle was the most significant ($p = 5.29 \times 10^{-10}$; MAF=0.22; Figure 3.9); it mapped to a 2.16 Mb region at 27.14-29.30 Mb and

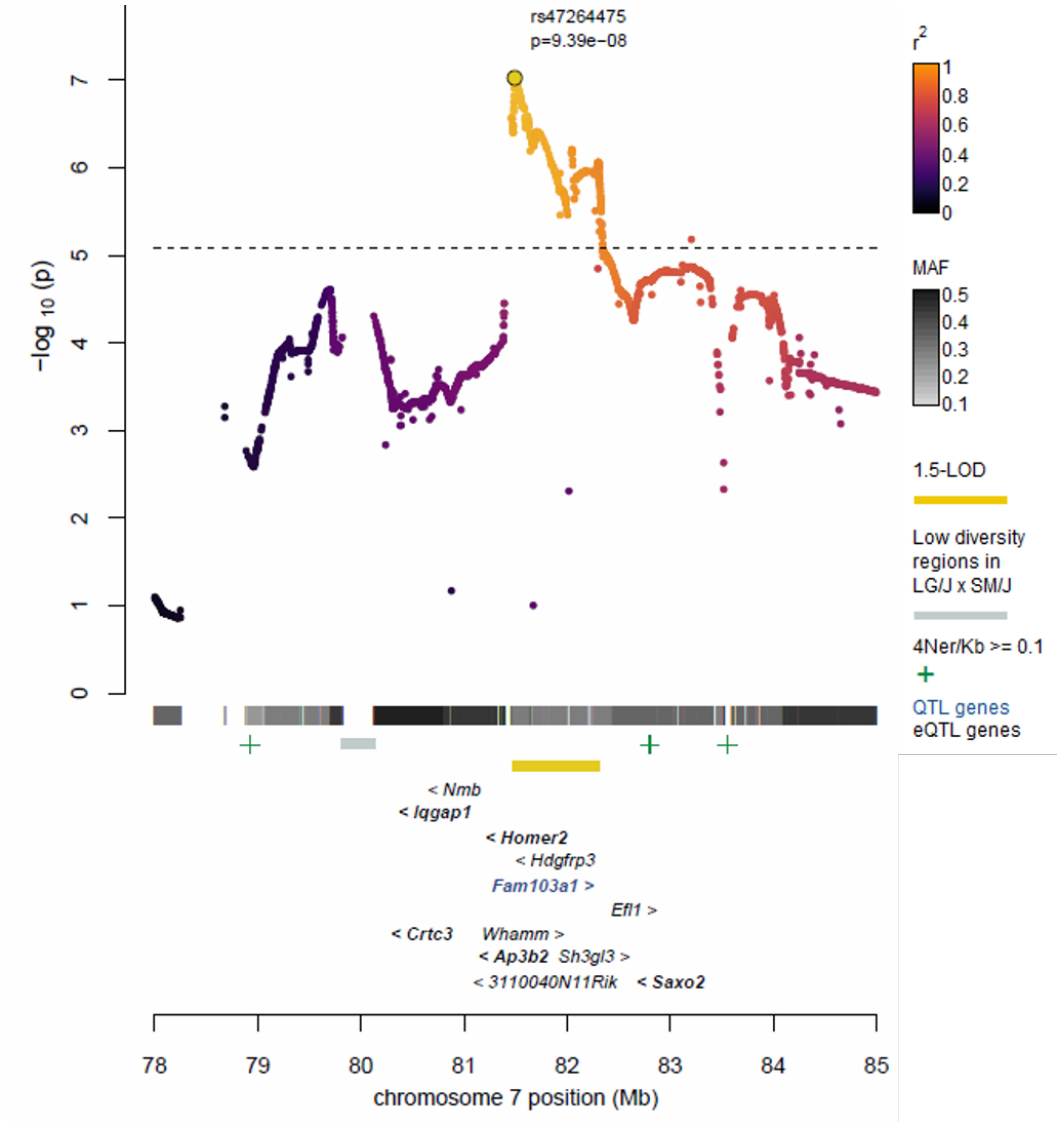

Figure 3.8: **QTL for startle response (block 2) on chromosome 7**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

contained 51 QTL genes and 29 *cis*-eGenes. Its top SNP, located in an intron of *Anks1*, had an effect size of 0.020 (Supplementary Table 2.3). Five genes, *Anks1, Ip6k3, Kifc5b, Grm4* and *Uhrf1bp1*, have missense mutations in LG; *Neurl1b* and *Itpr3* have missense mutations in SM. *Grm4* (glutamate receptor, metabotropic 4) has been associated with impaired PPI and abnormal motor learning in transgenic mice [322]. *Ip6k3* (inositol hexaphosphate kinase 3) has also been linked to abnormalities in motor learning, neural cell morphology, and nervous system physiology [323]. *Itpr3* (inositol 1,4,5-triphosphate receptor 3) has been associated with anxiety, social behavior, and locomotor activation in mutant mice used to model autism-related traits [324, 325].

The remaining genes that had missense mutations were not associated with mouse phenotypes, but a scan of the MGI database revealed additional eGenes in the QTL that have been found to influence behavior and brain development. QTLs for methamphetamine and saline activity also mapped to this locus, as did QTLs for fear and anxiety identified in a previous study of G34 mice [208]. I found this exciting because it implies that further exploration of the QTL (e.g. using mutant animals) could be fruitful for understanding why this region appears to have such widespread effects on behavior in the LG × SM AIL.

The top SNP within the startle QTL on chromosome 18 had the largest effect size (pve=0.026) out of all startle QTLs that I identified ($p = 6.74 \times 10^{-7}$, MAF=0.41, Figure 3.10). Startle block 2 also mapped to this region, but its top SNP ($p = 9.22 \times 10^{-7}$; MAF=0.43; effect=0.022) was located downstream, in an intron of *Ctnna1* (catenin alpha 1), a gene important for brain development and cell adhesion [326]. Each QTL was 3.8 Mb wide and included 88 genes, three of which, *Pkd2l2, Fam13b*, and *Sil1*, contain missense SNPs. *Pkd2l2* (polycystic kidney disease 2-like 2) has been linked to hyperactivity (MGI) and mutations in *Sil1* (ER chaperone *SIL1* homolog) have been associated with neurodegeneration and ataxia in mice and humans [327]. One gene in the region, *Hbegf* (heparin-binding EGF-like growth factor), has been associated with decreased PPI and other phenotypes related

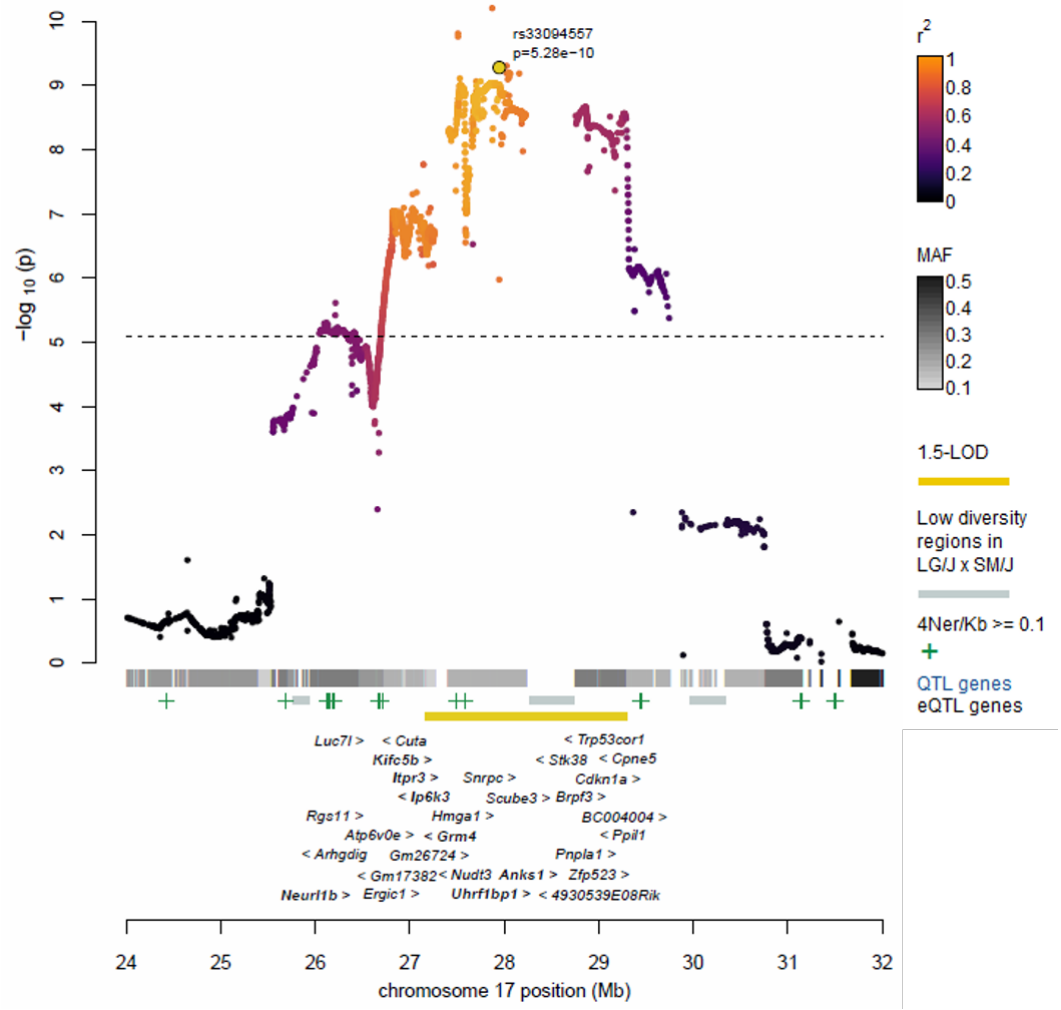

Figure 3.9: **QTL for mean startle response on chromosome 17**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

to psychiatric disease in mice [328].

High gene density within the QTL is due to overlap with the protocadherin α, β, and γ, gene clusters. Protocadherins are highly conserved signaling molecules whose functions are important for neuronal development, synaptic plasticity, and synapse assembly [329–331]. Furthermore, their genetic regulation is complex: alternative splicing of variable and constant exons can generate a variety of different protocadherin transcripts, similar to immunoglobulin genes in the immune system [332]. They may also be expressed mono-allelically from individual neurons [333]. Thus, regulation of protocadherins has been proposed as a mechanism underlying regional and functional diversity in the mammalian brain [329]. If the protocadherin cluster causes the QTL, this could suggest that variation in the startle response is due to differences in neuronal signaling networks. However, the mechanisms underlying the startle response remain uncertain. The abundance of genes within QTLs for startle that are related to central nervous system development, neuronal activity, and organism-level behavioral abnormalities could suggest that a network of multiple different factors is involved in regulating the startle response.

3.3.3 Integration of eQTLs and metabolic QTLs

Robust QTLs for body weight on chromosomes 2, 4, and 7

I identified 46 QTLs for body weight measured at 10 time points during behavioral testing, many of which converge on a single locus. Eight traits mapped to a region on chromosome 2 between 56.18-57.90. In each case, the SM allele was associated with higher body mass (Supplementary Table 2.3). Individual QTLs were 0.08-1.71 Mb wide, containing up to 8 QTL genes and 3 *cis*-eGenes (Supplementary Table 2.3). The strongest QTL was for weight on D5 of the CPP test ($p = 5.61 \times 10^{-7}$; MAF=0.47), when the median age was 58 days. This QTL did not contain any genes, but the top *cis*-eQTL SNP for *Nr4a2* (nuclear receptor subfamily 4, group A, member 2), a *cis*-eGene in PFC ($p = 6.96 \times 10^{-3}$), was located inside

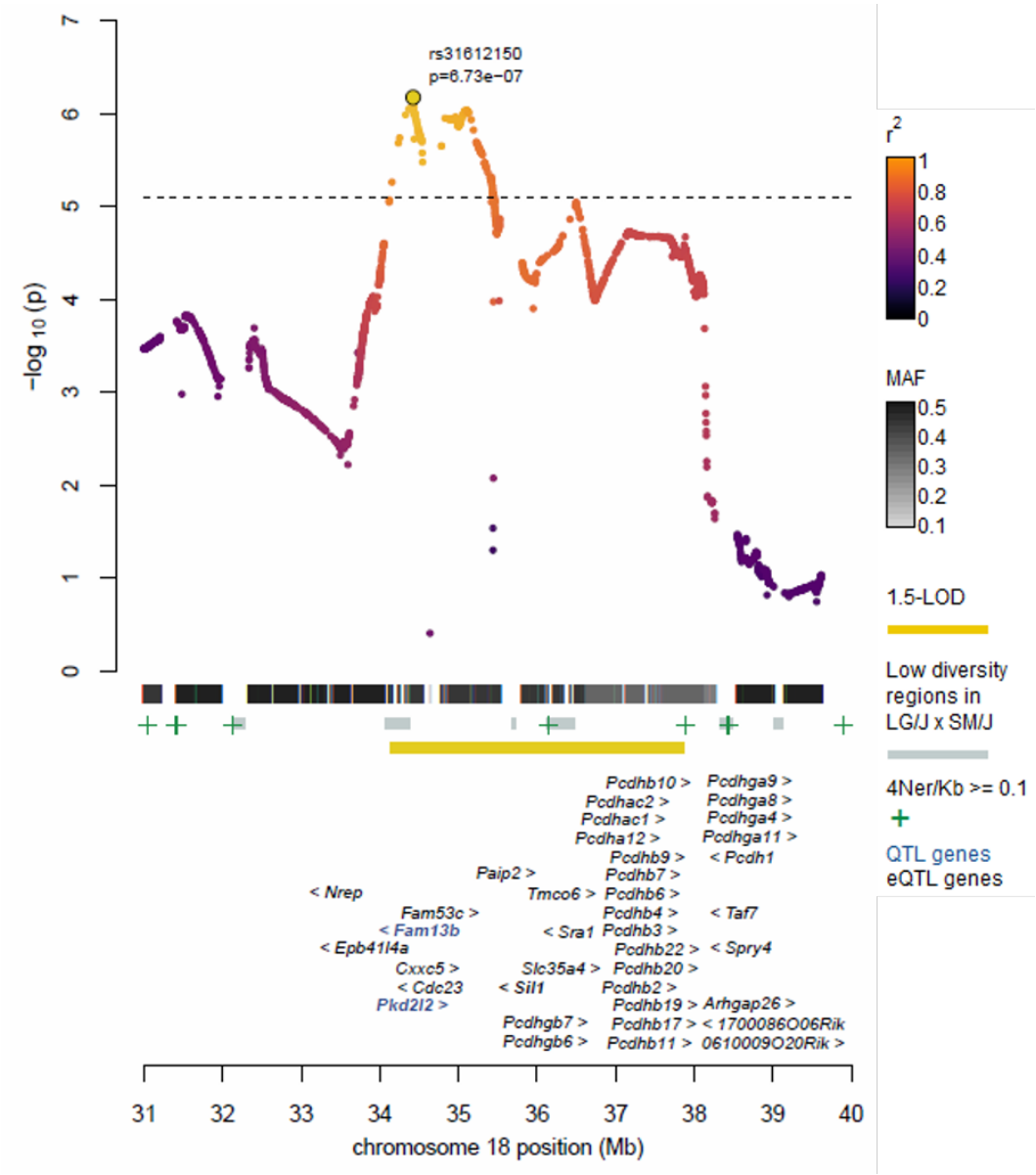

Figure 3.10: **QTL for mean startle response on chromosome 18**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

all body weight QTLs in the region, including the QTL for CPP D4 body weight, shown in Figure 3.11.

Nr4a2 has been linked to neurological and immune system abnormalities in mice: *in vivo* suppression of *Nr4a2* in T-cells resulted in accelerated weight loss after an immune challenge [334] and mice lacking *Nr4a2* mRNA in midbrain dopamine neurons are 40% smaller than their wild-type littermates at 8 weeks of age [**?**]. It has also been associated with decreased locomotor activity [**?**] which could suggest a behavioral component to body weight regulation.

However, other genes in the larger QTL region may also be of interest. For example, *Upp2*, (uridine phosphorylase 2) which overlapped 3 body weight QTLs, had a strong *cis*-eQTL in HIP, STR and PFC ($p \leq 1.26 \times 10^{-18}$) and is known to affect lipid metabolism in the mouse [335]. The region was also discovered in an 8.8 Mb QTL for body weight identified in congenic mice derived from a cross of C57BL/6J and Pbwg1, a wild-derived mouse strain (B6.Cg-Pbwg1). Interestingly, the wild allele was associated with lower fat, body weight, and blood glucose levels, and its effects were observed in mice on high-fat and low-fat diets [336]. Furthermore, *Upp2* was downregulated in B6.Cg-Pbwg1 mouse liver (the study also reported five other genes that showed differential expression in liver or adipose tissue) [336].

Eight traits mapped to a region between 99.40-103.70 Mb on chromosome 4, ranging in size from 2.25-4.01 Mb (Supplementary Table 2.3). QTLs had up to 35 genes and 16 *cis*-eGenes. The strongest QTL (body weight on D8 of CPP; $p = 5.62 \times 10^{-8}$; MAF=0.30; Figure 3.12) contained 24 QTL genes and nine *cis*-eGenes. Six other QTLs had the same top SNP, and in each case, the SM allele was associated with lower body weight. Interestingly, only 9% of mice are homozygous LG (41% and 50% are heterozygous or homozygous SM, respectively). The top SNP is ~300 Kb upstream of *Lepr*. *Lepr* is not a *cis*-eGene, but it encodes the receptor for leptin, a fat-specific hormone whose effects on obesity and

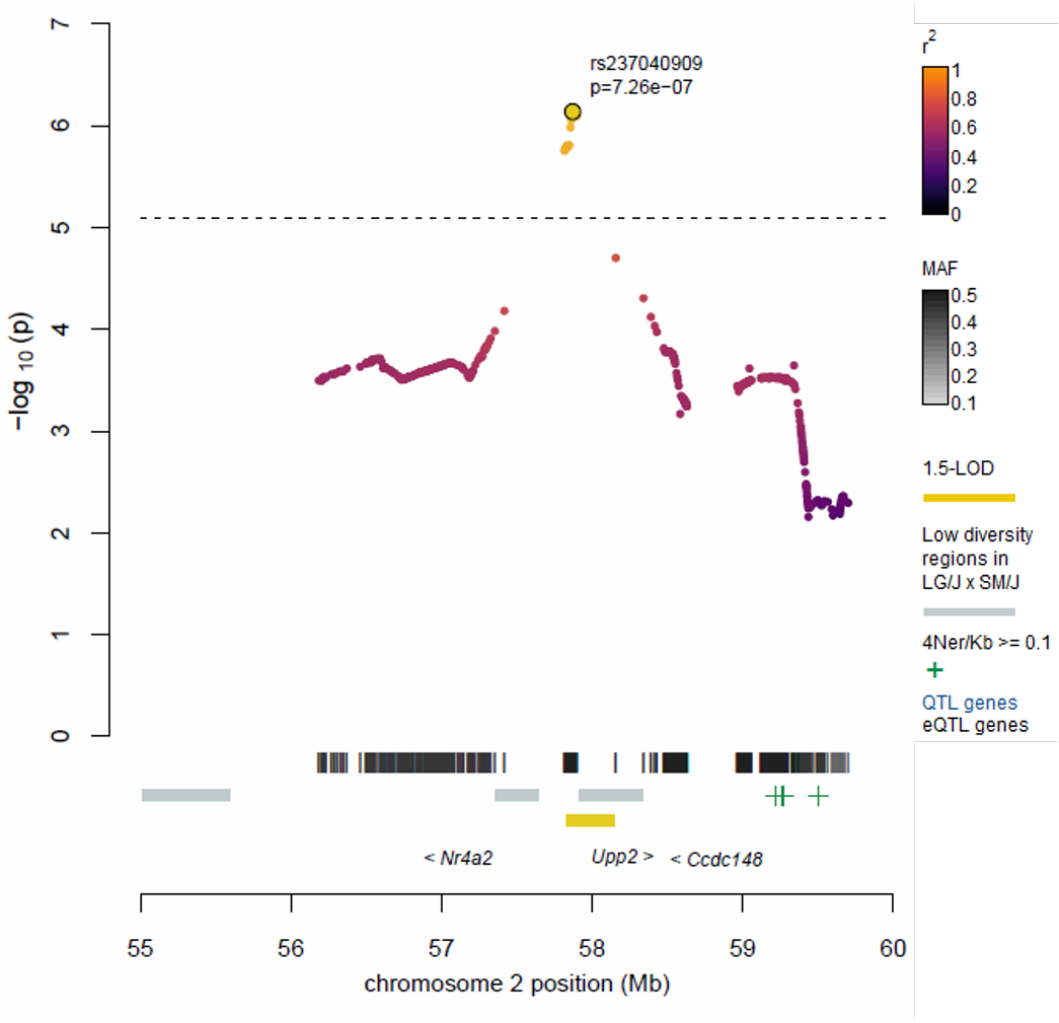

Figure 3.11: **QTL for body weight on chromosome 2**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

metabolism have been widely studied [337]. *Lepr* may account for some of the QTL signal, but it is also possible that other genes are involved.

A QTL at 105.67-110.81 Mb on chromosome 7 was associated with all body weight traits ($p \leq 1.47 \times 10^{-6}$). Individual QTLs were 3.41-5.14 Mb wide and contained up to 115 QTL genes and 31 *cis*-eGenes (Supplementary Table 2.3). Here, the LG allele was associated with increased weight. Eight QTLs had the same top SNP, a missense variant in *Tpp1* (MAF=0.49). *Tpp1* (tripeptidyl peptidase I) transgenic mice have been used to model late-infantile neuronal ceroid lipofuscinosis, a neurodegenerative disorder caused by mutations that prevent the *TPP1* protein from degrading certain lysosomal peptides, leading to an accumulation of lipid granules in various tissues [338, 339].

The human orthologs of four *cis*-eGenes overlapping the region (*Rpl27a, Stk33, Trim66*, and *Tub*) are located at or around SNPs that have been associated with obesity [340] and body mass index [341] in humans. *Tub* (tubby bipartite transcription factor) encodes a transcription factor, *TUBBY*, that cause late-onset obesity in mice that carry loss-of-function mutations [342]. *TUBBY* forms a complex with heterotrimeric G protein-coupled receptors [342], one of which is *Gnb1* (guanine nucleotide binding protein beta 1). *Gnb1* is located on chromosome 4 and has a *trans*-eQTL in HIP located 0.55 Mb upstream of *Tub* inside the body weight QTL. *Gnb1*, along with two other *trans*-eGenes regulated by SNPs within the body weight QTL, *Gucy1a2* (guanylate cyclase 1 soluble alpha 2, a PFC eGene on chromosome 9) and *Crebbp* (CREB binding protein, a HIP eGene on chromosome 6), have been associated with obesity in human GWAS [341, 343, 344]. Although the size of this QTL precludes selection of a single candidate gene, evidence for *trans* associations connecting the locus to other phenotypically relevant loci suggests that its relationship to obesity is complex.

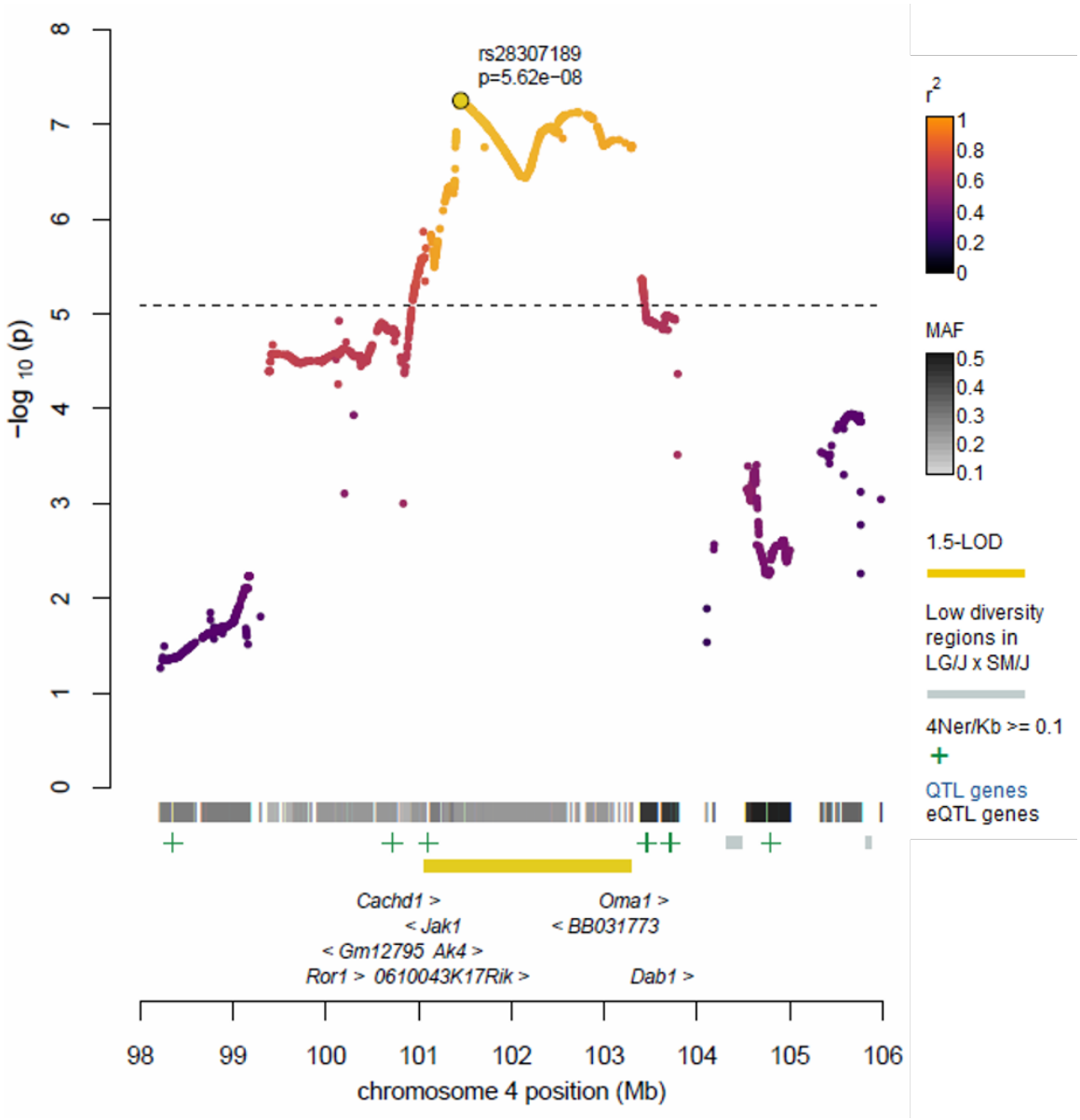

Figure 3.12: **QTL for body weight on chromosome 4**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

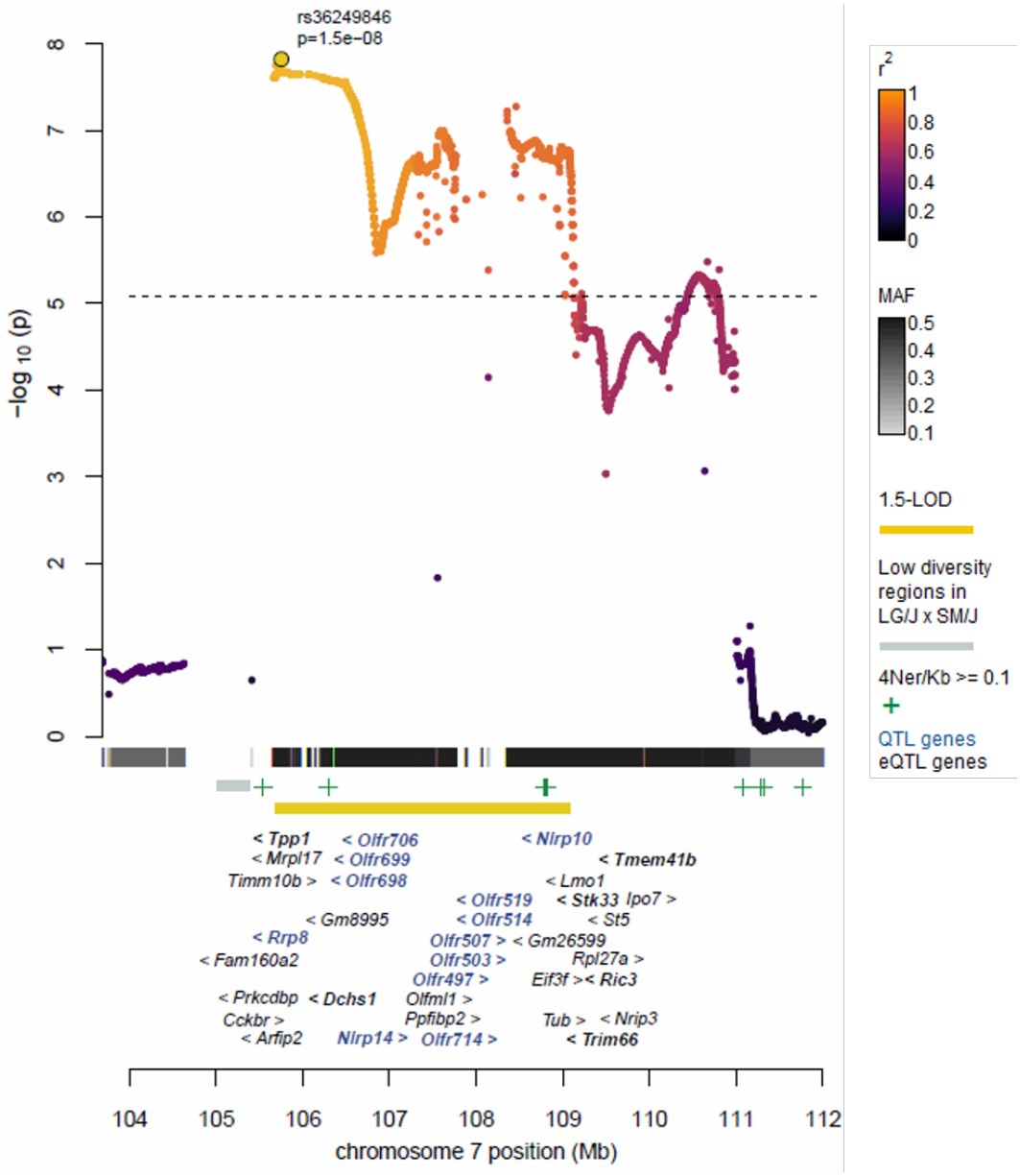

Figure 3.13: **QTL for body weight on D8 of the CPP test on chromosome 7**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

Strong signals for EDL are corroborated by other muscle weight QTLs and previous GWAS of muscle weight in G34.

I found 22 QTLs for leg muscle weight on 10 chromosomes (Supplementary Table 2.3). As for body weight, many muscle traits mapped to the same regions. I will focus on EDL muscle because it had the strongest QTLs.

I identified QTLs for EDL, plantaris and TA muscle weight on chromosome 4 at ~65 Mb. At each QTL, the LG allele was associated with greater muscle mass. The TA QTL ranges from 63.87-69.90 Mb ($p = 1.44 \times 10^{-6}$; MAF=0.29), whereas the others are much narrower, located at 65.32-65.84 Mb. The EDL association was one of the strongest I identified across all traits; its top SNP had an estimated effect size of 0.047. The top SNPs for EDL ($p = 7.88 \times 10^{-13}$; MAF=0.38; Figure 3.14) and plantaris ($p = 2.25 \times 10^{-6}$; MAF=0.39) are located in an intron of *Astn2* (astrotactin 2), a *cis*-eQTL gene that is differentially expressed in LG and SM soleus [181]. The amino acid sequence of *Astn2* is the same in LG and SM; therefore, the difference in expression between the LG and SM alleles could be due to SNPs within its regulatory region. In support of this hypothesis, the SM genome contains a regulatory region upstream of *Astn2* (~65.372-65.374 Mb) that is 16 bp longer than in LG [193]. Although other genes in the region, such as *Tlr4* (toll-like receptor 4) and *Trim32* (tripartite motif-containing 32) have been associated with muscle and bone phenotypes in the mouse (MGI), the results suggest that *Astn2* may also be important, yet its function has not been studied in muscle. Thus, I propose *Astn2* as a candidate gene for muscle mass.

Another strong EDL QTL mapped to ~104.33 Mb on chromosome 13 ($p = 2.03 \times 10^{-13}$, MAF=0.21; pve=0.045); this was the strongest QTL I identified across all traits. A second QTL for gastrocnemius provides additional support for the region ($p = 2.56 \times 10^{-7}$, MAF=0.21; pve=0.019). Each QTL is <0.5 Mb wide and flanked by large regions of low polymorphism between LG and SM. At both QTLs, the SM allele is associated with higher muscle mass. There are 5 QTL genes (*Trappc13, Trim32, Ppwd1, Cenpk, Adamts6*) and 6

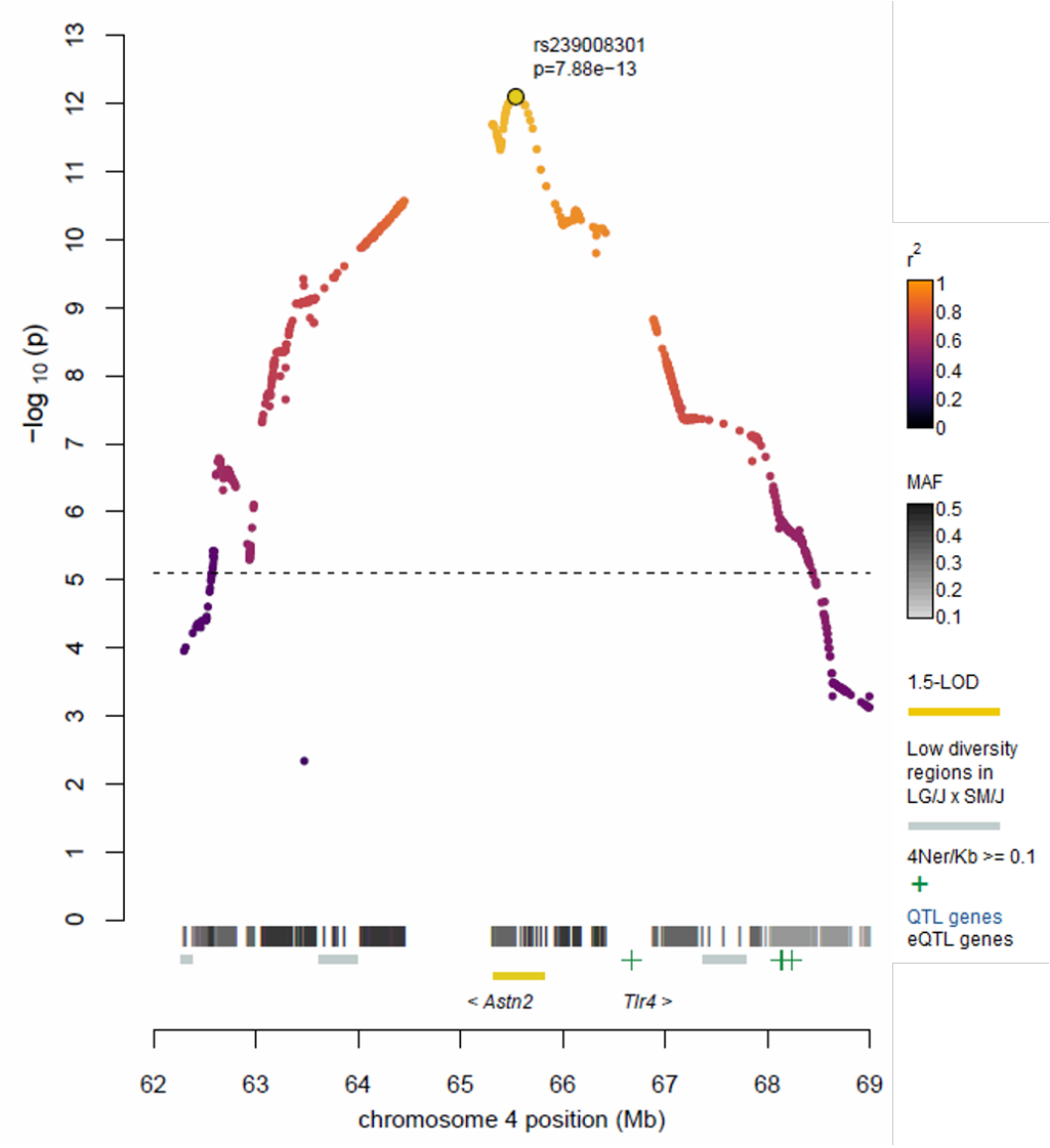

Figure 3.14: **QTL for EDL muscle weight on chromosome 4.**. The legend for this locus zoom figure is the same as before; a detailed description is provided in the caption for Figure 3.2.

cis-eGenes (*Nln, Ppwd1, Sgtb, Srek1, Trappc13, Erbb2ip*) within the overlapping QTLs. To my knowledge, only *Nln* (neurolysin, an oligopeptidase) has been implicated in muscle traits. However, *Trappc13* (trafficking protein particle complex 13) is differentially expressed in TA of LG and SM, with SM having lower expression of the gene [293]. SM also has lower expression of *Nln* in soleus muscle compared to LG [181], which I found interesting because *Nln* exhibits enzymatic activity in mouse skeletal muscle, where its effects are more pronounced in slow-twitch fibers, the primary type found in soleus [345]. With this information, I suggest that both *Trappc13* and *Nln* are promising candidates to study in greater detail.

3.3.4 Discussion

I identified thousands of *cis*-eQTLs and *trans*-eQTLs in the LG × SM AIL, including a regulatory *trans* hotspot that affected the activity of 85 genes. I used overlap of QTLs, eQTLs, differentially expressed genes, and functional annotations from existing databases to identify candidate genes.

There were two QTLs that I found particularly intriguing. The first was an association with D1 and D2 locomotor activity on chromosome 17 that contained *Crim1, Qpct,* and *Vit*. It would be reasonable to explore the effects of all three genes in a follow-up study, but if I could only choose one, it would be *Qpct*. Its expression pattern in HIP is the most distinct of the three eGenes when stratified by eQTL genotype, and its role in Alzheimer's disease and schizophrenia suggest that further study of *Qpct* could have translational relevance. The presence of a missense SNP in *Crim1* is also suggestive, but it is not clear if the mutation is deleterious. Moreover, I suspect that it could be challenging to study *Crim1* in a knockout mouse due to its crucial role in development (to my knowledge, only conditional knockouts have been used to study its effects in the brain).

I was also interested in the startle QTL on chromosome 17, which overlapped QTLs for fear, anxiety-related behavior, and locomotor activity in the AIL. *What could cause the as-*

sociation between this QTL and such a diverse set of behaviors? Due to LD within the ~2 Mb region and the observation that several QTL genes appear to have important roles in the brain and behavior, I was unable to resolve these questions using the data from G50-56. I would be very interested to know if any of the QTLs for different behaviors are caused by the same gene. In mice, fear is measured by the amount of time spent frozen in place after experiencing a negative stimulus (i.e. an electrical shock to the foot). This behavior is reflexive, which leads me to speculate that it could be more closely related to the startle response than to the other behaviors at the genetic level. I have also wondered if further investigation of these QTLs would reveal a dysfunctional network of genes with a variety of different effects on behavior. Of course, these hypotheses will remain speculative without additional data, and there are multiple strategies that could be used to address them. An obvious approach would be to fine-map the QTLs in future generations of the AIL. However, if I could repeat the study in a new sample, I would sequence mRNA at greater depth so that I would have enough power to detect eQTLs at the transcript level rather than summarizing mRNA abundance over the entire gene. Gene or transcript expression data could also be used to construct transcriptional networks, which could then be integrated with QTL and phenotype data to identify pathways and interactions important for behavior. Protein-protein interaction networks could provide useful information about the mechanisms that influence locomotor behavior, startle, and anxiety. These experiments would be useful for understanding the mechanisms that underlie other QTLs in addition to the ones on chromosome 17. They would also be useful for interpreting *trans*-eQTLs, particularly those that regulate the expression of many genes, such as the hotspot eQTL identified on chromosome 12.

Overall, my research with the Palmer Lab was successful in that it generated a variety of intriguing hypotheses that can be tested in future experiments. In fact, there were so many interesting results that it was impractical for me to explore all of them in detail (and I hope

that future scientists in the Palmer Lab will be interested in continuing my efforts). More broadly, my work demonstrated that integrating phenotype, genotype, and gene expression data collected from the same individuals is a valuable approach for studying complex traits. Additional discussion of the work presented in Chapters 2-3 is provided in Chapter 4.

3.4 Contributions

Dr. Abraham A. Palmer designed the experiment and provided financial support for this project (R01 DA021336). Additional funding came from F31 DA03635803 and T32 GM007197. I wrote in first-person for clarity, but this work was a team effort. Celine St. Pierre, Emily Leung, Yeonhee (Jenny) Chen, Matthew Cheung, Alisha Ho and I extracted more RNA than we were able to sequence. I prepared cDNA libraries for submission to the core Genomics Facility at the University of Chicago. Sequencing was performed by Dr. Pieter Faber and core staff. Jung Kyun Seo analyzed RNA sequencing data (alignment, expression quantification, quality control) with help from Dr. Stefan Canzar. I mapped eQTLs and analyzed the data. Mike Jarsulic at the Center for Research Informatics at the University Chicago was extremely helpful while I was submitting eQTL jobs to the cluster.

3.5 Supplementary Materials

3.5.1 Supplementary Methods

Identification and correction of RNA sample mix-ups

Sex swaps I applied multiple approaches to identify and resolve RNA sample mix-ups. First, I examined *Xist* expression to identify apparent sex mix-ups. *Xist* regulates dosage compensation and is only expressed in females. Male samples expressing *Xist* were changed to female, and female samples lacking *Xist* expression were changed to male.

Tissue swaps I merged expression data from HIP, PFC and STR and plotted the first two PCs of the data to identify tissue mix-ups. As expected, most samples clustered into distinct groups by tissue. I reassigned 12 mismatched samples to new tissues and removed 35 apparently contaminated samples that did not cluster with the rest of the data. The first two PCs correctly separated samples by tissue after error correction, implying that I resolved all cross-tissue mix-ups.

RNA genotype concordance with GBS data Next, I used RNA genotype concordance to identify apparent sample mix-ups. Genotypes were created from RNA sequencing data by retrieving allele counts at sites with at least 25 reads and no more than one mismatched base. A mouse may have up to 3 sets of genotypes created separately from each tissue. I measured RNA genotype concordance for all pairs of samples with the expectation that the best match for each sample would be from a different tissue belonging to the same mouse. If the best match belonged to a different mouse, the samples were flagged as potential mix-ups. I then examined concordance between RNA and GBS genotypes to reassign mixed-up sample IDs. If I could not resolve the identity of a sample, I discarded its expression data. A total of 108 samples were discarded during this process (33 HIP, 39 STR and 36 PFC). I also discarded expression data for 29 samples whose genotype data was removed during GBS quality control (11 HIP, 9 STR and 9 PFC). The average concordance rate among RNA genotypes derived from the same mouse was 94.6% after error correction, indicating that my approach was successful. Finally, I used correlation-based statistics to identify and remove 12 additional outliers (2 HIP, 3 STR and 7 PFC).

3.5.2 Supplementary Figures

3.5.3 Supplementary Tables

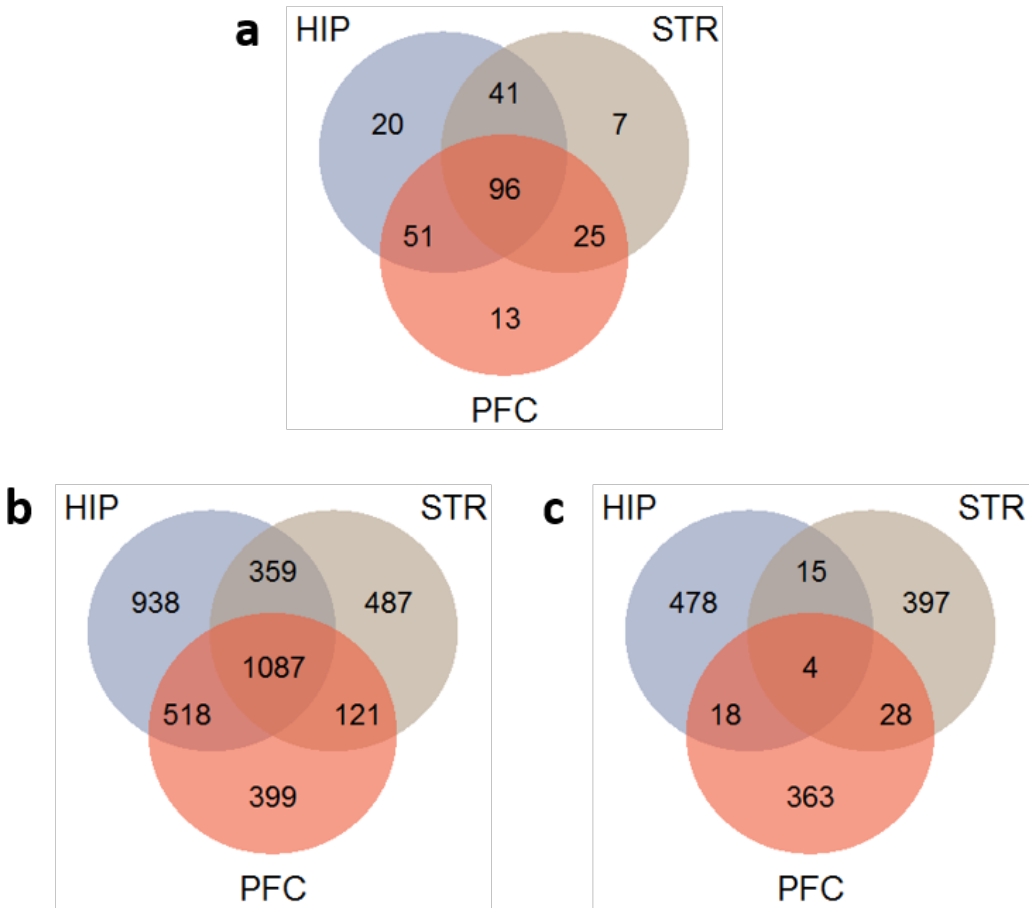

Figure 3.15: **Summary of eQTLs by brain region.**. **(a)** Number of overlapping samples sequenced from each tissue (208 HIP, 169 STR and 185 PFC) after quality control. **(b)** *cis*-eQTLs and **(c)** *trans*-eQTLs identified in each tissue (FDR=0.05).

Table 3.1: **Master *trans* regulators and their target genes.** (The supplementary file associated with this dissertation contains information on master *trans* regulators, or hotspots, including chromosome and position of the eQTL and its target gene, p-values for the strongest *trans*-eQTL SNP, the eQTL tissue, and if the target was also a *cis*-eQTL, its strongest *cis*-eQTL SNP, position, and chromosome is listed along with its *cis* region (1 Mb surrounding the gene's transcription start and end sites)).

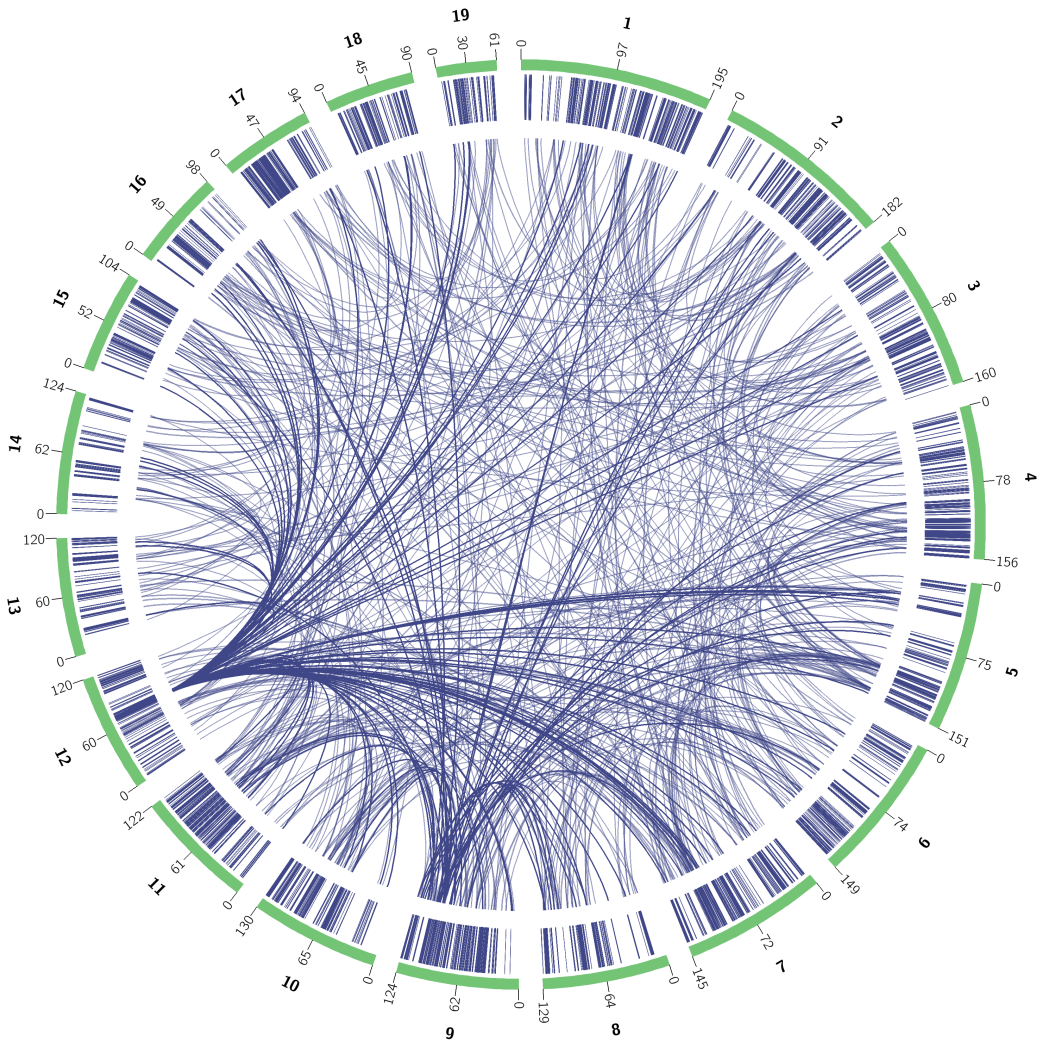

Figure 3.16: ***cis*-eQTLs and *trans*-eQTLs in HIP.**. Chromosomes are delineated by green bars on the outer edge of the plot. Each ring inside the circle shows locations of *cis*-eQTLs for HIP, and *trans*-eQTLs (FDR=0.05) are shown inside the circle.

Figure 3.17: ***cis*-eQTLs and *trans*-eQTLs in PFC.**. Chromosomes are delineated by green bars on the outer edge of the plot. Each ring inside the circle shows locations of *cis*-eQTLs for PFC, and *trans*-eQTLs (FDR=0.05) are shown inside the circle.

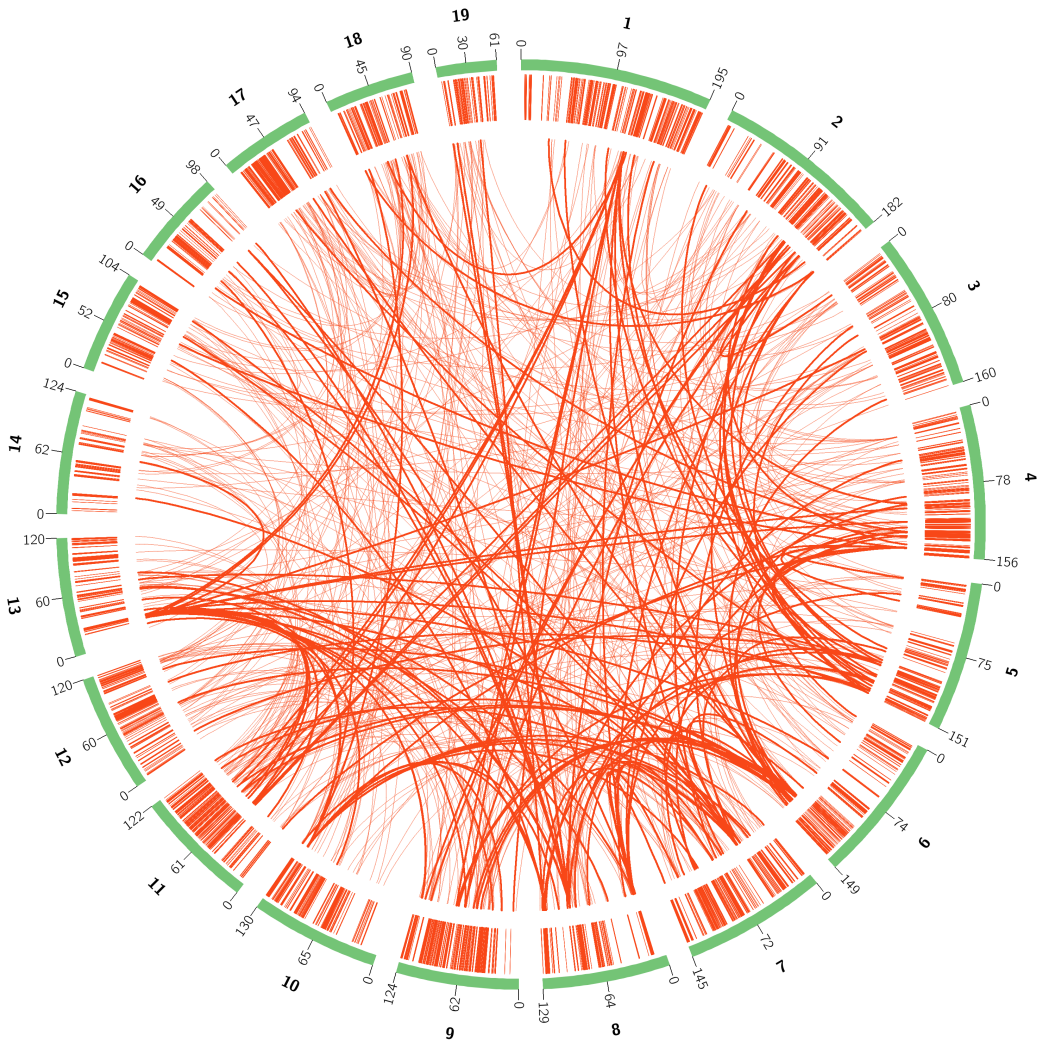

Figure 3.18: ***cis*-eQTLs and *trans*-eQTLs in STR.**. Chromosomes are delineated by green bars on the outer edge of the plot. Each ring inside the circle shows locations of *cis*-eQTLs for STR, and *trans*-eQTLs (FDR=0.05) are shown inside the circle.

CHAPTER 4

CONCLUSIONS

The broad goal of my work was to identify genes important for mouse behavior and physiology that might also be related to human psychiatric and metabolic disease. I used a simple mouse outbred population, the LG × SM AIL, to map QTLs and eQTLs and integrated these data to identify candidate genes. The Palmer Lab had previously used the AIL as a tool for fine-mapping F2 QTLs, which was its intended purpose [65]. This study demonstrates the utility of the LG × SM AIL as an *independent* mapping population that can be used to study a wide variety of traits.

I identified 126 QTLs for various behavioral and physiological traits using 1,063 male and female AIL mice from G50-56 with >500,000 genotypes obtained using GBS. I also identified thousands of eQTLs in HIP, PFC and STR in ~200 mice that were phenotyped and genotyped. This was a key advantage of using the mouse: although it is possible to impute gene expression data [39], obtain *post-mortem* brain tissue, or use blood or cell lines to map eQTLs in humans, obtaining healthy primary tissue from the same individuals that were phenotyped and genotyped in human GWAS is not currently possible.

First, I will briefly list advantages and limitations of using GBS to genotype the AILs. Then I will discuss some of our findings and their implications, including replication of an association between *Csmd1* and locomotor activity identified in G34. In Chapter 1, I discussed how the balance between power and precision in an AIL could be used to circumvent some of the challenges inherent to GWAS in unrelated humans. Here I discuss the differences I observed between the LG × SM AIL and human genetic architectures and list reasons that might explain why QTL effect sizes in AILs (and model organisms in general) tend to be higher than those identified in humans.

4.1 GBS as a genotyping strategy in large samples of outbred mice

The use of the LG × SM AIL as an independent mapping population was made possible by GBS, which allowed me to inexpensively genotype a large sample of mice. The availability of the LG and SM genomes [193] also enabled me to obtain a large number of SNPs (4.4 million before quality control and pruning for LD) with half the sequencing depth used in a previous study of commercially outbred mice [62]. Even before imputation, GBS yielded double the number of informative SNPs typed on the GigaMUGA (the most comprehensive genotyping array available for the mouse) [191] at approximately half the cost. Processing GBS data is not trivial: even with the simple genetic background of an AIL, GBS required greater bioinformatic expertise and computational effort than would have been required by an array. However, for many researchers (particularly those studying mouse strains with SNPs that are not well-represented on the GigaMUGA), the advantages of a high-throughput, low-cost genotyping strategy may outweigh the computational expense of GBS.

4.2 An integrative analysis of QTLs and eQTLs identified several candidate genes

I expected to have high mapping resolution in the AIL, but recognized that proximity to a QTL would be insufficient to determine causality even if I had achieved gene-level resolution. Gene expression studies provide a way to connect QTLs to molecular mechanisms. I used overlap of eQTLs and QTLs to identify candidate genes for behavior, body weight and muscle weight, demonstrating the utility of integrating gene expression data with QTLs in the mouse. I also identified thousands of *cis*-eQTLs and *trans*-eQTLs, including a *trans*-eQTL hotspot on chromosome 12 that regulated the expression of 85 genes.

For some candidate QTL genes, there was prior evidence of association with the trait of

interest, as for *Tub* (Figure 3.13) and *Upp2* (Figure 3.11), which were associated with body weight, or for *Oprd1* (Figure 3.2) and locomotor behavior. In other cases, eQTLs helped me to identify genes with less obvious connections to my traits of interest. For example, I found eight body weight QTLs that mapped to a missense mutation in *Tpp1* (Figure 3.13). In humans, mutations in *Tpp1* prevent its protein from degrading lipid molecules, causing them to accumulate in various tissues; mutations in *Tpp1* also cause late-infantile neuronal ceroid lipofuscinosis, a neurological disorder [338, 339]. These results could suggest that the role of *Tpp1* in regulating lipid accumulation is also important for obesity.

I also demonstrated that integrating QTLs and eQTLs is useful for developing mechanistic hypotheses about how genes might function to influence organism-level phenotypes. For example, I used overlap of our data and differentially expressed genes identified in LG and SM soleus [215] and TA [181] to identify three new candidate genes for muscle growth, *Astn2, Nln*, and *Trappc13*. An regulatory region upstream of *Astn2* is 16 bp longer in SM than in LG [193], suggesting that the difference in its expression could be due to a difference in gene regulation. Targeted mutations could be used in future studies to test the hypothesis that this regulatory region affects the expression of *Astn2* and determine whether or not it has a functional impact on skeletal muscle.

I also identified QTLs that contained several interesting genes, as for QTLs on chromosomes 7, 17, and 18 for startle. Several genes within these QTLs were involved in functions relevant to the brain, making it difficult to select candidates. However, I speculate that the presence of genes related to glutamatergic transmission, cell adhesion and signaling, neuron development, and neurotransmission within the startle QTLs (e.g. *Grm4, Grm5, Ctnna1, Homer2, Cdkn1a, Ip6k3, Pacsin1, Lemd2,* and the *Pcdh* gene cluster) could suggest that a network of genes is regulating the phenotype. A dysregulated network might also explain why many different behavioral QTLs were identified at the same locus on chromosome 17.

Similarly, the convergence of *cis*-eGenes, *trans*-eGenes, and *trans*-eQTLs at the body

weight locus on chromosome 7 (Figure 3.13) was intriguing because several of the associated factors (e.g. *Tub, Gnb1, Gucy1a2, Crebbp*) have been identified in human GWAS for obesity and body mass index [341, 343, 344]. It would be interesting to disentangle the effects underlying the QTLs on chromosomes 17 and 7 in future experiments. The chromosome 17 QTL could potentially provide insight into a variety of different behaviors, and the body fat locus on chromosome 7 may be useful for understanding human obesity. I suggest that a network-based approach integrating gene expression and proteomic data could be useful to explore these associations in future studies. I collected liver and kidney samples from the mice in this study, but to date, these have not been studied. These tissues could provide a starting point for exploring LG × SM physiology in greater detail.

4.3 Replication of an association between *Csmd1* and locomotor activity identified in G34

I replicated several QTLs identified in G34, one of which was *Csmd1*, a large protein-coding gene on chromosome 8 that I identified in association with saline activity [75]. Replication of GWAS is considered a necessary step in the field of human genetics. In model organisms, replication can be accomplished using the same standard, or by directly manipulating the implicated gene and observing the effects on the relevant phenotype. I performed both types of replication studies: I established that the same association is observed in G50-56 mice, and data from *Csmd1* knockout mice demonstrates that the null allele recapitulates the phenotype [313]. Thus, I have replicated our previous GWAS finding using two complimentary approaches, which together reflect the power of a model system.

Interestingly, neither Distler *et al. distler2012a* nor Steen *et al.* [314] found an effect of *Csmd1* on PPI in knockdown mice. PPI is a robust intermediate phenotype for schizophrenia [46]. Thus, this result was somewhat disappointing, as identifying an effect of *Csmd1* on PPI might have provided a system to study how the alterations in sensorimotor gating influences

psychiatric disease. However, the identification of a QTL for PPI in this sample suggests that the LG and SM inbred strains might provide a more fruitful genetic background for future studies of *Csmd1* in mutant mice.

4.4 Genetic architecture of complex traits in model organisms and humans

Why are QTL effect sizes in the LG × SM AIL (and model organisms in general) larger than the effect sizes typically observed in humans? One explanation is that the higher effect sizes observed model organisms are due to the combined effects of multiple factors in LD. While this could be true, it assumes that the causal alleles within the QTL affect the trait in the same direction, which is impossible to know without fine-mapping or performing additional experiments.

Instead, I argue that the discrepancy in QTL effect sizes observed in humans and model organisms can be explained by major differences in the genetic background of the two systems. For example, model organism populations do not experience the same selective pressures as human populations. That is, the traits in this study are not expected to be under positive or negative selection (recall that for the LG × SM AIL, population size and mate pairs are predetermined in each generation, and mice are provided food and shelter). However, this is not necessarily true for mutations related to disease in human populations, which would tend to remain rare if their individual effects on fitness were sufficiently large. Indeed, the ability of GWAS to identify causal loci for a given trait depends on genetic diversity, allele frequencies, and environmental factors. All of these variables can be controlled to a greater extent in model organisms than they can in humans, which can influence the number and magnitude of effects that can possibly be detected by GWAS.

In human populations, which have high genetic diversity, there are a very large number of variants that could possibly affect a given trait. The number of loci identified in human

GWAS increases linearly with sample size, but the proportion of trait heritability explained by individual variants has increased at a decelerating rate, since variants identified in larger studies tend to have smaller effects [346]. Because an AIL is derived from an F2 cross, its genetic diversity is limited to polymorphisms segregating between the parental strains. Thus, it may not be possible to detect the effects of loci with the potential to influence a trait if they were not polymorphic in the founders. However, the number of variants that could possibly influence a trait in an AIL is limited the to the number of SNPs segregating between the founders (~4 million in LG and SM), all of which are biallelic. Therefore, the proportion of phenotypic variance caused by genetic factors may be distributed across a smaller number of alleles, and individual effect sizes may be be larger. Of course, if trait heritability is low and the genetic architecture is highly polygenic, it will remain difficult to detect QTLs, but in general, restricting the search space to a few million variants will tend to increase power.

Another important corollary of low genetic diversity in an AIL is that rare alleles are virtually nonexistent. Having many observations for each allele or genotype class at a locus enhances power to detect phenotypic differences. This is because more observations leads to less variance around the mean phenotype for each allele or genotype, thereby allowing larger genetic effects to be observed (Figure 2.3).

A greater ability to control environmental variables in model organisms could also explain why smaller effect sizes are observed in humans. In addition, crosses between inbred strains are designed to maximize phenotypic variability (much like GWAS that sample individuals at the opposite extremes of the phenotypic spectrum). Thus, for certain traits, such as body size in LG $\times$ SM, it may be reasonable to expect larger effects, which could explain our observation that bone length, body weight, and muscle weight tended to have higher heritabilities and stronger QTL effects than the behavioral traits in this study.

4.5 Model organisms and human disease

Model organism research tends to use the gene as the unit of analysis because we assume that this will allow us to generalize from animal to human. However, if there are thousands of genes that all account for some fraction of the phenotype, then this approach may be limited. Accordingly, many have proposed a shift in focus from disease genes to disease networks [11, 262, 346].

It is difficult to predict what impact a candidate gene identified in the mouse might have for human disease, but it is also difficult to predict what impact a candidate gene identified in *humans* might have for human disease. Importantly, genetic studies in humans are essentially descriptive: they generate hypotheses. This is true of GWAS in general and applies equally to model organisms, but there is a crucial difference: model organisms can be used to generate hypotheses about disease mechanisms *and* they provide a system for testing those hypotheses experimentally.

Debates about the role of model organisms in human genetics can continue without end, but they will not enhance our understanding of biology. To learn about biology, we study model organisms, and we study genes; an improved understanding of disease is one consequence of this. Model organisms like the mouse allow us to study gene function within a biological context, the value of which will not cease to be relevant. I agree that a shift in focus from genes to networks is the right approach for thinking about disease, but I argue that in practice, the study of genes, transcripts, and proteins *in vivo* remains the best strategy for exploring disease mechanisms.

References

[1] Teri A. Manolio, Francis S. Collins, Nancy J. Cox, David B. Goldstein, Lucia A. Hindorff, David J. Hunter, Mark I. McCarthy, Erin M. Ramos, Lon R. Cardon, Aravinda Chakravarti, Judy H. Cho, Alan E. Guttmacher, Augustine Kong, Leonid Kruglyak, Elaine Mardis, Charles N. Rotimi, Montgomery Slatkin, David Valle, Alice S. Whittemore, Michael Boehnke, Andrew G. Clark, Evan E. Eichler, Greg Gibson, Jonathan L. Haines, Trudy F. C. Mackay, Steven A. McCarroll, and Peter M. Visscher. Finding the missing heritability of complex diseases. 461(7265):747–753, October 2009.

[2] Andrew DeWan, Mugen Liu, Stephen Hartman, Samuel Shao-Min Zhang, David T. L. Liu, Connie Zhao, Pancy O. S. Tam, Wai Man Chan, Dennis S. C. Lam, Michael Snyder, Colin Barnstable, Chi Pui Pang, and Josephine Hoh. HTRA1 Promoter Polymorphism in Wet Age-Related Macular Degeneration. 314(5801):989–992, November 2006.

[3] Robert J. Klein, Caroline Zeiss, Emily Y. Chew, Jen-Yue Tsai, Richard S. Sackler, Chad Haynes, Alice K. Henning, John Paul SanGiovanni, Shrikant M. Mane, Susan T. Mayne, Michael B. Bracken, Frederick L. Ferris, Jurg Ott, Colin Barnstable, and Josephine Hoh. Complement factor H polymorphism in age-related macular degeneration. *Science (New York, N.Y.)*, 308(5720):385–389, April 2005.

[4] Richard H. Duerr, Kent D. Taylor, Steven R. Brant, John D. Rioux, Mark S. Silverberg, Mark J. Daly, A. Hillary Steinhart, Clara Abraham, Miguel Regueiro, Anne Griffiths, Themistocles Dassopoulos, Alain Bitton, Huiying Yang, Stephan Targan, Lisa Wu Datta, Emily O. Kistner, L. Philip Schumm, Annette T. Lee, Peter K. Gregersen, M. Michael Barmada, Jerome I. Rotter, Dan L. Nicolae, and Judy H. Cho. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science (New York, N.Y.)*, 314(5804):1461–1463, December 2006.

[5] Peter M. Visscher, Matthew A. Brown, Mark I. McCarthy, and Jian Yang. Five Years of GWAS Discovery. *American Journal of Human Genetics*, 90(1):7–24, January 2012.

[6] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. 447(7145):661–678, June 2007.

[7] Joris A. Veltman and Han G. Brunner. De novo mutations in human genetic disease. *Nature Reviews. Genetics*, 13(8):565–575, July 2012.

[8] Ronald Aylmer Fisher. The Correlation Between Relatives on the Supposition of Mendelian Inheritance. 1918.

[9] Jian Yang, Beben Benyamin, Brian P. McEvoy, Scott Gordon, Anjali K. Henders, Dale R. Nyholt, Pamela A. Madden, Andrew C. Heath, Nicholas G. Martin, Grant W. Montgomery, Michael E. Goddard, and Peter M. Visscher. Common SNPs explain a

large proportion of the heritability for human height. *Nature Genetics*, 42(7):565–569, July 2010.

[10] Hilary K. Finucane, Brendan Bulik-Sullivan, Alexander Gusev, Gosia Trynka, Yakir Reshef, Po-Ru Loh, Verneri Anttila, Han Xu, Chongzhi Zang, Kyle Farh, Stephan Ripke, Felix R. Day, ReproGen Consortium, Schizophrenia Working Group of the Psychiatric Genomics Consortium, RACI Consortium, Shaun Purcell, Eli Stahl, Sara Lindstrom, John R. B. Perry, Yukinori Okada, Soumya Raychaudhuri, Mark J. Daly, Nick Patterson, Benjamin M. Neale, and Alkes L. Price. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, 47(11):1228–1235, November 2015.

[11] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An Expanded View of Complex Traits: From Polygenic to Omnigenic. 169(7):1177–1186, June 2017.

[12] Nitin S. Baliga, Johan L. M. Björkegren, Jef D. Boeke, Michael Boutros, Nigel P. S. Crawford, Aimée M. Dudley, Charles R. Farber, Allan Jones, Allan I. Levey, Aldons J. Lusis, H. Craig Mak, Joseph H. Nadeau, Marcus B. Noyes, Enrico Petretto, Nicholas T. Seyfried, Lars M. Steinmetz, and Sibylle C. Vonesch. The State of Systems Genetics in 2017. *Cell Systems*, 4(1):7–15, January 2017.

[13] Lenore Cowen, Trey Ideker, Benjamin J. Raphael, and Roded Sharan. Network propagation: A universal amplifier of genetic associations. *Nature Reviews. Genetics*, 18(9):551–562, September 2017.

[14] Joseph K. Pickrell, John C. Marioni, Athma A. Pai, Jacob F. Degner, Barbara E. Engelhardt, Everlyne Nkadori, Jean-Baptiste Veyrieras, Matthew Stephens, Yoav Gilad, and Jonathan K. Pritchard. Understanding mechanisms underlying human gene expression variation with RNA sequencing. 464(7289):768–772, April 2010.

[15] Heather E. Wheeler, Kaanan P. Shah, Jonathon Brenner, Tzintzuni Garcia, Keston Aquino-Michaels, GTEx Consortium, Nancy J. Cox, Dan L. Nicolae, and Hae Kyung Im. Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues. *PLOS Genetics*, 12(11):e1006423, November 2016.

[16] Mouse ENCODE Consortium, John A. Stamatoyannopoulos, Michael Snyder, Ross Hardison, Bing Ren, Thomas Gingeras, David M. Gilbert, Mark Groudine, Michael Bender, Rajinder Kaul, Theresa Canfield, Erica Giste, Audra Johnson, Mia Zhang, Gayathri Balasundaram, Rachel Byron, Vaughan Roach, Peter J. Sabo, Richard Sandstrom, A. Sandra Stehling, Robert E. Thurman, Sherman M. Weissman, Philip Cayting, Manoj Hariharan, Jin Lian, Yong Cheng, Stephen G. Landt, Zhihai Ma, Barbara J. Wold, Job Dekker, Gregory E. Crawford, Cheryl A. Keller, Weisheng Wu, Christopher Morrissey, Swathi A. Kumar, Tejaswini Mishra, Deepti Jain, Marta Byrska-Bishop, Daniel Blankenberg, Bryan R. Lajoie, Gaurav Jain, Amartya Sanyal, Kaun-Bei Chen, Olgert Denas, James Taylor, Gerd A. Blobel, Mitchell J. Weiss, Max Pimkin, Wulan

Deng, Georgi K. Marinov, Brian A. Williams, Katherine I. Fisher-Aylor, Gilberto Desalvo, Anthony Kiralusha, Diane Trout, Henry Amrhein, Ali Mortazavi, Lee Edsall, David McCleary, Samantha Kuan, Yin Shen, Feng Yue, Zhen Ye, Carrie A. Davis, Chris Zaleski, Sonali Jha, Chenghai Xue, Alex Dobin, Wei Lin, Meagan Fastuca, Huaien Wang, Roderic Guigo, Sarah Djebali, Julien Lagarde, Tyrone Ryba, Takayo Sasaki, Venkat S. Malladi, Melissa S. Cline, Vanessa M. Kirkup, Katrina Learned, Kate R. Rosenbloom, W. James Kent, Elise A. Feingold, Peter J. Good, Michael Pazin, Rebecca F. Lowdon, and Leslie B. Adams. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biology*, 13(8):418, August 2012.

[17] Andrew B. Stergachis, Shane Neph, Richard Sandstrom, Eric Haugen, Alex P. Reynolds, Miaohua Zhang, Rachel Byron, Theresa Canfield, Sandra Stelhing-Sun, Kristen Lee, Robert E. Thurman, Shinny Vong, Daniel Bates, Fidencio Neri, Morgan Diegel, Erika Giste, Douglas Dunn, Jeff Vierstra, R. Scott Hansen, Audra K. Johnson, Peter J. Sabo, Matthew S. Wilken, Thomas A. Reh, Piper M. Treuting, Rajinder Kaul, Mark Groudine, M. A. Bender, Elhanan Borenstein, and John A. Stamatoyannopoulos. Conservation of trans-acting circuitry during mammalian regulatory evolution. 515(7527):365–370, November 2014.

[18] Olgert Denas, Richard Sandstrom, Yong Cheng, Kathryn Beal, Javier Herrero, Ross C. Hardison, and James Taylor. Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. *BMC Genomics*, 16:87, 2015.

[19] Scott Smemo, Juan J. Tena, Kyoung-Han Kim, Eric R. Gamazon, Noboru J. Sakabe, Carlos Gómez-Marín, Ivy Aneas, Flavia L. Credidio, Débora R. Sobreira, Nora F. Wasserman, Ju Hee Lee, Vijitha Puviindran, Davis Tam, Michael Shen, Joe Eun Son, Niki Alizadeh Vakili, Hoon-Ki Sung, Silvia Naranjo, Rafael D. Acemel, Miguel Manzanares, Andras Nagy, Nancy J. Cox, Chi-Chung Hui, Jose Luis Gomez-Skarmeta, and Marcelo A. Nóbrega. Obesity-associated variants within FTO form long-range functional connections with IRX3. 507(7492):371–375, March 2014.

[20] Geet Duggal, Hao Wang, and Carl Kingsford. Higher-order chromatin domains link eQTLs with the expression of far-away genes. *Nucleic Acids Research*, 42(1):87–96, January 2014.

[21] Hyejung Won, Luis de la Torre-Ubieta, Jason L. Stein, Neelroop N. Parikshak, Jerry Huang, Carli K. Opland, Michael J. Gandal, Gavin J. Sutton, Farhad Hormozdiari, Daning Lu, Changhoon Lee, Eleazar Eskin, Irina Voineagu, Jason Ernst, and Daniel H. Geschwind. Chromosome conformation elucidates regulatory relationships in developing human brain. 538(7626):523–527, 10 27, 2016.

[22] Adaikalavan Ramasamy, Daniah Trabzuni, Sebastian Guelfi, Vibin Varghese, Colin Smith, Robert Walker, Tisham De, UK Brain Expression Consortium, North American Brain Expression Consortium, Lachlan Coin, Rohan de Silva, Mark R. Cookson,

Andrew B. Singleton, John Hardy, Mina Ryten, and Michael E. Weale. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nature Neuroscience*, 17(10):1418–1428, October 2014.

[23] Dan L. Nicolae, Eric Gamazon, Wei Zhang, Shiwei Duan, M. Eileen Dolan, and Nancy J. Cox. Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS. *PLoS genetics*, 6(4):e1000888, April 2010.

[24] Richard C. Davis, Atila van Nas, Lawrence W. Castellani, Yi Zhao, Zhiqiang Zhou, Pingzi Wen, Suzanne Yu, Hongxiu Qi, Melenie Rosales, Eric E. Schadt, Karl W. Broman, Miklós Péterfy, and Aldons J. Lusis. Systems genetics of susceptibility to obesity-induced diabetes in mice. *Physiological Genomics*, 44(1):1–13, January 2012.

[25] Xin He, Chris K. Fuller, Yi Song, Qingying Meng, Bin Zhang, Xia Yang, and Hao Li. Sherlock: Detecting gene-disease associations by matching patterns of expression QTL and GWAS. *American Journal of Human Genetics*, 92(5):667–680, May 2013.

[26] Darren A. Cusanovich, Minal Caliskan, Christine Billstrand, Katelyn Michelini, Claudia Chavarria, Sherryl De Leon, Amy Mitrano, Noah Lewellyn, Jack A. Elias, Geoffrey L. Chupp, Roberto M. Lang, Sanjiv J. Shah, Jeanne M. Decara, Yoav Gilad, and Carole Ober. Integrated analyses of gene expression and genetic association studies in a founder population. *Human Molecular Genetics*, 25(10):2104–2112, May 2016.

[27] Mete Civelek, Ying Wu, Calvin Pan, Chelsea K. Raulerson, Arthur Ko, Aiqing He, Charles Tilford, Niyas K. Saleem, Alena Stančáková, Laura J. Scott, Christian Fuchsberger, Heather M. Stringham, Anne U. Jackson, Narisu Narisu, Peter S. Chines, Kerrin S. Small, Johanna Kuusisto, Brian W. Parks, Päivi Pajukanta, Todd Kirchgessner, Francis S. Collins, Peter S. Gargalovic, Michael Boehnke, Markku Laakso, Karen L. Mohlke, and Aldons J. Lusis. Genetic Regulation of Adipose Gene Expression and Cardio-Metabolic Traits. *American Journal of Human Genetics*, 100(3):428–443, March 2017.

[28] Jennifer M. Whitehead Pavlides, Zhihong Zhu, Jacob Gratten, Allan F. McRae, Naomi R. Wray, and Jian Yang. Predicting gene targets from integrative analyses of summary data from GWAS and eQTL studies for 28 human complex traits. *Genome Medicine*, 8(1):84, 08 09, 2016.

[29] Daniel H. Geschwind and Jonathan Flint. Genetics and genomics of psychiatric disease. *Science (New York, N.Y.)*, 349(6255):1489–1494, September 2015.

[30] Ioanna Tachmazidou, Dániel Süveges, Josine L. Min, Graham R. S. Ritchie, Julia Steinberg, Klaudia Walter, Valentina Iotchkova, Jeremy Schwartzentruber, Jie Huang, Yasin Memari, Shane McCarthy, Andrew A. Crawford, Cristina Bombieri, Massimiliano Cocca, Aliki-Eleni Farmaki, Tom R. Gaunt, Pekka Jousilahti, Marjolein N. Kooijman, Benjamin Lehne, Giovanni Malerba, Satu Männistö, Angela Matchan, Carolina

Medina-Gomez, Sarah J. Metrustry, Abhishek Nag, Ioanna Ntalla, Lavinia Paternoster, Nigel W. Rayner, Cinzia Sala, William R. Scott, Hashem A. Shihab, Lorraine Southam, Beate St Pourcain, Michela Traglia, Katerina Trajanoska, Gialuigi Zaza, Weihua Zhang, María S. Artigas, Narinder Bansal, Marianne Benn, Zhongsheng Chen, Petr Danecek, Wei-Yu Lin, Adam Locke, Jian'an Luan, Alisa K. Manning, Antonella Mulas, Carlo Sidore, Anne Tybjaerg-Hansen, Anette Varbo, Magdalena Zoledziewska, Chris Finan, Konstantinos Hatzikotoulas, Audrey E. Hendricks, John P. Kemp, Alireza Moayyeri, Kalliope Panoutsopoulou, Michal Szpak, Scott G. Wilson, Michael Boehnke, Francesco Cucca, Emanuele Di Angelantonio, Claudia Langenberg, Cecilia Lindgren, Mark I. McCarthy, Andrew P. Morris, Børge G. Nordestgaard, Robert A. Scott, Martin D. Tobin, Nicholas J. Wareham, SpiroMeta Consortium, GoT2D Consortium, Paul Burton, John C. Chambers, George Davey Smith, George Dedoussis, Janine F. Felix, Oscar H. Franco, Giovanni Gambaro, Paolo Gasparini, Christopher J. Hammond, Albert Hofman, Vincent W. V. Jaddoe, Marcus Kleber, Jaspal S. Kooner, Markus Perola, Caroline Relton, Susan M. Ring, Fernando Rivadeneira, Veikko Salomaa, Timothy D. Spector, Oliver Stegle, Daniela Toniolo, André G. Uitterlinden, arcOGEN Consortium, Understanding Society Scientific Group, UK10K Consortium, Inês Barroso, Celia M. T. Greenwood, John R. B. Perry, Brian R. Walker, Adam S. Butterworth, Yali Xue, Richard Durbin, Kerrin S. Small, Nicole Soranzo, Nicholas J. Timpson, and Eleftheria Zeggini. Whole-Genome Sequencing Coupled to Imputation Discovers Genetic Signals for Anthropometric Traits. *American Journal of Human Genetics*, 100(6):865–884, June 2017.

[31] Jian Yang, Andrew Bakshi, Zhihong Zhu, Gibran Hemani, Anna A. E. Vinkhuyzen, Sang Hong Lee, Matthew R. Robinson, John R. B. Perry, Ilja M. Nolte, Jana V. van Vliet-Ostaptchouk, Harold Snieder, LifeLines Cohort Study, Tonu Esko, Lili Milani, Reedik Mägi, Andres Metspalu, Anders Hamsten, Patrik K. E. Magnusson, Nancy L. Pedersen, Erik Ingelsson, Nicole Soranzo, Matthew C. Keller, Naomi R. Wray, Michael E. Goddard, and Peter M. Visscher. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics*, 47(10):1114–1120, October 2015.

[32] Bogdan Pasaniuc, Nadin Rohland, Paul J. McLaren, Kiran Garimella, Noah Zaitlen, Heng Li, Namrata Gupta, Benjamin M. Neale, Mark J. Daly, Pamela Sklar, Patrick F. Sullivan, Sarah Bergen, Jennifer L. Moran, Christina M. Hultman, Paul Lichtenstein, Patrik Magnusson, Shaun M. Purcell, David W. Haas, Liming Liang, Shamil Sunyaev, Nick Patterson, Paul I. W. de Bakker, David Reich, and Alkes L. Price. Extremely low-coverage sequencing and imputation increases power for genome-wide association studies. *Nature Genetics*, 44(6):631–635, May 2012.

[33] Stephan J. Sanders, Michael T. Murtha, Abha R. Gupta, John D. Murdoch, Melanie J. Raubeson, A. Jeremy Willsey, A. Gulhan Ercan-Sencicek, Nicholas M. DiLullo, Neelroop N. Parikshak, Jason L. Stein, Michael F. Walker, Gordon T. Ober, Nicole A. Teran, Youeun Song, Paul El-Fishawy, Ryan C. Murtha, Murim Choi, John D. Overton, Robert D. Bjornson, Nicholas J. Carriero, Kyle A. Meyer, Kaya Bilguvar,

Shrikant M. Mane, Nenad Sestan, Richard P. Lifton, Murat Günel, Kathryn Roeder, Daniel H. Geschwind, Bernie Devlin, and Matthew W. State. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. 485(7397):237–241, April 2012.

[34] Michelle M. Stein, Cara L. Hrusch, Justyna Gozdz, Catherine Igartua, Vadim Pivniouk, Sean E. Murray, Julie G. Ledford, Mauricius Marques dos Santos, Rebecca L. Anderson, Nervana Metwali, Julia W. Neilson, Raina M. Maier, Jack A. Gilbert, Mark Holbreich, Peter S. Thorne, Fernando D. Martinez, Erika von Mutius, Donata Vercelli, Carole Ober, and Anne I. Sperling. Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *The New England Journal of Medicine*, 375(5):411–421, August 2016.

[35] T. B. Bigdeli, S. Ripke, R. E. Peterson, M. Trzaskowski, S.-A. Bacanu, A. Abdellaoui, T. F. M. Andlauer, A. T. F. Beekman, K. Berger, D. H. R. Blackwood, D. I. Boomsma, G. Breen, H. N. Buttenschøn, E. M. Byrne, S. Cichon, T.-K. Clarke, B. Couvy-Duchesne, N. Craddock, E. J. C. de Geus, F. Degenhardt, E. C. Dunn, A. C. Edwards, A. H. Fanous, A. J. Forstner, J. Frank, M. Gill, S. D. Gordon, H. J. Grabe, S. P. Hamilton, O. Hardiman, C. Hayward, A. C. Heath, A. K. Henders, S. Herms, I. B. Hickie, P. Hoffmann, G. Homuth, J.-J. Hottenga, M. Ising, R. Jansen, S. Kloiber, J. A. Knowles, M. Lang, Q. S. Li, S. Lucae, D. J. MacIntyre, P. a. F. Madden, N. G. Martin, P. J. McGrath, P. McGuffin, A. M. McIntosh, S. E. Medland, D. Mehta, C. M. Middeldorp, Y. Milaneschi, G. W. Montgomery, O. Mors, B. Müller-Myhsok, M. Nauck, D. R. Nyholt, M. M. Nöthen, M. J. Owen, B. W. J. H. Penninx, M. L. Pergadia, R. H. Perlis, W. J. Peyrot, D. J. Porteous, J. B. Potash, J. P. Rice, M. Rietschel, B. P. Riley, M. Rivera, R. Schoevers, T. G. Schulze, J. Shi, S. I. Shyn, J. H. Smit, J. W. Smoller, F. Streit, J. Strohmaier, A. Teumer, J. Treutlein, S. Van der Auwera, G. van Grootheest, A. M. van Hemert, H. Völzke, B. T. Webb, M. M. Weissman, J. Wellmann, G. Willemsen, S. H. Witt, D. F. Levinson, C. M. Lewis, N. R. Wray, J. Flint, P. F. Sullivan, and K. S. Kendler. Genetic effects influencing risk for major depressive disorder in China and Europe. *Translational Psychiatry*, 7(3):e1074, March 2017.

[36] Yanni Zeng, Pau Navarro, Ana M. Fernandez-Pujals, Lynsey S. Hall, Toni-Kim Clarke, Pippa A. Thomson, Blair H. Smith, Lynne J. Hocking, Sandosh Padmanabhan, Caroline Hayward, Donald J. MacIntyre, Naomi R. Wray, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, Ian J. Deary, David J. Porteous, Chris S. Haley, and Andrew M. McIntosh. A Combined Pathway and Regional Heritability Analysis Indicates NETRIN1 Pathway Is Associated With Major Depressive Disorder. *Biological Psychiatry*, 81(4):336–346, February 2017.

[37] Xiang Zhou, Peter Carbonetto, and Matthew Stephens. Polygenic modeling with bayesian sparse linear mixed models. *PLoS genetics*, 9(2):e1003264, 2013.

[38] Brendan K. Bulik-Sullivan, Po-Ru Loh, Hilary K. Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patter-

son, Mark J. Daly, Alkes L. Price, and Benjamin M. Neale. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, 47(3):291–295, March 2015.

[39] Eric R. Gamazon, Heather E. Wheeler, Kaanan P. Shah, Sahar V. Mozaffari, Keston Aquino-Michaels, Robert J. Carroll, Anne E. Eyler, Joshua C. Denny, GTEx Consortium, Dan L. Nicolae, Nancy J. Cox, and Hae Kyung Im. A gene-based association method for mapping traits using reference transcriptome data. *Nature Genetics*, 47(9):1091–1098, September 2015.

[40] Judith A. Blake, Janan T. Eppig, James A. Kadin, Joel E. Richardson, Cynthia L. Smith, Carol J. Bult, and the Mouse Genome Database Group. Mouse Genome Database (MGD)-2017: Community knowledge resource for the laboratory mouse. *Nucleic Acids Research*, 45(D1):D723–D729, January 2017.

[41] Arthur Liberzon, Chet Birger, Helga Thorvaldsdóttir, Mahmoud Ghandi, Jill P. Mesirov, and Pablo Tamayo. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Systems*, 1(6):417–425, December 2015.

[42] 1000 Genomes Project Consortium, Adam Auton, Lisa D. Brooks, Richard M. Durbin, Erik P. Garrison, Hyun Min Kang, Jan O. Korbel, Jonathan L. Marchini, Shane McCarthy, Gil A. McVean, and Gonçalo R. Abecasis. A global reference for human genetic variation. 526(7571):68–74, October 2015.

[43] Sarah A. Gagliano. It's All in the Brain: A Review of Available Functional Genomic Annotations. *Biological Psychiatry*, 81(6):478–483, March 2017.

[44] Roadmap Epigenomics Consortium, Anshul Kundaje, Wouter Meuleman, Jason Ernst, Misha Bilenky, Angela Yen, Alireza Heravi-Moussavi, Pouya Kheradpour, Zhizhuo Zhang, Jianrong Wang, Michael J. Ziller, Viren Amin, John W. Whitaker, Matthew D. Schultz, Lucas D. Ward, Abhishek Sarkar, Gerald Quon, Richard S. Sandstrom, Matthew L. Eaton, Yi-Chieh Wu, Andreas R. Pfenning, Xinchen Wang, Melina Claussnitzer, Yaping Liu, Cristian Coarfa, R. Alan Harris, Noam Shoresh, Charles B. Epstein, Elizabeta Gjoneska, Danny Leung, Wei Xie, R. David Hawkins, Ryan Lister, Chibo Hong, Philippe Gascard, Andrew J. Mungall, Richard Moore, Eric Chuah, Angela Tam, Theresa K. Canfield, R. Scott Hansen, Rajinder Kaul, Peter J. Sabo, Mukul S. Bansal, Annaick Carles, Jesse R. Dixon, Kai-How Farh, Soheil Feizi, Rosa Karlic, Ah-Ram Kim, Ashwinikumar Kulkarni, Daofeng Li, Rebecca Lowdon, GiNell Elliott, Tim R. Mercer, Shane J. Neph, Vitor Onuchic, Paz Polak, Nisha Rajagopal, Pradipta Ray, Richard C. Sallari, Kyle T. Siebenthall, Nicholas A. Sinnott-Armstrong, Michael Stevens, Robert E. Thurman, Jie Wu, Bo Zhang, Xin Zhou, Arthur E. Beaudet, Laurie A. Boyer, Philip L. De Jager, Peggy J. Farnham, Susan J. Fisher, David Haussler, Steven J. M. Jones, Wei Li, Marco A. Marra, Michael T. McManus, Shamil Sunyaev, James A. Thomson, Thea D. Tlsty, Li-Huei Tsai, Wei Wang, Robert A. Waterland, Michael Q. Zhang, Lisa H. Chadwick, Bradley E. Bernstein, Joseph F. Costello,

Joseph R. Ecker, Martin Hirst, Alexander Meissner, Aleksandar Milosavljevic, Bing Ren, John A. Stamatoyannopoulos, Ting Wang, and Manolis Kellis. Integrative analysis of 111 reference human epigenomes. 518(7539):317–330, February 2015.

[45] Carlos D. Bustamante, Esteban González Burchard, and Francisco M. De la Vega. Genomics for the world. 475(7355):163–165, July 2011.

[46] Neal R. Swerdlow, Martin Weber, Ying Qu, Gregory A. Light, and David L. Braff. Realistic expectations of prepulse inhibition in translational models for schizophrenia research. 199(3):331–388, August 2008.

[47] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: Author., 5th edition, 2013.

[48] Cross-Disorder Group of the Psychiatric Genomics Consortium, S. Hong Lee, Stephan Ripke, Benjamin M. Neale, Stephen V. Faraone, Shaun M. Purcell, Roy H. Perlis, Bryan J. Mowry, Anita Thapar, Michael E. Goddard, John S. Witte, Devin Absher, Ingrid Agartz, Huda Akil, Farooq Amin, Ole A. Andreassen, Adebayo Anjorin, Richard Anney, Verneri Anttila, Dan E. Arking, Philip Asherson, Maria H. Azevedo, Lena Backlund, Judith A. Badner, Anthony J. Bailey, Tobias Banaschewski, Jack D. Barchas, Michael R. Barnes, Thomas B. Barrett, Nicholas Bass, Agatino Battaglia, Michael Bauer, Mònica Bayés, Frank Bellivier, Sarah E. Bergen, Wade Berrettini, Catalina Betancur, Thomas Bettecken, Joseph Biederman, Elisabeth B. Binder, Donald W. Black, Douglas H. R. Blackwood, Cinnamon S. Bloss, Michael Boehnke, Dorret I. Boomsma, Gerome Breen, René Breuer, Richard Bruggeman, Paul Cormican, Nancy G. Buccola, Jan K. Buitelaar, William E. Bunney, Joseph D. Buxbaum, William F. Byerley, Enda M. Byrne, Sian Caesar, Wiepke Cahn, Rita M. Cantor, Miguel Casas, Aravinda Chakravarti, Kimberly Chambert, Khalid Choudhury, Sven Cichon, C. Robert Cloninger, David A. Collier, Edwin H. Cook, Hilary Coon, Bru Cormand, Aiden Corvin, William H. Coryell, David W. Craig, Ian W. Craig, Jennifer Crosbie, Michael L. Cuccaro, David Curtis, Darina Czamara, Susmita Datta, Geraldine Dawson, Richard Day, Eco J. De Geus, Franziska Degenhardt, Srdjan Djurovic, Gary J. Donohoe, Alysa E. Doyle, Jubao Duan, Frank Dudbridge, Eftichia Duketis, Richard P. Ebstein, Howard J. Edenberg, Josephine Elia, Sean Ennis, Bruno Etain, Ayman Fanous, Anne E. Farmer, I. Nicol Ferrier, Matthew Flickinger, Eric Fombonne, Tatiana Foroud, Josef Frank, Barbara Franke, Christine Fraser, Robert Freedman, Nelson B. Freimer, Christine M. Freitag, Marion Friedl, Louise Frisén, Louise Gallagher, Pablo V. Gejman, Lyudmila Georgieva, Elliot S. Gershon, Daniel H. Geschwind, Ina Giegling, Michael Gill, Scott D. Gordon, Katherine Gordon-Smith, Elaine K. Green, Tiffany A. Greenwood, Dorothy E. Grice, Magdalena Gross, Detelina Grozeva, Weihua Guan, Hugh Gurling, Lieuwe De Haan, Jonathan L. Haines, Hakon Hakonarson, Joachim Hallmayer, Steven P. Hamilton, Marian L. Hamshere, Thomas F. Hansen, Annette M. Hartmann, Martin Hautzinger, Andrew C. Heath, Anjali K. Henders, Stefan Herms, Ian B. Hickie, Maria Hipolito, Susanne Hoefels, Peter A. Holmans, Florian Holsboer, Witte J. Hoogendijk, Jouke-Jan Hottenga, Christina M. Hultman, Vanessa

Hus, Andrés Ingason, Marcus Ising, Stéphane Jamain, Edward G. Jones, Ian Jones, Lisa Jones, Jung-Ying Tzeng, Anna K. Kähler, René S. Kahn, Radhika Kandaswamy, Matthew C. Keller, James L. Kennedy, Elaine Kenny, Lindsey Kent, Yunjung Kim, George K. Kirov, Sabine M. Klauck, Lambertus Klei, James A. Knowles, Martin A. Kohli, Daniel L. Koller, Bettina Konte, Ania Korszun, Lydia Krabbendam, Robert Krasucki, Jonna Kuntsi, Phoenix Kwan, Mikael Landén, Niklas Långström, Mark Lathrop, Jacob Lawrence, William B. Lawson, Marion Leboyer, David H. Ledbetter, Phil H. Lee, Todd Lencz, Klaus-Peter Lesch, Douglas F. Levinson, Cathryn M. Lewis, Jun Li, Paul Lichtenstein, Jeffrey A. Lieberman, Dan-Yu Lin, Don H. Linszen, Chunyu Liu, Falk W. Lohoff, Sandra K. Loo, Catherine Lord, Jennifer K. Lowe, Susanne Lucae, Donald J. MacIntyre, Pamela A. F. Madden, Elena Maestrini, Patrik K. E. Magnusson, Pamela B. Mahon, Wolfgang Maier, Anil K. Malhotra, Shrikant M. Mane, Christa L. Martin, Nicholas G. Martin, Manuel Mattheisen, Keith Matthews, Morten Mattingsdal, Steven A. McCarroll, Kevin A. McGhee, James J. McGough, Patrick J. McGrath, Peter McGuffin, Melvin G. McInnis, Andrew McIntosh, Rebecca McKinney, Alan W. McLean, Francis J. McMahon, William M. McMahon, Andrew McQuillin, Helena Medeiros, Sarah E. Medland, Sandra Meier, Ingrid Melle, Fan Meng, Jobst Meyer, Christel M. Middeldorp, Lefkos Middleton, Vihra Milanova, Ana Miranda, Anthony P. Monaco, Grant W. Montgomery, Jennifer L. Moran, Daniel Moreno-De-Luca, Gunnar Morken, Derek W. Morris, Eric M. Morrow, Valentina Moskvina, Pierandrea Muglia, Thomas W. Mühleisen, Walter J. Muir, Bertram Müller-Myhsok, Michael Murtha, Richard M. Myers, Inez Myin-Germeys, Michael C. Neale, Stan F. Nelson, Caroline M. Nievergelt, Ivan Nikolov, Vishwajit Nimgaonkar, Willem A. Nolen, Markus M. Nöthen, John I. Nurnberger, Evaristus A. Nwulia, Dale R. Nyholt, Colm O'Dushlaine, Robert D. Oades, Ann Olincy, Guiomar Oliveira, Line Olsen, Roel A. Ophoff, Urban Osby, Michael J. Owen, Aarno Palotie, Jeremy R. Parr, Andrew D. Paterson, Carlos N. Pato, Michele T. Pato, Brenda W. Penninx, Michele L. Pergadia, Margaret A. Pericak-Vance, Benjamin S. Pickard, Jonathan Pimm, Joseph Piven, Danielle Posthuma, James B. Potash, Fritz Poustka, Peter Propping, Vinay Puri, Digby J. Quested, Emma M. Quinn, Josep Antoni Ramos-Quiroga, Henrik B. Rasmussen, Soumya Raychaudhuri, Karola Rehnström, Andreas Reif, Marta Ribasés, John P. Ric. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nature Genetics*, 45(9):984–994, September 2013.

[49] Andreas Meyer-Lindenberg and Daniel R. Weinberger. Intermediate phenotypes and genetic mechanisms of psychiatric disorders. *Nature Reviews. Neuroscience*, 7(10):818–827, October 2006.

[50] Robert A. Power, Stacy Steinberg, Gyda Bjornsdottir, Cornelius A. Rietveld, Abdel Abdellaoui, Michel M. Nivard, Magnus Johannesson, Tessel E. Galesloot, Jouke J. Hottenga, Gonneke Willemsen, David Cesarini, Daniel J. Benjamin, Patrik K. E. Magnusson, Fredrik Ullén, Henning Tiemeier, Albert Hofman, Frank J. A. van Rooij, G. Bragi Walters, Engilbert Sigurdsson, Thorgeir E. Thorgeirsson, Andres Ingason, Agnar Helgason, Augustine Kong, Lambertus A. Kiemeney, Philipp Koellinger, Dorret I.

Boomsma, Daniel Gudbjartsson, Hreinn Stefansson, and Kari Stefansson. Polygenic risk scores for schizophrenia and bipolar disorder predict creativity. *Nature Neuroscience*, 18(7):953–955, July 2015.

[51] Dean Carcone and Anthony C. Ruocco. Six Years of Research on the National Institute of Mental Health's Research Domain Criteria (RDoC) Initiative: A Systematic Review. *Frontiers in Cellular Neuroscience*, 11:46, 2017.

[52] Clarissa C. Parker, Hao Chen, Shelly B. Flagel, Aron M. Geurts, Jerry B. Richards, Terry E. Robinson, Leah C. Solberg Woods, and Abraham A. Palmer. Rats are the smart choice: Rationale for a renewed focus on rats in behavioral genetics. 76 Pt B:250–258, January 2014.

[53] Rat Genome Sequencing and Mapping Consortium, Amelie Baud, Roel Hermsen, Victor Guryev, Pernilla Stridh, Delyth Graham, Martin W. McBride, Tatiana Foroud, Sophie Calderari, Margarita Diez, Johan Ockinger, Amennai D. Beyeen, Alan Gillett, Nada Abdelmagid, Andre Ortlieb Guerreiro-Cacais, Maja Jagodic, Jonatan Tuncel, Ulrika Norin, Elisabeth Beattie, Ngan Huynh, William H. Miller, Daniel L. Koller, Imranul Alam, Samreen Falak, Mary Osborne-Pellegrin, Esther Martinez-Membrives, Toni Canete, Gloria Blazquez, Elia Vicens-Costa, Carme Mont-Cardona, Sira Diaz-Moran, Adolf Tobena, Oliver Hummel, Diana Zelenika, Kathrin Saar, Giannino Patone, Anja Bauerfeind, Marie-Therese Bihoreau, Matthias Heinig, Young-Ae Lee, Carola Rintisch, Herbert Schulz, David A. Wheeler, Kim C. Worley, Donna M. Muzny, Richard A. Gibbs, Mark Lathrop, Nico Lansu, Pim Toonen, Frans Paul Ruzius, Ewart de Bruijn, Heidi Hauser, David J. Adams, Thomas Keane, Santosh S. Atanur, Tim J. Aitman, Paul Flicek, Tomas Malinauskas, E. Yvonne Jones, Diana Ekman, Regina Lopez-Aumatell, Anna F. Dominiczak, Martina Johannesson, Rikard Holmdahl, Tomas Olsson, Dominique Gauguier, Norbert Hubner, Alberto Fernandez-Teruel, Edwin Cuppen, Richard Mott, and Jonathan Flint. Combined sequence-based and genetic mapping analysis of complex traits in outbred rats. *Nature Genetics*, 45(7):767–775, July 2013.

[54] Amelie Baud and Jonathan Flint. Identifying genes for neurobehavioural traits in rodents: Progress and pitfalls. *Disease Models & Mechanisms*, 10(4):373–383, April 2017.

[55] Alessandra Breschi, Thomas R. Gingeras, and Roderic Guigó. Comparative transcriptomics in human and mouse. *Nature Reviews. Genetics*, 18(7):425–440, July 2017.

[56] Steve D. M. Brown, Wolfgang Wurst, Ralf Kühn, and John M. Hancock. The Functional Annotation of Mammalian Genomes: The Challenge of Phenotyping. *Annual Review of Genetics*, 43(1):305–333, 2009.

[57] Janan T. Eppig, Joel E. Richardson, James A. Kadin, Martin Ringwald, Judith A. Blake, and Carol J. Bult. Mouse Genome Informatics (MGI): Reflecting on 25 years. *Mammalian Genome*, 26(7-8):272–284, 2015.

[58] Kevin J. Mitchell, Z. Josh Huang, Bita Moghaddam, and Akira Sawa. Following the genes: A framework for animal modeling of psychiatric disorders. *BMC biology*, 9:76, November 2011.

[59] Bruno B. Averbeck and Matthew V. Chafee. Using model systems to understand errant plasticity mechanisms in psychiatric disorders. *Nature Neuroscience*, 19(11):1418–1425, October 2016.

[60] Elmira Anderzhanova, Thomas Kirmeier, and Carsten T. Wotjak. Animal models in psychiatric research: The RDoC system as a new framework for endophenotype-oriented translational neuroscience. *Neurobiology of Stress*, 7:47–56, December 2017.

[61] Jonathan Flint and Trudy F. C. Mackay. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Research*, 19(5):723–733, May 2009.

[62] Clarissa C. Parker, Shyam Gopalakrishnan, Peter Carbonetto, Natalia M. Gonzales, Emily Leung, Yeonhee J. Park, Emmanuel Aryee, Joe Davis, David A. Blizard, Cheryl L. Ackert-Bicknell, Arimantas Lionikas, Jonathan K. Pritchard, and Abraham A. Palmer. Genome-wide association study of behavioral, physiological and gene expression traits in outbred CFW mice. *Nature Genetics*, 48(8):919–926, August 2016.

[63] Jérôme Nicod, Robert W. Davies, Na Cai, Carl Hassett, Leo Goodstadt, Cormac Cosgrove, Benjamin K Yee, Vikte Lionikaite, Rebecca E McIntyre, Carol Ann Remme, Elisabeth M. Lodder, Jennifer S. Gregory, Tertius Hough, Russell Joynson, Hayley Phelps, Barbara Nell, Clare Rowe, Joe Wood, Alison Walling, Nasrin Bopp, Amarjit Bhomra, Polinka Hernandez-Pliego, Jacques Callebert, Richard M. Aspden, Nick P Talbot, Peter A Robbins, Mark Harrison, Martin Fray, Jean-Marie Launay, Yigal M. Pinto, David A. Blizard, Connie R. Bezzina, David J Adams, Paul Franken, Tom Weaver, Sara Wells, Steve DM Brown, Paul K Potter, Paul Klenerman, Arimantas Lionikas, Richard Mott, and Jonathan Flint. Genome-wide association of multiple complex traits in outbred mice by ultra low-coverage sequencing. *Nature genetics*, 48(8):912–918, August 2016.

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

[65] A. Darvasi and M. Soller. Advanced intercross lines, an experimental population for fine genetic mapping. 141(3):1199–1207, November 1995.

[66] Ruth Chia, Francesca Achilli, Michael F. W. Festing, and Elizabeth M. C. Fisher. The origins and uses of mouse outbred stocks. *Nature Genetics*, 37(11):1181–1186, November 2005.

[67] Riyan Cheng, Clarissa C. Parker, Mark Abney, and Abraham A. Palmer. Practical considerations regarding the use of genotype and pedigree data to model relatedness in

the context of genome-wide association studies. *G3 (Bethesda, Md.)*, 3(10):1861–1867, October 2013.

[68] Natalia M. Gonzales and Abraham A. Palmer. Fine-mapping QTLs in advanced intercross lines and other outbred populations. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 25(7-8):271–292, August 2014.

[69] Jennifer Listgarten, Christoph Lippert, Carl M. Kadie, Robert I. Davidson, Eleazar Eskin, and David Heckerman. Improved linear mixed models for genome-wide association studies. *Nature Methods*, 9(6):525–526, May 2012.

[70] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. 155(2):945–959, June 2000.

[71] John Novembre, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, Karen S. King, Sven Bergmann, Matthew R. Nelson, Matthew Stephens, and Carlos D. Bustamante. Genes mirror geography within Europe. 456(7218):98–101, November 2008.

[72] Jianming Yu, Gael Pressoir, William H. Briggs, Irie Vroh Bi, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, Brandon S. Gaut, Dahlia M. Nielsen, James B. Holland, Stephen Kresovich, and Edward S. Buckler. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2):203–208, February 2006.

[73] B. Devlin and K. Roeder. Genomic control for association studies. 55(4):997–1004, December 1999.

[74] Mark Abney, Carole Ober, and Mary Sara McPeek. Quantitative-trait homozygosity and association mapping and empirical genomewide significance in large, complex pedigrees: Fasting serum-insulin level in the Hutterites. *American Journal of Human Genetics*, 70(4):920–934, April 2002.

[75] Riyan Cheng, Jackie E. Lim, Kaitlin E. Samocha, Greta Sokoloff, Mark Abney, Andrew D. Skol, and Abraham A. Palmer. Genome-wide association studies and the problem of relatedness among advanced intercross lines and other highly recombinant populations. 185(3):1033–1044, July 2010.

[76] Mete Civelek and Aldons J. Lusis. Systems genetics approaches to understand complex traits. *Nature Reviews. Genetics*, 15(1):34–48, January 2014.

[77] Abraham A. Palmer and Harriet de Wit. Translational genetic approaches to substance use disorders: Bridging the gap between mice and humans. *Human Genetics*, 131(6):931–939, June 2012.

[78] Ben Lehner. Genotype to phenotype: Lessons from model organisms for human genetics. *Nature Reviews. Genetics*, 14(3):168–178, March 2013.

[79] Miguel A. Gama Sosa, Rita De Gasperi, and Gregory A. Elder. Modeling human neurodegenerative diseases in transgenic systems. *Human Genetics*, 131(4):535–563, April 2012.

[80] Aida Moreno-Moral and Enrico Petretto. From integrative genomics to systems genetics in the rat to link genotypes to phenotypes. *Disease Models & Mechanisms*, 9(10):1097–1110, October 2016.

[81] David C. Glahn, Emma E. M. Knowles, D. Reese McKay, Emma Sprooten, Henriette Raventós, John Blangero, Irving I. Gottesman, and Laura Almasy. Arguments for the sake of endophenotypes: Examining common misconceptions about the use of endophenotypes in psychiatric genetics. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics*, 165B(2):122–130, March 2014.

[82] Jonathan Flint, Nicholas Timpson, and Marcus Munafò. Assessing the utility of intermediate phenotypes for genetic mapping of psychiatric disease. *Trends in Neurosciences*, 37(12):733–741, December 2014.

[83] David Goldman and Francesca Ducci. Deconstruction of vulnerability to complex diseases: Enhanced effect sizes and power of intermediate phenotypes. 7:124–130, November 2007.

[84] Tanya M. Teslovich, Kiran Musunuru, Albert V. Smith, Andrew C. Edmondson, Ioannis M. Stylianou, Masahiro Koseki, James P. Pirruccello, Samuli Ripatti, Daniel I. Chasman, Cristen J. Willer, Christopher T. Johansen, Sigrid W. Fouchier, Aaron Isaacs, Gina M. Peloso, Maja Barbalic, Sally L. Ricketts, Joshua C. Bis, Yurii S. Aulchenko, Gudmar Thorleifsson, Mary F. Feitosa, John Chambers, Marju Orho-Melander, Olle Melander, Toby Johnson, Xiaohui Li, Xiuqing Guo, Mingyao Li, Yoon Shin Cho, Min Jin Go, Young Jin Kim, Jong-Young Lee, Taesung Park, Kyunga Kim, Xueling Sim, Rick Twee-Hee Ong, Damien C. Croteau-Chonka, Leslie A. Lange, Joshua D. Smith, Kijoung Song, Jing Hua Zhao, Xin Yuan, Jian'an Luan, Claudia Lamina, Andreas Ziegler, Weihua Zhang, Robert Y. L. Zee, Alan F. Wright, Jacqueline C. M. Witteman, James F. Wilson, Gonneke Willemsen, H.-Erich Wichmann, John B. Whitfield, Dawn M. Waterworth, Nicholas J. Wareham, Gérard Waeber, Peter Vollenweider, Benjamin F. Voight, Veronique Vitart, Andre G. Uitterlinden, Manuela Uda, Jaakko Tuomilehto, John R. Thompson, Toshiko Tanaka, Ida Surakka, Heather M. Stringham, Tim D. Spector, Nicole Soranzo, Johannes H. Smit, Juha Sinisalo, Kaisa Silander, Eric J. G. Sijbrands, Angelo Scuteri, James Scott, David Schlessinger, Serena Sanna, Veikko Salomaa, Juha Saharinen, Chiara Sabatti, Aimo Ruokonen, Igor Rudan, Lynda M. Rose, Robert Roberts, Mark Rieder, Bruce M. Psaty, Peter P. Pramstaller, Irene Pichler, Markus Perola, Brenda W. J. H. Penninx, Nancy L. Pedersen, Cristian Pattaro, Alex N. Parker, Guillaume Pare, Ben A. Oostra, Christopher J. O'Donnell, Markku S. Nieminen, Deborah A. Nickerson, Grant W.

Montgomery, Thomas Meitinger, Ruth McPherson, Mark I. McCarthy, Wendy McArdle, David Masson, Nicholas G. Martin, Fabio Marroni, Massimo Mangino, Patrik K. E. Magnusson, Gavin Lucas, Robert Luben, Ruth J. F. Loos, Marja-Liisa Lokki, Guillaume Lettre, Claudia Langenberg, Lenore J. Launer, Edward G. Lakatta, Reijo Laaksonen, Kirsten O. Kyvik, Florian Kronenberg, Inke R. König, Kay-Tee Khaw, Jaakko Kaprio, Lee M. Kaplan, Asa Johansson, Marjo-Riitta Jarvelin, A. Cecile J. W. Janssens, Erik Ingelsson, Wilmar Igl, G. Kees Hovingh, Jouke-Jan Hottenga, Albert Hofman, Andrew A. Hicks, Christian Hengstenberg, Iris M. Heid, Caroline Hayward, Aki S. Havulinna, Nicholas D. Hastie, Tamara B. Harris, Talin Haritunians, Alistair S. Hall, Ulf Gyllensten, Candace Guiducci, Leif C. Groop, Elena Gonzalez, Christian Gieger, Nelson B. Freimer, Luigi Ferrucci, Jeanette Erdmann, Paul Elliott, Kenechi G. Ejebe, Angela Döring, Anna F. Dominiczak, Serkalem Demissie, Panagiotis Deloukas, Eco J. C. de Geus, Ulf de Faire, Gabriel Crawford, Francis S. Collins, Yii-der I. Chen, Mark J. Caulfield, Harry Campbell, Noel P. Burtt, Lori L. Bonnycastle, Dorret I. Boomsma, S. Matthijs Boekholdt, Richard N. Bergman, Inês Barroso, Stefania Bandinelli, Christie M. Ballantyne, Themistocles L. Assimes, Thomas Quertermous, David Altshuler, Mark Seielstad, Tien Y. Wong, E.-Shyong Tai, Alan B. Feranil, Christopher W. Kuzawa, Linda S. Adair, Herman A. Taylor, Ingrid B. Borecki, Stacey B. Gabriel, James G. Wilson, Hilma Holm, Unnur Thorsteinsdottir, Vilmundur Gudnason, Ronald M. Krauss, Karen L. Mohlke, Jose M. Ordovas, Patricia B. Munroe, Jaspal S. Kooner, Alan R. Tall, Robert A. Hegele, John J. P. Kastelein, Eric E. Schadt, Jerome I. Rotter, Eric Boerwinkle, David P. Strachan, Vincent Mooser, Kari Stefansson, Muredach P. Reilly, Nilesh J. Samani, Heribert Schunkert, L. Adrienne Cupples, Manjinder S. Sandhu, Paul M. Ridker, Daniel J. Rader, Cornelia M. van Duijn, Leena Peltonen, Gonçalo R. Abecasis, Michael Boehnke, and Sekar Kathiresan. Biological, clinical and population relevance of 95 loci for blood lipids. 466(7307):707–713, August 2010.

[85] Sumeet A. Khetarpal, Katrine T. Schjoldager, Christina Christoffersen, Avanthi Raghavan, Andrew C. Edmondson, Heiko M. Reutter, Bouhouche Ahmed, Reda Ouazzani, Gina M. Peloso, Cecilia Vitali, Wei Zhao, Amritha Varshini Hanasoge Somasundara, John S. Millar, YoSon Park, Gayani Fernando, Valentin Livanov, Seungbum Choi, Eric Noé, Pritesh Patel, Siew Peng Ho, Myocardial Infarction Exome Sequencing Study, Todd G. Kirchgessner, Hans H. Wandall, Lars Hansen, Eric P. Bennett, Sergey Y. Vakhrushev, Danish Saleheen, Sekar Kathiresan, Christopher D. Brown, Rami Abou Jamra, Eric LeGuern, Henrik Clausen, and Daniel J. Rader. Loss of Function of GALNT2 Lowers High-Density Lipoproteins in Humans, Nonhuman Primates, and Rodents. *Cell Metabolism*, 24(2):234–245, August 2016.

[86] Cristen J Willer, Serena Sanna, Anne U Jackson, Angelo Scuteri, Lori L Bonnycastle, Robert Clarke, Simon C Heath, Nicholas J Timpson, Samer S Najjar, Heather M Stringham, James Strait, William L Duren, Andrea Maschio, Fabio Busonero, Antonella Mulas, Giuseppe Albai, Amy J Swift, Mario A Morken, Narisu Narisu, Derrick Bennett, Sarah Parish, Haiqing Shen, Pilar Galan, Pierre Meneton, Serge Hercberg,

Diana Zelenika, Wei-Min Chen, Yun Li, Laura J Scott, Paul A Scheet, Jouko Sundvall, Richard M Watanabe, Ramaiah Nagaraja, Shah Ebrahim, Debbie A Lawlor, Yoav Ben-Shlomo, George Davey-Smith, Alan R Shuldiner, Rory Collins, Richard N Bergman, Manuela Uda, Jaakko Tuomilehto, Antonio Cao, Francis S Collins, Edward Lakatta, G Mark Lathrop, Michael Boehnke, David Schlessinger, Karen L Mohlke, and Gonçalo R Abecasis. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nature genetics*, 40(2):161–169, February 2008.

[87] Nilesh J. Samani, Jeanette Erdmann, Alistair S. Hall, Christian Hengstenberg, Massimo Mangino, Bjoern Mayer, Richard J. Dixon, Thomas Meitinger, Peter Braund, H.-Erich Wichmann, Jennifer H. Barrett, Inke R. König, Suzanne E. Stevens, Silke Szymczak, David-Alexandre Tregouet, Mark M. Iles, Friedrich Pahlke, Helen Pollard, Wolfgang Lieb, Francois Cambien, Marcus Fischer, Willem Ouwehand, Stefan Blankenberg, Anthony J. Balmforth, Andrea Baessler, Stephen G. Ball, Tim M. Strom, Ingrid Braenne, Christian Gieger, Panos Deloukas, Martin D. Tobin, Andreas Ziegler, John R. Thompson, Heribert Schunkert, and WTCCC and the Cardiogenics Consortium. Genomewide association analysis of coronary artery disease. *The New England Journal of Medicine*, 357(5):443–453, August 2007.

[88] Kiran Musunuru, Alanna Strong, Maria Frank-Kamenetsky, Noemi E. Lee, Tim Ahfeldt, Katherine V. Sachs, Xiaoyu Li, Hui Li, Nicolas Kuperwasser, Vera M. Ruda, James P. Pirruccello, Brian Muchmore, Ludmila Prokunina-Olsson, Jennifer L. Hall, Eric E. Schadt, Carlos R. Morales, Sissel Lund-Katz, Michael C. Phillips, Jamie Wong, William Cantley, Timothy Racie, Kenechi G. Ejebe, Marju Orho-Melander, Olle Melander, Victor Koteliansky, Kevin Fitzgerald, Ronald M. Krauss, Chad A. Cowan, Sekar Kathiresan, and Daniel J. Rader. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. 466(7307):714–719, August 2010.

[89] Olga Jarinova, Alexandre F. R. Stewart, Robert Roberts, George Wells, Paulina Lau, Thet Naing, Christine Buerki, Bradley W. McLean, Richard C. Cook, Joel S. Parker, and Ruth McPherson. Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 29(10):1671–1677, October 2009.

[90] Mads Kjolby, Olav M. Andersen, Tilman Breiderhoff, Anja W. Fjorback, Karen Marie Pedersen, Peder Madsen, Pernille Jansen, Joerg Heeren, Thomas E. Willnow, and Anders Nykjaer. Sort1, encoded by the cardiovascular risk locus 1p13.3, is a regulator of hepatic lipoprotein export. *Cell Metabolism*, 12(3):213–223, September 2010.

[91] Sathya Prakash and Rajesh Sagar. Psychiatric classification: Current debate and future directions. *Asian Journal of Psychiatry*, 20:15–21, April 2016.

[92] Marc V. Fuccillo, Patrick E. Rothwell, and Robert C. Malenka. From Synapses to Behavior: What Rodent Models Can Tell Us About Neuropsychiatric Disease. *Biological Psychiatry*, 79(1):4–6, January 2016.

[93] United Nations Office on Drugs and Crime. World Drug Report 2016. https://www.unodc.org/wdr2016/, 2016.

[94] Nora D. Volkow, George F. Koob, and A. Thomas McLellan. Neurobiologic Advances from the Brain Disease Model of Addiction. *The New England Journal of Medicine*, 374(4):363–371, January 2016.

[95] Kenneth S. Kendler, Xiangning Chen, Danielle Dick, Hermine Maes, Nathan Gillespie, Michael C. Neale, and Brien Riley. Recent advances in the genetic epidemiology and molecular genetics of substance use disorders. *Nature Neuroscience*, 15(2):181–189, January 2012.

[96] Laura E. Kwako, Reza Momenan, Raye Z. Litten, George F. Koob, and David Goldman. Addictions Neuroclinical Assessment: A Neuroscience-Based Framework for Addictive Disorders. *Biological Psychiatry*, 80(3):179–189, August 2016.

[97] Jeffrey W. Dalley and Trevor W. Robbins. Fractionating impulsivity: Neuropsychiatric implications. *Nature Reviews. Neuroscience*, 18(3):158–171, February 2017.

[98] Joel Swendsen and Michel Le Moal. Individual vulnerability to addiction. *Annals of the New York Academy of Sciences*, 1216:73–85, January 2011.

[99] James MacKillop, Jessica Weafer, Joshua C Gray, Assaf Oshri, Abraham Palmer, and Harriet de Wit. The latent structure of impulsivity: Impulsive choice, impulsive action, and impulsive personality traits. 233(18):3361–3370, September 2016.

[100] Jill B. Becker, Adam N. Perry, and Christel Westenbroek. Sex differences in the neural mechanisms mediating addiction: A new synthesis and hypothesis. *Biology of Sex Differences*, 3:14, June 2012.

[101] Inger Burnett-Zeigler, Maureen A. Walton, Mark Ilgen, Kristen L. Barry, Stephen T. Chermack, Robert A. Zucker, Marc A. Zimmerman, Brenda M. Booth, and Frederic C. Blow. Prevalence and Correlates of Mental Health Problems and Treatment Among Adolescents Seen in Primary Care. *Journal of Adolescent Health*, 50(6):559–564, June 2012.

[102] Jill B. Becker and George F. Koob. Sex Differences in Animal Models: Focus on Addiction. *Pharmacological Reviews*, 68(2):242–263, April 2016.

[103] Amy B. Hart, Harriet de Wit, and Abraham A. Palmer. Genetic factors modulating the response to stimulant drugs in humans. *Current Topics in Behavioral Neurosciences*, 12:537–577, 2012.

[104] Jeffrey F. Scherrer, Julia D. Grant, Alexis E. Duncan, Carolyn E. Sartor, Jon R. Haber, Theodore Jacob, and Kathleen K. Bucholz. Subjective effects to cannabis are associated with use, abuse and dependence after adjusting for genetic and environmental influences. *Drug and Alcohol Dependence*, 105(1-2):76–82, November 2009.

[105] John C. Churchwell, Paul D. Carey, Helen L. Ferrett, Dan J. Stein, and Deborah A. Yurgelun-Todd. Abnormal striatal circuitry and intensified novelty seeking among adolescents who abuse methamphetamine and cannabis. *Developmental Neuroscience*, 34(4):310–317, 2012.

[106] Harriet de Wit and Tamara J. Phillips. Do initial responses to drugs predict future use or abuse? *Neuroscience and Biobehavioral Reviews*, 36(6):1565–1576, July 2012.

[107] T. E. Robinson and K. C. Berridge. The neural basis of drug craving: An incentive-sensitization theory of addiction. *Brain Research. Brain Research Reviews*, 18(3):247–291, 1993 Sep-Dec.

[108] Nora D. Volkow, Gene-Jack Wang, Joanna S. Fowler, and Dardo Tomasi. Addiction circuitry in the human brain. *Annual Review of Pharmacology and Toxicology*, 52:321–336, 2012.

[109] Tamara J. Phillips, Helen M. Kamens, and Jeanna M. Wheeler. Behavioral genetic contributions to the study of addiction-related amphetamine effects. *Neuroscience and Biobehavioral Reviews*, 32(4):707–759, 2008.

[110] Scott Edwards and George F. Koob. Experimental psychiatric illness and drug abuse models: From human to animal, an overview. *Methods in Molecular Biology (Clifton, N.J.)*, 829:31–48, 2012.

[111] Christian Heidbreder. Advances in animal models of drug addiction. *Current Topics in Behavioral Neurosciences*, 7:213–250, 2011.

[112] Thomas M. Tzschentke. Measuring reward with the conditioned place preference (CPP) paradigm: Update of the last decade. *Addiction Biology*, 12(3-4):227–462, September 2007.

[113] H. de Wit, E. H. Uhlenhuth, and C. E. Johanson. Individual differences in the reinforcing and subjective effects of amphetamine and diazepam. *Drug and Alcohol Dependence*, 16(4):341–360, February 1986.

[114] Leah M. Mayo, Diana Fraser, Emma Childs, Reza Momenan, Daniel W. Hommer, Harriet de Wit, and Markus Heilig. Conditioned preference to a methamphetamine-associated contextual cue in humans. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 38(6):921–929, May 2013.

[115] Minae Niwa, Yijin Yan, and Toshitaka Nabeshima. Genes and molecules that can potentiate or attenuate psychostimulant dependence: Relevance of data from animal models to human addiction. *Annals of the New York Academy of Sciences*, 1141:76–95, October 2008.

[116] Heng Li. Improving SNP discovery by base alignment quality. *Bioinformatics (Oxford, England)*, 27(8):1157–1158, April 2011.

[117] George R. Uhl, Tomas Drgon, Catherine Johnson, Chuan-Yun Li, Carlo Contoreggi, Judith Hess, Daniel Naiman, and Qing-Rong Liu. Molecular genetics of addiction and related heritable phenotypes: Genome-wide association approaches identify "connectivity constellation" and drug target genes with pleiotropic effects. *Annals of the New York Academy of Sciences*, 1141:318–381, October 2008.

[118] George R. Uhl, Tomas Drgon, Catherine Johnson, and Qing-Rong Liu. Addiction genetics and pleiotropic effects of common haplotypes that make polygenic contributions to vulnerability to substance dependence. *Journal of Neurogenetics*, 23(3):272–282, 2009.

[119] S. Shabani, C. S. McKinnon, C. Reed, C. L. Cunningham, and T. J. Phillips. Sensitivity to rewarding or aversive effects of methamphetamine determines methamphetamine intake. *Genes, Brain, and Behavior*, 10(6):625–636, August 2011.

[120] Ezio Tirelli, Giovanni Laviola, and Walter Adriani. Ontogenesis of behavioral sensitization and conditioned place preference induced by psychostimulants in laboratory rodents. *Neuroscience and Biobehavioral Reviews*, 27(1-2):163–178, 2003 Jan-Mar.

[121] Christopher L. Cunningham, Christina M. Gremel, and Peter A. Groblewski. Drug-induced conditioned place preference and aversion in mice. *Nature Protocols*, 1(4):1662–1670, 2006.

[122] Ximena Kedikian, Maria Paula Faillace, and Ramón Bernabeu. Behavioral and molecular analysis of nicotine-conditioned place preference in zebrafish. *PloS One*, 8(7):e69453, 2013.

[123] Adam D. Collier, Kanza M. Khan, Erika M. Caramillo, Richard S. Mohn, and David J. Echevarria. Zebrafish and conditioned place preference: A translational model of drug reward. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, 55:16–25, December 2014.

[124] Alistair James Brock, Susan M. G. Goody, Andrew N. Mead, Ari Sudwarts, Matthew O. Parker, and Caroline H. Brennan. Assessing the value of the zebrafish conditioned place preference model for predicting human abuse potential. *The Journal of Pharmacology and Experimental Therapeutics*, August 2017.

[125] Adebobola Imeh-Nathaniel, Adekunle Adedeji, Robert Huber, and Thomas I. Nathaniel. The rewarding properties of methamphetamine in an invertebrate model of drug addiction. *Physiology & Behavior*, 153:40–46, January 2016.

[126] Antonio Alcaro, Jaak Panksepp, and Robert Huber. D-Amphetamine stimulates unconditioned exploration/approach behaviors in crayfish: Towards a conserved evolutionary function of ancestral drug reward. *Pharmacology, Biochemistry, and Behavior*, 99(1):75–80, July 2011.

[127] Leah M. Mayo and Harriet de Wit. Acquisition of responses to a methamphetamine-associated cue in healthy humans: Self-report, behavioral, and psychophysiological measures. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 40(7):1734–1741, June 2015.

[128] Leah M. Mayo and Harriet de Wit. Acquisition of Conditioned Responses to a Novel Alcohol-Paired Cue in Social Drinkers. *Journal of Studies on Alcohol and Drugs*, 77(2):317–326, March 2016.

[129] Joel S. Cavallo, Leah M. Mayo, and Harriet de Wit. Acquisition of Conditioning between Methamphetamine and Cues in Healthy Humans. *PloS One*, 11(8):e0161541, 2016.

[130] Emma Childs and Harriet de Wit. Amphetamine-induced place preference in humans. *Biological Psychiatry*, 65(10):900–904, May 2009.

[131] Emma Childs and Harriet de Wit. Contextual conditioning enhances the psychostimulant and incentive properties of d-amphetamine in humans. *Addiction Biology*, 18(6):985–992, November 2013.

[132] Lee Hogarth, Bernard W. Balleine, Laura H. Corbit, and Simon Killcross. Associative learning mechanisms underpinning the transition from recreational drug use to addiction. *Annals of the New York Academy of Sciences*, 1282:12–24, April 2013.

[133] C. Orsini, A. Bonito-Oliva, D. Conversi, and S. Cabib. Susceptibility to conditioned place preference induced by addictive drugs in mice of the C57BL/6 and DBA/2 inbred strains. 181(2):327–336, September 2005.

[134] Jana Drgonova, Donna Walther, Sulabh Singhal, Kennedy Johnson, Brice Kessler, Juan Troncoso, and George R. Uhl. Altered CSMD1 Expression Alters Cocaine-Conditioned Place Preference: Mutual Support for a Complex Locus from Human and Mouse Models. *PloS One*, 10(7):e0120908, 2015.

[135] Jana Drgonova, Donna Walther, Katherine J. Wang, G. Luke Hartstein, Bryson Lochte, Juan Troncoso, Noriko Uetani, Yoichiro Iwakura, and George R. Uhl. Mouse model for PTPRD associations with WED/RLS and addiction: Reduced expression alters locomotion, sleep behaviors and cocaine-conditioned place preference. *Molecular Medicine (Cambridge, Mass.)*, July 2015.

[136] Philip M. Newton, Christine J. Orr, Melisa J. Wallace, Chanki Kim, Hee-Sup Shin, and Robert O. Messing. Deletion of N-type calcium channels alters ethanol reward and reduces ethanol consumption in mice. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(44):9862–9869, November 2004.

[137] Carrie L. Heusner and Richard D. Palmiter. Expression of mutant NMDA receptors in dopamine D1 receptor-containing cells prevents cocaine sensitization and decreases cocaine preference. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(28):6651–6657, July 2005.

[138] Tresa M. McGranahan, Natalie E. Patzlaff, Sharon R. Grady, Stephen F. Heinemann, and T. K. Booker. $a4\beta2$ nicotinic acetylcholine receptors on dopaminergic neurons mediate nicotine reward and anxiety relief. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(30):10891–10902, July 2011.

[139] V. M. Philip, S. Duvvuru, B. Gomero, T. A. Ansah, C. D. Blaha, M. N. Cook, K. M. Hamre, W. R. Lariviere, D. B. Matthews, G. Mittleman, D. Goldowitz, and E. J. Chesler. High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. *Genes, Brain, and Behavior*, 9(2):129–159, March 2010.

[140] Laura N. Smith, Jakub P. Jedynak, Miles R. Fontenot, Carly F. Hale, Karen C. Dietz, Makoto Taniguchi, Feba S. Thomas, Benjamin C. Zirlin, Shari G. Birnbaum, Kimberly M. Huber, Mark J. Thomas, and Christopher W. Cowan. Fragile X mental retardation protein regulates synaptic and behavioral plasticity to repeated cocaine administration. 82(3):645–658, May 2014.

[141] I. Sora, F. S. Hall, A. M. Andrews, M. Itokawa, X. F. Li, H. B. Wei, C. Wichems, K. P. Lesch, D. L. Murphy, and G. R. Uhl. Molecular mechanisms of cocaine reward: Combined dopamine and serotonin transporter knockouts eliminate cocaine place preference. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):5300–5305, April 2001.

[142] Karen Brami-Cherrier, Emmanuel Valjent, Denis Hervé, Joanne Darragh, Jean-Christophe Corvol, Christiane Pages, Simon J. Arthur, Arthur J. Simon, Jean-Antoine Girault, and Jocelyne Caboche. Parsing molecular and behavioral effects of cocaine in mitogen- and stress-activated protein kinase-1-deficient mice. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(49):11444–11454, December 2005.

[143] Pierre F. Durieux, Bertrand Bearzatto, Stefania Guiducci, Thorsten Buch, Ari Waisman, Michele Zoli, Serge N. Schiffmann, and Alban de Kerchove d'Exaerde. D2R striatopallidal neurons inhibit both locomotor and drug reward processes. *Nature Neuroscience*, 12(4):393–395, April 2009.

[144] Karen K. Szumlinski, Marlin H. Dehoff, Shin H. Kang, Kelly A. Frys, Kevin D. Lominac, Matthias Klugmann, Jason Rohrer, William Griffin, Shigenobu Toda, Nicolas P. Champtiaux, Thomas Berry, Jian C. Tu, Stephanie E. Shealy, Matthew J. During, Lawrence D. Middaugh, Paul F. Worley, and Peter W. Kalivas. Homer proteins regulate sensitivity to cocaine. 43(3):401–413, August 2004.

[145] David C. Martinelli, Kylie S. Chew, Astrid Rohlmann, Matthew Y. Lum, Susanne Ressl, Samer Hattar, Axel T. Brunger, Markus Missler, and Thomas C. Südhof. Expression of C1ql3 in Discrete Neuronal Populations Controls Efferent Synapse Numbers and Diverse Behaviors. 91(5):1034–1051, September 2016.

[146] Camron D. Bryant, Loren A. Kole, Michael A. Guido, Riyan Cheng, and Abraham A. Palmer. Methamphetamine-induced conditioned place preference in LG/J and SM/J mouse strains and an F45/F46 advanced intercross line. *Frontiers in Genetics*, 3:126, 2012.

[147] Clarissa C. Parker, Riyan Cheng, Greta Sokoloff, Jackie E. Lim, Andrew D. Skol, Mark Abney, and Abraham A. Palmer. Fine-mapping alleles for body weight in LG/J X SM/J F2 and F(34) advanced intercross lines. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 22(9-10):563–571, October 2011.

[148] Paul Vezina and Marco Leyton. Conditioned cues and the expression of stimulant sensitization in animals and humans. 56 Suppl 1:160–168, 2009.

[149] Camron D. Bryant, Michael A. Guido, Loren A. Kole, and Riyan Cheng. The heritability of oxycodone reward and concomitant phenotypes in a LG/J X SM/J mouse advanced intercross line. *Addiction Biology*, 19(4):552–561, July 2014.

[150] Hidetoshi Takahashi, Masao Iwase, Ryouhei Ishii, Kazutaka Ohi, Motoyuki Fukumoto, Michiyo Azechi, Koji Ikezawa, Ryu Kurimoto, Leonides Canuet, Takayuki Nakahachi, Naomi Iike, Shinji Tagami, Takashi Morihara, Masayasu Okochi, Toshihisa Tanaka, Hiroaki Kazui, Tetsuhiko Yoshida, Hitoshi Tanimukai, Yuka Yasuda, Takashi Kudo, Ryota Hashimoto, and Masatoshi Takeda. Impaired prepulse inhibition and habituation of acoustic startle response in Japanese patients with schizophrenia. *Neuroscience Research*, 62(3):187–194, November 2008.

[151] A. Parwani, E. J. Duncan, E. Bartlett, S. H. Madonick, T. R. Efferen, R. Rajan, M. Sanfilipo, P. B. Chappell, S. Chakravorty, S. Gonzenbach, G. N. Ko, and J. P. Rotrosen. Impaired prepulse inhibition of acoustic startle in schizophrenia. *Biological Psychiatry*, 47(7):662–669, April 2000.

[152] Neal R. Swerdlow, Gregory A. Light, Michael L. Thomas, Joyce Sprock, Monica E. Calkins, Michael F. Green, Tiffany A. Greenwood, Raquel E. Gur, Ruben C. Gur, Laura C. Lazzeroni, Keith H. Nuechterlein, Allen D. Radant, Larry J. Seidman, Larry J. Siever, Jeremy M. Silverman, William S. Stone, Catherine A. Sugar, Debby W. Tsuang, Ming T. Tsuang, Bruce I. Turetsky, and David L. Braff. Deficient prepulse inhibition in schizophrenia in a multi-site cohort: Internal replication and extension. *Schizophrenia Research*, May 2017.

[153] Tiffany A. Greenwood, Gregory A. Light, Neal R. Swerdlow, Monica E. Calkins, Michael F. Green, Raquel E. Gur, Ruben C. Gur, Laura C. Lazzeroni, Keith H. Nuechterlein, Ann Olincy, Allen D. Radant, Larry J. Seidman, Larry J. Siever, Jeremy M. Silverman, William S. Stone, Catherine A. Sugar, Debby W. Tsuang, Ming T. Tsuang, Bruce I. Turetsky, Robert Freedman, and David L. Braff. Gating Deficit Heritability and Correlation With Increased Clinical Severity in Schizophrenia Patients With Positive Family History. *The American Journal of Psychiatry*, 173(4):385–391, April 2016.

[154] Klaus Hoenig, Andrea Hochrein, Boris B. Quednow, Wolfgang Maier, and Michael Wagner. Impaired prepulse inhibition of acoustic startle in obsessive-compulsive disorder. *Biological Psychiatry*, 57(10):1153–1158, May 2005.

[155] Susanne E. Ahmari. Using mice to model Obsessive Compulsive Disorder: From genes to circuits. 321:121–137, May 2016.

[156] N. R. Swerdlow, J. Paulsen, D. L. Braff, N. Butters, M. A. Geyer, and M. R. Swenson. Impaired prepulse inhibition of acoustic and tactile startle response in patients with Huntington's disease. *Journal of Neurology, Neurosurgery, and Psychiatry*, 58(2):192–200, February 1995.

[157] D. L. Braff, M. A. Geyer, and N. R. Swerdlow. Human studies of prepulse inhibition of startle: Normal subjects, patient groups, and pharmacological studies. 156(2-3):234–258, July 2001.

[158] Grainne M. McAlonan, Eileen Daly, Veena Kumari, Hugo D. Critchley, Therese van Amelsvoort, John Suckling, Andrew Simmons, Thordur Sigmundsson, Kathyrn Greenwood, Ailsa Russell, Nicole Schmitz, Francesca Happe, Patricia Howlin, and Declan G. M. Murphy. Brain anatomy and sensorimotor gating in Asperger's syndrome. *Brain: A Journal of Neurology*, 125(Pt 7):1594–1606, July 2002.

[159] Juan P. Vargas, Estrella Díaz, Manuel Portavella, and Juan C. López. Animal Models of Maladaptive Traits: Disorders in Sensorimotor Gating and Attentional Quantifiable Responses as Possible Endophenotypes. *Frontiers in Psychology*, 7:206, 2016.

[160] Tomoya Nakamura, Jumpei Matsumoto, Yusaku Takamura, Yoko Ishii, Masakiyo Sasahara, Taketoshi Ono, and Hisao Nishijo. Relationships among parvalbumin-immunoreactive neuron density, phase-locked gamma oscillations, and autistic/schizophrenic symptoms in PDGFR-β knock-out and control mice. *PloS One*, 10(3):e0119258, 2015.

[161] Tsuyoshi Nakai, Taku Nagai, Motoki Tanaka, Norimichi Itoh, Naoya Asai, Atsushi Enomoto, Masato Asai, Shinnosuke Yamada, Ali Bin Saifullah, Masahiro Sokabe, Masahide Takahashi, and Kiyofumi Yamada. Girdin phosphorylation is crucial for synaptic plasticity and memory: A potential role in the interaction of BDNF/TrkB/Akt signaling with NMDA receptor. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 34(45):14995–15008, November 2014.

[162] Francesco Papaleo, Feng Yang, Clare Paterson, Sara Palumbo, Gregory V. Carr, Yanhong Wang, Kirsten Floyd, Wenwei Huang, Craig J. Thomas, Jingshan Chen, Daniel R. Weinberger, and Amanda J. Law. Behavioral, Neurophysiological, and Synaptic Impairment in a Transgenic Neuregulin1 (NRG1-IV) Murine Schizophrenia Model. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 36(17):4859–4875, April 2016.

[163] Hui-Feng Jiao, Xiang-Dong Sun, Ryan Bates, Lei Xiong, Lei Zhang, Fang Liu, Lei Li, Hong-Sheng Zhang, Shun-Qi Wang, Ming-Tao Xiong, Mihir Patel, Alexis M. Stranahan, Wen-Cheng Xiong, Bao-Ming Li, and Lin Mei. Transmembrane protein 108 is required for glutamatergic transmission in dentate gyrus. *Proceedings of the National Academy of Sciences of the United States of America*, 114(5):1177–1182, January 2017.

[164] Tiffany A. Greenwood, Neal R. Swerdlow, Raquel E. Gur, Kristin S. Cadenhead, Monica E. Calkins, Dorcas J. Dobie, Robert Freedman, Michael F. Green, Ruben C. Gur, Laura C. Lazzeroni, Keith H. Nuechterlein, Ann Olincy, Allen D. Radant, Amrita Ray, Nicholas J. Schork, Larry J. Seidman, Larry J. Siever, Jeremy M. Silverman, William S. Stone, Catherine A. Sugar, Debby W. Tsuang, Ming T. Tsuang, Bruce I. Turetsky, Gregory A. Light, and David L. Braff. Genome-wide linkage analyses of 12 endophenotypes for schizophrenia from the Consortium on the Genetics of Schizophrenia. *The American Journal of Psychiatry*, 170(5):521–532, May 2013.

[165] M. Koch. The neurobiology of startle. *Progress in Neurobiology*, 59(2):107–128, October 1999.

[166] F. K. Graham. Presidential Address, 1974. The more or less startling effects of weak prestimulation. 12(3):238–248, May 1975.

[167] K. E. Samocha, J. E. Lim, R. Cheng, G. Sokoloff, and A. A. Palmer. Fine mapping of QTL for prepulse inhibition in LG/J and SM/J mice using F(2) and advanced intercross lines. *Genes, Brain, and Behavior*, 9(7):759–767, October 2010.

[168] Abraham A. Palmer, Laura L. Breen, Pamela Flodman, Lisa H. Conti, M. Anne Spence, and Morton P. Printz. Identification of quantitative trait loci for prepulse inhibition in rats. 165(3):270–279, January 2003.

[169] L. J. Sittig, P. Carbonetto, K. A. Engel, K. S. Krauss, and A. A. Palmer. Integration of genome-wide association and extant brain expression QTL identifies candidate genes influencing prepulse inhibition in inbred F1 mice. *Genes, Brain, and Behavior*, 15(2):260–270, February 2016.

[170] Neal R. Swerdlow, David L. Braff, and Mark A. Geyer. Sensorimotor gating of the startle reflex: What we said 25 years ago, what has happened since then, and what comes next. *Journal of Psychopharmacology (Oxford, England)*, 30(11):1072–1081, November 2016.

[171] Trine Bjørg Hammer, Bob Oranje, Birgitte Fagerlund, Hannah Bro, and Birte Y. Glenthøj. Stability of prepulse inhibition and habituation of the startle reflex in schizophrenia: A 6-year follow-up study of initially antipsychotic-naive, first-episode schizophrenia patients. *The International Journal of Neuropsychopharmacology*, 14(7):913–925, August 2011.

[172] Albert H. C. Wong and Sheena A. Josselyn. Caution When Diagnosing Your Mouse With Schizophrenia: The Use and Misuse of Model Animals for Understanding Psychiatric Disorders. *Biological Psychiatry*, 79(1):32–38, January 2016.

[173] Rocio Moreno-Paublete, Barbara Canlon, and Christopher R. Cederroth. Differential Neural Responses Underlying the Inhibition of the Startle Response by Pre-Pulses or Gaps in Mice. *Frontiers in Cellular Neuroscience*, 11:19, 2017.

[174] J. Deckert, H. Weber, C. Villmann, T. B. Lonsdorf, J. Richter, M. Andreatta, A. Arias-Vasquez, L. Hommers, L. Kent, C. Schartner, S. Cichon, C. Wolf, N. Schaefer, C. R. von Collenberg, B. Wachter, R. Blum, D. Schümann, R. Scharfenort, J. Schumacher, A. J. Forstner, C. Baumann, M. A. Schiele, S. Notzon, P. Zwanzger, J. G. E. Janzing, T. Galesloot, L. A. Kiemeney, A. Gajewska, E. Glotzbach-Schoon, A. Mühlberger, G. Alpers, T. Fydrich, L. Fehm, A. L. Gerlach, T. Kircher, T. Lang, A. Ströhle, V. Arolt, H.-U. Wittchen, R. Kalisch, C. Büchel, A. Hamm, M. M. Nöthen, M. Romanos, K. Domschke, P. Pauli, and A. Reif. GLRB allelic variation associated with agoraphobic cognitions, increased startle response and fear network activation: A potential neurogenetic pathway to panic disorder. *Molecular Psychiatry*, February 2017.

[175] Heather A. Lawson and James M. Cheverud. Metabolic syndrome components in murine models. *Endocrine, Metabolic & Immune Disorders Drug Targets*, 10(1):25–40, March 2010.

[176] Aldi T. Kraja, Michael A. Province, Pinchia Huang, Joseph P. Jarvis, Treva Rice, James M. Cheverud, and D. C. Rao. Trends in metabolic syndrome and gene networks in human and rodent models. *Endocrine, Metabolic & Immune Disorders Drug Targets*, 8(3):198–207, September 2008.

[177] J. A. Beck, S. Lloyd, M. Hafezparast, M. Lennon-Pierce, J. T. Eppig, M. F. Festing, and E. M. Fisher. Genealogies of mouse inbred strains. *Nature Genetics*, 24(1):23–25, January 2000.

[178] J. M. Cheverud, L. S. Pletscher, T. T. Vaughn, and B. Marshall. Differential response to dietary fat in large (LG/J) and small (SM/J) inbred mouse strains. *Physiological Genomics*, 1(1):33–39, July 1999.

[179] T. T. Vaughn, L. S. Pletscher, A. Peripato, K. King-Ellison, E. Adams, C. Erikson, and J. M. Cheverud. Mapping quantitative trait loci for murine growth: A closer look at genetic architecture. *Genetical Research*, 74(3):313–322, December 1999.

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

[181] A. M. Carroll, R. Cheng, E. S. R. Collie-Duguid, C. Meharg, M. E. Scholz, S. Fiering, J. L. Fields, A. A. Palmer, and A. Lionikas. Fine-mapping of genes determining

extrafusal fiber properties in murine soleus muscle. *Physiological Genomics*, 49(3):141–150, March 2017.

[182] Youngwon Kim, Katrien Wijndaele, Duck-Chul Lee, Stephen J. Sharp, Nick Wareham, and Soren Brage. Independent and joint associations of grip strength and adiposity with all-cause and cardiovascular disease mortality in 403,199 adults: The UK Biobank study. *The American Journal of Clinical Nutrition*, August 2017.

[183] Simon Timpka, Ingemar F. Petersson, Caddie Zhou, and Martin Englund. Muscle strength in adolescent men and risk of cardiovascular disease events and mortality in middle age: A prospective cohort study. *BMC medicine*, 12:62, April 2014.

[184] Eleonora Poggiogalle, Lorenzo Maria Donini, Andrea Lenzi, Claudio Chiesa, and Lucia Pacifico. Non-alcoholic fatty liver disease connections with fat-free tissues: A focus on bone and skeletal muscle. *World Journal of Gastroenterology*, 23(10):1747–1757, March 2017.

[185] Darin Bloemberg and Joe Quadrilatero. Rapid determination of myosin heavy chain expression in rat, mouse, and human skeletal muscle using multicolor immunofluorescence analysis. *PloS One*, 7(4):e35273, 2012.

[186] Graziella Messina and Giulio Cossu. The origin of embryonic and fetal myoblasts: A role of Pax3 and Pax7. *Genes & Development*, 23(8):902–905, April 2009.

[187] Thomas H. Ehrich, Tomas Hrbek, Jane P. Kenney-Hunt, L. Susan Pletscher, Bing Wang, Clay F. Semenkovich, and James M. Cheverud. Fine-mapping gene-by-diet interactions on chromosome 13 in a LG\/J X SM\/J murine model of obesity. 54(6):1863–1872, June 2005.

[188] Robert J. Elshire, Jeffrey C. Glaubitz, Qi Sun, Jesse A. Poland, Ken Kawamoto, Edward S. Buckler, and Sharon E. Mitchell. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5):e19379, May 2011.

[189] John W. Davey, Paul A. Hohenlohe, Paul D. Etter, Jason Q. Boone, Julian M. Catchen, and Mark L. Blaxter. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews. Genetics*, 12(7):499–510, June 2011.

[190] Christopher J. Fitzpatrick, Shyam Gopalakrishnan, Elizabeth S. Cogan, Lindsay M. Yager, Paul J. Meyer, Vedran Lovic, Benjamin T. Saunders, Clarissa C. Parker, Natalia M. Gonzales, Emmanuel Aryee, Shelly B. Flagel, Abraham A. Palmer, Terry E. Robinson, and Jonathan D. Morrow. 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, 2013.

[191] Andrew P. Morgan, Chen-Ping Fu, Chia-Yu Kao, Catherine E. Welsh, John P. Didion, Liran Yadgary, Leeanna Hyacinth, Martin T. Ferris, Timothy A. Bell, Darla R.

Miller, Paola Giusti-Rodriguez, Randal J. Nonneman, Kevin D. Cook, Jason K. Whitmire, Lisa E. Gralinski, Mark Keller, Alan D. Attie, Gary A. Churchill, Petko Petkov, Patrick F. Sullivan, Jennifer R. Brennan, Leonard McMillan, and Fernando Pardo-Manuel de Villena. The Mouse Universal Genotyping Array: From Substrains to Subspecies. *G3 (Bethesda, Md.)*, 6(2):263–279, December 2015.

[192] Y. Li, R. Breitling, L. B. Snoek, K. J. Velde, M. A. Swertz, and J. Riksen. Global genetic robustness of the alternative splicing machinery in Caenorhabditis elegans. 186, 2010.

[193] Igor Nikolskiy, Donald F. Conrad, Sung Chun, Justin C. Fay, James M. Cheverud, and Heather A. Lawson. Using whole-genome sequences of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides. *BMC genomics*, 16:415, May 2015.

[194] Mark A. DePristo, Eric Banks, Ryan Poplin, Kiran V. Garimella, Jared R. Maguire, Christopher Hartl, Anthony A. Philippakis, Guillermo del Angel, Manuel A. Rivas, Matt Hanna, Aaron McKenna, Tim J. Fennell, Andrew M. Kernytsky, Andrey Y. Sivachenko, Kristian Cibulskis, Stacey B. Gabriel, David Altshuler, and Mark J. Daly. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498, May 2011.

[195] Thorfinn S. Korneliussen, Anders Albrechtsen, and Rasmus Nielsen. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, 15(1):356, November 2014.

[196] Brian L. Browning and Sharon R. Browning. Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics*, 98(1):116–126, January 2016.

[197] Sharon R. Browning and Brian L. Browning. 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, November 2007.

[198] Xiang Zhou and Matthew Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–824, June 2012.

[199] Xiang Zhou. A unifed framework for variance component estimation with summary statistics in genome-wide association studies. 042846, 2016.

[200] Jong Wha J. Joo, Farhad Hormozdiari, Buhm Han, and Eleazar Eskin. Multiple testing correction in linear mixed models. *Genome Biology*, 17:62, April 2016.

[201] Buhm Han, Hyun Min Kang, and Eleazar Eskin. Rapid and accurate multiple testing correction and power estimation for millions of correlated markers. *PLoS genetics*, 5(4):e1000456, April 2009.

[202] W. K. Silvers. *The Coat Colors of Mice: A Model for Mammalian Gene Action and Interaction*. Springer, New York, softcover reprint of the original 1st ed. 1979 edition edition, October 2013.

[203] Clarissa C. Parker, Greta Sokoloff, Riyan Cheng, and Abraham A. Palmer. Genome-wide association for fear conditioning in an advanced intercross mouse line. *Behavior Genetics*, 42(3):437–448, May 2012.

[204] Shelly B. Flagel, Maria Waselus, Sarah M. Clinton, Stanley J. Watson, and Huda Akil. Antecedents and consequences of drug abuse in rats selectively bred for high and low response to novelty. 76 Pt B:425–436, January 2014.

[205] Anna Delprato, Marie-Paule Algéo, Brice Bonheur, Jason A. Bubier, Lu Lu, Robert W. Williams, Elissa J. Chesler, and Wim E. Crusio. QTL and Systems Genetics Analysis of Mouse Grooming and Behavioral Responses to Novelty in an Open Field. *Genes, Brain, and Behavior*, May 2017.

[206] Bjarte Håvik, Stephanie Le Hellard, Marcella Rietschel, Helle Lybæk, Srdjan Djurovic, Manuel Mattheisen, Thomas W. Mühleisen, Franziska Degenhardt, Lutz Priebe, Wolfgang Maier, Rene Breuer, Thomas G. Schulze, Ingrid Agartz, Ingrid Melle, Thomas Hansen, Clive R. Bramham, Markus M. Nöthen, Beth Stevens, Thomas Werge, Ole A. Andreassen, Sven Cichon, and Vidar M. Steen. The complement control-related genes CSMD1 and CSMD2 associate to schizophrenia. *Biological Psychiatry*, 70(1):35–42, July 2011.

[207] Wei Xu, Sarah Cohen-Woods, Qian Chen, Abdul Noor, Jo Knight, Georgina Hosang, Sagar V. Parikh, Vincenzo De Luca, Federica Tozzi, Pierandrea Muglia, Julia Forte, Andrew McQuillin, Pingzhao Hu, Hugh M. D. Gurling, James L. Kennedy, Peter McGuffin, Anne Farmer, John Strauss, and John B. Vincent. Genome-wide association study of bipolar disorder in Canadian and UK populations corroborates disease loci including SYNE1 and CSMD1. *BMC medical genetics*, 15:2, January 2014.

[208] Clarissa C. Parker, Peter Carbonetto, Greta Sokoloff, Yeonhee J. Park, Mark Abney, and Abraham A. Palmer. High-resolution genetic mapping of complex traits from a combined analysis of F2 and advanced intercross mice. 198(1):103–116, September 2014.

[209] Thomas H. Ehrich, Jane P. Kenney, Ty T. Vaughn, L. Susan Pletscher, and James M. Cheverud. Diet, obesity, and hyperglycemia in LG/J and SM/J mice. *Obesity Research*, 11(11):1400–1410, November 2003.

[210] Heather A. Lawson, Janet E. Cady, Charlyn Partridge, Jason B. Wolf, Clay F. Semenkovich, and James M. Cheverud. Genetic effects at pleiotropic loci are context-dependent with consequences for the maintenance of genetic variation in populations. *PLoS genetics*, 7(9):e1002256, September 2011.

[211] Anlu Chen, Yang Liu, Scott M. Williams, Nathan Morris, and David A. Buchner. Widespread Inter-Chromosomal Epistasis Regulates Glucose Homeostasis And Gene Expression. page 132175, May 2017.

[212] Heather A. Lawson, Arthur Lee, Gloria L. Fawcett, Bing Wang, L. Susan Pletscher, Taylor J. Maxwell, Thomas H. Ehrich, Jane P. Kenney-Hunt, Jason B. Wolf, Clay F. Semenkovich, and James M. Cheverud. The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J X SM/J murine model. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 22(3-4):197–208, April 2011.

[213] J. M. Cheverud, T. T. Vaughn, L. S. Pletscher, A. C. Peripato, E. S. Adams, C. F. Erikson, and K. J. King-Ellison. Genetic architecture of adiposity in the cross of LG/J and SM/J inbred mice. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 12(1):3–12, January 2001.

[214] Heather A. Lawson, Kathleen M. Zelle, Gloria L. Fawcett, Bing Wang, L. Susan Pletscher, Taylor J. Maxwell, Thomas H. Ehrich, Jane P. Kenney-Hunt, Jason B. Wolf, Clay F. Semenkovich, and James M. Cheverud. Genetic, epigenetic, and gene-by-diet interaction effects underlie variation in serum lipids in a LG/J X SM/J murine model. *Journal of Lipid Research*, 51(10):2976–2984, October 2010.

[215] Andrew M. Carroll, Abraham A. Palmer, and Arimantas Lionikas. QTL Analysis of Type I and Type IIA Fibers in Soleus Muscle in a Cross between LG/J and SM/J Mouse Strains. *Frontiers in Genetics*, 2:99, 2011.

[216] Tomas Hrbek, Reinaldo Alves de Brito, B. Wang, L. Susan Pletscher, and James M. Cheverud. Genetic characterization of a new set of recombinant inbred lines (LGXSM) formed from the inter-cross of SM/J and LG/J inbred mouse strains. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 17(5):417–429, May 2006.

[217] Julie Le Merrer, Xavier Rezai, Grégory Scherrer, Jérôme A. J. Becker, and Brigitte L. Kieffer. Impaired Hippocampus-Dependent and Facilitated Striatum-Dependent Behaviors in Mice Lacking the Delta Opioid Receptor. 38(6):1050–1059, May 2013.

[218] A. Delprato, B. Bonheur, M.-P. Algéo, P. Rosay, L. Lu, R. W. Williams, and W. E. Crusio. Systems genetic analysis of hippocampal neuroanatomy and spatial learning in mice. *Genes, Brain, and Behavior*, 14(8):591–606, November 2015.

[219] Mark Abney. Identity-by-descent estimation and mapping of qualitative traits in large, complex pedigrees. 179(3):1577–1590, July 2008.

[220] Lide Han and Mark Abney. Identity by descent estimation with dense genome-wide genotype data. *Genetic Epidemiology*, 35(6):557–567, September 2011.

[221] Benjamin N. Bimber, Michael J. Raboin, John Letaw, Kimberly A. Nevonen, Jennifer E. Spindel, Susan R. McCouch, Rita Cervera-Juanes, Eliot Spindel, Lucia Carbone, Betsy Ferguson, and Amanda Vinson. Whole-genome characterization in pedigreed non-human primates using genotyping-by-sequencing (GBS) and imputation. *BMC Genomics*, 17(1), August 2016.

[222] Jonathan Marchini and Bryan Howie. Genotype imputation for genome-wide association studies. *Nature Reviews. Genetics*, 11(7):499–511, July 2010.

[223] Lucy Huang, Chaolong Wang, and Noah A. Rosenberg. The relationship between imputation error and statistical power in genetic association studies in diverse populations. *American Journal of Human Genetics*, 85(5):692–698, November 2009.

[224] M. Li, X.-J. Luo, M. Rietschel, C. M. Lewis, M. Mattheisen, B. Müller-Myhsok, S. Jamain, M. Leboyer, M. Landén, P. M. Thompson, S. Cichon, M. M. Nöthen, T. G. Schulze, P. F. Sullivan, S. E. Bergen, G. Donohoe, D. W. Morris, A. Hargreaves, M. Gill, A. Corvin, C. Hultman, A. W. Toga, L. Shi, Q. Lin, H. Shi, L. Gan, A. Meyer-Lindenberg, D. Czamara, C. Henry, B. Etain, J. C. Bis, M. A. Ikram, M. Fornage, S. Debette, L. J. Launer, S. Seshadri, S. Erk, H. Walter, A. Heinz, F. Bellivier, J. L. Stein, S. E. Medland, A. Arias Vasquez, D. P. Hibar, B. Franke, N. G. Martin, M. J. Wright, MooDS Bipolar Consortium, Swedish Bipolar Study Group, Alzheimer's Disease Neuroimaging Initiative, ENIGMA Consortium, CHARGE Consortium, and B. Su. Allelic differences between Europeans and Chinese for CREB1 SNPs and their implications in gene expression regulation, hippocampal structure and function, and bipolar disorder susceptibility. *Molecular Psychiatry*, 19(4):452–461, April 2014.

[225] Allison Cox, Cheryl L. Ackert-Bicknell, Beth L. Dumont, Yueming Ding, Jordana Tzenova Bell, Gudrun A. Brockmann, Jon E. Wergedal, Carol Bult, Beverly Paigen, Jonathan Flint, Shirng-Wern Tsaih, Gary A. Churchill, and Karl W. Broman. A new standard genetic map for the laboratory mouse. 182(4):1335–1344, August 2009.

[226] Riyan Cheng, Mark Abney, Abraham A. Palmer, and Andrew D. Skol. QTLRel: An R package for genome-wide association studies in which relatedness is a concern. *BMC genetics*, 12:66, July 2011.

[227] Jan Graffelman. Exploring Diallelic genetic markers: The Hardy Weinberg package. *Journal of statistical software*, 64(3):1–23, March 2015.

[228] Christopher C. Chang, Carson C. Chow, Laurent CAM Tellier, Shashaank Vattikuti, Shaun M. Purcell, and James J. Lee. Second-generation PLINK: Rising to the challenge of larger and richer datasets. 4:7, 2015.

[229] Frank W. Albert and Leonid Kruglyak. The role of regulatory variation in complex traits and disease. *Nature Reviews. Genetics*, 16(4):197–212, April 2015.

[230] Aswin Sekar, Allison R. Bialas, Heather de Rivera, Avery Davis, Timothy R. Hammond, Nolan Kamitaki, Katherine Tooley, Jessy Presumey, Matthew Baum, Vanessa Van Doren, Giulio Genovese, Samuel A. Rose, Robert E. Handsaker, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Mark J. Daly, Michael C. Carroll, Beth Stevens, and Steven A. McCarroll. Schizophrenia risk from complex variation of complement component 4. 530(7589):177–183, February 2016.

[231] Catherine Morrissey, Ian C. Grieve, Matthias Heinig, Santosh Atanur, Enrico Petretto, Michal Pravenec, Norbert Hubner, and Timothy J. Aitman. Integrated genomic approaches to identification of candidate genes underlying metabolic and cardiovascular phenotypes in the spontaneously hypertensive rat. *Physiological Genomics*, 43(21):1207–1218, November 2011.

[232] Athma A. Pai, Jonathan K. Pritchard, and Yoav Gilad. The genetic and mechanistic basis for variation in gene regulation. *PLoS genetics*, 11(1):e1004857, January 2015.

[233] M. Liu, S. M. Malone, U. Vaidyanathan, M. C. Keller, G. Abecasis, M. McGue, W. G. Iacono, and S. I. Vrieze. Psychophysiological endophenotypes to characterize mechanisms of known schizophrenia genetic loci. *Psychological Medicine*, 47(6):1116–1125, April 2017.

[234] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science (New York, N.Y.)*, 348(6235):648–660, May 2015.

[235] Marta Melé, Pedro G. Ferreira, Ferran Reverter, David S. DeLuca, Jean Monlong, Michael Sammeth, Taylor R. Young, Jakob M. Goldmann, Dmitri D. Pervouchine, Timothy J. Sullivan, Rory Johnson, Ayellet V. Segrè, Sarah Djebali, Anastasia Niarchou, GTEx Consortium, Fred A. Wright, Tuuli Lappalainen, Miquel Calvo, Gad Getz, Emmanouil T. Dermitzakis, Kristin G. Ardlie, and Roderic Guigó. Human genomics. The human transcriptome across tissues and individuals. *Science (New York, N.Y.)*, 348(6235):660–665, May 2015.

[236] Alexandra C. Nica, Leopold Parts, Daniel Glass, James Nisbet, Amy Barrett, Magdalena Sekowska, Mary Travers, Simon Potter, Elin Grundberg, Kerrin Small, Asa K. Hedman, Veronique Bataille, Jordana Tzenova Bell, Gabriela Surdulescu, Antigone S. Dimas, Catherine Ingle, Frank O. Nestle, Paola di Meglio, Josine L. Min, Alicja Wilk, Christopher J. Hammond, Neelam Hassanali, Tsun-Po Yang, Stephen B. Montgomery, Steve O'Rahilly, Cecilia M. Lindgren, Krina T. Zondervan, Nicole Soranzo, Inês Barroso, Richard Durbin, Kourosh Ahmadi, Panos Deloukas, Mark I. McCarthy, Emmanouil T. Dermitzakis, Timothy D. Spector, and MuTHER Consortium. The architecture of gene regulatory variation across multiple human tissues: The MuTHER study. *PLoS genetics*, 7(2):e1002003, February 2011.

[237] Chen Yao, Roby Joehanes, Andrew D. Johnson, Tianxiao Huan, Chunyu Liu, Jane E. Freedman, Peter J. Munson, David E. Hill, Marc Vidal, and Daniel Levy. Dynamic

Role of trans Regulation of Gene Expression in Relation to Complex Traits. *American Journal of Human Genetics*, 100(4):571–580, April 2017.

[238] R. B. Brem, G. Yvert, R. Clinton, and L. Kruglyak. Genetic dissection of transcriptional regulation in budding yeast. 296, 2002.

[239] Elissa J. Chesler, Lu Lu, Siming Shou, Yanhua Qu, Jing Gu, Jintao Wang, Hui Chen Hsu, John D. Mountz, Nicole E. Baldwin, Michael A. Langston, David W. Threadgill, Kenneth F. Manly, and Robert W. Williams. Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nature Genetics*, 37(3):233–242, March 2005.

[240] Mathieu Clément-Ziza, Francesc X. Marsellach, Sandra Codlin, Manos A. Papadakis, Susanne Reinhardt, María Rodríguez-López, Stuart Martin, Samuel Marguerat, Alexander Schmidt, Eunhye Lee, Christopher T. Workman, Jürg Bähler, and Andreas Beyer. Natural genetic variation impacts expression levels of coding, non-coding, and antisense transcripts in fission yeast. *Molecular Systems Biology*, 10:764, November 2014.

[241] Yehudit Hasin-Brumshtein, Arshad H. Khan, Farhad Hormozdiari, Calvin Pan, Brian W. Parks, Vladislav A. Petyuk, Paul D. Piehowski, Anneke Brümmer, Matteo Pellegrini, Xinshu Xiao, Eleazar Eskin, Richard D. Smith, Aldons J. Lusis, and Desmond J. Smith. Hypothalamic transcriptomes of 99 mouse strains reveal trans eQTL hotspots, splicing QTLs and novel non-coding genes. 5, September 2016.

[242] Boel Brynedal, JinMyung Choi, Towfique Raj, Robert Bjornson, Barbara E. Stranger, Benjamin M. Neale, Benjamin F. Voight, and Chris Cotsapas. Large-Scale trans-eQTLs Affect Hundreds of Transcripts and Mediate Patterns of Transcriptional Co-regulation. *American Journal of Human Genetics*, 100(4):581–591, April 2017.

[243] Stephanie McGarrity, Rob Mason, Kevin C. Fone, Marie Pezze, and Tobias Bast. Hippocampal Neural Disinhibition Causes Attentional and Memory Deficits. *Cerebral Cortex (New York, N.Y.: 1991)*, August 2016.

[244] Christoph Anacker and René Hen. Adult hippocampal neurogenesis and cognitive flexibility — linking memory and mood. *Nature Reviews Neuroscience*, 18(6):335–346, June 2017.

[245] Howard Eichenbaum. Prefrontal-hippocampal interactions in episodic memory. *Nature Reviews Neuroscience*, advance online publication, June 2017.

[246] Alison R. Preston and Howard Eichenbaum. Interplay of Hippocampus and Prefrontal Cortex in Memory. *Current Biology*, 23(17):R764–R773, September 2013.

[247] Rita Z. Goldstein and Nora D. Volkow. Dysfunction of the prefrontal cortex in addiction: Neuroimaging findings and clinical implications. *Nature Reviews. Neuroscience*, 12(11):652–669, October 2011.

[248] James M. Otis, Vijay M. K. Namboodiri, Ana M. Matan, Elisa S. Voets, Emily P. Mohorn, Oksana Kosyk, Jenna A. McHenry, J. Elliott Robinson, Shanna L. Resendez, Mark A. Rossi, and Garret D. Stuber. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. 543(7643):103–107, 03 02, 2017.

[249] Tobias Bast, Marie Pezze, and Stephanie McGarrity. Cognitive deficits caused by prefrontal cortical and hippocampal neural disinhibition. *British Journal of Pharmacology*, May 2017.

[250] Marjan Jahanshahi, Ignacio Obeso, John C. Rothwell, and José A. Obeso. A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. *Nature Reviews Neuroscience*, 16(12):719–732, December 2015.

[251] L. M. Yager, A. F. Garcia, A. M. Wunsch, and S. M. Ferguson. The ins and outs of the striatum: Role in drug addiction. 301:529–541, August 2015.

[252] Anthony H. V. Schapira, K. Ray Chaudhuri, and Peter Jenner. Non-motor features of Parkinson disease. *Nature Reviews Neuroscience*, 18(7):435–450, July 2017.

[253] Daniela Vitucci, Annabella Di Giorgio, Francesco Napolitano, Barbara Pelosi, Giuseppe Blasi, Francesco Errico, Maria Teresa Attrotto, Barbara Gelao, Leonardo Fazio, Paolo Taurisano, Anna Di Maio, Valentina Marsili, Massimo Pasqualetti, Alessandro Bertolino, and Alessandro Usiello. Rasd2 Modulates Prefronto-Striatal Phenotypes in Humans and 'Schizophrenia-Like Behaviors' in Mice. *Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology*, 41(3):916–927, February 2016.

[254] Samanta Rodrigues, Cristiane Salum, and Tatiana L. Ferreira. Dorsal striatum D1-expressing neurons are involved with sensorimotor gating on prepulse inhibition test. *Journal of Psychopharmacology (Oxford, England)*, 31(4):505–513, April 2017.

[255] PsychENCODE Consortium, Schahram Akbarian, Chunyu Liu, James A. Knowles, Flora M. Vaccarino, Peggy J. Farnham, Gregory E. Crawford, Andrew E. Jaffe, Dalila Pinto, Stella Dracheva, Daniel H. Geschwind, Jonathan Mill, Angus C. Nairn, Alexej Abyzov, Sirisha Pochareddy, Shyam Prabhakar, Sherman Weissman, Patrick F. Sullivan, Matthew W. State, Zhiping Weng, Mette A. Peters, Kevin P. White, Mark B. Gerstein, Anahita Amiri, Chris Armoskus, Allison E. Ashley-Koch, Taejeong Bae, Andrea Beckel-Mitchener, Benjamin P. Berman, Gerhard A. Coetzee, Gianfilippo Coppola, Nancy Francoeur, Menachem Fromer, Robert Gao, Kay Grennan, Jennifer Herstein, David H. Kavanagh, Nikolay A. Ivanov, Yan Jiang, Robert R. Kitchen, Alexey Kozlenkov, Marija Kundakovic, Mingfeng Li, Zhen Li, Shuang Liu, Lara M. Mangravite, Eugenio Mattei, Eirene Markenscoff-Papadimitriou, Fábio C. P. Navarro, Nicole North, Larsson Omberg, David Panchision, Neelroop Parikshak, Jeremie Poschmann, Amanda J. Price, Michael Purcaro, Timothy E. Reddy, Panos Roussos, Shannon Schreiner, Soraya Scuderi, Robert Sebra, Mikihito Shibata, Annie W. Shieh, Mario Skarica, Wenjie Sun, Vivek Swarup, Amber Thomas, Junko Tsuji, Harm van Bakel,

Daifeng Wang, Yongjun Wang, Kai Wang, Donna M. Werling, A. Jeremy Willsey, Heather Witt, Hyejung Won, Chloe C. Y. Wong, Gregory A. Wray, Emily Y. Wu, Xuming Xu, Lijing Yao, Geetha Senthil, Thomas Lehner, Pamela Sklar, and Nenad Sestan. The PsychENCODE project. *Nature Neuroscience*, 18(12):1707–1712, December 2015.

[256] Jennifer Harrow, Adam Frankish, Jose M. Gonzalez, Electra Tapanari, Mark Diekhans, Felix Kokocinski, Bronwen L. Aken, Daniel Barrell, Amonida Zadissa, Stephen Searle, If Barnes, Alexandra Bignell, Veronika Boychenko, Toby Hunt, Mike Kay, Gaurab Mukherjee, Jeena Rajan, Gloria Despacio-Reyes, Gary Saunders, Charles Steward, Rachel Harte, Michael Lin, Cédric Howald, Andrea Tanzer, Thomas Derrien, Jacqueline Chrast, Nathalie Walters, Suganthi Balasubramanian, Baikang Pei, Michael Tress, Jose Manuel Rodriguez, Iakes Ezkurdia, Jeltje van Baren, Michael Brent, David Haussler, Manolis Kellis, Alfonso Valencia, Alexandre Reymond, Mark Gerstein, Roderic Guigó, and Tim J. Hubbard. GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Research*, 22(9):1760–1774, September 2012.

[257] Richard C. Davis, Atila van Nas, Brian Bennett, Luz Orozco, Calvin Pan, Christoph D. Rau, Eleazar Eskin, and Aldons J. Lusis. Genome-wide association mapping of blood cell traits in mice. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 24(3-4):105–118, April 2013.

[258] Blanca E. Himes, Keith Sheppard, Annerose Berndt, Adriana S. Leme, Rachel A. Myers, Christopher R. Gignoux, Albert M. Levin, W. James Gauderman, James J. Yang, Rasika A. Mathias, Isabelle Romieu, Dara G. Torgerson, Lindsey A. Roth, Scott Huntsman, Celeste Eng, Barbara Klanderman, John Ziniti, Jody Senter-Sylvia, Stanley J. Szefler, Robert F. Lemanske, Robert S. Zeiger, Robert C. Strunk, Fernando D. Martinez, Homer Boushey, Vernon M. Chinchilli, Elliot Israel, David Mauger, Gerard H. Koppelman, Dirkje S. Postma, Maartje A. E. Nieuwenhuis, Judith M. Vonk, John J. Lima, Charles G. Irvin, Stephen P. Peters, Michiaki Kubo, Mayumi Tamari, Yusuke Nakamura, Augusto A. Litonjua, Kelan G. Tantisira, Benjamin A. Raby, Eugene R. Bleecker, Deborah A. Meyers, Stephanie J. London, Kathleen C. Barnes, Frank D. Gilliland, L. Keoki Williams, Esteban G. Burchard, Dan L. Nicolae, Carole Ober, Dawn L. DeMeo, Edwin K. Silverman, Beverly Paigen, Gary Churchill, Steve D. Shapiro, and Scott T. Weiss. Integration of mouse and human genome-wide association data identifies KCNIP4 as an asthma gene. *PloS One*, 8(2):e56179, 2013.

[259] J.M. Welch, J. Lu, R.M. Rodriguiz, N.C. Trotta, J. Peca, J.-D. Ding, C. Feliciano, M. Chen, J.P. Adams, J. Luo, S.M. Dudek, R.J. Weinberg, N. Calakos, W.C. Wetsel, and G. Feng. Cortico-striatal synaptic defects and OCD-like behaviours in Sapap3-mutant mice. 448(7156):894–900, 2007.

[260] Eric Burguière, Patrícia Monteiro, Guoping Feng, and Ann M. Graybiel. Optogenetic Stimulation of Lateral Orbitofronto-Striatal Pathway Suppresses Compulsive Behaviors. 340(6137):1243–1246, June 2013.

[261] Susanne E. Ahmari, Timothy Spellman, Neria L. Douglass, Mazen A. Kheirbek, H. Blair Simpson, Karl Deisseroth, Joshua A. Gordon, and René Hen. Repeated cortico-striatal stimulation generates persistent OCD-like behavior. *Science (New York, N.Y.)*, 340(6137):1234–1239, June 2013.

[262] Peter J. Uhlhaas and Wolf Singer. Neuronal Dynamics and Neuropsychiatric Disorders: Toward a Translational Paradigm for Dysfunctional Large-Scale Networks. 75(6):963–980, September 2012.

[263] Mary Shimoyama, Stanley J. F. Laulederkind, Jeff De Pons, Rajni Nigam, Jennifer R. Smith, Marek Tutaj, Victoria Petri, G. Thomas Hayman, Shur-Jen Wang, Omid Ghiasvand, Jyothi Thota, and Melinda R. Dwinell. Exploring human disease using the Rat Genome Database. *Disease Models & Mechanisms*, 9(10):1089–1095, October 2016.

[264] Yoav Gilad and Orna Mizrahi-Man. A reanalysis of mouse ENCODE comparative gene expression data. 4:121, 2015.

[265] Shin Lin, Yiing Lin, Joseph R. Nery, Mark A. Urich, Alessandra Breschi, Carrie A. Davis, Alexander Dobin, Christopher Zaleski, Michael A. Beer, William C. Chapman, Thomas R. Gingeras, Joseph R. Ecker, and Michael P. Snyder. Comparison of the transcriptional landscapes between human and mouse tissues. *Proceedings of the National Academy of Sciences*, 111(48):17224–17229, February 2014.

[266] Keizo Takao, Hideo Hagihara, and Tsuyoshi Miyakawa. Reply to Warren et al. and Shay et al.: Commonalities across species do exist and are potentially important. *Proceedings of the National Academy of Sciences of the United States of America*, 112(4):E347–348, January 2015.

[267] Tal Shay, James A. Lederer, and Christophe Benoist. Genomic responses to inflammation in mouse models mimic humans: We concur, apples to oranges comparisons won't do. *Proceedings of the National Academy of Sciences of the United States of America*, 112(4):E346, January 2015.

[268] Junhee Seok, H. Shaw Warren, Alex G. Cuenca, Michael N. Mindrinos, Henry V. Baker, Weihong Xu, Daniel R. Richards, Grace P. McDonald-Smith, Hong Gao, Laura Hennessy, Celeste C. Finnerty, Cecilia M. López, Shari Honari, Ernest E. Moore, Joseph P. Minei, Joseph Cuschieri, Paul E. Bankey, Jeffrey L. Johnson, Jason Sperry, Avery B. Nathens, Timothy R. Billiar, Michael A. West, Marc G. Jeschke, Matthew B. Klein, Richard L. Gamelli, Nicole S. Gibran, Bernard H. Brownstein, Carol Miller-Graziano, Steve E. Calvano, Philip H. Mason, J. Perren Cobb, Laurence G. Rahme, Stephen F. Lowry, Ronald V. Maier, Lyle L. Moldawer, David N. Herndon, Ronald W. Davis, Wenzhong Xiao, Ronald G. Tompkins, and Inflammation and Host Response to Injury, Large Scale Collaborative Research Program. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proceedings of the National Academy of Sciences of the United States of America*, 110(9):3507–3512, February 2013.

[269] Irene Gallego Romero, Athma A. Pai, Jenny Tung, and Yoav Gilad. RNA-seq: Impact of RNA degradation on transcript quantification. *BMC biology*, 12:42, May 2014.

[270] Zong Hong Zhang, Dhanisha J. Jhaveri, Vikki M. Marshall, Denis C. Bauer, Janette Edson, Ramesh K. Narayanan, Gregory J. Robinson, Andreas E. Lundberg, Perry F. Bartlett, Naomi R. Wray, and Qiong-Yi Zhao. A comparative study of techniques for differential expression analysis on RNA-Seq data. *PloS One*, 9(8):e103207, 2014.

[271] Mouse Genome Sequencing Consortium, Robert H. Waterston, Kerstin Lindblad-Toh, Ewan Birney, Jane Rogers, Josep F. Abril, Pankaj Agarwal, Richa Agarwala, Rachel Ainscough, Marina Alexandersson, Peter An, Stylianos E. Antonarakis, John Attwood, Robert Baertsch, Jonathon Bailey, Karen Barlow, Stephan Beck, Eric Berry, Bruce Birren, Toby Bloom, Peer Bork, Marc Botcherby, Nicolas Bray, Michael R. Brent, Daniel G. Brown, Stephen D. Brown, Carol Bult, John Burton, Jonathan Butler, Robert D. Campbell, Piero Carninci, Simon Cawley, Francesca Chiaromonte, Asif T. Chinwalla, Deanna M. Church, Michele Clamp, Christopher Clee, Francis S. Collins, Lisa L. Cook, Richard R. Copley, Alan Coulson, Olivier Couronne, James Cuff, Val Curwen, Tim Cutts, Mark Daly, Robert David, Joy Davies, Kimberly D. Delehaunty, Justin Deri, Emmanouil T. Dermitzakis, Colin Dewey, Nicholas J. Dickens, Mark Diekhans, Sheila Dodge, Inna Dubchak, Diane M. Dunn, Sean R. Eddy, Laura Elnitski, Richard D. Emes, Pallavi Eswara, Eduardo Eyras, Adam Felsenfeld, Ginger A. Fewell, Paul Flicek, Karen Foley, Wayne N. Frankel, Lucinda A. Fulton, Robert S. Fulton, Terrence S. Furey, Diane Gage, Richard A. Gibbs, Gustavo Glusman, Sante Gnerre, Nick Goldman, Leo Goodstadt, Darren Grafham, Tina A. Graves, Eric D. Green, Simon Gregory, Roderic Guigó, Mark Guyer, Ross C. Hardison, David Haussler, Yoshihide Hayashizaki, LaDeana W. Hillier, Angela Hinrichs, Wratko Hlavina, Timothy Holzer, Fan Hsu, Axin Hua, Tim Hubbard, Adrienne Hunt, Ian Jackson, David B. Jaffe, L. Steven Johnson, Matthew Jones, Thomas A. Jones, Ann Joy, Michael Kamal, Elinor K. Karlsson, Donna Karolchik, Arkadiusz Kasprzyk, Jun Kawai, Evan Keibler, Cristyn Kells, W. James Kent, Andrew Kirby, Diana L. Kolbe, Ian Korf, Raju S. Kucherlapati, Edward J. Kulbokas, David Kulp, Tom Landers, J. P. Leger, Steven Leonard, Ivica Letunic, Rosie Levine, Jia Li, Ming Li, Christine Lloyd, Susan Lucas, Bin Ma, Donna R. Maglott, Elaine R. Mardis, Lucy Matthews, Evan Mauceli, John H. Mayer, Megan McCarthy, W. Richard McCombie, Stuart McLaren, Kirsten McLay, John D. McPherson, Jim Meldrim, Beverley Meredith, Jill P. Mesirov, Webb Miller, Tracie L. Miner, Emmanuel Mongin, Kate T. Montgomery, Michael Morgan, Richard Mott, James C. Mullikin, Donna M. Muzny, William E. Nash, Joanne O. Nelson, Michael N. Nhan, Robert Nicol, Zemin Ning, Chad Nusbaum, Michael J. O'Connor, Yasushi Okazaki, Karen Oliver, Emma Overton-Larty, Lior Pachter, Genís Parra, Kymberlie H. Pepin, Jane Peterson, Pavel Pevzner, Robert Plumb, Craig S. Pohl, Alex Poliakov, Tracy C. Ponce, Chris P. Ponting, Simon Potter, Michael Quail, Alexandre Reymond, Bruce A. Roe, Krishna M. Roskin, Edward M. Rubin, Alistair G. Rust, Ralph Santos, Victor Sapojnikov, Brian Schultz, Jörg Schultz, Matthias S. Schwartz, Scott Schwartz, Carol Scott, Steven Seaman, Steve Searle, Ted Sharpe, Andrew Sheri-

dan, Ratna Shownkeen, Sarah Sims, Jonathan B. Singer, Guy Slater, Arian Smit, Douglas R. Smith, Brian Spencer, Arne Stabenau, Nicole Stange-Thomann, Charles Sugnet, Mikita Suyama, Glenn Tesler, Johanna Thompson, David Torrents, Evanne Trevaskis, John Tromp, Catherine Ucla, Abel Ureta-Vidal, Jade P. Vinson, Andrew C. Von Niederhausern, Claire M. Wade, Melanie Wall, Ryan J. Weber, Robert B. Weiss, Michael C. Wendl, Anthony P. West, Kris Wetterstrand, Raymond Wheeler, Simon Whelan, Jamey Wierzbowski, David Willey, Sophie Williams, Richard K. Wilson, Eitan Winter, Kim C. Worley, Dudley Wyman, Shan Yang, Shiaw-Pyng Yang, Evgeny M. Zdobnov, Michael C. Zody, and Eric S. Lander. Initial sequencing and comparative analysis of the mouse genome. 420(6915):520–562, December 2002.

[272] C. Amid, A. Bahr, A. Mujica, N. Sampson, S. E. Bikar, A. Winterpacht, B. Zabel, T. Hankeln, and E. R. Schmidt. Comparative genomic sequencing reveals a strikingly similar architecture of a conserved syntenic region on human chromosome 11p15.3 (including gene ST5) and mouse chromosome 7. *Cytogenetics and Cell Genetics*, 93(3-4):284–290, 2001.

[273] Tao Yu, Zhongyou Li, Zhengping Jia, Steven J. Clapcote, Chunhong Liu, Shaomin Li, Suhail Asrar, Annie Pao, Rongqing Chen, Ni Fan, Sandra Carattini-Rivera, Allison R. Bechard, Shoshana Spring, R. Mark Henkelman, George Stoica, Sei-Ichi Matsui, Norma J. Nowak, John C. Roder, Chu Chen, Allan Bradley, and Y. Eugene Yu. A mouse model of Down syndrome trisomic for all human chromosome 21 syntenic regions. *Human Molecular Genetics*, 19(14):2780–2791, July 2010.

[274] Javier Herrero, Matthieu Muffato, Kathryn Beal, Stephen Fitzgerald, Leo Gordon, Miguel Pignatelli, Albert J. Vilella, Stephen M. J. Searle, Ridwan Amode, Simon Brent, William Spooner, Eugene Kulesha, Andrew Yates, and Paul Flicek. Ensembl comparative genomics resources. *Database: The Journal of Biological Databases and Curation*, 2016, 2016.

[275] Dmitri D. Pervouchine, Sarah Djebali, Alessandra Breschi, Carrie A. Davis, Pablo Prieto Barja, Alex Dobin, Andrea Tanzer, Julien Lagarde, Chris Zaleski, Lei-Hoon See, Meagan Fastuca, Jorg Drenkow, Huaien Wang, Giovanni Bussotti, Baikang Pei, Suganthi Balasubramanian, Jean Monlong, Arif Harmanci, Mark Gerstein, Michael A. Beer, Cedric Notredame, Roderic Guigó, and Thomas R. Gingeras. Enhanced transcriptome maps from multiple mouse tissues reveal evolutionary constraint in gene expression. *Nature Communications*, 6:5903, January 2015.

[276] Gioele La Manno, Daniel Gyllborg, Simone Codeluppi, Kaneyasu Nishimura, Carmen Salto, Amit Zeisel, Lars E. Borm, Simon R. W. Stott, Enrique M. Toledo, J. Carlos Villaescusa, Peter Lönnerberg, Jesper Ryge, Roger A. Barker, Ernest Arenas, and Sten Linnarsson. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. 167(2):566–580.e19, October 2016.

[277] Alex Bayés, Mark O. Collins, Mike D. R. Croning, Louie N. van de Lagemaat, Jyoti S. Choudhary, and Seth G. N. Grant. Comparative study of human and mouse postsy-

naptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PloS One*, 7(10):e46683, 2012.

[278] Chongyuan Luo, Christopher L. Keown, Laurie Kurihara, Jingtian Zhou, Yupeng He, Junhao Li, Rosa Castanon, Jacinta Lucero, Joseph R. Nery, Justin P. Sandoval, Brian Bui, Terrence J. Sejnowski, Timothy T. Harkins, Eran A. Mukamel, M. Margarita Behrens, and Joseph R. Ecker. Single-cell methylomes identify neuronal subtypes and regulatory elements in mammalian cortex. *Science (New York, N.Y.)*, 357(6351):600–604, August 2017.

[279] Xiangqun Zheng-Bradley, Johan Rung, Helen Parkinson, and Alvis Brazma. Large scale comparison of global gene expression patterns in human and mouse. *Genome Biology*, 11(12):R124, 2010.

[280] Alessandra Breschi, Sarah Djebali, Jesse Gillis, Dmitri D. Pervouchine, Alex Dobin, Carrie A. Davis, Thomas R. Gingeras, and Roderic Guigó. Gene-specific patterns of expression variation across organs and species. *Genome Biology*, 17(1):151, July 2016.

[281] Yong Cheng, Zhihai Ma, Bong-Hyun Kim, Weisheng Wu, Philip Cayting, Alan P. Boyle, Vasavi Sundaram, Xiaoyun Xing, Nergiz Dogan, Jingjing Li, Ghia Euskirchen, Shin Lin, Yiing Lin, Axel Visel, Trupti Kawli, Xinqiong Yang, Dorrelyn Patacsil, Cheryl A. Keller, Belinda Giardine, mouse ENCODE Consortium, Anshul Kundaje, Ting Wang, Len A. Pennacchio, Zhiping Weng, Ross C. Hardison, and Michael P. Snyder. Principles of regulatory information conservation between mouse and human. 515(7527):371–375, November 2014.

[282] Mitchell Guttman, Ido Amit, Manuel Garber, Courtney French, Michael F. Lin, David Feldser, Maite Huarte, Or Zuk, Bryce W. Carey, John P. Cassady, Moran N. Cabili, Rudolf Jaenisch, Tarjei S. Mikkelsen, Tyler Jacks, Nir Hacohen, Bradley E. Bernstein, Manolis Kellis, Aviv Regev, John L. Rinn, and Eric S. Lander. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. 458(7235):223–227, March 2009.

[283] Olga A. Vakhrusheva, Georgii A. Bazykin, and Alexey S. Kondrashov. Genome-Level Analysis of Selective Constraint without Apparent Sequence Conservation. *Genome Biology and Evolution*, 5(3):532–541, 2013.

[284] Anne-Ruxandra Carvunis, Tina Wang, Dylan Skola, Alice Yu, Jonathan Chen, Jason F. Kreisberg, and Trey Ideker. Evidence for a common evolutionary rate in metazoan transcriptional networks. 4, December 2015.

[285] Feng Yue, Yong Cheng, Alessandra Breschi, Jeff Vierstra, Weisheng Wu, Tyrone Ryba, Richard Sandstrom, Zhihai Ma, Carrie Davis, Benjamin D. Pope, Yin Shen, Dmitri D. Pervouchine, Sarah Djebali, Robert E. Thurman, Rajinder Kaul, Eric Rynes, Anthony Kirilusha, Georgi K. Marinov, Brian A. Williams, Diane Trout, Henry Amrhein, Katherine Fisher-Aylor, Igor Antoshechkin, Gilberto DeSalvo, Lei-Hoon See, Meagan

Fastuca, Jorg Drenkow, Chris Zaleski, Alex Dobin, Pablo Prieto, Julien Lagarde, Giovanni Bussotti, Andrea Tanzer, Olgert Denas, Kanwei Li, M. A. Bender, Miaohua Zhang, Rachel Byron, Mark T. Groudine, David McCleary, Long Pham, Zhen Ye, Samantha Kuan, Lee Edsall, Yi-Chieh Wu, Matthew D. Rasmussen, Mukul S. Bansal, Manolis Kellis, Cheryl A. Keller, Christapher S. Morrissey, Tejaswini Mishra, Deepti Jain, Nergiz Dogan, Robert S. Harris, Philip Cayting, Trupti Kawli, Alan P. Boyle, Ghia Euskirchen, Anshul Kundaje, Shin Lin, Yiing Lin, Camden Jansen, Venkat S. Malladi, Melissa S. Cline, Drew T. Erickson, Vanessa M. Kirkup, Katrina Learned, Cricket A. Sloan, Kate R. Rosenbloom, Beatriz Lacerda de Sousa, Kathryn Beal, Miguel Pignatelli, Paul Flicek, Jin Lian, Tamer Kahveci, Dongwon Lee, W. James Kent, Miguel Ramalho Santos, Javier Herrero, Cedric Notredame, Audra Johnson, Shinny Vong, Kristen Lee, Daniel Bates, Fidencio Neri, Morgan Diegel, Theresa Canfield, Peter J. Sabo, Matthew S. Wilken, Thomas A. Reh, Erika Giste, Anthony Shafer, Tanya Kutyavin, Eric Haugen, Douglas Dunn, Alex P. Reynolds, Shane Neph, Richard Humbert, R. Scott Hansen, Marella De Bruijn, Licia Selleri, Alexander Rudensky, Steven Josefowicz, Robert Samstein, Evan E. Eichler, Stuart H. Orkin, Dana Levasseur, Thalia Papayannopoulou, Kai-Hsin Chang, Arthur Skoultchi, Srikanta Gosh, Christine Disteche, Piper Treuting, Yanli Wang, Mitchell J. Weiss, Gerd A. Blobel, Xiaoyi Cao, Sheng Zhong, Ting Wang, Peter J. Good, Rebecca F. Lowdon, Leslie B. Adams, Xiao-Qiao Zhou, Michael J. Pazin, Elise A. Feingold, Barbara Wold, James Taylor, Ali Mortazavi, Sherman M. Weissman, John A. Stamatoyannopoulos, Michael P. Snyder, Roderic Guigo, Thomas R. Gingeras, David M. Gilbert, Ross C. Hardison, Michael A. Beer, Bing Ren, and Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. 515(7527):355–364, November 2014.

[286] Jacob F. Degner, John C. Marioni, Athma A. Pai, Joseph K. Pickrell, Everlyne Nkadori, Yoav Gilad, and Jonathan K. Pritchard. Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics (Oxford, England)*, 25(24):3207–3212, December 2009.

[287] X. Wang and A. G. Clark. Using next-generation RNA sequencing to identify imprinted genes. 113(2):156–166, August 2014.

[288] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 12(4):357–360, April 2015.

[289] Bronwen L. Aken, Sarah Ayling, Daniel Barrell, Laura Clarke, Valery Curwen, Susan Fairley, Julio Fernandez Banet, Konstantinos Billis, Carlos García Girón, Thibaut Hourlier, Kevin Howe, Andreas Kähäri, Felix Kokocinski, Fergal J. Martin, Daniel N. Murphy, Rishi Nag, Magali Ruffier, Michael Schuster, Y. Amy Tang, Jan-Hinnerk Vogel, Simon White, Amonida Zadissa, Paul Flicek, and Stephen M. J. Searle. The Ensembl gene annotation system. *Database: The Journal of Biological Databases and Curation*, 2016, 2016.

[290] Michael Lawrence, Wolfgang Huber, Hervé Pagès, Patrick Aboyoun, Marc Carlson, Robert Gentleman, Martin T. Morgan, and Vincent J. Carey. Software for computing and annotating genomic ranges. *PLoS computational biology*, 9(8):e1003118, 2013.

[291] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biology*, 11(10):R106, 2010.

[292] Joe R. Davis, Laure Fresard, David A. Knowles, Mauro Pala, Carlos D. Bustamante, Alexis Battle, and Stephen B. Montgomery. An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *American Journal of Human Genetics*, 98(1):216–224, January 2016.

[293] Arimantas Lionikas, Caroline Meharg, Jonathan Mj Derry, Aivaras Ratkevicius, Andrew M. Carroll, David J. Vandenbergh, and David A. Blizard. Resolving candidate genes of mouse skeletal muscle QTL via RNA-Seq and expression network analyses. *BMC genomics*, 13:592, November 2012.

[294] Jacqueline MacArthur, Emily Bowler, Maria Cerezo, Laurent Gil, Peggy Hall, Emma Hastings, Heather Junkins, Aoife McMahon, Annalisa Milano, Joannella Morales, Zoe May Pendlington, Danielle Welter, Tony Burdett, Lucia Hindorff, Paul Flicek, Fiona Cunningham, and Helen Parkinson. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, 45(D1):D896–D901, January 2017.

[295] Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, and Avi Ma'ayan. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–97, July 2016.

[296] Edward Y. Chen, Christopher M. Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R. Clark, and Avi Ma'ayan. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC bioinformatics*, 14:128, April 2013.

[297] Sarah R. Langley, Leonardo Bottolo, Jaroslav Kunes, Josef Zicha, Vaclav Zidek, Norbert Hubner, Stuart A. Cook, Michal Pravenec, Timothy J. Aitman, and Enrico Petretto. Systems-level approaches reveal conservation of trans-regulated genes in the rat and genetic determinants of blood pressure in humans. *Cardiovascular Research*, 97(4):653–665, March 2013.

[298] Xue Li, Valentina Perissi, Forrest Liu, David W. Rose, and Michael G. Rosenfeld. Tissue-specific regulation of retinal and pituitary precursor cell proliferation. *Science (New York, N.Y.)*, 297(5584):1180–1183, August 2002.

[299] Jinshu Xu and Pin-Xian Xu. Eya-six are necessary for survival of nephrogenic cord progenitors and inducing nephric duct development before ureteric bud formation.

Developmental Dynamics: An Official Publication of the American Association of Anatomists, 244(7):866–873, July 2015.

[300] Ray-Chang Wu, Ming Jiang, Arthur L. Beaudet, and Mei-Yi Wu. ARID4A and ARID4B regulate male fertility, a functional link to the AR and RB pathways. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12):4616–4621, March 2013.

[301] Mei-Yi Wu, Ting-Fen Tsai, and Arthur L. Beaudet. Deficiency of Rbbp1/Arid4a and Rbbp1l1/Arid4b alters epigenetic modifications and suppresses an imprinting defect in the PWS/AS domain. *Genes & Development*, 20(20):2859–2870, October 2006.

[302] Shuhei Tomita, Masaki Ueno, Masami Sakamoto, Yuki Kitahama, Masaaki Ueki, Nobuhiro Maekawa, Haruhiko Sakamoto, Max Gassmann, Ryoichiro Kageyama, Natsuo Ueda, Frank J. Gonzalez, and Yousuke Takahama. Defective brain development in mice lacking the Hif-1alpha gene in neural cells. *Molecular and Cellular Biology*, 23(19):6739–6749, October 2003.

[303] Nina Korsisaari, Derrick J. Rossi, Anders Paetau, Patrick Charnay, Mark Henkemeyer, and Tomi P. Mäkelä. Conditional ablation of the Mat1 subunit of TFIIH in Schwann cells provides evidence that Mat1 is not required for general transcription. *Journal of Cell Science*, 115(Pt 22):4275–4284, November 2002.

[304] Fangyan Dai, Tao Shen, Zhaoyong Li, Xia Lin, and Xin-Hua Feng. PPM1A dephosphorylates RanBP3 to enable efficient nuclear export of Smad2 and Smad3. *EMBO reports*, 12(11):1175–1181, October 2011.

[305] D. Filliol, S. Ghozland, J. Chluba, M. Martin, H. W. Matthes, F. Simonin, K. Befort, C. Gavériaux-Ruff, A. Dierich, M. LeMeur, O. Valverde, R. Maldonado, and B. L. Kieffer. Mice deficient for delta- and mu-opioid receptors exhibit opposing alterations of emotional responses. *Nature Genetics*, 25(2):195–200, June 2000.

[306] Sandra Seeher, Bradley A. Carlson, Angela C. Miniard, Eva K. Wirth, Yassin Mahdi, Dolph L. Hatfield, Donna M. Driscoll, and Ulrich Schweizer. Impaired selenoprotein expression in brain triggers striatal neuronal loss leading to co-ordination defects in mice. *The Biochemical Journal*, 462(1):67–75, August 2014.

[307] Yassin Mahdi, Xue-Ming Xu, Bradley A. Carlson, Noelia Fradejas, Paul Günter, Doreen Braun, Eileen Southon, Lino Tessarollo, Dolph L. Hatfield, and Ulrich Schweizer. Expression of Selenoproteins Is Maintained in Mice Carrying Mutations in SECp43, the tRNA Selenocysteine 1 Associated Protein (Trnau1ap). *PloS One*, 10(6):e0127349, 2015.

[308] Hadassa Brunschwig, Liat Levi, Eyal Ben-David, Robert W. Williams, Benjamin Yakir, and Sagiv Shifman. Fine-scale maps of recombination rates and hotspots in the mouse genome. 191(3):757–764, July 2012.

[309] Jhih-Rong Lin, Ying Cai, Quanwei Zhang, Wen Zhang, Rubén Nogales-Cadenas, and Zhengdong D. Zhang. Integrated Post-GWAS Analysis Sheds New Light on the Disease Mechanisms of Schizophrenia. 204(4):1587–1600, December 2016.

[310] Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. 511(7510):421–427, July 2014.

[311] Matt McGue, Yiwei Zhang, Michael B. Miller, Saonli Basu, Scott Vrieze, Brian Hicks, Steve Malone, William S. Oetting, and William G. Iacono. A genome-wide association study of behavioral disinhibition. *Behavior Genetics*, 43(5):363–373, September 2013.

[312] Elizabeth T. Cirulli, Dalia Kasperaviciūte, Deborah K. Attix, Anna C. Need, Dongliang Ge, Greg Gibson, and David B. Goldstein. Common genetic variation and performance on standardized cognitive tests. *European journal of human genetics: EJHG*, 18(7):815–820, July 2010.

[313] Margaret G. Distler, Mark D. Opal, Stephanie C. Dulawa, and Abraham A. Palmer. Assessment of behaviors modeling aspects of schizophrenia in Csmd1 mutant mice. *PloS One*, 7(12):e51235, 2012.

[314] Vidar M. Steen, Chirag Nepal, Kari M. Ersland, Rita Holdhus, Marianne Nævdal, Siri M. Ratvik, Silje Skrede, and Bjarte Håvik. Neuropsychological Deficits in Mice Depleted of the Schizophrenia Susceptibility Gene CSMD1. *PLoS ONE*, 8(11), November 2013.

[315] G. Kolle, K. Georgas, G. P. Holmes, M. H. Little, and T. Yamada. CRIM1, a novel gene encoding a cysteine-rich repeat protein, is developmentally regulated and implicated in vertebrate CNS development and organogenesis. *Mechanisms of Development*, 90(2):181–193, February 2000.

[316] Anca Alexandru, Wolfgang Jagla, Sigrid Graubner, Andreas Becker, Christoph Bäuscher, Stephanie Kohlmann, Reinhard Sedlmeier, Kerstin A. Raber, Holger Cynis, Raik Rönicke, Klaus G. Reymann, Elisabeth Petrasch-Parwez, Maike Hartlage-Rübsamen, Alexander Waniek, Steffen Rossner, Stephan Schilling, Alexander P. Osmand, Hans-Ulrich Demuth, and Stephan von Hörsten. Selective hippocampal neurodegeneration in transgenic mice expressing small amounts of truncated Aβ is induced by pyroglutamate-Aβ formation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(36):12790–12801, September 2011.

[317] Sadim Jawhar, Oliver Wirths, Stephan Schilling, Sigrid Graubner, Hans-Ulrich Demuth, and Thomas A. Bayer. Overexpression of glutaminyl cyclase, the enzyme responsible for pyroglutamate A{beta} formation, induces behavioral deficits, and glutaminyl cyclase knock-out rescues the behavioral phenotype in 5XFAD mice. *The Journal of Biological Chemistry*, 286(6):4454–4460, February 2011.

[318] Felemban Athary M. Abdulhaleem, Xiaohong Song, Rie Kawano, Naohiro Uezono, Ayako Ito, Giasuddin Ahmed, Mahmud Hossain, Kinichi Nakashima, Hideaki Tanaka,

and Kunimasa Ohta. Akhirin regulates the proliferation and differentiation of neural stem cells in intact and injured mouse spinal cord. *Developmental Neurobiology*, 75(5):494–504, May 2015.

[319] E. Seong, B. H. Wainer, E. D. Hughes, T. L. Saunders, M. Burmeister, and V. Faundez. Genetic analysis of the neuronal and ubiquitous AP-3 adaptor complexes reveals divergent functions in brain. *Molecular Biology of the Cell*, 16(1):128–140, January 2005.

[320] Heather Fairfield, Anuj Srivastava, Guruprasad Ananda, Rangjiao Liu, Martin Kircher, Anuradha Lakshminarayana, Belinda S. Harris, Son Yong Karst, Louise A. Dionne, Coleen C. Kane, Michelle Curtain, Melissa L. Berry, Patricia F. Ward-Bailey, Ian Greenstein, Candice Byers, Anne Czechanski, Jocelyn Sharp, Kristina Palmer, Polyxeni Gudis, Whitney Martin, Abby Tadenev, Laurent Bogdanik, C. Herbert Pratt, Bo Chang, David G. Schroeder, Gregory A. Cox, Paul Cliften, Jeffrey Milbrandt, Stephen Murray, Robert Burgess, David E. Bergstrom, Leah Rae Donahue, Hanan Hamamy, Amira Masri, Federico A. Santoni, Periklis Makrythanasis, Stylianos E. Antonarakis, Jay Shendure, and Laura G. Reinholdt. Exome sequencing reveals pathogenic mutations in 91 strains of mice with Mendelian disorders. *Genome Research*, 25(7):948–957, July 2015.

[321] Can Gao, Shanti F. Frausto, Anita L. Guedea, Natalie C. Tronson, Vladimir Jovasevic, Katie Leaderbrand, Kevin A. Corcoran, Yomayra F. Guzmán, Geoffrey T. Swanson, and Jelena Radulovic. IQGAP1 regulates NR2A signaling, spine density, and cognitive processes. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 31(23):8533–8542, June 2011.

[322] R. Pekhletski, R. Gerlai, L. S. Overstreet, X. P. Huang, N. Agopyan, N. T. Slater, W. Abramow-Newerly, J. C. Roder, and D. R. Hampson. Impaired cerebellar synaptic plasticity and motor performance in mice lacking the mGluR4 subtype of metabotropic glutamate receptor. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 16(20):6364–6373, October 1996.

[323] Chenglai Fu, Jing Xu, Ruo-Jing Li, Joshua A. Crawford, A. Basit Khan, Ting Martin Ma, Jiyoung Y. Cha, Adele M. Snowman, Mikhail V. Pletnikov, and Solomon H. Snyder. Inositol Hexakisphosphate Kinase-3 Regulates the Morphology and Synapse Formation of Cerebellar Purkinje Cells via Spectrin/Adducin. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 35(31):11056–11067, August 2015.

[324] H. G. McFarlane, G. K. Kusek, M. Yang, J. L. Phoenix, V. J. Bolivar, and J. N. Crawley. Autism-like behavioral phenotypes in BTBR T+tf/J mice. *Genes, Brain, and Behavior*, 7(2):152–163, March 2008.

[325] Laura R. Fenlon, Sha Liu, Ilan Gobius, Nyoman D. Kurniawan, Skyle Murphy, Randal X. Moldrich, and Linda J. Richards. Formation of functional areas in the cerebral

cortex is disrupted in a mouse model of autism spectrum disorder. *Neural Development*, 10:10, April 2015.

[326] Adam M. Stocker and Anjen Chenn. Focal Reduction of αE-catenin Causes Premature Differentiation and Reduction of β-catenin Signaling During Cortical Development. *Developmental biology*, 328(1):66–77, April 2009.

[327] Lihong Zhao, Christine Rosales, Kevin Seburn, David Ron, and Susan L. Ackerman. Alteration of the unfolded protein response modifies neurodegeneration in a mouse model of Marinesco-Sjögren syndrome. *Human Molecular Genetics*, 19(1):25–35, January 2010.

[328] Atsushi Oyagi, Yasuhisa Oida, Kenichi Kakefuda, Masamitsu Shimazawa, Norifumi Shioda, Shigeki Moriguchi, Kiyoyuki Kitaichi, Daisuke Nanba, Kazumasa Yamaguchi, Yasuhide Furuta, Kohji Fukunaga, Shigeki Higashiyama, and Hideaki Hara. Generation and characterization of conditional heparin-binding EGF-like growth factor knockout mice. *PloS One*, 4(10):e7461, October 2009.

[329] Weisheng V. Chen, Francisco J. Alvarez, Julie L. Lefebvre, Brad Friedman, Chiamaka Nwakeze, Eric Geiman, Courtney Smith, Chan Aye Thu, Juan Carlos Tapia, Bosiljka Tasic, Joshua R. Sanes, and Tom Maniatis. Functional significance of isoform diversification in the protocadherin gamma gene cluster. 75(3):402–409, August 2012.

[330] Sonoko Hasegawa, Hiroaki Kobayashi, Makiko Kumagai, Hiroshi Nishimaru, Etsuko Tarusawa, Hiro Kanda, Makoto Sanbo, Yumiko Yoshimura, Masumi Hirabayashi, Takahiro Hirabayashi, and Takeshi Yagi. Clustered Protocadherins Are Required for Building Functional Neural Circuits. *Frontiers in Molecular Neuroscience*, 10:114, 2017.

[331] Michael J. Molumby, Rachel M. Anderson, Dillan J. Newbold, Norah K. Koblesky, Andrew M. Garrett, Dietmar Schreiner, Jason J. Radley, and Joshua A. Weiner. γ-Protocadherins Interact with Neuroligin-1 and Negatively Regulate Dendritic Spine Morphogenesis. *Cell Reports*, 18(11):2702–2714, March 2017.

[332] Polina Kehayova, Kevin Monahan, Weisheng Chen, and Tom Maniatis. Regulatory elements required for the activation and repression of the protocadherin-α gene cluster. *Proceedings of the National Academy of Sciences*, 108(41):17195–17200, November 2011.

[333] Rotem Rubinstein, Chan Aye Thu, Kerry Marie Goodman, Holly Noelle Wolcott, Fabiana Bahna, Seetha Mannepalli, Goran Ahlsen, Maxime Chevee, Adnan Halim, Henrik Clausen, Tom Maniatis, Lawrence Shapiro, and Barry Honig. Molecular logic of neuronal self-recognition through protocadherin domain interactions. 163(3):629–642, October 2015.

[334] Takashi Sekiya, Ikkou Kashiwagi, Naoko Inoue, Rimpei Morita, Shohei Hori, Herman Waldmann, Alexander Y. Rudensky, Hiroshi Ichinose, Daniel Metzger, Pierre Chambon, and Akihiko Yoshimura. The nuclear orphan receptor Nr4a2 induces Foxp3 and regulates differentiation of CD4+ T cells. *Nature Communications*, 2:269, 2011.

[335] Yuan Zhang, Joyce J. Repa, Yusuke Inoue, Graham P. Hayhurst, Frank J. Gonzalez, and David J. Mangelsdorf. Identification of a liver-specific uridine phosphorylase that is regulated by multiple lipid-sensing nuclear receptors. *Molecular Endocrinology (Baltimore, Md.)*, 18(4):851–862, April 2004.

[336] Md Bazlur R. Mollah and Akira Ishikawa. A wild derived quantitative trait locus on mouse chromosome 2 prevents obesity. *BMC genetics*, 11:84, September 2010.

[337] Jeffrey S. Flier and Eleftheria Maratos-Flier. Leptin's Physiologic Role: Does the Emperor of Energy Balance Have No Clothes? *Cell Metabolism*, June 2017.

[338] David E. Sleat, Mukarram El-Banna, Istvan Sohar, Kwi-Hye Kim, Kostantin Dobrenis, Steven U. Walkley, and Peter Lobel. Residual levels of tripeptidyl-peptidase I activity dramatically ameliorate disease in late-infantile neuronal ceroid lipofuscinosis. *Molecular Genetics and Metabolism*, 94(2):222–233, June 2008.

[339] R. M. Gardiner. The molecular genetic basis of the neuronal ceroid lipofuscinoses. *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, 21(3 Suppl):S15–19, 2000.

[340] Sonja I. Berndt, Stefan Gustafsson, Reedik Mägi, Andrea Ganna, Eleanor Wheeler, Mary F. Feitosa, Anne E. Justice, Keri L. Monda, Damien C. Croteau-Chonka, Felix R. Day, Tõnu Esko, Tove Fall, Teresa Ferreira, Davide Gentilini, Anne U. Jackson, Jian'an Luan, Joshua C. Randall, Sailaja Vedantam, Cristen J. Willer, Thomas W. Winkler, Andrew R. Wood, Tsegaselassie Workalemahu, Yi-Juan Hu, Sang Hong Lee, Liming Liang, Dan-Yu Lin, Josine L. Min, Benjamin M. Neale, Gudmar Thorleifsson, Jian Yang, Eva Albrecht, Najaf Amin, Jennifer L. Bragg-Gresham, Gemma Cadby, Martin den Heijer, Niina Eklund, Krista Fischer, Anuj Goel, Jouke-Jan Hottenga, Jennifer E. Huffman, Ivonne Jarick, Åsa Johansson, Toby Johnson, Stavroula Kanoni, Marcus E. Kleber, Inke R. König, Kati Kristiansson, Zoltán Kutalik, Claudia Lamina, Cecile Lecoeur, Guo Li, Massimo Mangino, Wendy L. McArdle, Carolina Medina-Gomez, Martina Müller-Nurasyid, Julius S. Ngwa, Ilja M. Nolte, Lavinia Paternoster, Sonali Pechlivanis, Markus Perola, Marjolein J. Peters, Michael Preuss, Lynda M. Rose, Jianxin Shi, Dmitry Shungin, Albert Vernon Smith, Rona J. Strawbridge, Ida Surakka, Alexander Teumer, Mieke D. Trip, Jonathan Tyrer, Jana V. Van Vliet-Ostaptchouk, Liesbeth Vandenput, Lindsay L. Waite, Jing Hua Zhao, Devin Absher, Folkert W. Asselbergs, Mustafa Atalay, Antony P. Attwood, Anthony J. Balmforth, Hanneke Basart, John Beilby, Lori L. Bonnycastle, Paolo Brambilla, Marcel Bruinenberg, Harry Campbell, Daniel I. Chasman, Peter S. Chines, Francis S. Collins, John M. Connell, William O. Cookson, Ulf de Faire, Femmie de Vegt, Mariano Dei, Maria Dimitriou,

Sarah Edkins, Karol Estrada, David M. Evans, Martin Farrall, Marco M. Ferrario, Jean Ferrières, Lude Franke, Francesca Frau, Pablo V. Gejman, Harald Grallert, Henrik Grönberg, Vilmundur Gudnason, Alistair S. Hall, Per Hall, Anna-Liisa Hartikainen, Caroline Hayward, Nancy L. Heard-Costa, Andrew C. Heath, Johannes Hebebrand, Georg Homuth, Frank B. Hu, Sarah E. Hunt, Elina Hyppönen, Carlos Iribarren, Kevin B. Jacobs, John-Olov Jansson, Antti Jula, Mika Kähönen, Sekar Kathiresan, Frank Kee, Kay-Tee Khaw, Mika Kivimäki, Wolfgang Koenig, Aldi T. Kraja, Meena Kumari, Kari Kuulasmaa, Johanna Kuusisto, Jaana H. Laitinen, Timo A. Lakka, Claudia Langenberg, Lenore J. Launer, Lars Lind, Jaana Lindström, Jianjun Liu, Antonio Liuzzi, Marja-Liisa Lokki, Mattias Lorentzon, Pamela A. Madden, Patrik K. Magnusson, Paolo Manunta, Diana Marek, Winfried März, Irene Mateo Leach, Barbara McKnight, Sarah E. Medland, Evelin Mihailov, Lili Milani, Grant W. Montgomery, Vincent Mooser, Thomas W. Mühleisen, Patricia B. Munroe, Arthur W. Musk, Narisu Narisu, Gerjan Navis, George Nicholson, Ellen A. Nohr, Ken K. Ong, Ben A. Oostra, Colin N. A. Palmer, Aarno Palotie, John F. Peden, Nancy Pedersen, Annette Peters, Ozren Polasek, Anneli Pouta, Peter P. Pramstaller, Inga Prokopenko, Carolin Pütter, Aparna Radhakrishnan, Olli Raitakari, Augusto Rendon, Fernando Rivadeneira, Igor Rudan, Timo E. Saaristo, Jennifer G. Sambrook, Alan R. Sanders, Serena Sanna, Jouko Saramies, Sabine Schipf, Stefan Schreiber, Heribert Schunkert, So-Youn Shin, Stefano Signorini, Juha Sinisalo, Boris Skrobek, Nicole Soranzo, Alena Stančáková, Klaus Stark, Jonathan C. Stephens, Kathleen Stirrups, Ronald P. Stolk, Michael Stumvoll, Amy J. Swift, Eirini V. Theodoraki, Barbara Thorand, David-Alexandre Tregouet, Elena Tremoli, Melanie M. Van der Klauw, Joyce B. J. van Meurs, Sita H. Vermeulen, Jorma Viikari, Jarmo Virtamo, Veronique Vitart, Gérard Waeber, Zhaoming Wang, Elisabeth Widén, Sarah H. Wild, Gonneke Willemsen, Bernhard R. Winkelmann, Jacqueline C. M. Witteman, Bruce H. R. Wolffenbuttel, Andrew Wong, Alan F. Wright, M. Carola Zillikens, Philippe Amouyel, Bernhard O. Boehm, Eric Boerwinkle, Dorret I. Boomsma, Mark J. Caulfield, Stephen J. Chanock, L. Adrienne Cupples, Daniele Cusi, George V. Dedoussis, Jeanette Erdmann, Johan G. Eriksson, Paul W. Franks, Philippe Froguel, Christian Gieger, Ulf Gyllensten, Anders Hamsten, Tamara B. Harris, Christian Hengstenberg, Andrew A. Hicks, Aroon Hingorani, Anke Hinney, Albert Hofman, Kees G. Hovingh, Kristian Hveem, Thomas Illig, Marjo-Riitta Jarvelin, Karl-Heinz Jöckel, Sirkka M. Keinanen-Kiukaanniemi, Lambertus A. Kiemeney, Diana Kuh, Markku Laakso, Terho Lehtimäki, Douglas F. Levinson, Nicholas G. Martin, Andres Metspalu, Andrew D. Morris, Markku S. Nieminen, Inger Njølstad, Claes Ohlsson, Albertine J. Oldehinkel, Willem H. Ouwehand, Lyle J. Palmer, Brenda Penninx, Chris Power, Michael A. Province, Bruce M. Psaty, Lu Qi, Rainer Rauramaa, Paul M. Ridker, Samuli Ripatti, Veikko Salomaa, Nilesh J. Samani, Harold Snieder, Thorkild I. A. Sørensen, Timothy D. Spector, Kari Stefansson, Anke Tönjes, Jaakko Tuomilehto, André G. Uitterlinden, Matti Uusitupa, Pim van der Harst, Peter Vollenweider, Henri Wallaschofski, Nicholas J. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nature Genetics*, 45(5):501–512, May 2013.

[341] Adam E. Locke, Bratati Kahali, Sonja I. Berndt, Anne E. Justice, Tune H. Pers, Felix R. Day, Corey Powell, Sailaja Vedantam, Martin L. Buchkovich, Jian Yang, Damien C. Croteau-Chonka, Tonu Esko, Tove Fall, Teresa Ferreira, Stefan Gustafsson, Zoltán Kutalik, Jian'an Luan, Reedik Mägi, Joshua C. Randall, Thomas W. Winkler, Andrew R. Wood, Tsegaselassie Workalemahu, Jessica D. Faul, Jennifer A. Smith, Jing Hua Zhao, Wei Zhao, Jin Chen, Rudolf Fehrmann, Åsa K. Hedman, Juha Karjalainen, Ellen M. Schmidt, Devin Absher, Najaf Amin, Denise Anderson, Marian Beekman, Jennifer L. Bolton, Jennifer L. Bragg-Gresham, Steven Buyske, Ayse Demirkan, Guohong Deng, Georg B. Ehret, Bjarke Feenstra, Mary F. Feitosa, Krista Fischer, Anuj Goel, Jian Gong, Anne U. Jackson, Stavroula Kanoni, Marcus E. Kleber, Kati Kristiansson, Unhee Lim, Vaneet Lotay, Massimo Mangino, Irene Mateo Leach, Carolina Medina-Gomez, Sarah E. Medland, Michael A. Nalls, Cameron D. Palmer, Dorota Pasko, Sonali Pechlivanis, Marjolein J. Peters, Inga Prokopenko, Dmitry Shungin, Alena Stančáková, Rona J. Strawbridge, Yun Ju Sung, Toshiko Tanaka, Alexander Teumer, Stella Trompet, Sander W. van der Laan, Jessica van Setten, Jana V. Van Vliet-Ostaptchouk, Zhaoming Wang, Loïc Yengo, Weihua Zhang, Aaron Isaacs, Eva Albrecht, Johan Ärnlöv, Gillian M. Arscott, Antony P. Attwood, Stefania Bandinelli, Amy Barrett, Isabelita N. Bas, Claire Bellis, Amanda J. Bennett, Christian Berne, Roza Blagieva, Matthias Blüher, Stefan Böhringer, Lori L. Bonnycastle, Yvonne Böttcher, Heather A. Boyd, Marcel Bruinenberg, Ida H. Caspersen, Yii-Der Ida Chen, Robert Clarke, E. Warwick Daw, Anton J. M. de Craen, Graciela Delgado, Maria Dimitriou, Alex S. F. Doney, Niina Eklund, Karol Estrada, Elodie Eury, Lasse Folkersen, Ross M. Fraser, Melissa E. Garcia, Frank Geller, Vilmantas Giedraitis, Bruna Gigante, Alan S. Go, Alain Golay, Alison H. Goodall, Scott D. Gordon, Mathias Gorski, Hans-Jörgen Grabe, Harald Grallert, Tanja B. Grammer, Jürgen Gräßler, Henrik Grönberg, Christopher J. Groves, Gaëlle Gusto, Jeffrey Haessler, Per Hall, Toomas Haller, Goran Hallmans, Catharina A. Hartman, Maija Hassinen, Caroline Hayward, Nancy L. Heard-Costa, Quinta Helmer, Christian Hengstenberg, Oddgeir Holmen, Jouke-Jan Hottenga, Alan L. James, Janina M. Jeff, Åsa Johansson, Jennifer Jolley, Thorhildur Juliusdottir, Leena Kinnunen, Wolfgang Koenig, Markku Koskenvuo, Wolfgang Kratzer, Jaana Laitinen, Claudia Lamina, Karin Leander, Nanette R. Lee, Peter Lichtner, Lars Lind, Jaana Lindström, Ken Sin Lo, Stéphane Lobbens, Roberto Lorbeer, Yingchang Lu, François Mach, Patrik K. E. Magnusson, Anubha Mahajan, Wendy L. McArdle, Stela McLachlan, Cristina Menni, Sigrun Merger, Evelin Mihailov, Lili Milani, Alireza Moayyeri, Keri L. Monda, Mario A. Morken, Antonella Mulas, Gabriele Müller, Martina Müller-Nurasyid, Arthur W. Musk, Ramaiah Nagaraja, Markus M. Nöthen, Ilja M. Nolte, Stefan Pilz, Nigel W. Rayner, Frida Renstrom, Rainer Rettig, Janina S. Ried, Stephan Ripke, Neil R. Robertson, Lynda M. Rose, Serena Sanna, Hubert Scharnagl, Salome Scholtens, Fredrick R. Schumacher, William R. Scott, Thomas Seufferlein, Jianxin Shi, Albert Vernon Smith, Joanna Smolonska, Alice V. Stanton, Valgerdur Steinthorsdottir, Kathleen Stirrups, Heather M. Stringham, Johan Sundström, Morris A. Swertz, Amy J. Swift, Ann-Christine Syvänen, Sian-Tsung Tan, Bamidele O. Tayo, Barbara Thorand, Gudmar Thorleifsson, Jonathan P. Tyrer, Hae-Won Uh, Lies-

beth Vandenput, Frank C. Verhulst, Sita H. Vermeulen, Niek Verweij, Judith M. Vonk, Lindsay L. Waite, Helen R. Warren, Dawn Waterworth, Michael N. Weedon, Lynne R. Wilkens, Christina Willenborg, Tom Wilsgaard, Mary K. Wojczynski, Andrew Wong, Alan F. Wright, Qunyuan Zhang, LifeLines Cohort Study, Eoin P. Brennan, Murim Choi, Zari Dastani, Alexander W. Drong, Per Eriksson, Anders Franco-Cereceda, Jesper R. Gådin, Ali G. Gharavi, Michael E. Goddard, Robert E. Handsaker, Jinyan Huang, Fredrik Karpe, Sekar Kathiresan, Sarah Keildson, Krzysztof Kiryluk, Michiaki Kubo, Jong-Young Lee, Liming Liang, Richard P. Lifton, Baoshan Ma, Steven A. McCarroll, Amy J. McKnight, Josine L. Min, Miriam F. Moffatt, Grant W. Montgomery, Joanne M. Murabito, George Nicholson, Dale R. Nyholt, Yukinori Okada, John R. B. Perry, Rajkumar Dorajoo, Eva Reinmaa, Rany M. Salem, Niina Sandholm, Robert A. Scott, Lisette Stolk, Atsushi Takahashi, Toshihiro Tanaka, Ferdinand M. Van't Hooft, Anna A. E. Vinkhuyzen, Harm-Jan Westra, Wei Zheng, Krina T. Zondervan, ADIPOGen Consortium, AGEN-BMI Working Group, CARDIOGRAMplusC4D Consortium, CKDGen Consortium, GLGC, ICBP, MAGIC Investigators, MuTHER Consortium, MIGen Consortium, PAGE Consortium, ReproGen Consortium, GENIE Consortium, International Endogene Consortium, Andrew C. Heath, Dominique Arveiler, Stephan J. L. Bakker, John Beilby, Richard. Genetic studies of body mass index yield new insights for obesity biology. 518(7538):197–206, February 2015.

[342] Sandro Santagata, Titus J. Boggon, Cheryl L. Baird, Carlos A. Gomez, Jin Zhao, Wei Song Shan, David G. Myszka, and Lawrence Shapiro. G-Protein Signaling Through Tubby Proteins. 292(5524):2041–2050, June 2001.

[343] Anthony G. Comuzzie, Shelley A. Cole, Sandra L. Laston, V. Saroja Voruganti, Karin Haack, Richard A. Gibbs, and Nancy F. Butte. Novel genetic loci identified for the pathophysiology of childhood obesity in the Hispanic population. *PloS One*, 7(12):e51954, 2012.

[344] Caroline S. Fox, Yongmei Liu, Charles C. White, Mary Feitosa, Albert V. Smith, Nancy Heard-Costa, Kurt Lohman, GIANT Consortium, MAGIC Consortium, GLGC Consortium, Andrew D. Johnson, Meredith C. Foster, Danielle M. Greenawalt, Paula Griffin, Jinghong Ding, Anne B. Newman, Fran Tylavsky, Iva Miljkovic, Stephen B. Kritchevsky, Lenore Launer, Melissa Garcia, Gudny Eiriksdottir, J. Jeffrey Carr, Vilmunder Gudnason, Tamara B. Harris, L. Adrienne Cupples, and Ingrid B. Borecki. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS genetics*, 8(5):e1002695, 2012.

[345] Diogo M. L. P. Cavalcanti, Leandro M. Castro, José C. Rosa Neto, Marilia Seelaender, Rodrigo X. Neves, Vitor Oliveira, Fábio L. Forti, Leo K. Iwai, Fabio C. Gozzo, Mihail Todiras, Ines Schadock, Carlos C. Barros, Michael Bader, and Emer S. Ferro. Neurolysin knockout mice generation and initial phenotype characterization. *The Journal of Biological Chemistry*, 289(22):15426–15440, May 2014.

[346] Peter M. Visscher, Naomi R. Wray, Qian Zhang, Pamela Sklar, Mark I. McCarthy, Matthew A. Brown, and Jian Yang. 10 Years of GWAS Discovery: Biology, Function, and Translation. *American Journal of Human Genetics*, 101(1):5–22, July 2017.