

**An integrative functional genomics approach towards
quantitative trait gene nomination in existing and emerging
mouse genetic reference populations**

**A Dissertation
Presented for the
Doctor of Philosophy
Degree**

The University of Tennessee, Knoxville

**Vivek M. Philip
May 2012**

Copyright © 2012 by Vivek M. Philip

All rights reserved.

DEDICATION

To my loving wife, Lucy Reynolds Philip, who has been with me and unconditionally supported me through both the joyful and difficult times.

To our son Nicholas Vivek Philip, a constant source of joyful entertainment and a reminder of what matters most in life.

ACKNOWLEDGEMENTS

I would like to thank my parents, Commander (Retd.) Mohan and Annamma Philip, my brother, Vinod M. Philip, my sister-in-law Ira S. Philip, and my father-in-law and mother-in-law, Michael and Charlene Reynolds for their support and encouragement during my years in graduate school. The unwavering support of my family and their confidence in my abilities was immeasurably beneficial in keeping me focused and motivated throughout the duration of my graduate studies.

I would like to express my gratitude towards my advisor, Dr. Elissa J. Chesler for providing me with the opportunity of pursuing my graduate studies under her tutelage. I also thank her for giving me the opportunity to attend multiple scientific conferences, on a yearly basis. These conferences helped in improving upon and in learning new concepts directly related to my field of research. Furthermore, it provided me with the opportunity to meet and network with other graduate students and faculty.

I also thank Drs. Brynn H. Voy, Arnold M. Saxton, Matthew A. Cooper, Wayne Frankel and Mary Ann Handel for their support, advice and feedback on my research. It has truly been an extremely enjoyable experience. Last but not the least, I thank the Genome Science and Technology program for giving me the opportunity to pursue my graduate studies.

ABSTRACT

An approach that has been widely applied for the genetic dissection of complex traits is Quantitative Trait Locus (QTL) mapping. QTL mapping identifies genomic regions that harbor polymorphisms, responsible for the observed variation in a complex trait. If these polymorphisms are located within a gene, then these genes are called Quantitative Trait Genes (QTG). Prior to advancements in QTL mapping populations, QTL mapping resolution was often poor, resulting in large QTL intervals. Therefore, after mapping a QTL, fine mapping was initiated to further reduce the QTL interval and to identify the QTG. While successful, fine mapping using genetic approaches have been extremely time and resource intensive, making it the rate-limiting step in QTG discovery. Thus far, only a few QTGs have been successfully identified and validated. The disproportionate ratio of QTLs mapped to QTGs identified has been a cause of concern.

Successful QTG discovery relies on the power and resolution with which QTLs are mapped and the genetic architecture of the underlying QTL mapping population. Here, QTL mapping performance in two recently developed QTL mapping populations, namely the expanded BXD Recombinant Inbred (RI) strain panel and the collaborative cross (CC) are assessed. Results indicate that while both the expanded BXD RI strain panel and the CC improve QTL mapping resolution, the CC is able to achieve greater precision and resolution in QTL

mapping. However, neither the BXD RI nor the CC facilitates gene level resolution in QTL mapping.

Recent studies have used the integration and convergence of evidence among functional genomics studies as a successful strategy towards the efficient and rapid nomination of QTG. Here, the complementary *in silico* approach of integrative functional genomics, using GeneWeaver (www.geneweaver.org), is applied towards the reduction of two cocaine-induced locomotor activation QTLs, mapped in the expanded BXD RI strain panel. Integrative functional genomic analyses of these QTLs led to the nomination of *Rab3b* as a putative QTG. Functional assessment of *Rab3b* using *Rab3bcd* knockout mice reveals its role in acute habituation mediated cocaine response, serving as evidence of the efficiency and utility of integrative functional genomics for the identification of highly relevant QTG.

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION	1
Background	2
Quantitative trait locus mapping.....	4
Methods for mapping QTL.....	4
Single-marker analysis	5
Interval mapping	6
From QTL to quantitative trait genes	9
Genetic approaches	10
Existing mouse genetic resources	11
Congenic strain pairs and interval specific congenics	11
Chromosome substitution strains	13
Advanced intercross lines	14
Heterogeneous stock	15
Emerging mouse genetic resources	16
Collaborative cross.....	16
Diversity outbred	17
Bioinformatic approaches.....	17
Comparative genomics	17
Combined cross analysis.....	18
Haplotype association mapping	19
Interval-specific haplotype analysis.....	19
Genome-wide haplotype analysis	20
Sequence and gene expression comparison	21
Characteristics of genes affecting quantitative traits.....	22
Integrative functional genomics	23
CHAPTER II: QTL MAPPING AND QTG DISCOVERY IN EXISTING GENETIC	
REFERENCE POPULATIONS	26
Abstract	26
Introduction	27
Properties of a RI strain panel	28
Initial BXD RI strains and behavior genetics	30
Expanded BXD RI strain set	30
Methods	32
High-throughput behavioral phenotyping	32
Statistical genetics in the expanded BXD RI strain panel	33
Strain effects and sex differences	33
Sub-populations effects.....	34
Comparison of data across time and phenotypes	35
QTL mapping and genetic correlations	36
Multi-trait QTL mapping.....	37
Combinatorial analysis of gene to phenotype associations	38
Results	39

Strain effects and sex differences	39
Sub-populations effects	41
Comparison of phenotypes across time and environments	42
Saline-induced locomotor response	42
Thermal nociception	43
QTL mapping and genetic correlations	45
Multi-trait QTL mapping	49
Combinatorial analysis of gene to phenotype associations	53
Discussion	56
CHAPTER III: PROMISE OF EMERGING GENETIC REFERENCE	
POPULATIONS	62
Abstract	62
Introduction	63
Breeding design	64
Methods	66
Subjects	66
General phenotyping methods	67
Genotyping and SNP selection	68
Data analysis	68
QTL mapping	69
Results	71
Phenotypic variation	71
Analysis of inbreeding depression	76
Intercrossing, inbreeding and their effect on allele frequencies	79
QTL mapping in the CC	83
Discussion	101
CHAPTER IV: INTEGRATIVE FUNCTIONAL GENOMICS AND QTG	
IDENTIFICATION	107
Abstract	107
Introduction	108
Methods	112
QTL mapping in the BXD RI	112
Integrative functional genomic analysis of the candidate genes	113
Assessment of <i>Rab3b</i> polymorphisms	114
Animals	116
Breeding and genotyping	116
Behavioral analysis	117
Assessment of acute and sensitized response to cocaine induced locomotor activity	117
Assessment of habituation and its effect on cocaine induced locomotor activity.	118
Tissue retrieval	119
Statistical analysis	119
qRT-PCR	120

Results	121
QTL mapping	121
Integrative functional genomic analysis reveals <i>Rab3b</i> as a putative QTG acute cocaine response	125
Remapping of previously published QTLs	129
Assessment of <i>Rab3b</i> polymorphisms	130
Assessment of cocaine response in <i>Rab3bcd</i> triple null mice	133
Habituation	133
Acute response to cocaine	137
Response to repeated exposure to cocaine	137
Sensitized response to cocaine	142
Effect of habituation on cocaine induced locomotor activation	142
Differences in <i>Rab3b</i> transcript lengths between C57BL/6J and DBA/2J ..	144
CART transcript abundance in <i>Rab3bcd</i> ^{-/-} and C57BL/6J	146
Discussion	148
CHAPTER V: CONCLUSION	157
QTL mapping in the expanded BXD Recombinant Inbred and the emerging collaborative cross mouse strains	158
<i>In silico</i> reduction of QTL intervals	160
The future of QTG nomination and validation	163
REFERENCES	165
APPENDIX	180
VITA	193

LIST OF FIGURES

Figure 1-1. A typical QTL LOD curve.....	7
Figure 2-1. Construction of a RI strain panel.	29
Figure 2-2. Strain intraclass correlation for all measures.	40
Figure 2-3. Correlation of behavioral measures across environments and time.	43
Figure 2-4. QTL mapping of tail clip latency in old and recent BXD RI strain panels.	46
Figure 2-5. Effect plot by BXD RI Strain panel.	48
Figure 2-6. Equivalent F2 sample sizes for 58 RI strains for varying n and h^2	58
Figure 3-1. Collaborative cross breeding design (34).	65
Figure 3-2. Trait variation in the CC compared to the BXD.	77
Figure 3-3. Litter size and inbreeding depression.....	80
Figure 3-4. Changes in minor allele frequencies (MAF) due to inbreeding.	82
Figure 3-5. Allele loss and inbreeding.	84
Figure 3-6. QTL mapping in the CC.....	85
Figure 4-1. Donor, branch and acceptor consensus sequences (227).	115
Figure 4-2. QTL Mapping of cocaine-induced locomotor activation in the expanded BXD RI panel.	122
Figure 4-3. Main and interaction effect plots.....	123
Figure 4-4. Chromosome 4 and Chromosome 15 QTL LOD support intervals.	124
Figure 4-5. Integrative functional analysis in GeneWeaver (11).	126
Figure 4-6. Habituation deficits in the <i>Rab3bcd</i> ^{-/-}	135
Figure 4-7. Consistent increase locomotor activity in saline-treated <i>Rab3bcd</i> ^{-/-}	138
Figure 4-8. <i>Rab3bcd</i> ^{-/-} and acute cocaine response.	139
Figure 4-9. <i>Rab3bcd</i> ^{-/-} and lack of sensitized cocaine response.....	143

LIST OF TABLES

Table 2-1. Existing genetic reference populations.	31
Table 2-2. Multi-trait QTL mapping.	52
Table 2-3. Top 5 and 10 percent genes of the resultant gene sets and their KEGG pathway.	55
Table 3-1. Trait heritability and variation in the emerging CC.	72
Table 3-2. Genome-wide significant and suggestive QTLs, effect sizes and allele effects.	86
Table 3-3. Additional loci detected using composite interval mapping.	102
Table 4-1 Consequence of upstream and intronic polymorphisms between C57BL/6J and DBA/2J.	127
Table 4-2 Remapping of previously published QTLs.	131
Table 4-3. Assessment of baseline locomotor differences in <i>Rab3bcd</i> ^{-/-}	140
Table 4-4. ΔC_t values for probes targeting <i>Rab3b</i> -S and <i>Rab3b</i> -L transcripts. .	145
Table 4-5. <i>Cartpt</i> ΔC_t values for C57BL/6J and <i>Rab3bcd</i> ^{-/-} among naïve, saline and cocaine treated mice.	147
Table 4-6. <i>Cartpt</i> ΔC_t values for C57BL/6J and <i>Rab3bcd</i> ^{-/-} among naïve and challenged (saline and cocaine treated) mice.	147
Table A1. List of gene sets used in the integrative functional genomic analysis using GeneWeaver (11).	180

CHAPTER I: INTRODUCTION

This dissertation illustrates the progress that has been made to address the challenge of identifying quantitative trait genes (QTG) underlying complex traits. Identification of QTG has relied heavily on the technique of quantitative trait locus (QTL) mapping. However, the power and resolution of detecting QTL and thereby QTG, depends on the nature of the experimental crosses or genetic reference populations (GRPs) used. To improve upon the prospects of identifying QTGs using QTL mapping, advances in GRPs have been made. The studies presented here help assess the performance of QTL mapping and prospects of QTG discovery in two GRPs, namely the expanded BXD Recombinant Inbred (RI) strain panel (*Chapter II*) and the collaborative cross (CC) GRP (*Chapter III*). While advances in GRPs have improved the performance of QTL mapping, gene level resolution is still not attainable. Hence, after a QTL has been mapped additional fine mapping is required to identify the QTG. Even though, fine mapping has been successful in identifying QTG, it is an extremely time and resource intensive process. To expedite the discovery of QTGs, complementary bioinformatic and computational approaches have been proposed, with limited success. In *Chapter IV*, the *in silico* approach of integrative functional genomics is proposed as a complementary approach to fine mapping and existing bioinformatic approaches and is applied towards the behavior of cocaine-induced activation for the nomination and validation of a putative QTG.

Background

Tremendous success has been achieved in understanding the genetic basis of a variety of complex human diseases and disorders. Part of this success is attributable to the field of behavioral quantitative trait genetics. Utilizing induced organismal models of complex human disorders, behavioral genetics aims to understand the genetic basis of quantitative traits and mechanisms that facilitate interactions between organisms and their environments. Among the variety of model organisms available, *Mus musculus* has been widely used to study both simple, Mendelian diseases (e.g. cystic fibrosis (174), Hermansky-Pudlak syndrome (69) and Waardenburg syndrome (141)) and complex multigenic disorders (e.g. alcohol preference (75, 177, 225), diabetes (207) and bipolar disorders (65, 128)). The high degree of homology between human and mouse genomes, availability of basic and advanced genomic tools, resources and decades of research findings are some of the reasons for the widespread use of *Mus musculus* as a model organism.

Studies involving mice as models for human disorders dates back to the early 1900's. Around this time, mouse geneticists started to develop inbred strains of mice and document spontaneously arising mutations. Since then inbred strains of mice have been the mainstay of mouse genetics. They have been used as founders for the derivation of quantitative trait mapping populations (e.g. F2, BC and genetic reference populations), which have been utilized in complex trait analysis. Despite the availability of inbred strains, mapping populations and

genomic resources, it was not until the early 1990's that complex traits were first mapped using mice. This occurred as a consequence of the development of statistical methodologies along with the availability of large collections of molecular tools and/or genetic markers facilitating the association of trait variation to genetic variation. Furthermore, these methodologies permitted, albeit crudely, the estimation of genomic regions responsible for complex traits. This led to the development of the field of quantitative trait locus (QTL) mapping.

Advances in genomic resources and statistical methodologies, even though extremely beneficial, have resulted in the current situation that plagues the field of QTL mapping, i.e., far more QTLs than QTG. The low turnover ratio of QTL to QTG has raised doubts about the utility of QTL mapping with proposals of alternate approaches such as ENU mutagenesis (148) in identifying genes underlying complex traits. While such concerns are reasonable, the rate-limiting step in the identification of QTG is not the QTL mapping process, but rather the lack of an efficient and rapid methodology that facilitates reduction of a QTL interval to nominate a QTG. While expansions (e.g. recent expansion of the BXD RI strain panel (162)) and advances (e.g. collaborative cross (42)) in genetic reference populations will improve precision and resolution in QTL mapping, complementary computationally efficient approaches that facilitate QTG discovery must be developed. One such approach is integrative functional genomics, which provides a robust, efficient and scalable framework within which the reduction of a QTL interval to highly relevant QTG can be achieved. Furthermore, integrative functional genomics allows analyses to be performed

across multiple species and experimental data types, thereby implicitly facilitating translation research.

Quantitative trait locus mapping

Behavioral quantitative trait genetics aims to achieve a better understanding of the genetic basis of behavior. To do so one must associate phenotypic variation in behavior to genetic or sequence variation. Genetic variation comes in the form of segregating polymorphisms. A widely used statistical method that facilitates detection of genomic regions that are significantly associated with phenotypic variation is QTL mapping. The concept of QTL mapping was first defined by K. Sax in 1923 (181). QTL mapping belongs to the class of genome-wide scanning approaches rather than a candidate gene approach. Unlike the candidate gene approach, genome-wide scanning approaches make no presumptions regarding the importance of any particular gene or genes for the phenotype being investigated. As a result, QTL mapping is a truly unbiased approach that helps facilitate the discovery of polymorphisms associated with a given phenotype.

Methods for mapping QTL

QTL mapping is often carried out in populations that originate from crosses of inbred lines that differ phenotypically for the trait of interest. As these lines are inbred, they are homozygous at all loci (markers and QTL) of their respective genomes. The first filial (F1) generation involving these inbred lines will be heterozygous and genetically identical across all loci. F1 progeny can then

be either backcrossed to one of the two parents generating a backcross (BC) population or sib mated to generate a second filial (F₂) population. Both the BC and the F₂ populations consist of genetically diverse individuals, who also exhibit phenotypic variation and are consequently extremely useful in QTL mapping. Another valuable and widely used population for QTL mapping is the panel of recombinant inbred lines (RILs). RILs are derived from twenty or more generations of inbreeding via full sibling matings or selfing in plants, starting with an F₂ population. During the process of inbreeding, the genomes of the F₂ progeny undergo recombination, resulting in the creation of RILs that are genomic mosaics of their founders. Historically, RILs have proven to be extremely useful in mapping QTL in multiple model organisms, including, *M. musculus*, *C. elegans* and *D. melanogaster*. Once a QTL mapping population has been derived, the next steps involves phenotyping and genotyping all individuals within the population. This is followed by application of one of two statistical methods for mapping QTL, namely, *single marker analysis* or *interval mapping*.

Single-marker analysis

Single-marker analysis is the simplest approach to QTL mapping (15, 223). This approach utilizes an analysis of variance (ANOVA) methodology, also referred to as marker regression. In this method, the genotypes at a QTL marker are separated into their respective groups. If one were to consider a BC mapping population, individuals within this mapping population can be divided

into one of two groups, namely, homozygotes or heterozygotes. Statistical tests for strength of evidence of a QTL are then carried out either using a t-test or an F-test depending upon the type of cross. The primary advantage of single marker analysis is its simplicity. However, this approach has multiple limitations. First, this approach fails to account for recombination between a marker and a QTL. Second, incorrect estimates of QTL location can be obtained, especially when a marker is far away from the true QTL. Third, individuals with missing genotypes have to be omitted from the analysis.

Interval mapping

Interval mapping was developed in order to overcome the simplistic nature of single marker analysis (114). Interval mapping incorporates the genetic map of the markers and interrogates the genome at regular intervals for the evidence of a QTL. By calculating a LOD (logarithm of odds favoring the presence of a QTL) score across the genome, interval mapping generates a LOD curve (Figure 1-1). In Figure 1-1, the LOD curve at its maximum point indicates the most likely location of a QTL. Using either 1-, 1.5- or 2-LOD drops from the peak marker, an estimated location or confidence interval (CI) of a QTL is obtained. The 1.5-LOD drop is the most widely used, as it corresponds to an approximately 95% QTL CI (61). However, careful attention should be paid towards the selection of LOD drop based CI as it depends largely on the QTL effect size and marker density. The main advantage of interval mapping over single marker analysis is that in the light of incomplete or missing genotypic data,

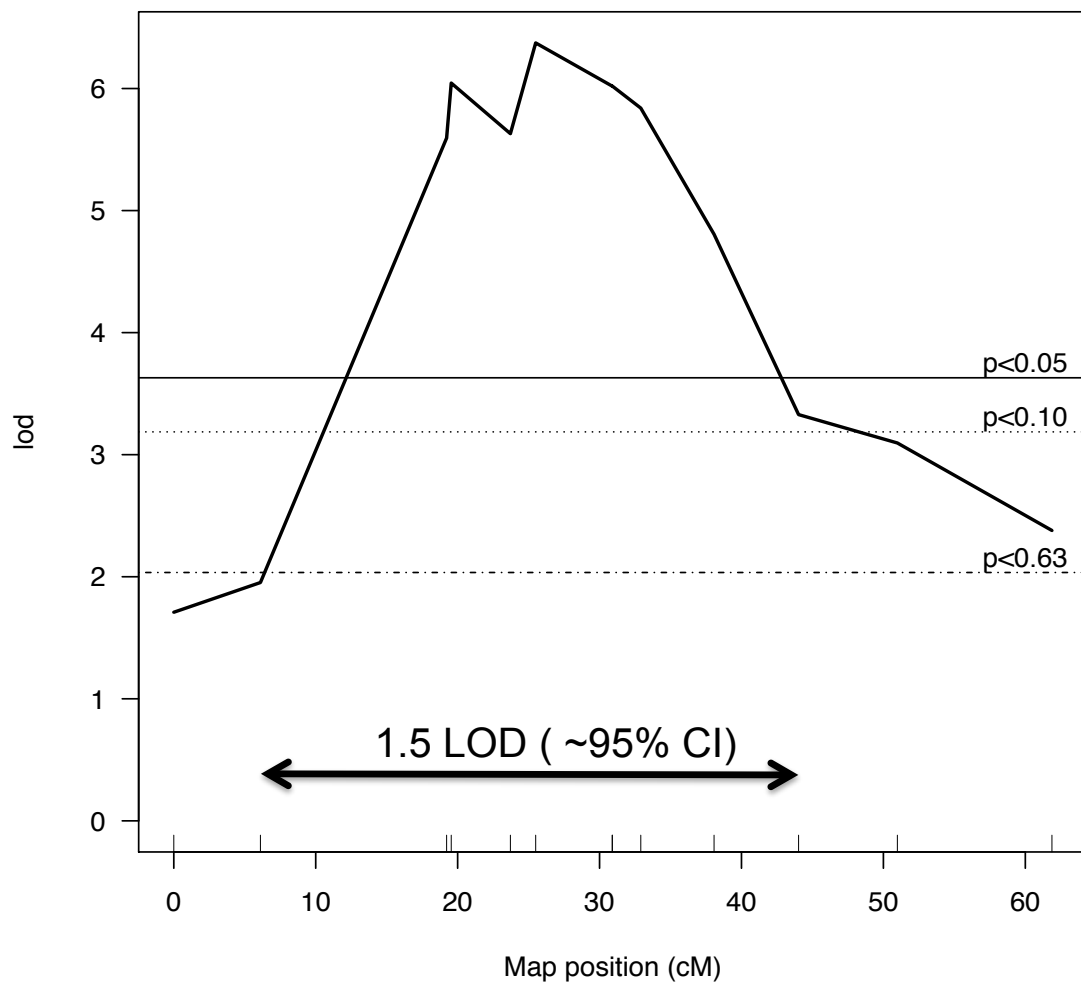


Figure 1-1. A typical QTL LOD curve.

Data plotted here are a significant Chr 5 QTL from the experiment of Boyartchuk and colleagues on the susceptibility to *Listeria monocytogenes* infection in the mouse using R/QTL (22, 27). The solid, dotted and dashed lines represent permutation based significance and suggestive thresholds. Vertical ticks on the x-axis represents genotyped marker on Chr 5 in this study.

interval mapping imputes the missing genotype using the flanking markers rather than discard the individual. The observed LOD score (obtained using the actual mapping dataset) is then compared to the distribution of peak LOD scores obtained following interval mapping on each permuted dataset. QTL significance is then determined based on the proportion of permuted LOD scores exceeding the observed LOD score, giving an estimated p value. It should be noted that the distribution of a LOD scores obtained during a permutation analysis is influenced by the type of cross, number of markers, the inter-marker interval, missing genotypic data patterns and the original or observed phenotypic distribution (26). This was shown by Lander and Botstein (114), who performed a variety of simulation estimating the LOD threshold as a function of cross type, genome sizes and marker density. Depending on the cross type, a map expansion factor of 1 (BC), 2 (F2) or 4 (RIL) needs to be incorporated into the calculation of LOD scores (114). The permutation approach put forth by Churchill and Doerge can be applied to any mapping dataset, regardless of the distribution of the underlying data. However, this approach must be used appropriately, taking into account the underlying population or family structure of the mapping population. Failing to do so will result in incorrect estimates of significance thresholds leading to inflated type I error rates in QTL mapping. Specifically, simple permutations described above cannot be applied without modifications to mapping populations such as advanced intercross lines (AIL) and heterogeneous stock (HS). In such mapping populations, due to their inherent population structure, approaches such

as mixed model association (103) or model averaging approaches (214) are employed to map QTL after taking into account the population structure.

From QTL to quantitative trait genes

Advances in genotyping and phenotyping technologies have resulted in an exponential increase in the number of published QTLs. However, very little progress has been made in the positional cloning or identification of the relevant gene(s) affecting complex traits. As of today over 4,600 QTLs exists. Of these QTLs only 20 genes have been nominated and validated as QTG (74, 138). The current miniscule ratio of QTG to QTL has raised concerns about the utility of QTL mapping for dissecting complex traits (148). The reasons why little progress has been made in the identification of QTG include:

- The majority of the cloned QTLs to date have been QTL with unusually large effect sizes (74). However, QTLs typically have small effect sizes, thereby making them harder to dissect and clone. It is not reasonable to expect that experimental procedures used to clone large effect QTL will also be both applicable and successful for small effect QTL. Unlike large effect QTL, small effect QTL are driven by multiple linked loci, thereby requiring many more recombinants to successfully isolate a single gene. For example, a genetic strategy employing congenic strain pairs will not successfully clone small effect QTLs. Interval specific congenics could probably facilitate cloning of small effect QTLs, but would require large numbers of recombinant progeny.

- It has been reported that the underlying molecular basis of QTL are often due to non-coding regulatory variants compared to coding regulatory variants (107, 129). Unlike coding regulatory variants that would result in a missense, nonsense or some other form of easily detectable polymorphism, non-coding regulatory variations are harder to detect and validate.
- Lastly, a typical QTL interval width can span tens of cM. A typical QTL interval in mice spans 10 – 30 cM or approximately 20 to 60 Mb (78). Multiple factors affect the QTL interval, including, marker density, phenotypic variance accounted for by the QTL, and the type of cross. Large QTL intervals will invariably contain hundreds of genes (also called candidate genes), making it tedious and painstaking to identify the causative gene or genes.

Methodologies that currently exist for the nomination of QTGs can be classified into two main categories, namely, genetic and bioinformatic approaches.

Genetic approaches

Genetic approaches facilitate QTG nomination by making use of the concept of homologous recombination. Using homologous recombination, the original QTL region is reduced to a narrower region consisting of a handful of potential QTG. Classical approaches include *congenic strain pairs*, *chromosome substitution strains*, *advanced intercross lines* and *heterogeneous stock*. In

addition to classical approaches, newer genetic approaches have been developed that include *collaborative cross* and *Diversity outbred*.

Existing mouse genetic resources

Congenic strain pairs and interval specific congenics

Congenic strain pairs or congenics have been the most widely used resource for fine-mapping QTL. Congenics are derived from two inbred strains, a donor and recipient strain, which differ on the trait of interest. The progenitors strains are crossed to produce the F1 generation progeny. F1 progeny carrying the donor region of interest are selected for and backcrossed to the recipient strain. At least ten generations of backcrossing to the recipient strain is performed to reduce the amounts of donor strain genome, such that 99.9% of the genome in the progeny is recipient derived. At this generation, the residual amount of unlinked donor strain genome is $< 0.01\%$. This is done in order to reduce the amount of “background” or hitchhiking genes that are closely linked to the donor strain region of interest. The presence of hitchhiking genes in the recipient genome would very easily lead us to the wrong conclusion and in the failure to capture the true QTL. At the end of backcrossing, heterozygous congenics are intercrossed to obtain homozygous congenics. The primary drawback of classical congenics is that the donor strain interval tends to be too large, often around 50 Mb, and is often dissected into several linked QTLs (169, 170). If these linked QTLs exert opposite effects on the phenotype, then the identification of the causative gene(s) involved becomes even more difficult (60,

237). Furthermore, isolating a QTL onto a recipient strain can result in a reduction in the effect size compared to the originally mapped QTL. This would further complicate the identification of the causative gene(s). In order to overcome these limitations, interval specific congenics (ISC) were developed. Backcrossing the congenics or consomics to the background strain of interest serves as the first step in the derivation of ISC. Some among the first generation (N1) offspring will be recombinant within the original interval of interest, i.e. the QTL interval. These offspring are identified using markers that reside within or flank the QTL interval. Once these individuals are identified, further refinement of the interval of interest is achieved by backcrossing to the background strain. Once the maximum amount of interval refinement has been achieved, backcrossing generates a panel of ISC lines. At this stage the ISCLs are heterozygous at the region of interest. In order to obtain ISCLs that are homozygous for the donor regions, the heterozygous mice are intercrossed. Phenotyping the ISCL facilitates detection of narrow genomic regions positively associated with or manifesting the QTL of interest.

The use of ISCLs often achieves fine mapping of the original QTL to 1-2 Mb. This results in only a handful of potentially causative genes to follow-up, thereby facilitating QTG discovery. *Mpdz* (189), *Il-2* (55) and *Kcnj10* (71) are QTGs that have been identified using the ISC approach in mice, while *Cyp11b1* (44) is an example of application of the ISC approach in rats. Despite its theoretical simplicity and feasibility, this approach has multiple challenges and drawbacks. First, there is a good chance that the QTL of interest might not be

captured due to loss of an epistatic interaction or closely linked QTL that exists with the original QTL. Secondly, it takes a considerable amount of time and resources to generate sufficient numbers of ISCs to successfully identify a single candidate gene or gene variant that defines the QTL.

Chromosome substitution strains

Chromosome substitution strains (CSSs) or consomics strains have been proposed as a means to achieve the genetic dissection of a QTL. Transferring an entire chromosome from a donor strain onto a background strain by repeated backcrossing generates CSS strains (134, 149). The first complete set of CSS created was from the two inbred strains A/J (donor) and C57BL/6J (background) (192, 193). While CSS have been used to dissect QTL, they have also been utilized to map QTLs. This is because a CSS panel has the advantage of mapping QTLs that explain a greater proportion of the total phenotypic variance or QTL effect size and the ability to employ lower significance thresholds towards the detection of significant QTLs. Their utility in high-resolution mapping is made possible in two ways, through rapid creation of interval specific congenic lines and recombinant progeny testing (RPT) (50). By backcrossing the CSS that harbors the QTL followed by intercrossing the F1 progeny, a congenic (smaller introgressed intervals in comparison to the original CSS) F2 population is created. A new set of ISC is created from the congenic F2 population followed by high-resolution mapping. In RPT, animals that are recombinant in the QTL region of interest are backcrossed to a background strain

that helps in the identification of the QTL location relative to the site of recombination (19, 68).

Despite the ease with which one can map and generate a CSS panel, they suffer from a major limitation, i.e. their inability to facilitate the dissection of large effect QTLs into many loci smaller of effects. This hinders the ability with which smaller effects QTLs can be fine-mapped using this resource.

Advanced intercross lines

Advanced intercross lines (AIL) facilitate refinement in QTL mapping by introducing new recombinations via intercrossing (51). Starting with a F2 population, intercrossing is typically performed for eight or more generations (to arrive at F10), by randomly intercrossing progeny in each generation to produce the next. Careful attention is paid towards the selection of mice that produce the next generation by avoiding sib-pairs and cousins. This is done in an attempt to maintain a randomly mating population and avoid fixation (due to inbreeding) of alleles so as to maintain maximum genotypic diversity. Increased numbers of recombinant progeny within an AIL facilitates high-resolution mapping of QTLs. It has been shown that for a QTL with large effect mapped in a large population ($N = 1500$), an AIL (greater than F10) achieves a QTL CI of approximately 5 cM compared to 14.8 cM in an F2 population (51). Furthermore, AILs can help resolve closely linked QTLs. For example, if two QTLs are closely linked (approximately 5 cM apart), an F2 population will fail to dissect them. However, in an AIL at F10 these loci are predicted to be 26 cM apart facilitating

detection of both loci and thereby resolving linked QTLs (51). Numerous studies have made use of AILs for fine-mapping QTLs (67, 94, 180, 199, 220, 221, 234, 238). Despite the advantages offered by an AIL, the primary disadvantage is the inherent presence of family structure that exists at the final generation of AILs. Family structure or non-independence among individuals due to the complex breeding history of AILs, when unaccounted for, results in detection of spurious QTLs (160). Statistical methods have been devised to account for family structure and must be utilized when an AIL is used for fine-mapping a QTL (104, 153, 214).

Heterogeneous stock

A series of repeated intercrosses involving inbred strains of mice results in the derivation of the heterogeneous stock (HS). Existing HS have been derived from an initial cross involving eight progenitor strains. Currently, two HS panels exist: the Boulder IBG heterogeneous stock (136) has been in existence for over 60 generations and is derived from the C57BL/6, BALB/c, RIII, AKR, DBA/2, I, A/J and C3/HeJ strains and the Northport heterogeneous stock (53) that has been in existence for over 40 generations and is derived from the A/J, AKR/J, BALB/cJ, C3H/HeJ, CBA/J, DBA/2J and LP/J strains. New recombinations are introduced by using a pseudo-random breeding protocol of circle mating. The increase in numbers of recombinations results in higher-resolution mapping. In order to map traits using the HS, specialized statistical methodologies are utilized that incorporate the probability that an allele descends

from one of the eight progenitor strains followed by association of these alleles to the phenotype of interest (145). The fine-grained mosaic of the HS along with specialized statistical methods has facilitated high-resolution QTL to within 1 cM for several behavioral measures (90, 131, 145, 204, 205, 230). The primary disadvantages of the HS lie in the complexity of the mapping analysis, requirement of high-density genotyping and large mapping sample sizes required to fine-map QTLs (144).

Emerging mouse genetic resources

Collaborative cross

The collaborative cross (CC) is an emerging mouse genetic recombinant inbred (RI) panel that is derived combining the genomes of eight genetically diverse founder strains, namely, A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ, and WSB/EiJ (42). When completed the CC will be the largest RI panel, vastly exceeding the genotypic and phenotypic variation captured in currently existing RI panel such as the BXD RI. Like the HS, the CC is a fine-grained mosaic of genomes of eight inbred strains with increased recombinations facilitating high-resolution QTL mapping. Studies involving the interim CC lines have validated increased precision and resolution in QTL mapping and vastly increased phenotypic variation across multiple phenotypes compared to existing RI panels (8, 62, 165). The limitations of this strategy are primarily due to the complexity of mapping and interpretation of

mapping results. Furthermore, despite improvements in resolution of QTLs, single gene identification is yet to be achieved.

Diversity outbred

Originating from a set of 160 breeding lines that founded the CC population, the DO is produced by an outbreeding rather than an inbreeding strategy (41). The DO breeding design ensures a balanced contribution of the founder genomes and at the same time is shielded from the consequences of inbreeding, such as allele purging. As the DO is outbred, the mice within this population will be heterozygous, thereby incorporating high levels of allelic variation and genetic diversity. The high levels of genetic diversity in the DO are comparable to that in humans, making the DO an extremely useful resource for the study of complex human disorders.

Bioinformatic approaches

Challenges in transitioning from QTL to QTG have led to the development of complementary bioinformatic approaches that facilitate QTG nomination. Bioinformatic approaches incorporate publicly available genomic resources that include sequence and gene expression databases to refine and narrow a QTL interval (57).

Comparative genomics

Comparative genomics takes advantage of the structural conservation of genomes across mammals, specifically those that have been

used in QTL mapping. Human and mouse genomes have similar genomic contents and linear arrangement of genes (222). Furthermore, over 300 genomic segments with conserved synteny have been identified in whole-genome comparison of humans and mouse (163). Comparison of human and rat genomes have revealed similar results (155). By searching for overlapping QTL regions for the same phenotype across multiple species, comparative genomics has been used successfully to narrow large QTL intervals. Identification of *Rf-1* (renal failure-1) is an example of the successful application of this approach for the reduction in the QTL interval. Using synteny between the rat and human genome, this QTL was reduced to a set of 66 highly likely candidate diseases genes. This strategy has also been used for the identification of candidate gene loci for male fertility (156) and reduction of QTL interval for HDL cholesterol levels (30, 47). The biggest challenge with this strategy is when a rodent QTL maps to multiple syntenic blocks along the human genome instead of one contiguous block, making it harder to detect the true homologous gene.

Combined cross analysis

Combined cross analysis works on the assumption that when the same causal polymorphism and gene underlies a QTL that has been mapped in multiple crosses, by combining these crosses one increases the number of recombination events in the combined cross panel thereby achieving increased resolution or reduction in the QTL interval. Furthermore, combining multiple crosses results in an increase in the mapping sample size, facilitating increased

power in the detection and resolution of QTLs with small effect sizes (121). This strategy has been successfully applied in achieving better QTL mapping resolution across diverse traits, such as HDL cholesterol (47, 201, 224) and ear hole punch healing (39) in mice and experimental autoimmune encephalomyelitis in rats (96). Limitations of this strategy stem primarily from the fact that not all QTL have been mapped in multiple crosses. Most of the multiple cross studies involve the analysis of cholesterol levels, making it the most feasible trait to be employed in a combined cross analysis. Care should also be taken to ensure that the phenotype or trait mapped across the multiple crosses is the same. Combining diverse phenotypic data across multiple crosses will produce spurious results.

Haplotype association mapping

Haplotype association mapping (HAM) stems from the concept of genomic segments sharing polymorphisms that in turn are associated with common ancestral origins. These genomic segments are often referred to as haplotype blocks. Two methods of HAM exist, namely, *interval-specific haplotype analysis* and *genome-wide haplotype analysis*.

Interval-specific haplotype analysis

Here, an underlying QTL detected in a two strain mapping study is mapped back to their respective ancestral subspecies of origin. If the QTL region is identical by descent (IBD) i.e., originating from a common ancestral subspecies, then it is unlikely to contain the causal genetic

polymorphism. This strategy has been applied towards the genetic dissection of rodent QTLs. A metastasis QTL on Chr 19 was reduced from over 400 genes to around 23 genes (158). Another example is of a blood pressure QTL on Chr 1 that was originally around 42 cM but was refined to an interval of 2.3 cM (56). A more recent example of a behavioral trait that was refined is of open field behavior, wherein, by employing multiple crosses and interval specific haplotype analysis, an approximately 50 Mb region was refined to a 7.8 Mb region on Chr 9 (66).

Genome-wide haplotype analysis

Contrary to interval-specific haplotype analysis, genome wide haplotype analysis is performed in the absence of the *a priori* knowledge that is obtained from a QTL mapping study. Also known as *in silico* QTL mapping, genome-wide haplotype analysis seeks to associate phenotype variation across a panel of inbred strains to their genotypes rather than to a mapping population derived from inbred strains. Grupe and colleagues (83) were the first to develop the methodology behind *in silico* QTL mapping. Using this methodology Smith and colleagues (195) were the first to perform *in silico* QTL mapping of atherosclerosis susceptibility across six inbred strains.

The concept and use of HAM in QTL mapping has been rather controversial. Questions have been raised regarding the actual sample sizes or number of strains and the basis of the statistical approach (35, 49). Even in the light of these controversies, HAM has been utilized for narrowing QTL intervals of

HDL-C concentration on Chr 8 (167) and Chr 5 (221). Pletcher and colleagues used HAM to identify a single candidate gene containing a non-synonymous polymorphism, starting from a QTL that spanned 5 Mb (167). Wang and colleagues used a combination of AIL first refine the QTL interval from 8.6 Mb to 2.6 Mb. Using HAM, the authors were able to further refine this QTL to a 0.4 Mb region (221). HAM has also been used for the identification of candidate lung tumor susceptibility genes by refine the original QTL interval to < 0.5 Mb (123) and in the detection of *Cer1* as a potential candidate for bone mineral density (BMD) regulation across a panel of inbred strains. By studying *CER1* in Chinese populations with high and low BMD, the authors were able to detect a non-synonymous polymorphism within *CER1* that conferred an increased risk of low BMD in premenopausal women and an increased risk of vertebral fractures in postmenopausal women (206).

Sequence and gene expression comparison

Variation present within a QTL interval could either be due to differences in the gene sequence or expression. As a result, sequence comparison and gene expression analyses have been performed in order to detect the most plausible QTG. The recently completed next-generation sequencing (NGS) by the Sanger Center Mouse Genomes Project has made available the genomes of 17 inbred mouse strains commonly used in QTL mapping at very high resolutions (105, 231). This will result in the detection of large insertion/deletions, copy number variations (CNVs) and other types of

sequence variants that will help facilitate QTG discovery (105, 217, 231). The advantage of sequence analysis is in its ability to identify sequence differences that may not lie within the coding region of the gene (229).

In addition to sequence comparisons, gene expression analyses serves as a complementary strategy for the identification of QTGs. The idea behind gene expression analyses is that a QTL will contain genes that harbor sequence polymorphisms that effect gene function, expression or both. The availability of public gene expression databases, such as The Gene Expression Omnibus (64), Array Express (23), The Alan Brain Atlas (1, 118), BioGPS (200), ExQuest (28) and The Institute for Genome Research (2), facilitate tissue specific gene expression and differential expression analyses. However, due to the large numbers of gene expression differences exhibited by murine strains, fine-mapping will be necessary (138).

Characteristics of genes affecting quantitative traits

Glazier and colleagues (78) and The Complex Trait Consortium (CTC) (5), have proposed a sets of guidelines or standards to characterize a gene as the QTG. Listed below is a summarization of the characteristics of a QTG:

- The QTL that contains the potential QTG should meet statistically significant genome-wide thresholds (58, 113). In the absence of clear genome-wide significance from a single study, the QTL should meet consistent suggestive genome-wide thresholds across multiple independent studies.

- Polymorphisms within a QTG, either in the coding or non-coding region, should translate into changes in either the structure or expression of the gene product. Coding sequence non-synonymous polymorphisms are the easiest class of polymorphisms to detect and validate. Synonymous coding region polymorphisms have been linked to differential mRNA stability. Intronic (non-coding) or promoter region polymorphisms could be detected as changes in mRNA splicing and transcription factor binding (130). Even though salient, these changes could potentially affect the amount, timing and tissue specific pattern of gene expression
- *In vivo* confirmation of a QTG via alteration of its function should be obtained. This can be performed using one of many strategies, namely, targeted gene mutations such as knockouts or deficiency-complementation tests, transgenesis, viral-mediated RNA interference (RNAi) and knock-ins.

For a gene to be proven to be the QTG it must possess more than one of the above stated QTG characteristics. In the absence of a clear and significant empirical result that proves a gene as the QTG, preponderance of circumstantial evidence will cumulatively work towards the nomination of a QTG.

Integrative functional genomics

To truly expedite the process of QTG discovery, the vast repertoire of genomic resources and datasets currently available, need to be harnessed and utilized efficiently. Currently two popular approaches exist that help integrate

diverse genomic data sets, namely, meta-analysis and convergent functional genomics. Meta-analysis is a statistical approach that analyzes large collections of experimental data with the goal of integrating experimental results and findings (77). Meta-analysis has been applied to a wide array of genomic data sets, such as, microarrays (112, 143, 146, 147, 172, 173) and QTL mapping studies (17, 185, 228). Convergent functional genomics (CFG) is a Bayesian strategy that facilitates integration of gene expression data across human linkage analysis and relevant animal models (150). CFG has been successfully applied towards the identification of candidate genes and blood biomarkers for bipolar disorder (111, 116, 159). Despite the successes of meta-analysis and CFG, both approaches lack computationally efficient infrastructures in the form of databases, algorithms and user interfaces that facilitates genome scale integrative analyses in a seamless and scalable manner. Furthermore, while CFG integrates human and rodent data sets, meta-analysis fails to attain cross-species analysis. Given that many model organisms have their genomes sequenced, it would be beneficial for integrative approaches to encompass more than just two species. Additionally, CFG is tailored towards the needs of a specific laboratory and/or specific research area.

To truly expedite the process of QTG discovery, computationally efficient algorithms, databases and user-interfaces need to be developed and utilized. One such approach is integrative functional genomics. Integrative functional genomics facilitates convergent analyses across multiple genomic experiments, platforms and species (11), thereby readily lending itself towards the discovery of QTGs.

Candidate genes within a QTL interval can be integrated with other relevant empirically derived genomic data sets or gene sets. The gene sets to be used in the analysis are selected based on the trait in which the QTL was detected. For example, an integrative functional genomic analysis for a cocaine-induced locomotor response QTL will include gene sets spanning multiple species for basal expression across brain reward regions, differential gene expression across brain reward regions for stimulants that influence brain reward regions (such as methamphetamines and nicotine), previously mapped QTL studies involving cocaine-induced locomotion and polymorphism information for the genes across the founder strains. A ranked list (based on occurrence count across genomic data sources) of candidate genes can be obtained to prioritize genes for future validation studies. Such an analysis approach will yield highly relevant polymorphic genes that can be functionally validated and eventually nominated as a QTG.

CHAPTER II: QTL MAPPING AND QTG DISCOVERY IN EXISTING GENETIC REFERENCE POPULATIONS

This chapter is adapted from the following publication:

Philip VM, Duvvuru S, Gomero B, Ansah TA, Blaha CD, Cook MN, Hamre KM, Lariviere WR, Matthews DB, Mittleman G, Goldowitz D, Chesler EJ.

Genes, Brain and Behavior (2010) 9: 129-59.

My contributions to this paper were heritability analysis, QTL mapping, gene co-expression analysis, gene-phenotype combinatorial analysis, KEGG and GO pathway analysis, interpretation of results, and part of the writing.

Abstract

Genetic reference populations, particularly the BXD recombinant inbred (BXD RI) strains derived from C57BL/6J and DBA/2J mice, are a valuable resource for the discovery of the bio-molecular substrates and genetic drivers that mirror complex human disorders such as drug addiction and alcoholism. The initial set of BXD RI strain panel consisted of only a limited number of strains (approximately 30). The combination of a limited number of strains and sparse marker maps resulted in low QTL mapping precision and resolution. In 2004, the BXD RI strain panel underwent a large expansion resulting in the addition of 46 new BXD RI strains, making it the largest genetic reference population. Furthermore, the expanded BXD RI strain panel has increased numbers of recombination events resulting in improved precision and resolution in QTL mapping. Additionally, the presence of approximately 1.8 million SNPs between C57BL/6J and DBA/2J strains greatly improves the prospects of QTG nomination and validation.

Using the recently expanded BXD RI strain panel, members of the Tennessee Mouse Genome Consortium (TMGC) have obtained phenotype data from over 250 measures related to multiple behavioral assays across several batteries: response to, and withdrawal from cocaine, MDMA, morphine and alcohol; novelty seeking; behavioral despair and related neurological phenomena; pain sensitivity; stress sensitivity; anxiety; hyperactivity; and sleep/wake cycles. All traits have been measured in both sexes in approximately 70 strains of the recently expanded panel of BXD recombinant inbred strains. Sex differences and heritability estimates were obtained for each trait, and a comparison of early ($n = 32$) and recent ($n = 37$) BXD RI lines was performed. Primary data is publicly available for heritability, sex difference and genetic analyses using the MouseTrack database, and is also available in GeneNetwork.org for QTL detection and genetic analysis of gene expression. Together with the results of related studies, these data form a public resource for integrative systems genetic analysis of neurobehavioral traits.

Introduction

In the late 1950's, Donald Bailey was the first to realize the potential of genetic reference populations (GRP) for linkage analysis (191). Once established, a GRP can be used repeatedly over time and across laboratory environments primarily due to their genetic stability. Recombinant inbred (RI) strains are an example of a GRP. RI strains were initially used to map single loci associated with Mendelian traits. However, over the last few decades, RI strains

have also become an invaluable tool for the genetic dissection of complex traits (168). The construction of an RI strain panel is initiated by selecting two highly inbred progenitor strains that typically vary for a trait of interest. These progenitors are intercrossed to form the F1 progeny. At the F1 generation, all offspring are heterozygous and genetically identical. F1 progeny are crossed to produce the F2 generation in which, due to meiotic recombination, random rearrangement of the progenitor genomes occur. At the F2 generation, randomly chosen pairs of animals are used as founders for subsequent derivation of RI strains of mice. Performing brother-sister mating for at least 20 generations eventually produces RI strains. After 20 sequential rounds of brother-sister mating, the genomes of these RI strains are essentially fixed (approximately 99% homozygosity) (162). This creates panels of RI strains wherein individuals within any given strain are genetically identical to each other. Figure 2-1 illustrates the process of creation of the BXD RI strain panel, the largest and most widely used RI panel.

Properties of a RI strain panel

Due to the process of inbreeding, RI strains are fixed and homozygous at almost all loci. This makes them an “immortal” resource as they can be maintained at the inbred homozygous state by continuation of brother-sister mating as long as required. RI strains have three primary advantages over an F2 or BC. First, as they are genetically stable, experimental results obtained across these panels can be compared across time and laboratories (82). Second,

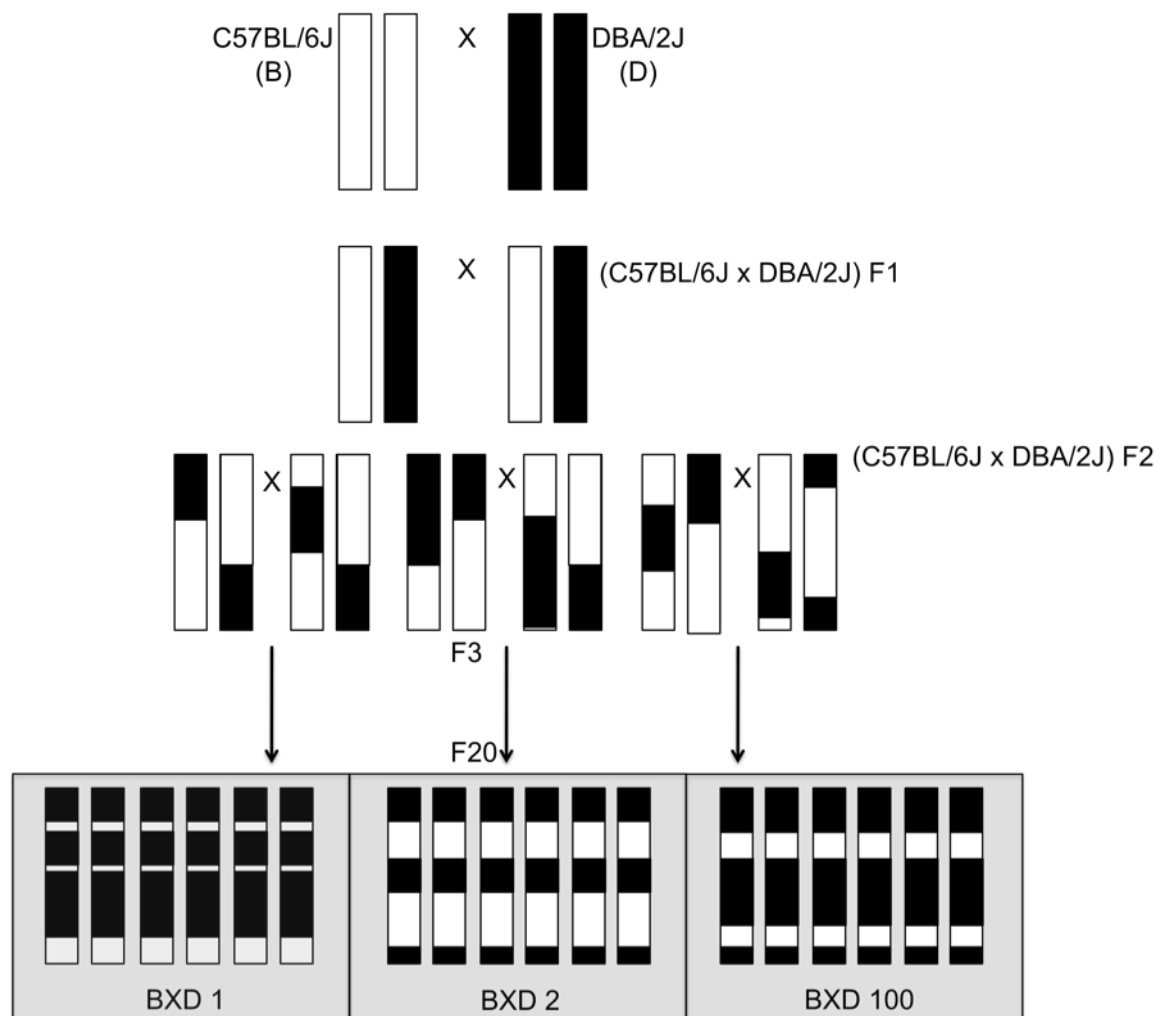


Figure 2-1. Construction of a RI strain panel.

Derivation of the BXD RI strain panel starts with intercrossing C57BL/6J (B) and DBA/2J (D) inbred strains of mice. Three pairs of F2 siblings are randomly chosen as founders for three resultant BXD RI strains called BXD 1, BXD 2 and BXD 100.

multiple experimental observations can be obtained per strain therefore facilitating replication and increased statistical power (16). Thirdly, due to their fixed genomes, genotyping is required only once, thereby greatly reducing genotyping costs.

Initial BXD RI strains and behavior genetics

Given the above-mentioned advantages of RI strains, it is not surprising that they have been often used in studying quantitative traits. Among the panels of RI strains currently available, BXD RI strains are the largest (Table 2-1). The BXD RI strain panel was first established by B. A. Taylor in 1977 (208) and expanded to 35 strains in 1999 (209). The earliest reports of RI strains, specifically the BXD RI strain panel, in the genetic dissection of behavioral traits dates back to the early 1980's (48, 135, 186). Even though early studies were limited in their power to detect QTLs, due to the limited numbers of strains utilized (approximately 20 or fewer strains), the advantages and utility of a relatively small RI strain panel called for the generation of larger RI strain panels (168).

Expanded BXD RI strain set

In 2004, Pierce and colleagues expanded the existing BXD RI strain set to a total of 79 strains (162). These strains were derived from an advanced intercross (AI) between C57BL/6J (B) and DBA/2J (D) progenitor strains. As these new BXD strains were derived from AI lines (also termed BXD ARI), they have increased numbers of recombinations, facilitating improved mapping resolution (162, 187). Pierce and colleagues have estimated that the number of

Table 2-1. Existing genetic reference populations.

Name	Progenitor Strains	Strains	Known recombinations	Recombinations / strain
AKXD	AKR/J (AK) x DBA/2J (D)	24	1027	42.8
AXB/BXA	A/J (A) x C57BL/6J (B)	34	1600	54.4
BXD	C57BL/6J (B) x DBA/2J (D)	89	1848 (36 older BXD) 4366 (53 newer BXD)	48.1 82.4
BXH	C57BL/6J (B) x C3H/HeJ (H)	20	755 (16 core BXH strains)	48.5
CXB	BALB/cBy (C) x C57BL/6J (B)	13	694	53.4
LXS	Inbred Long-Sleep (L) x Inbred Short-Sleep (S)	77	3598	46.7

recombinations detectable in the 46 BXD ARI lines is equivalent to at least 90 BXD RI strains generated from an F2 population. Furthermore, the BXD ARI lines are able to detect QTLs accounting for half the amount of genetic variance compared to the BXD RI lines (24% compared to 47%) (162). Using this expanded strain set will facilitate detection of QTLs explaining small effects (at least 10% genetic variation) to 1 – 2 Mb intervals (162). Another advantageous property of the BXD RI strain panel is that the B6 and D2 progenitor strains have been deeply sequenced facilitating the discovery of 1.8 million SNPs. These progenitor strain differences used in concert with BXD RI genotypes permits the discovery of coding sequence variants, such as SNPs and insertion-deletions.

Despite the utility of RI strains in studying complex traits, their use has been limited to only a few areas, such as basal behaviors and ethanol related traits. Furthermore, in most cases, only one sex was studied. To truly harness the power of the largest RI strain population currently available, it is vital to employ high-throughput behavioral phenotyping techniques that facilitate quick and accurate measurements of behavior. This is particularly of great importance, given the derivation of RI strain panels, such as the CC, that are expected to be 4 – 5 times larger than the BXD RI strain panel.

Methods

High-throughput behavioral phenotyping

In order to rapidly develop a broad base of behavioral phenotyping data, phenotyping was performed in several multivariate test batteries. Each battery

consists of a set of tests administered serially to individual mice. This approach also allows for the examination of partial-correlation within strain (non-genetic correlation) but may upwardly bias estimates of genetic correlation obtained using strain means of the measures. A given mouse was assigned to one and only one battery and received all tests in that battery. The order of repeated testing was either fixed where logical and necessary or, in the case of the nociception battery, varied systematically using randomly generated Latin-square designs each applied to a different strain and sex. In general, for fixed-order batteries, the least stressful measures were obtained first, and all baseline measures were necessarily obtained before conditioning or drug exposures. The testing protocols were largely derivative of those developed in consultation with the external advisory board of the Tennessee Mouse Genome Consortium (TMGC) ENU-Neuromutagenesis Program (79). Detailed methods of the phenotyping protocols are presented elsewhere (164).

Statistical genetics in the expanded BXD RI strain panel

Strain effects and sex differences

Sex, strain and strain $\times$ sex interaction effects were tested using non-sequential sums of squares estimation in a general linear model. The model used for testing these effects was

$$y_{ij} = \mu + \text{Strain}_i + \text{Sex}_j + \text{Strain}_i \times \text{Sex}_j + \varepsilon_{ij}, \quad (1)$$

Where, y_{ij} is the phenotype being measured for Strain_i and Sex_j . In addition to testing these effects, we estimated the magnitude of strain, sex and

strain × sex effects by estimating the intra-class correlation, a partial ω^2 , for each effect per trait. Partial ω^2 was estimated as the proportion of variance accounted for by the main or interaction effects relative to the total phenotypic variance. All variance components were estimated using the REML option of SAS PROC VARCOMP. This method can be biased due to departures from normality, a common phenomenon for behavioral traits. The percent variance accounted for by strain is considered by some to be an estimate of broad-sense heritability (89, 127) for clones, and is formulated as the strain intra-class correlation:

$$h^2 = \sigma^2_{\text{Between Strain}} / (\sigma^2_{\text{Between Strain}} + \sigma^2_{\text{Within Strain}}). \quad (2)$$

Standard errors were obtained using an adjustment for unbalanced data (202). This calculation was performed for the two sexes separately, and for the data combining both sexes. It should be noted that this measure provides an estimate of the resemblance among relatives in a population in which segregation had occurred and does not reflect the transmission of genetic material in a randomly mating population. It is nonetheless an indication of suitability and plausibility of genetic analysis for a given phenotype.

Sub-populations effects

A limited number of systematic differences exist in the genotypes of the old and new BXD RI strains such that loci have been identified that segregate in only one of the two sub-populations (187). To test for global phenotypic effects of these differences, we use the nested model,

$$y_{ij} = \mu + \text{Strain}_i (\text{Sub-Population}_j) + \text{Sub-Population}_j + \epsilon_{ij}$$

An alternate F-test is applied. The F ratio is $MS_{\text{Sub-Population}} / MS_{\text{Strain(Sub-Population)}}$. This test is used because the individual mice in the study are not independent, but rather are replicates within a strain. Therefore, we use the appropriate error term and degrees of freedom for the random effect of strain nested within the sub-populations. To account for multiple testing, the false discovery rate (FDR) was controlled using q-value estimation for the sex and strain main effects and sex $\times$ strain interaction effects. The R/qvalue software developed by Storey (198) was used for the analyses. A q-value threshold of 0.05 was used to identify significant results. Quantitative trait locus mapping was also performed separately for sub-populations.

Comparison of data across time and phenotypes

To determine if behavioral measures collected across time and laboratories can be integrated we correlated saline induced locomotor activity and thermal nociception. Saline induced locomotor activity was measured in two laboratories in the current study (n = 62), Memphis (locomotion in 15 mins) and UTHSC (locomotion in 20 mins). These data were compared to saline-induced locomotion measured by Demarest and colleagues in 1999 (n = 25). Thermal nociception measured in two laboratories in the current study (n = 62), Memphis and ORNL, were compared to Mogil and colleagues (140) and Brigman and colleagues (24).

QTL mapping and genetic correlations

GeneNetwork (GeneNetwork.org) serves as a database and analysis engine for QTL mapping and genetic correlations among strain phenotypic means obtained in the BXD RI reference population. Strain means were computed in MouseTrack by sex and other cofactors. For pooled sexes, the mean values were the least-squares means with strain, sex and strain $\times$ sex in the model. Male and female strain means for each phenotyping battery were exported from MouseTrack for submission to GeneNetwork. This enables gene expression correlation and interval mapping, candidate gene searches and multi-trait analyses.

Each exported dataset was subject to an interval mapping analysis, which uses GeneNetwork's embedded MapManager software, WebQTL, (132) to perform Haley–Knott regression. Empirical p values were derived using 1000 permutations using the incorporated permutation feature of WebQTL. The peak of each statistically significant ($p < 0.05$) or suggestive ($p < 0.63$) (113) QTL was determined based on empirical p values (58). A one-LOD drop-off was used to determine the QTL confidence interval about each peak. Positional candidates residing within a one-LOD drop from the peak of each statistically significant and suggestive QTL were identified.

Trait data were correlated against the following GeneNetwork gene expression tissue databases: whole brain [INIA Brain mRNA 430 (June 2006) RMA (161), neocortex [HQF BXD NeoCortex ILM6v1.1 (February 2008) RankInv (76)], striatum [HBP Rosen Striatum M430V2 (April 2005) RMA (178)],

cerebellum [SJUT Cerebellum mRNA M430 (March 2005) RMA (32)] and hippocampus [Hippocampus Consortium M430V2 (June 2006) RMA (106)]. Lists of genes were generated based on their correlation to each phenotype using a correlation p value < 0.001 .

Multi-trait QTL mapping

Multi-trait QTL analysis can be performed by extracting common underlying factors from multiple behavioral phenotypes and generating strain-specific factor scores. The underlying hypothesis of this type of analysis is that behavioral measures for stress, anxiety, pain and addiction to drugs of abuse are under common genetic control and should share some degree of correlation. Brigman and colleagues performed a similar decomposition of anxiety and fear behavior in BXD RI mice (24). This approach to multi-trait QTL mapping has also been performed by Trullas and Skolnick (212), wherein factor analysis was used to report that elevated plus maze behavior predicts anxiety-like behavior, and more recently by Henderson and colleagues, who performed QTL mapping on multiple phenotypic assays of anxiety-like behavior (87). The value of performing such studies in the BXD RI population is that the data can be expanded indefinitely with additional independent phenotypic profiling which adds depth and detail to the multi-dimensional analysis.

As the phenotypic data contained missing observations, any strain with more than 25% of trait data missing was removed from this analysis. This resulted in the elimination of 32 of the 95 BXD strains, giving us an effective

sample size of 63 BXD strains. The resultant data set was subjected to column mean imputation in order to fill in the missing trait values, given that the data were missing at random, i.e. not missing over all measures within a battery. Another issue that was encountered during this analysis was that of dimensionality. The dataset used in the factor analysis had more variables/traits (p) than observations/strains (n) (i.e. $n < p$). This results in incorrect estimation of the covariance matrix and thereby leads to singularity of the estimated covariance matrix. This issue was addressed by using the James–Stein-type shrinkage estimator (182) of the covariance matrix. R packages e1071 (for missing data imputation), corpcor (for covariance shrinkage), nFACTOR and factanal were used for the purposes of factor analysis. Factor loadings were analyzed to obtain factor interpretations. Factor scores were obtained for all interpreted latent factors. Quantitative trait locus mapping was performed on these latent factors to identify common genetic drivers of variability in factor scores.

Combinatorial analysis of gene to phenotype associations

Each set of gene expression correlates with p value < 0.001 for gene–phenotype association was subject to combinatorial analysis to identify those genes that were directly correlated to multiple phenotypes. Positional candidates defined as those that reside within a one-LOD support interval of significant ($p < 0.05$) or suggestive ($p < 0.63$) QTL. The top 5 and 10 percent of the highly connected genes were analyzed for over-representation of genes in the Kyoto

Encyclopedia of Genes and Genomes (KEGG) pathways (101) using the analysis tool, WebGestalt (235), which performs a hyper-geometric test for the similarity of the list of highly connected genes and the members of curated pathways in the KEGG database. The resulting list of enriched categories of genes represents those processes, functions and molecular classes that are most involved in genetic variation in behavior.

Results

Strain effects and sex differences

Generalized linear models were used to test for sex and strain main effects, and sex $\times$ strain interaction effects. False discovery rate analysis was used to control the family-wise error rate at 0.05 and showed significant sex effects for 97 of the 257 measures with five expected false positives, significant strain effects for all 257 measures, and significant sex $\times$ strain interactions for 144 of the 257 measures with seven expected false positives. In the event that there was no trend toward a significant sex difference or interaction, male and female data were combined for subsequent QTL mapping and genetic correlations. Effect sizes were estimated as partial ω^2 for the main effects of strain and sex, and for the sex $\times$ strain interaction. Consistent with statistically significant strain effects, 146 measures under consideration had large effect sizes ($\omega^2 > 0.30$), 93 had intermediate effect sizes ($0.30 > \omega^2 > 0.10$) and 19 had small effect sizes ($\omega^2 < 0.10$) (Figure 2-2). Locomotor activity in a variety of apparatus and under different drug exposure conditions typically had large strain

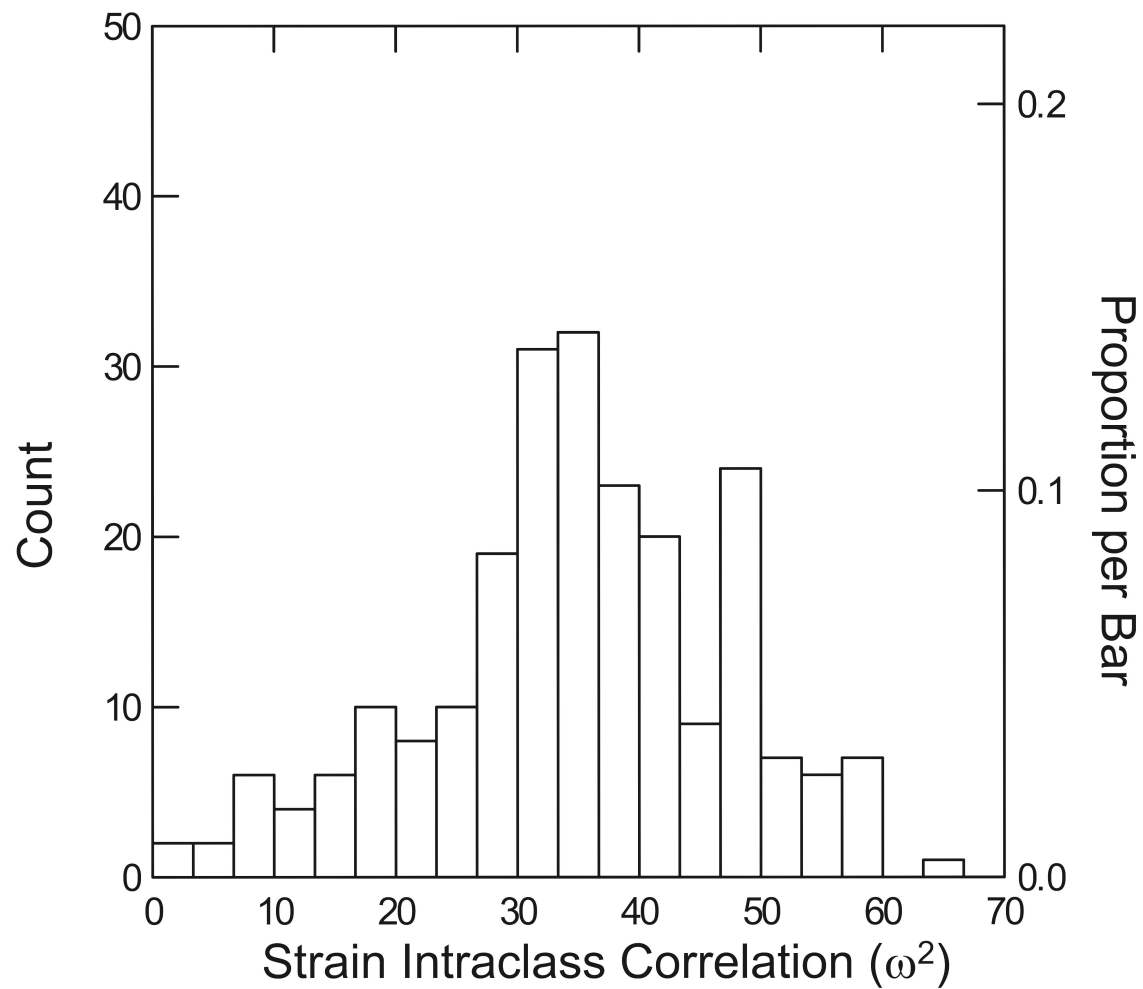


Figure 2-2. Strain intraclass correlation for all measures.

A histogram of strain intra-class correlations for all behavioral measures (sexes combined). A majority of the behavioral measures (146 of 257) have $\omega^2 > 0.30$, making them amenable to QTL mapping (from (164)).

effect sizes, consistent with the previous reports (216). Other traits with large effect sizes include acoustic startle response, pre-pulse inhibition, morphine withdrawal and alcohol withdrawal. Intermediate effect sizes were observed for nociception-related traits, morphine side-effects, ethanol-induced ataxia, baseline HIC, blood ethanol concentration, cocaine-conditioned place preference and anxiety. Only a small number of measures had effect sizes below 0.1. These were most often measures derived from linear combinations of two measures, for which the expected variance of the derived scores is at least four times the variance of either measure alone (98). The 30% effect size is a conventional guide for traits amenable to genetic analysis. The proportion of variance accounted for by strain varies between the sexes for many traits. For some traits, females have a higher proportion of non-genetic variation and for others males exhibit more non-genetic variation. Across test batteries, except for adrenal weights, the main effect of strain consistently accounts for most of the variation. The partial ω^2 for the main effects of sex were considerably lower than the strain main effects, except for left and right adrenal weight measures. However, sex $\times$ strain interaction effects were of greater magnitude than the main effects of sex, though not as large as strain effects.

Sub-populations effects

In our comparison of historical strains with the newly expanded BXD lines (BXD Sub-population effects), we evaluated differences among the three groups of BXD RI lines tested simultaneously for all 257 measures. These were BXD 1–

32 (208), BXD 33–42 (209) and BXD 43–100 (162). Only the adrenal weight measures show any significant differences among the three sub-populations for the left ($p = 3.23\text{E-}04$) and right ($p = 1.99\text{E-}04$) adrenal weights, when sub-population means are considered. Below we address the issue of mapping in the combined sub-populations.

Comparison of phenotypes across time and environments

Saline-induced locomotor response

Locomotor activity in an open field has been commonly used as a measure for exploration, novelty seeking, anxiety and predisposition to addiction to drugs of abuse. In our study, two laboratories, Memphis (cocaine) and UTHSC (ethanol), collected data on open field locomotion following saline in 15 and 20 min on 64 BXD RI strains, respectively. We compared these data with data collected on saline-induced locomotor response by Demarest and colleagues on 25 BXD RI strains (54). Results indicate that there is good correlation among data on saline-induced locomotor response across the three sets (Figure 2-3). Correlations among the $r_{\text{Memphis,UTHSC}} = 0.556$, $r_{\text{Memphis,Demarest}} = 0.781$ and $r_{\text{Demarest,UTHSC}} = 0.695$. These results confirm previous reports that locomotor activity is a highly stable trait and that saline-induced locomotor response is not only robust across laboratories but also across previously published studies.

Thermal nociception

Sensitivity to thermal stimuli is a complex trait that has been used as a

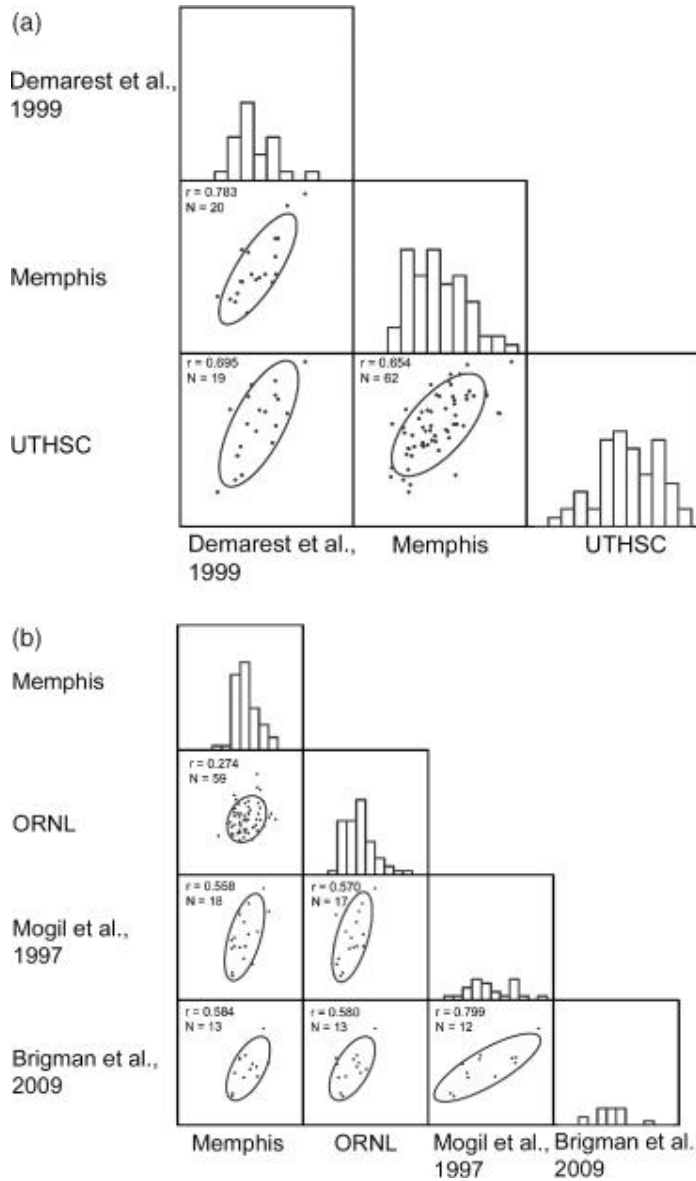


Figure 2-3. Correlation of behavioral measures across environments and time.

(a) Saline-induced locomotor activity is well correlated across test sites and time periods. (b) Another measure for which positive correlations existed across test sites and time periods is of thermal nociception (from (164)).

predictor of sensitivity to analgesic drugs. Thermal nociception (hot plate latency) was performed in two laboratories, namely, ORNL and Memphis in this study. Data collected in this study were compared with a previously published study of thermal nociception by Mogil and colleagues (140) and Brigman and colleagues (24) (GeneNetwork RecordID: 10897). Correlations ranged from 0.274 and 0.799 (Figure 2-3). There is poor correlation among the data collected at the Memphis and ORNL laboratories and can likely be attributed to the two different hot plate temperatures used. The hot plate test was performed at 52°C at Memphis, whereas at ORNL the hot plate temperature was 54°C. Specifically, ORNL hot plate data have a slightly higher correlation to the Mogil study compared with the Memphis hot plate data. This is likely due to the 54°C that was employed in both the ORNL hot plate study and the Mogil study. The range of values obtained at ORNL and Memphis was lower than those of the other testing laboratories, which produced some range restriction in the observed correlation coefficients. This may be due to more consistent laboratory environmental conditions in these assays or to experimenter effects on stress-induced analgesia, both of which have been previously reported to influence thermal nociception (37, 38). These correlations are lower than the inter-laboratory correlations than those reported by Wahlsten and colleagues (216). This may be due to the limited genetic variability in the BXD RI panel relative to a survey of the common inbred strains, but could also be due to the lower heritability of hot plate measures relative to other measures for which this type of analysis has been performed.

QTL mapping and genetic correlations

As an example of a single-trait analysis in the newly expanded strain set, we performed an analysis of mechanical nociception, the latency to respond to a plastic-coated smooth alligator clip placed on the tail (Figure 2-4). Previous studies of this trait in inbred mouse strains show high heritability of 0.69 (115). Analysis in MouseTrack using the RI analysis tool showed that this measure has statistically significant main effects of sex and strain (Sex: $\omega^2 = 0.0029$, $p < 0.04$; Strain: $\omega^2 = 0.26$, $p < 0.001$), and interaction effects (Sex $\times$ Strain: $\omega^2 = 0.021$, $p < 0.03$). The variances accounted for by strain are 0.32 and 0.27 in females and males, respectively. BXD sub-population effects on females ($p < 0.68$) and males ($p < 0.93$) were non-significant. Because the main effect of sex was significant, male and female data were exported separately for analysis in GeneNetwork, which uses a single vector of strain means as input. Permutation thresholds for significant and suggestive LOD scores were 3.78 and 2.21, respectively, in males and 3.6 and 2.27, respectively, in females. Interval mapping of tail clip latency for males of the full BXD RI panel shows suggestive QTL on Chr 2 (LOD = 2.8) and 9 (LOD = 3.5), whereas suggestive QTL for females of the full panel are located on Chr 1 (LOD = 3.5), 11 (LOD = 3.0) and 17 (LOD = 2.5). Genes within the one-LOD confidence interval from QTL peaks for males (Chr 2: 77–97 Mb; Chr 9: 44–48 Mb) and females (Chr 1: 90–95 Mb and 97–107 Mb; Chr 11: 53–55 Mb; Chr 17: 78–84 Mb) were exported as positional candidate gene lists. These results, generated in the full RI panel (BXD 1–100) were compared with interval mapping

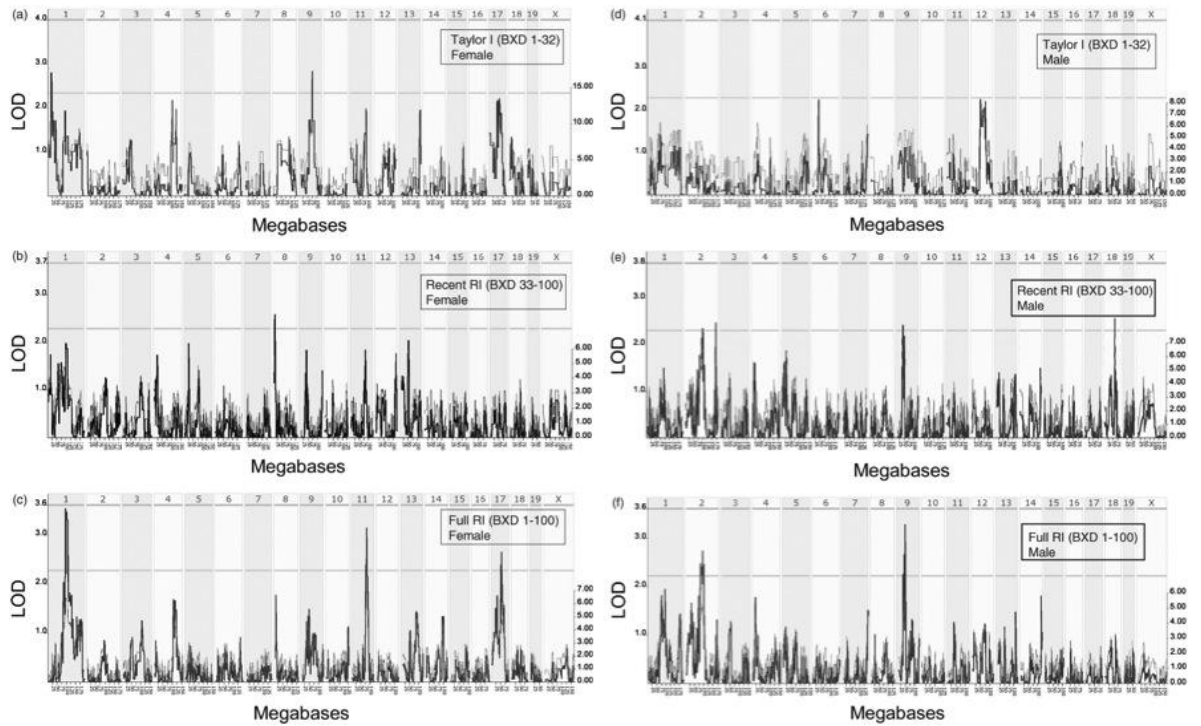


Figure 2-4. QTL mapping of tail clip latency in old and recent BXD RI strain panels.

QTL analysis of a single-trait, tail clip latency was performed across three BXD panels, namely, Taylor I (BXD 1–32), Recent BXD RI (BXD 33–100) and Full BXD RI (BXD 1–100). In females, suggestive QTL were detected on Chr 1 and 9 for Taylor I (a), Chr 8 for the Recent BXD RI (BXD 33–100) (b) and Chr 1, 11 and 17 for the Full RI Lines (c). Mapping of tail clip latency in males showed no suggestive QTL for Taylor I (d), suggestive QTL on Chr 2, 9 and 18 for Recent BXD RI (BXD 33–100) (e), and two suggestive QTL on Chr 2 and 9 for the Full RI panel (f) (from (164)).

results generated from previously available BXD RI lines including just the original set (Taylor I, BXD 1–32), and with exclusive use of the recent RI lines by Taylor (Taylor II, BXD 33–42) and by Peirce and colleagues (BXD 43–100). Quantitative trait locus mapping for each of the four sets (two male and two female sets) was performed. In females, suggestive QTL were detected on Chr 1 and 9 for Taylor I (Figure 2-4a), and Chr 8 for the recent RI set (Figure 2-4b). Mapping of tail clip latency in males showed no suggestive QTL for Taylor I (Figure 2-4d) and three suggestive QTL on Chr 2, 9 and 18 for the recent RI set (Figure 2-4e). The marker present at the peak of the suggestive QTL for females is gnf01.018.340 (Chr 1 at 21.43 Mb). This single nucleotide polymorphism (SNP) lies within the *Kcnq5* potassium channel gene (KCNQ5) (UCSC genome browser, <http://genome.ucsc.edu/index.html>, July 7, 2009). KCNQ channel (K(V)7.2-5) genes have been implicated in neuropathic pain (81). The effect at this marker across the Taylor I and recent BXD RI set indicates that while mice with the D allele have consistent phenotypic means across both sets, mice with the B allele have higher means in the Taylor I lines and lower means in the recent lines relative to the D2 allele. However, in the expanded RI set, the D2 allele is the allele with the higher effect and consequently results in a non-significant QTL (Figure 2-5). Genetic correlation of tail clip latency to genome-wide gene expression was performed for each of five brain tissues for which expression profiles across the BXD RI lines are in GeneNetwork, namely, whole brain, neocortex, striatum, cerebellum and hippocampus. Correlation results from each of the five brain regions were filtered using a *p* value threshold of 0.001 to

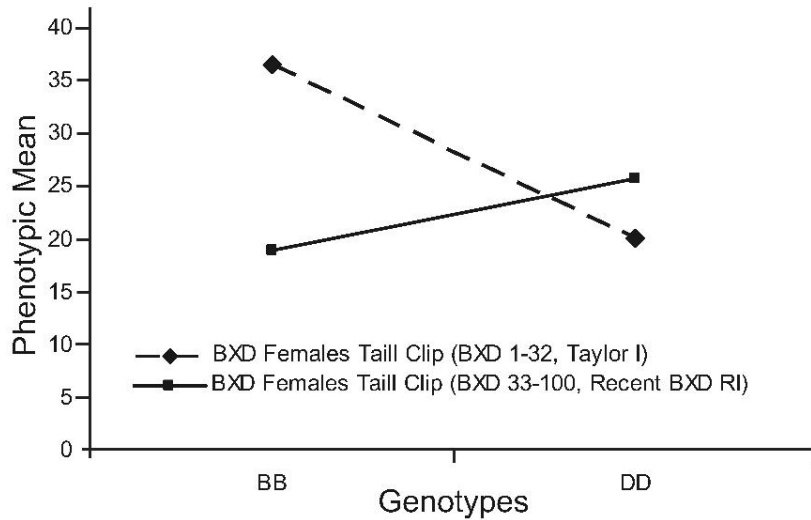


Figure 2-5. Effect plot by BXD RI Strain panel.

Effect plot of marker gnf01.018.340 (Chr 1at 21.43 MB) in the old (BXD 1 - 32) and recent BXD RI strain panel (BXD 33 - 100) for tail clip latency in females. The marker present at the suggestive QTL in Taylor 1 has a higher phenotypic means for the B allele when compared with the D allele. This difference is significant and therefore results in the presence of a QTL. The same marker in the Recent BXD RI lines has a higher phenotypic mean for the D allele, but no QTL is present within this strain set (from (164)).

create coexpression candidate lists. All candidate gene lists obtained from interval mapping and genetic correlations were then uploaded to GeneWeaver.org for combinatorial analysis. Maximum gene to measure (phenotype) connectivity was observed for *Slfn5* or schlafen 5 (associated with positional candidates on Chr 11 for females, male cerebellum gene expression correlates and male striatum gene expression correlates) and *Ankrd12* or ankyrin repeat domain 12 (associated with female cerebellum gene expression correlates, male cerebellum gene expression correlates and male hippocampus gene expression correlates). Other candidate genes in the interval were not found on this list of correlates, but include genes belonging to voltage-gated sodium (e.g. *Scn2b* and *Scn4b*, Tail Clip-male positional candidates on Chr 4) or calcium channels class of genes (13), cholinergic receptors class of genes (e.g. *Chrm4*, Tail Clip-male positional candidates on Chr 2, (63)) and calcitonin receptor-like genes or calcitonin gene-related peptides (e.g. *Calcr1*, Tail Clip Male Positional candidates on Chr 2, (120)), all of which have substantial literature support for involvement in nociception.

Multi-trait QTL mapping

A total of 27 factors accounting for 73% of the total variance for the 234 variables were obtained using a maximum likelihood factor analysis of behavioral traits spanning multiple test batteries. A parallel analysis and an optimal coordinates analysis from the R/nFACTOR package both show eight informative factors. The top 15 factors together account for 63.5% of the total variance

among the trait measures. Factors accounting for trivial amount of variances (<1.3%) have been excluded, as they would contribute little to the overall interpretation of the factors.

Factor interpretations were obtained by examining high and low factor loadings of behavioral measures on factors. Significant correlations among factors exist and are expected as no factor rotation procedure was applied, as factor rotation procedures are often employed to identify distinct factors. Results indicate that measures within test batteries load onto multiple factors, thereby supporting the hypothesis that each latent factor is associated with multiple behavioral measures and each behavioral phenotype measure is determined by variation along multiple trait dimensions. Factor 1 has high factor loadings for measures almost entirely from the Cocaine test battery and can be interpreted as injection stress-induced locomotor activity. Factor 2 represents morphine withdrawal measures such as jumps, defecation and urination. Factor 3 represents morphine activity/response. Factor 4 represents reactivity measures such as startle response and anxiety-like measures. Factor 5 predominantly consists of measures pertaining to locomotor activity in a novel environment. Factor 6 contains measures pertaining to conflict avoidance. Factor 7, like Factor 3, consists of measures with high factor loadings pertaining to morphine activity, but instead of response these measures are related to duration of activity. Factor 8 consists of measures related to cocaine sensitization. Factor 9 consists of measures related to stress and anxiety such as mechanical sensitivity (von Frey threshold) and zero-maze measures and adrenal weights. Factor 10 consists of

measures related to vertical activity spanning general behavior and cocaine test batteries. Factor 11 represents measures related to startle response. Factor 12 consists of measures pertaining to drug environment conditioning. Factors 13 and 14 both consist of responsivity measures, with Factor 13 describing response to sensitivity, whereas Factor 14 describes response to novelty. Lastly, Factor 15 represents anxiety related to acute stress.

Quantitative trait locus mapping of these 15 latent factors showed the presence of suggestive QTL on Chr 1, 4, 5, 6, 8 9, 10, 11, 13 and significant QTL on Chr 9 (Factor 3), Chr 10 (Factor 2) and Chr 13 (Factors 1 and 13) (Table 2-2). Results indicate that multiple factors map onto a single chromosome, for example, distal Chr 13, has suggestive QTL for Factors 1 (76–86 Mbp), 5 (78–84 Mbp, 94–98 Mbp), 6 (75–98 Mbp) and 10 (91–98 Mbp) and significant QTL for Factors 2 (78–80, 88–92 and 94–98 Mbp) and 13 (75–84 Mbp). Factor 15 also maps to Chr 13 but maps to the proximal region rather than the distal region. The Chr 13 genomic region that encompasses the above-mentioned QTL spans 23 Mbp (75–98 Mbp) with a total of 124 genes in the interval. Of these genes ninety-six are polymorphic, among the compelling candidates is *Mctp1* (multiple C2 domains, transmembrane 1). Interestingly, *Mctp1* has previously been reported to be strongly associated with bipolar disorder in individuals of European ancestry (Scott et al. 2009). The large number of polymorphic genes within Chr 13 QTL intervals could be one possible explanation for the fact that multiple factors map onto the same genomic region. To determine the relationship of these factors to a substance use-related phenotype, we correlated the factor

Table 2-2. Multi-trait QTL mapping.

Factor	Interpretation	Chromosome											
		1	4	5	6	8	9	10	11	13	15	18	19
1	Locomotor (Injection stress-induced activation)									3.4*			
2	Morphine Withdrawal				3.2*					4.2**		2.8*	
3	Morphine Activity (Initial/Peak response)			3.0*				4.9**	3.2*				
4	Anxiety/Reactivity		2.9*				4.5**		3.0*				
5	Locomotor response to a novel environment						2.5*			2.7*			
6	Activity (Conflict-Avoidance)					2.3*				3.0*			
7	Morphine (Activity:Duration)												
8	Cocaine Sensitization						2.8*				3.4*		
9	Stress-Anxiety												
10	Vertical Activity Performance (Exploration)	2.6*			2.7*		2.7*			3.5*			
11	Startle												2.6*
12	Drug-Environment Conditioning												
13	Responsivity-Sensitivity									5.0**			
14	Response to Novelty	3.0*	2.9*										
15	Anxiety-Acute Stress									2.8*			

scores to phenotypic values from a previous study of ethanol self-administration (166). This analysis illustrates an approach to determine which dimensions of heritable behavioral variation are most related to a specific behavioral phenotype. Correlations of alcohol preference to factor scores ranged from $|r| = 0.076$ to $|r| = 0.674$, for Factor 4 for which the correlation was negative such that mice with increased startle response and anxiety-like phenotypes had higher alcohol preference values and vice-versa. This factor is regulated by a QTL on chromosome 9 that contains *Scn4b*, which is an expression correlate of both ethanol withdrawal and Factor 4.

Combinatorial analysis of gene to phenotype associations

GeneNetwork analysis of each trait resulted in the generation of over 3,500 sets of genes that are associated with trait values either via genetic correlation of gene expression in one of the five target tissues or through positional candidacy in significant or suggestive QTL. These gene–phenotype associations were integrated via a combinatorial analysis, which represents the data as a bi-partite graph consisting of gene vertices and phenotype vertices. Here, a ‘phenotype’ vertex represents each list of correlates or candidates, and the gene vertices represent each list member. Edges connecting these vertices represent significant gene – phenotype associations. The union of all lists contained occurrences of approximately 33,000 genes and ESTs. However, many of these genes are represented only once, and through convergent analysis can be eliminated. By analyzing the number of phenotypes to which

each gene was associated, we were able to identify those genes that most broadly and reliably related to the behavioral functions assessed. A ranked ordering of genes was generated based on each gene's connectivity to phenotypes. These rankings were made a part of the NeuroSNP database for supplementing addiction arrays (179). The most highly connected gene was *Mef2c*; interestingly this gene also resides in the Chromosome 13 QTL interval for several behavioral factors. It should be noted that the composition of the behavioral phenotypes in this analysis influences the results obtained. For example, some aspects of behavior such as locomotor behavior and anxiety-like behaviors have many measures in this study. As such, they are over-weighted relative to other aspects of behavior. Despite the overweighting of certain measures, we conclude that the most highly connected genes are still broadly connected with approximately 150 unique traits. These traits include traits measuring anxiety, stress and pain sensitivity and span multiple test batteries namely cocaine, general behavior, morphine, ethanol and pain.

Genes that were highly connected to many of the trait measures were entered into gene set over-representation analysis using WebGestalt system, which performs a hyper-geometric test for the enrichment of Gene Ontology and KEGG pathway annotations among sets of genes. Among the top most highly connected 5% of genes, significant over-representation ($p < 1E - 05$) was observed in 30 pathways of relevance to brain and behavior. The top most highly connected 10% of genes showed significant over-representation ($p < 1E - 05$) of 32 pathways related to brain and behavior (Table 2-3). These results indicate the

Table 2-3. Top 5 and 10 percent genes of the resultant gene sets and their KEGG pathway.

KEGG Pathways	Hypergeometric Test for Enrichment (<i>p</i> values)	
	Top 5% of Genes (1684)	Top 10% of Genes (3368)
Adherens junction	5.42E-08	8.18E-07
Adipocytokine signaling pathway	4.43E-03	3.38E-09
Axon guidance	2.74E-11	7.72E-08
B cell receptor signaling pathway	3.21E-06	1.47E-08
Calcium signaling pathway	1.02E-09	1.03E-11
Cell adhesion molecules (CAMs)	9.12E-09	5.02E-05
Cell cycle	1.67E-09	2.26E-04
Chronic myeloid leukemia	1.53E-05	4.53E-06
Colorectal cancer	6.16E-09	6.31E-08
Cytokine-cytokine receptor interaction	NE	1.05E-07
Focal adhesion	4.42E-09	4.96E-08
Gap junction	6.08E-08	4.57E-07
Huntington's disease	2.09E-07	8.18E-08
Insulin signaling pathway	8.06E-08	7.15E-08
Jak-STAT signaling pathway	7.35E-03	1.91E-06
Leukocyte transendothelial migration	1.38E-05	8.57E-08
Long-term depression	6.56E-08	7.13E-05
Long-term potentiation	9.80E-11	1.04E-05
MAPK signaling pathway	8.66E-15	1.43E-13
Natural killer cell mediated cytotoxicity	1.07E-07	3.60E-14
Neuroactive ligand-receptor interaction	NE	1.31E-07
Phosphatidylinositol signaling system	7.27E-06	1.31E-03
Regulation of actin cytoskeleton	1.13E-12	5.51E-10
T cell receptor signaling pathway	4.48E-08	9.56E-12
Tight junction	6.36E-07	1.45E-04
Toll-like receptor signaling pathway	7.98E-05	1.73E-09
Wnt signaling pathway	1.07E-15	2.50E-09

pathways that are either candidate causes or are under regulation of genetic polymorphisms that are commonly observed in the BXD population. The principle behind this approach is that genetic variation affects many genes associated with QTL candidates, and that the convergent evidence gained from multi-dimensional analyses of the same traits can highlight the most frequently associated pathways. Because these data have all been deposited in the GeneWeaver.org tool, users of these data can perform analyses of specific and roughly balanced subsets of the traits.

Discussion

BXD RI panels are currently the largest RI strain panel available. In addition to being the largest, the BXD RI panel have a rich history of use in behavioral genetics, increased numbers of recombinations and improved precision in QTL mapping (compared to a F2 or BC mapping population). However, effective utilization of this resource warrants improvements in behavioral phenotyping protocols that facilitate rapid, efficient and accurate phenotypic measurements across large numbers of strains. By taking advantage of the high-throughput behavioral cores established by the NIMG ENU-Neuromutagenesis program of the TMGC, we were able to measure over 250 behavioral phenotypes across four sites in the expanded BXD RI strain panel. These measures were obtained, on average, across 58 BXD RI strains with at least 10 males and 10 females per strain. 64.61% of these phenotypes have genetic effect sizes of at least 30% making them amenable to QTL mapping. A

genetic effect size of 30% with a within-strain sample size of 20, allows us to detect QTLs where the correlations between the molecular marker and the behavioral phenotype is approximately 0.97. It should be noted that these calculations were performed using two assumptions: First, the molecular marker is directly linked to the QTL and second, only additive QTLs are considered. Violations of these assumptions will result in lower correlations between the molecular marker and phenotype, possibly leading to increases in false positives. Additionally, the sample sizes employed in our study enable us to detect significant QTLs ($p < 0.05$) with an effect size of 16% with 90% power. Increasing the number of strains will help in detecting QTLs with smaller effect sizes. For example, if one were to utilize the entire BXD RI strain panel ($n = 89$), one can map QTLs with an effect size of 11%. Furthermore, if one were to utilize a F2 population to map traits with heritability of 0.30, then the effective sample size would be approximately 295. This number increases to 760, when the heritability of the trait being mapped is low, around 0.10 (Figure 2-6). It would be extremely resource and cost intensive (genotyping) to perform QTL mapping population in a F2 populations with the above mentioned sample sizes. Even if methods such as selective genotyping are used to abate genotyping costs, this comes with a decrease in statistical power to detect QTLs.

The majority of the behavioral traits measured had significant main effects of sex and strain $\times$ sex interactions. This observation is consistent with previous

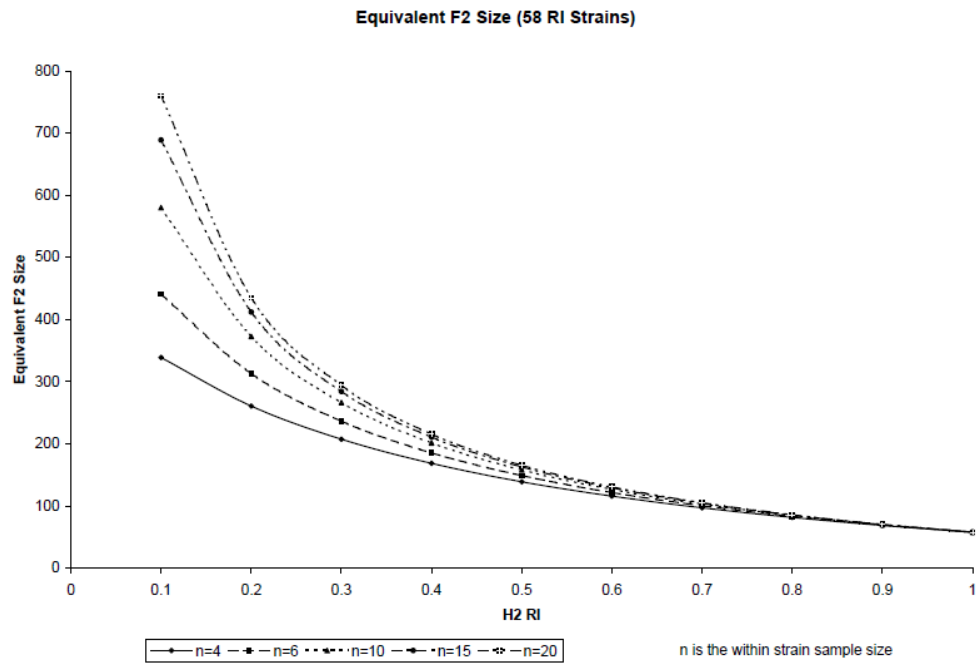


Figure 2-6. Equivalent F2 sample sizes for 58 RI strains for varying n and h^2 .

Estimated F2 sample sizes for 58 RI strains as a function of within strain sample size (n) and heritability (h^2). As h^2 decreases the equivalent F2 sample size required increases compared to an RI strain panel, resulting in increased cost (due to genotyping).

studies (38, 215). It is important to account for significant strain $\times$ sex interactions while mapping such traits. In GeneNetwork, this can be achieved by mapping the two sexes separately (as was done in our study). Another approach is to regress male data with female data and vice versa followed by single marker correlations using the resultant regression residuals (70). Both the above-mentioned approaches decrease the overall statistical power as they involve grouping the data by sex prior to running the analysis. By employing a nested model that appropriately accounts for sex across and within strain, better statistical power in mapping traits can be achieved. However, permuting such a model is complex due to the resulting non-independence within strains. Permutations methodologies have been proposed to address such issues (160). In our study, we found sex differences in trait heritability such that for certain traits female were higher than males and vice versa. This suggests increased within strain variance relative to between strain variance. Increased within strain variation is indicative of additional factors that influence the proportion of phenotypic variation that is genetic. In males such factors include, social dominance. In females, studies have revealed the influence of estrous cycle variation in behavior (14). Housing density and social isolation due to testing protocols also introduce additional variation that influences trait heritability (139).

Since the BXD RI strain panel has gone through two expansions after its initiation in 1979, we assessed sub-population effects across all traits measured. Only male adrenal weights had significant sub-population main effects. Since all

but one trait showed significant sub-populations effects, phenotypically it is safe to consider all three BXD RI sub-populations as one large BXD RI strain panel. However, genotype distribution patterns reveal a completely different picture. Shifman and colleagues (187) detected 52 of 13,367 SNPs wherein the first BXD RI sub-population showed no variation in their genotype distribution patterns. In the second BXD RI sub-population, only five of the 52 SNPs show no variation. All SNPs with fixed SDPs in the first two BXD RI sets show variation in the expanded BXD RI sub-population. The lack of variation in the SDPs of typed loci in the first and second sub-population will affect QTL mapping, as these loci will be non-detectable. Similarly using the entire BXD RI panel without accounting for sub-population structure could result in varying mapping results.

Despite the advantages offered by the BXD RI strain panel, it suffers from certain limitations. First, as it is derived from two progenitor strains, the breeding design is only able to capture a limited amount of genetic variation (approximately 16%). Second, centimorgan or sub-centimorgan precision in QTL mapping is not yet attainable. And lastly, the presence of linkage disequilibrium could lead to detection of false positive QTLs. Even in the light of these limitations, the ability to undertake integrative data analysis within the BXD RI strain panel makes them an ideal resource for the genetic analysis of quantitative traits. This is primarily due to its role as a genetic reference population. With the advent of advanced mapping populations, such as the CC, the limitations of RI panels such as the BXD RI will be alleviated. The CC is an emerging RI panel

that is being derived from eight phenotypically and genetically diverse founder strains (34, 42, 95, 142). The initial founder strain diversity facilitates increases in the total amount of genetic variation captured, numbers of recombination, and precision and resolution of QTLs. Application of high-throughput behavioral phenotyping techniques to the CC may help in validating QTL detected in two progenitor RI strain panels. Furthermore, the increased mapping resolution supported by the CC will facilitate and improve prospects of QTG discovery.

CHAPTER III: PROMISE OF EMERGING GENETIC REFERENCE POPULATIONS

This chapter is adapted from the following publication:

Philip VM, Sokoloff G, Ackert-Bicknell CL, Striz M, Branstetter L, Beckmann MA, Spence JS, Jackson BL, Galloway LD, Barker P, Wymore AM, Hunsicker PR, Durtschi DC, Shaw GS, Shinpock S, Manly KF, Miller DR, Donohue KD, Culiati CT, Churchill GA, Lariviere WR, Palmer AA, O'Hara BF, Voy BH, Chesler EJ. *Genome Res.* 2011 Aug;21(8):1223-38.

My contributions to this paper were QTL mapping, analysis of inbreeding depression, SNP selection for the custom CC genotyping array, interpretation of results and part of the writing.

Abstract

Over the last few decades, panels of genetic reference populations (GRP) have been utilized for the genetic dissection of complex traits. A lot has been learnt from existing GRP, however they are yet to be utilized to their full potential. Furthermore, current GRP fail to model complex human diseases due to limited heterogeneity as a consequence of derivation from two founder strains. The collaborative cross (CC) is a mouse reference population with high allelic diversity that is being constructed using a randomized breeding design that systematically out-crosses eight founder strains, followed by inbreeding to establish new highly recombinant inbred (RI) strains. Five of the eight founders are common laboratory strains and three are wild-derived. Since its inception, the partially inbred CC has been characterized for physiological, morphological and behavioral traits. The construction of this population provided a unique opportunity to observe phenotypic variation as new allelic combinations arose

through inter-crossing and inbreeding to create new stable genetic combinations. Processes including inbreeding depression and its impact on allelic and phenotypic diversity were assessed. Phenotypic variation in the CC breeding population exceeds that of any existing mouse genetic reference populations due to both high founder genetic diversity and novel epistatic combinations. However, some focal evidence of allele purging was detected including a suggestive QTL for litter size in a location of changing allele frequency. Despite these natural consequences of inbreeding, high diversity and precision for genetic mapping remains. These results demonstrate the potential of the CC population, once completed.

Introduction

Genetic reference populations (GRP) are an invaluable resource for the study of complex traits. An extremely valuable characteristic of these populations is their ability to facilitate data integration across multiple phenotypic domains and biological scales (210). However, currently existing reference populations are yet to be utilized to their full potential. This is primarily due to the derivation of existing panels from two founder genomes and limited number of RI strains available in an analysis. The presence of only two founder strains limits the amount of genetic variation that can be captured. Additionally, the founder strains (the classical inbred strains currently used) share large fractions of their genomes due to identity by descent (IBD), further limiting allelic diversity and variation. This results in non-detectable genetic variation causing “blind-spots” in

QTL mapping (175). These limitations have necessitated the development of new and advanced GRP. The collaborative cross (CC) is one such emerging advanced GRP. The CC is derived from eight genetically and phenotypically diverse inbred founder strains. The eight strains that were selected are – A/J, C57BL/6J, 129S1/SvImJ, NOD/LtJ, NZO/HILtJ, CAST/EiJ, PWK/PhJ and WSB/EiJ. Three of the eight founder strains, namely, CAST/EiJ, PWK/PhJ and WSB/EiJ are wild-derived strains representing, *M. m. castaneus*, *M. m. musculus* and *M. m. domesticus* mouse sub-species, respectively. The inclusion of three wild-derived strains introduces additional genetic diversity and in addition to the five inbred strains, helps capture approximately 90% of available genetic variation. In addition to increased genetic variation, the CC is expected to have higher numbers of recombinations and larger numbers of independent genetically diverse RI strains than any existing GRP. These features will make the CC a widely used resource for the genetic analysis of complex human disorders and diseases.

Breeding design

The breeding scheme employed in the derivation of the CC involves three generations of out-crossing followed by over 20 generations of inbreeding (Figure 3-1). The first generation of out-crossing (G0) results in 56 G1 (two-way) hybrids. Pairs of non-overlapping G1 hybrids are then out-crossed to generate four-way G2 progeny. Out-crossing pairs of four-way G2 yields an eight-way G2 progeny (G2:F1). The G2:F1 generation serves as the starting point for inbreeding, by

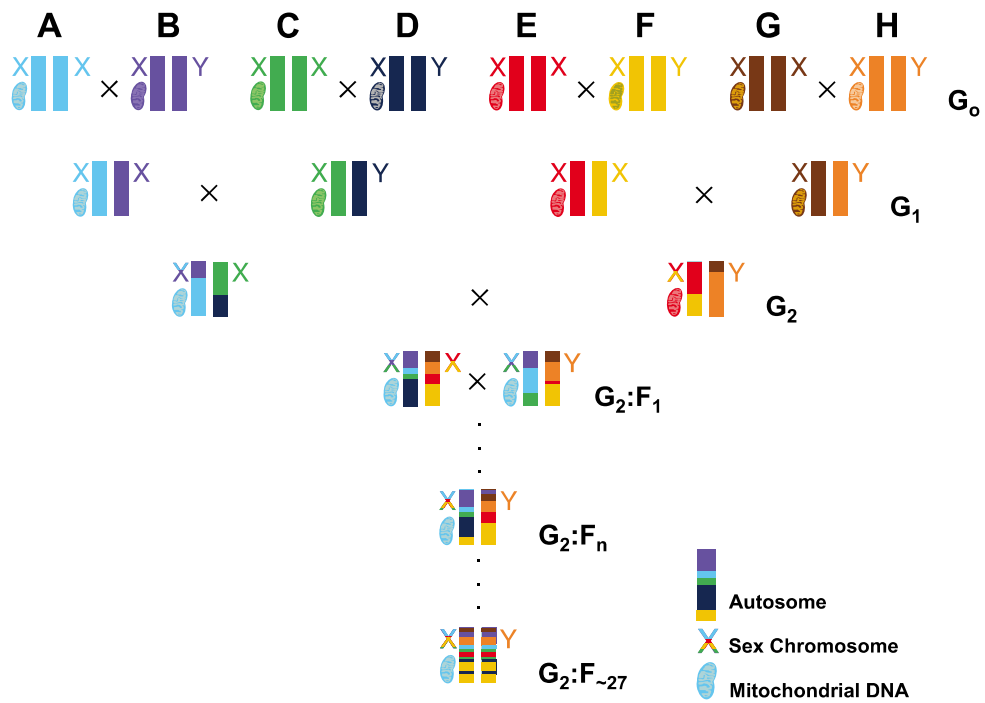


Figure 3-1. Collaborative cross breeding design (34).

Breeding design for one eight-way CC RI strain is shown. Letters A-H represent the eight CC founder strains. Positions of the eight founder strains determine contributors of mitochondrial DNA and Y chromosome. In above schematic, strain A contributes the mitochondrial DNA, while strain H contributes the Y chromosome (from (34)).

sibling mating for over 20 generations. The overall goal of breeding design described above is to generate hundreds of independent CC RI strains, each a fine-grained mosaic of the eight founder strains.

Methods

Subjects

A total of 650 lines were initiated according to the Complex Trait Consortium protocol (42). Of these, 414 lines with at least one male or one female survived to the G2:F5 generation, while the remaining lines were lost during the inbreeding process. Litters were weaned at approximately three weeks of age, into breeding pairs determined by the CCWORKS software assisted mating system developed by Kenneth F. Manly (34). Litter size, parental age at birth of first litter, and length of the interval from mating to the first litter were recorded as measures of fecundity. Mice were maintained in breeding pairs until they entered the phenotyping protocol. Retired breeders from the cross population were phenotyped on a variety of behavioral, physiological and morphological measures. Phenotyping was performed across generations to evaluate trait heritability. Phenotyping and genotyping were performed in at least one breeding pair per line from generation G2:F5-G2:F8 for QTL analysis. Most of the genotyped mice came from the G2:F5 generation. From this generation, 235 lines were tested. At this generation the population is estimated to be approximately 75% inbred at the genotyped loci (34), based on allelic identity by descent calculations performed using the R/ricalc package (25). Because assays

were added to the phenotyping pipeline throughout the course of the breeding project not all mice were evaluated on all measures. Approximately 7,500 mice of both sexes (except for gonadal fat pad mass and testes weight) from up to 626 lines were phenotyped for body weight, tail length, fasting plasma glucose level, and weights of kidney and heart, behavioral and physiological traits; of these, approximately 3,000 have been phenotyped for adiposity based on perigonadal fat pad weight.

Mice did not enter the phenotyping queue until the birth of their grand-progeny to ensure that lines could be propagated. Average age of testing was 38.5 weeks. Over 95% of the mice were phenotyped between the ages of 20 to 100 weeks. Of the total number of mice phenotyped, 52% were females and 48% were males.

General phenotyping methods

The phenotyping strategy was designed to capture a panel of complex traits amenable to high throughput analyses while broadly reflecting behavior, morphology and physiology. Upon birth of grand-progeny mice were separated into individual pens for phenotyping. The individuals were then moved to a holding room where they were housed for three weeks prior to testing where females were checked to ensure that they were not pregnant at the time of testing. Mice were housed in this room throughout testing. At the start of the week prior to the first test an elbow shaped PVC pipe was added to each cage as a standardized enrichment device. At this point in the queue, mice intended to

begin testing the following week remained in place on one side of one rack.

Detailed phenotyping and dissection methods are described in detail elsewhere (165)

Genotyping and SNP selection

A custom array using the Illumina iSelect platform for the Infinium system was developed for SNP genotyping. A subset of 11,969 SNPs were chosen from the NIEHS-Perlegen combined SNP panel (232). A sliding window was utilized in order to search for the minimum sets of SNPs that could discern all eight founder strains. The search was terminated either when a set of SNPs were identified that together could discriminate all eight founder haplotypes or the window length reached 1.86 Mbp. Upon termination a new search window was initiated. The resulting array enabled genotyping at over 1200 such windows. This algorithm ensured that for a given number of SNPs on the array, the maximum density of informative markers could be used, while still ensuring that at any locus, ancestry can be traced back to one of eight founders.

Data analysis

Statistical analyses were performed within the MouseTrack database environment (mouse.ornl.gov; (10)), which includes SAS v. 9.1.3 tools for univariate statistics, univariate and multivariate outlier detection, modeling of litter size, and calculation of parent-offspring regression coefficients (34). For modeling of cross type and other effects on litter size and inbreeding depression,

we employed mixed models with random effects of specific crosses within cross types. Heritability for the progenitor generation was estimated using strain intraclass correlations, and for subsequent generations, it was estimated using parent-offspring regression by calculating the slope of the regression between the mid-parent (mean of the two parents) and offspring trait values. We note that these estimates cannot be extrapolated to heritability defined as transmission of phenotypic values in wild randomly mating populations. In wild randomly mating populations it is customary to adjust regression coefficients by relatedness of the two samples, nor is that our objective. We use unadjusted parent-offspring relations for the monitoring of resemblance among relatives throughout the inbreeding process. For the majority of traits, untransformed data were used to estimate resemblance of relatives. To more closely reflect genetic contribution to tissue function rather than allometric relationship to body weight, organ weights were first adjusted to body weight using a regression model in which individual trait values were regressed to the population mean. Only data from extant lines in the terminal generation studied were used to calculate parent-offspring similarity estimates.

QTL mapping

QTL mapping was undertaken using the R/Happy package developed based on multipoint HAPPY dynamic-programming algorithm and regression models (145). Trait data from G2:F5 generations and greater were utilized in the mapping analysis. Sample sizes ranged from 160 to 293 lines with at most one

male and/or one female per line. Each trait was assessed for normality prior to its mapping. Appropriate data transformations were applied to trait data that deviated significantly from normality. For each trait an additive model genome scan was preformed. For each marker, R/Happy (145) returns a $-\log P$ value. Additionally, for each marker using the residual sums of square under the null hypothesis of no segregating QTL (RSS_0) and alternative hypothesis (RSS_1 , additive or full model) LOD scores were calculated using the following equation:

$$\text{LOD Score} = (n/2) * \log_{10}(RSS_0/RSS_1).$$

This statistic can be utilized to determine significant and/or suggestive QTLs. Genome-wide significance thresholds were obtained using a modified permutation algorithm rather than the one available in the R/Happy package. The R/Happy package treats each observation as an independent observation and permutes the phenotype between subjects while keeping the genotypes fixed. In the mapping data utilized in the collaborative cross, we use two subjects per line (one male and one female). This introduces a dependent relationship among subjects (sib-pairs). Therefore lines are permuted rather than subjects and sex is randomized within lines, while keeping the genotypes fixed. One thousand permutations were performed and permutation based genome-wide thresholds were used to identify significant (5%, 10%) and suggestive (63%) QTLs. Confidence intervals around significant and/or suggestive QTL peak was defined using a 1.5 LOD support interval. Genes within this interval are considered candidate genes.

Results

Phenotypic variation

Phenotypic variance for many traits increased in the G1 generation (Table 3-1) relative to progenitor lines. However, phenotypic variation decreased slightly in the G2 and beyond. For all morphometric measures, the distribution of individual scores showed greater dispersion in the CC lines relative to the progenitors (Kolmogorov-Smirnov $p < .05$). As expected from the wide range of body sizes among the progenitor strains, indicators of overall growth, vary widely among the CC lines. For example, among progenitors, body weight ranged from 25.56 ± 0.85 g (WSB/EiJ) to 29.66 ± 2.16 g (NZO/HILtJ), whereas in the inbreeding generations (G2:F5 – G2:F8) this measure ranged from 23.65 g to 32.00 g. Tail length, an index of linear growth, ranged from 67.92 ± 7.98 mm (PWK/PhJ) to 94 ± 7.94 mm (NOD/LtJ) among the progenitors, whereas it ranged from 59.04 mm to 112.00 mm among the inbreeding generations. Heart weight and adiposity index also showed considerable variation among the progenitors and in the CC offspring, but the CC range was within the progenitors.

While it is interesting to note the spread of phenotypic variation relative to progenitor lines, increased phenotypic variation among offspring is expected due to transgressive segregation, the random assortment of increasor and decreasor alleles that were once in a fixed assortment in the progenitor lines. Therefore, we assessed phenotypic variation in the CC in contrast to a relatively large and widely used genetic reference population, the BXD recombinant inbred lines, for

Table 3-1. Trait heritability and variation in the emerging CC.

	Gen.	G0	G1	G2	G2	G2	G2	G2	G2	G2	G2	G2
					F1	F2	F3	F4	F5	F6	F7	F8
Trait												
Body weight [g]	n _{mice}	70	240	557	731	1002	950	754	611	403	333	181
	μ	26.25	33.95	30.8	30.91	31.23	30.49	29.13	27.64	27.92	24.78	24.33
	S.D.	12.21	10.94	5.79	5.65	6.21	6.41	6.05	5.89	6.26	6.91	6.54
	σ ²	148.84	119.54	33.41	31.84	38.52	41.08	36.56	34.59	39.19	47.65	42.65
	Range	15 - 63.4	19.36 - 93.36	20.14 - 51.68	18.92 - 50.32	15.75 - 56.06	16.25 - 56.53	14.67 - 54.12	13.7 - 53.3	17.22 - 45.57	12.29 - 50.47	11.8 - 47.6
	C.V.	0.47	0.32	0.19	0.18	0.2	0.21	0.21	0.21	0.22	0.28	0.27
	h ²	0.5	NA	0.3	0.36	0.42	0.64	0.56	0.46	0.61	0.59	0.59
	S.E.	NA	NA	0.04	0.06	0.05	0.04	0.05	0.04	0.05	0.07	0.09
Tail length [cm]	n _{mice}	63	185	313	471	702	717	607	566	367	277	141
	μ	85.43	96.95	97.33	98.55	97.34	95.85	93.36	91.17	90.99	85.57	85.65
	S.D.	10.73	4.62	5.42	5.56	6.18	6.92	8.41	7.94	8.49	9.81	12.15
	σ ²	115.08	21.31	29.3	30.86	38.09	47.83	70.66	62.9	71.94	96.15	147.42
	Range	45 - 100	78 - 110	65 - 113	86 - 115	80 - 118	75 - 115	20 - 112	60 - 112	32 - 110	40 - 105	0 - 105
	C.V.	0.13	0.05	0.06	0.06	0.06	0.07	0.09	0.09	0.09	0.11	0.14
	h ²	0	0.55	0.28	0.44	0.58	0.66	0.7	0.72	0.69	0.67	0.72
	S.E.	NA	0.08	0.1	0.08	0.07	0.04	0.05	0.04	0.05	0.07	0.18

Table 3-1 continued.

Gen.		G0	G1	G2	G2	G2	G2	G2	G2	G2	G2	G2
					F1	F2	F3	F4	F5	F6	F7	F8
Trait												
Heart weight [g]	n _{mice}	58	179	274	371	536	670	605	566	367	277	141
	μ	0.17	0.19	0.19	0.19	0.21	0.22	0.2	0.19	0.19	0.17	0.16
	S.D.	0.05	0.05	0.06	0.05	0.07	0.06	0.07	0.06	0.06	0.06	0.06
	σ ²	0	0	0	0	0	0	0	0	0	0	0
	Range	0.11 - 0.33	0.12 - 0.32	0.11 - 0.7	0.12 - 0.39	0.1 - 0.9	0.09 - 0.5	0.1 - 0.59	0.1 - 0.8	0.1 - 0.44	0.07 - 0.64	0.08 - 0.31
	C.V.	0.29	0.26	0.32	0.26	0.33	0.27	0.35	0.32	0.32	0.35	0.38
	h ²	NA	0.59	0.07	0.2	0.34	0.21	0.34	0.47	0.49	0.29	0.66
	S.E.	NA	0.12	0.09	0.06	0.09	0.05	0.05	0.06	0.06	0.09	0.14
Hot plate latency [log s]	n _{mice}	23	271	221	202	120	156	273	437	317	185	70
	μ	1.74	1.57	1.67	1.6	1.62	1.65	1.71	1.7	1.65	1.67	1.62
	S.D.	0.43	0.28	0.33	0.32	0.35	0.43	0.41	0.42	0.41	0.38	0.34
	σ ²	0.18	0.08	0.1	0.1	0.12	0.18	0.17	0.17	0.17	0.14	0.11
	Range	0.96 - 2.79	0.97 - 2.54	0.97 - 2.63	0.63 - 2.41	0.83 - 2.62	0.62 - 3.1	0.53 - 3.2	0.79 - 3.31	0.86 - 3.21	0.79 - 2.62	0.91 - 2.52
	C.V.	0.25	0.18	0.2	0.2	0.22	0.26	0.24	0.25	0.25	0.23	0.21
	h ²	NA	0.26	0.02	0.29	0.53	NA	0.38	0.44	0.45	0.45	0.37
	S.E.	NA	0.07	0.1	0.1	0.15	NA	0.12	0.1	0.06	0.08	0.1
Captur	n _{mice}	221	804	964	1069	933	776	618	566	366	201	87

Table 3-1 continued.

Gen.		G0	G1	G2	G2	G2	G2	G2	G2	G2	G2	G2
					F1	F2	F3	F4	F5	F6	F7	F8
Trait												
e and hold combi ned wildne ss score	μ	0.52	0.48	0.52	0.62	0.59	0.52	0.52	0.47	0.44	0.46	0.41
	S.D.	1.18	1.05	1	0.88	0.93	0.79	0.8	0.79	0.77	0.68	0.66
	σ^2	1.38	1.1	0.99	0.76	0.86	0.61	0.63	0.61	0.59	0.46	0.43
	Range	0 - 7	0 - 8	0 - 10	0 - 9	0 - 10	0 - 6	0 - 6	0 - 8	0 - 5	0 - 3	0 - 4
	C.V.	2.27	2.19	1.92	1.42	1.58	1.52	1.54	1.68	1.75	1.48	1.61
	h ²	0.96	0.47	0.04	0.15	0.24	0.32	0.28	0.14	0.4	0.36	NA
	S.E.	0.17	0.07	0.07	0.05	0.06	0.04	0.05	0.07	0.1	0.18	NA
74 Distan ce from center of open field [cm]	n _{mice}	NA	173	59	103	83	83	60	134	166	117	56
	μ	NA	5.31	4.8	5.27	4.29	5.24	5.67	5.25	5.53	5.76	5.28
	S.D.	NA	2.85	1.94	3.3	2.65	2.82	3.59	3.17	3.65	3.68	4.25
	σ^2	NA	8.1	3.73	10.86	6.99	7.94	12.83	10.05	13.25	13.48	18.04
	Range	NA	1.02 - 14.46	1.83 - 9.9	1.29 - 17.26	1.26 - 14.24	1.1 - 11.72	1.54 - 13.87	1.32 - 16.44	0.85 - 16.98	0.81 - 16.05	1.17 - 16.43
	C.V.	NA	0.54	0.4	0.63	0.62	0.54	0.63	0.6	0.66	0.64	0.8
	h ²	NA	0.15	NA	0.07	0.07	0.49	0.01	0.36	0.22	0.32	0.7
	S.E.	NA	0.09	NA	0.34	0.2	0.34	0.32	0.21	0.24	0.16	0.41
Distan ce travele	n _{mice}	NA	173	60	107	86	92	65	136	168	120	56
	μ	NA	1298.5 6	1469.1 1	1410.5 6	1439.9 8	1566.3	1324.8 9	1215.5 1	1283.3 4	1242.5 6	1399.1 2

Table 3-1 continued.

Gen.		G0	G1	G2	G2	G2	G2	G2	G2	G2	G2	G2
					F1	F2	F3	F4	F5	F6	F7	F8
Trait												
d in an open field during the first 3 min [cm]	S.D.	NA	480.13	353.66	412.91	469.71	634.68	510.09	489.89	590.29	622.01	801.14
	σ^2	NA	23052 0.7	12507 2.1	17048 9	22062 4.7	40280 7	26018 4.1	23998 2.6	34844 0.4	38688 4.3	64181 2
	Range	NA	450.13	806.37	505.23	269.28	551.92	180.85	111.09	118.71	147.03	147.71
			- 2751.2 7	- 2433.7 9	- 2420.2	- 2665.2 2	- 4380.5 3	- 2189.8 1	- 3016.9 1	- 3003.9	- 2609.8 8	- 4385.2 1
	C.V.	NA	0.37	0.24	0.29	0.33	0.41	0.39	0.4	0.46	0.5	0.57
	h2	NA	0.08	NA	0.48	0.49	0.4	NA	0.58	0.22	0.58	0.89
	S.E.	NA	0.06	NA	0.26	0.24	0.46	NA	0.29	0.2	0.15	0.32

which we have previously reported phenotypic values for diverse behavioral measures (164) obtained under the same conditions as our collaborative cross phenotypic data. Phenotypic values observed in generations G2:F5 and beyond for thermal nociception, distance travelled in the open field and body length spanned a far greater range than those obtained in the BXD RI strain panel (Figure 3-2).

Analysis of inbreeding depression

During the generation of the CC, we were able to assess the impact of inbreeding in the emerging CC population using the measure of litter size in these mice, as previous studies have shown that litter size reflects the impact of inbreeding depression in mouse populations (88, 91). Among the G1 crosses, mean litter size was lower for crosses involving one or two wild-derived lines ($F_{\text{cross type (2, 51)}} = 6.07$; $p < 0.0043$). The pattern of mean differences reveals outcrossing depression in litter size for some crosses. As previously reported, NZO/HILtJ x PWK/PhJ and NZO/HILtJ x CAST/EiJ matings were avoided due to exceedingly poor mating performance, and PWK/PhJ x 129/SvImJ gave rise to infertile males (34). The mean litter size in the crosses was lower than both parental means for the 129/SvImJ x NZO/HILtJ, but not for the reciprocal NZO/HILtJ x 129/SvImJ cross. The mean litter size was between the two progenitors for C57BL6/J x WSB/EiJ and its reciprocal, 129/SvImJ x PWK/PhJ and its reciprocal, CAST/EiJ x NOD/HILtJ, WSB/EiJ x A/J and WSB/EiJ 129/SvImJ. For all other crosses, the majority of which were crosses between

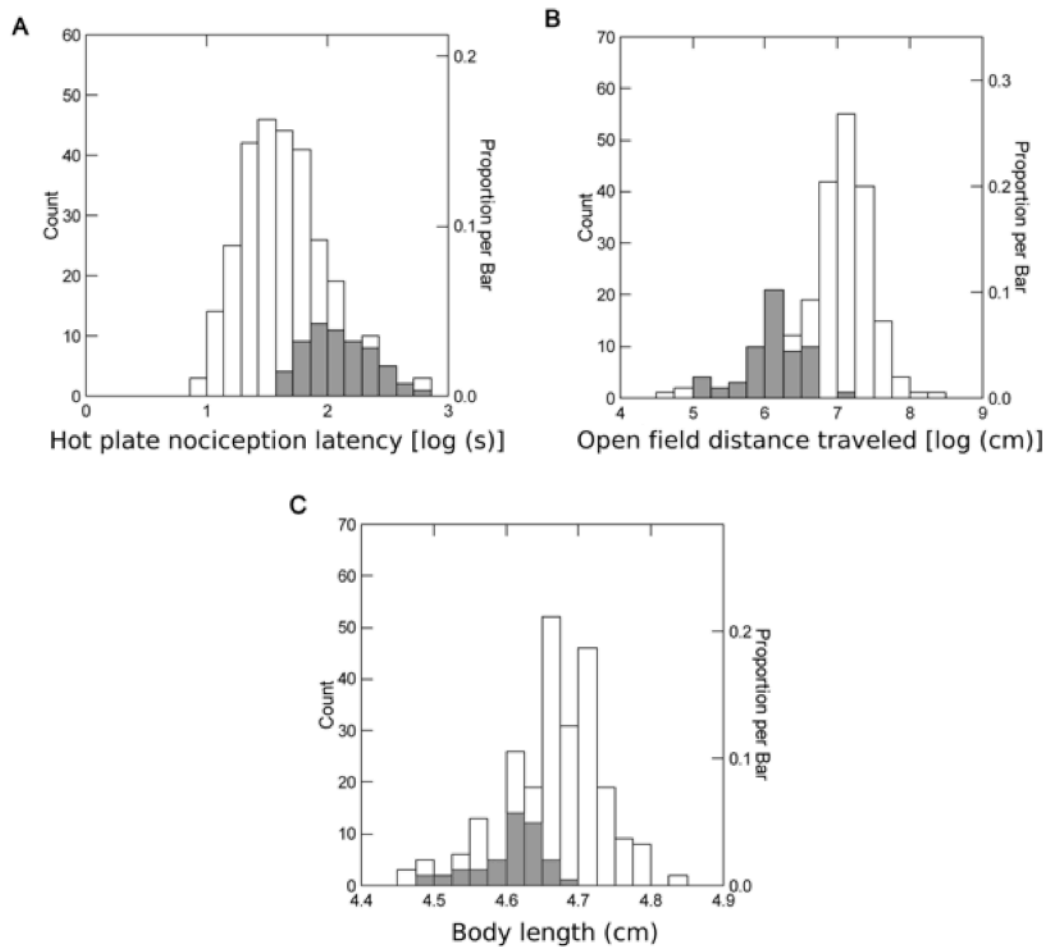


Figure 3-2. Trait variation in the CC compared to the BXD.

Shaded bars represent the phenotypic distribution of the BXD population (A) Distribution of thermal nociception across the CC and BXD populations. (B) Distribution of distance travelled in the open field across the CC (during the first 3 minutes) and BXD populations (during the first 5 minutes). (C) Distribution of body lengths across the CC and BXD populations (from (165)).

common laboratory strains, an increase in litter size was observed.

Among the G2 crosses, patterns of both depression and heterosis were again observed. G2 mean litter sizes were compared to G1 litter sizes for each of 576 four-way crosses (G1 x G1). Among these, 19 exhibited decreased litter size indicative of depression, 330 exhibited heterosis, and the remainder fell between their G1 parental means. All four-way crosses in the G2 generation were classified into 15 configurations based on whether the parents were offspring of common x common, common x wild-derived, or wild-derived x wild-derived strains. Analysis of change in litter sizes (δ_{LS}) from G1 to G2 reveal significant effects of configuration ($F(15,561) = 3.00$; $p < 0.0002$) and number of wild-derived lines in the four-way cross ($F(3,572) = 7.39$; $p < 0.0001$). With increasing numbers of wild-derived lines among the four grandparents, δ_{LS} is increasingly positive ($\delta_{LS}^0 < \delta_{LS}^1 < \delta_{LS}^2 < \delta_{LS}^3$).

Effects of genetic composition are dependent on the direction of the cross. When the common x common derived parent is female, the change in litter size from the previous generation is lower than for all other configurations ($F(1,561) = 6.29$; $p < 0.012$), and the litter size itself is higher. As a further test of the directionality of mating effects, we directly compared each comparable cross of common x common derived males or females to other configurations (e.g., [(common x common) x (wild x common)] vs. [(common x wild) x (common x common)]) which revealed significant effects only for the difference between CC x WW and WW x CC crosses such that wild derived males resulted in an increase

in litter size when crossed to a common background than do wild-derived females ($F(1,561) = 6.29$; $p < 0.012$). This was tested again by comparing the effect of crossing wild-derived males to that of common-derived males into common derived females, which again reveals a greater increase in litter size when the male is wild-derived ($F(1,561) = 5.04$ $p < 0.025$).

Fecundity and reproductive fitness among the advanced inbreeding lines typically varied outside the range of progenitors. Dispersion of CC fertility related phenotypes, obtained from G2:F5 and subsequent generations, exceeds the inbred progenitors for measures including mean litter size, parental age at first litter, and interval from mating to first litter (Kolmogorov-Smirnov $p < .05$). However, litter sizes in these lines cover a smaller range than the G2:F1 population (Figure 3-3). The shrinking range was unilateral such that the maximum litter size obtained among the lines in inbreeding generations was successively lower than that observed in the G2:F1, a phenomenon considered consistent with over-dominance related inbreeding depression.

Intercrossing, inbreeding and their effect on allele frequencies

Lines from G2:F1 and G2:F7 were genotyped to track changes in allele frequencies across generations. Progenitor alleles entered the population at a minimum minor allele frequency (MAF) of 0.125 for all autosomal locations and it is expected that allele frequencies remain stable throughout each generation and genome location. Efforts were made to balance alleles on the sex chromosomes during initial breeding (34). Expected X chromosome allele frequencies are

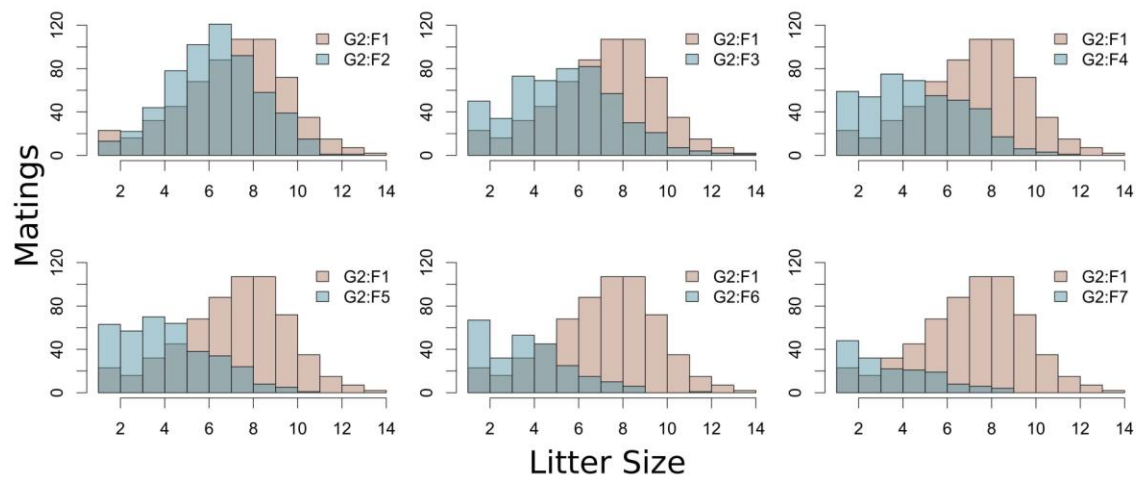


Figure 3-3. Litter size and inbreeding depression.

As a consequence of inbreeding, litter size diminishes at each successive generation of inbreeding. This reflects the phenomena of over-dominance as a mechanism of inbreeding (from (165)).

dependent on the position of the progenitor strains in each line's mating scheme and were adjusted accordingly. Distribution of strain specific alleles (MAF = 0.125) grouped by progenitor strain of origin indicates that there is no statistically significant deviation from expectation for the MAF at the majority of these loci using a χ^2 test for goodness of fit with a family-wise false discovery rate of 0.05. Only 95 strain-specific genotyped SNPs showed significant deviations when comparing progenitor allele frequencies to those of G2:F1 allele frequencies. Sixty-two were from PWK/PhJ, 15 of these were located on Chr 16 and 22 of which were on Chr X. Another 21 of these 95 loci were from CAST/EiJ, 18 were on Chr X. The remaining loci originated from WSB/EiJ (8 loci), A/J (1 locus), and C57BL/6J (3 loci). Genotyped loci from the three other progenitor strains, namely, 129S1/SvImJ, NZO/HILtJ and NOD/LtJ showed no significant deviation from expected segregation patterns. Comparison of G2:F1 strain specific alleles relative to those of G2:F7 resulted in only 13 loci with significant deviations from expected frequencies. Nine of these loci were from PWK/PhJ while the remaining four loci were from CAST/EiJ. Twelve loci show declining allele frequencies throughout both out-crossing and inbreeding. Of these, eight loci were from PWK/PhJ and four were from CAST/EiJ. The strain specific allele distributions spread from G2:F1 to G2:F7 indicating that overall alleles at some loci increase in frequency while other alleles decrease in frequency through the process of inbreeding (Figure 3-4). A comparison of observed local allele frequencies relative to expected frequencies in the G0 to G2:F1 and G2:F1 to G2:F7

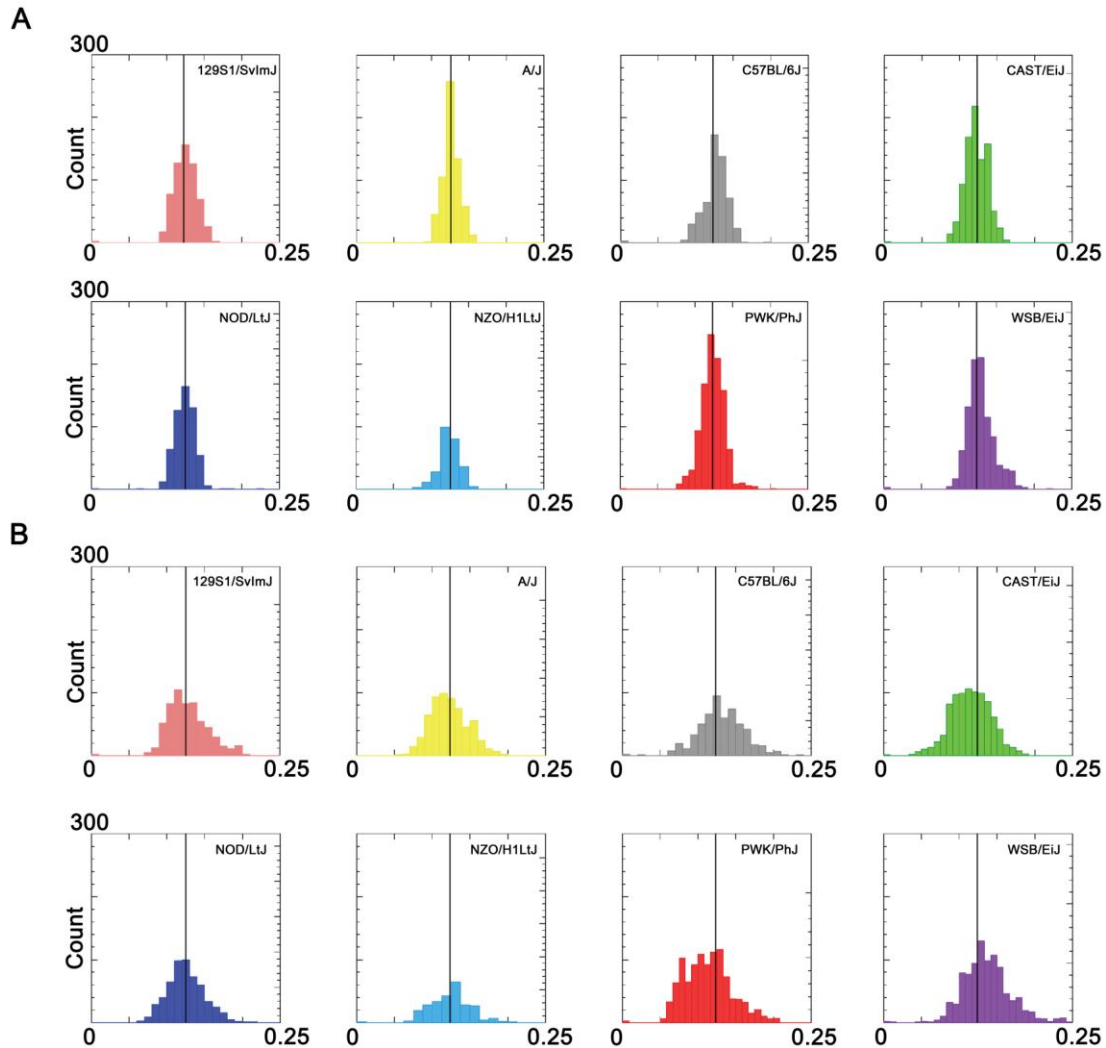


Figure 3-4. Changes in minor allele frequencies (MAF) due to inbreeding.

Strain specific MAF in the G2:F1 (A) generation depicts less variation than the G2:F7 (B) generation, with some spread of allele frequencies evident in the G2:F7 generation. Allele frequency distributions become asymmetrical for some strains. The distribution becomes right skewed PWK/PhJ indicating more loss than gain has occurred, and left skewed for WSB/EiJ (from (165)).

generations reveals that this increase in variance occurs on a genome wide level (Figure 3-5). In general, observed changes in allele frequencies relative to expected frequencies were less pronounced following out-crossing (Figure 3-5A) than during inbreeding (Figure 3-5B).

QTL mapping in the CC

A total of 102 reproductive, behavioral, physiological and morphological traits were subject to QTL mapping, though several of the measures were highly correlated, including 50 sleep related parameters and several similar blood measures. Approximately 72 unique, though not statistically independent phenotypes were considered. Among these, eight statistically significant ($p < 0.05$, Table 3-2) main effect QTLs were detected, including red cell distribution width, hotplate latency, average percent of sleep time in the dark, distance traveled in open field during the first three minutes, periosteal circumference of the femur, and peak activity onset after sleep deprivation, body length and cumulative distance from the center of the open field (Figure 3-6). An additional nine suggestive QTL were detected ($p < 0.10$, Table 3-2). Confidence intervals around the significant QTL peaks, estimated using 1.5-LOD drop-off, had an average width of 3.98 Mb and contained an average of 50 genes each, with fewer than 10 candidate coding genes for three of the traits. The narrowest 1.5 LOD drop interval detected is 530 Kbp for Peak activity time in hours from activity onset after sleep deprivation (Chr 9: 29.70 - 30.03 Mbp Build 36 between RefSNPs rs33767143 and rs6264816), which harbors coding sequence for a

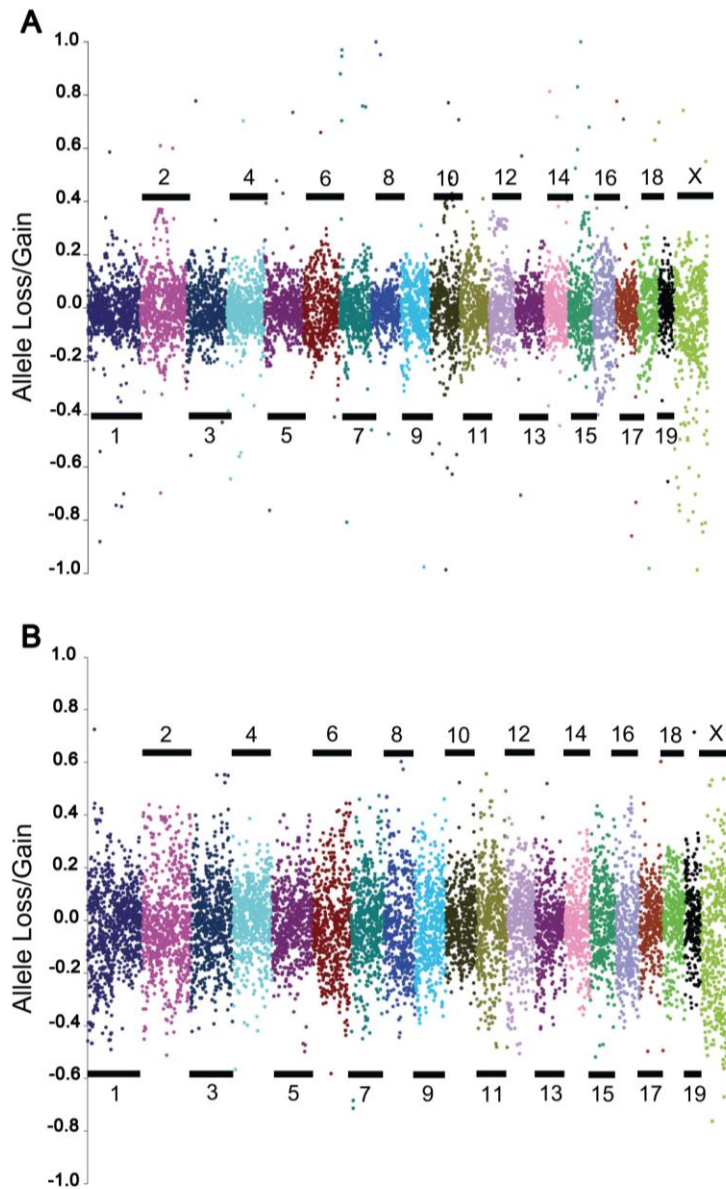


Figure 3-5. Allele loss and inbreeding.

Comparison of the percent allele loss between (A) progenitors (G0) and final out-cross generation (G2:F1). (B) G2:F1 and the seventh inbreeding generation (G2:F7). Positive values indicate SNPs with an increase in minor allele frequency, while negative values indicate a decrease (allele loss) in minor allele frequency. The y-axis indicates % change from G0 allele frequency (from (165)).

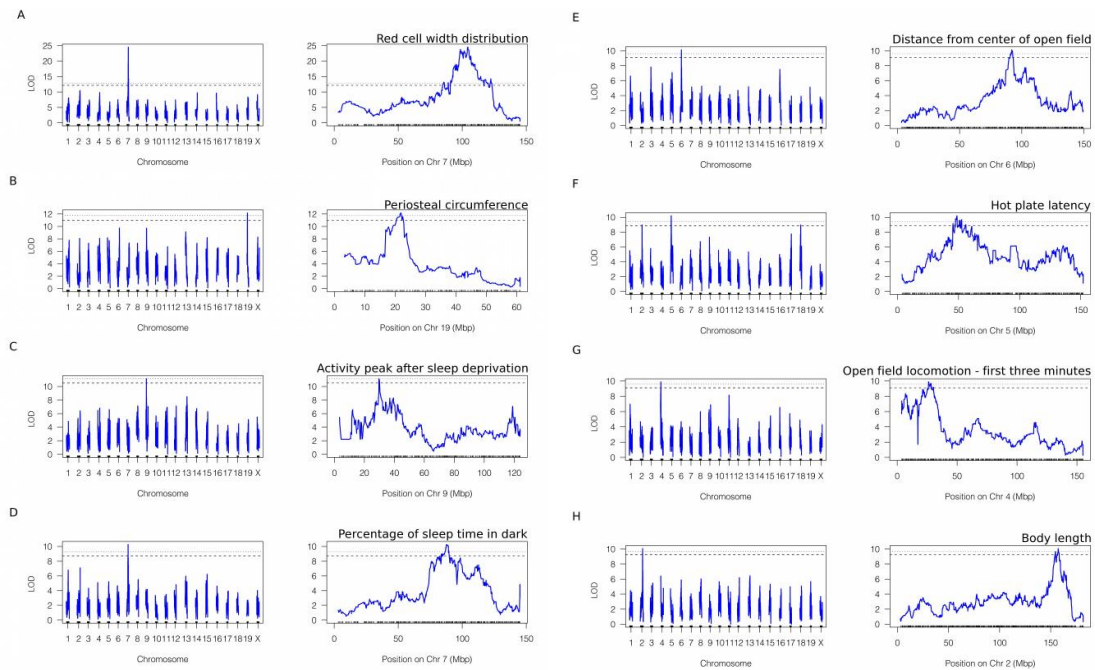


Figure 3-6. QTL mapping in the CC.

(A) Red blood cell width distribution. (B) Periosteal circumference. (C) Peak activity time in hours from dark onset after sleep deprivation. (D) Average percentage of sleep time over dark cycles for all baseline days. (E) Average minimum distance of the center of the mouse from the absolute center of the open field (cm). (F) Thermal nociception. (G) Open field locomotion in the first three minutes. (H) Body length. Horizontal lines indicate genome-wide significance thresholds based on 1,000 permutations. Dotted lines are genome-wide significant thresholds at $p \leq 0.05$; dashed lines indicate genome-wide suggestive thresholds at $p \leq 0.10$ (from (165))

Table 3-2. Genome-wide significant and suggestive QTLs, effect sizes and allele effects.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
SIGNIFICANT QTLs									
Body length	2	rs2733802 4	152.9- 158.5	70	6.93	106.24	129S1/SvJ mJ	104.61 $\pm$ 0.96	-1.63
							A/J	106.61 $\pm$ 1.07	0.37
							C57BL6/J	104.20 $\pm$ 0.97	-2.04
							CAST/EiJ	106.53 $\pm$ 1.02	0.29
							NOD/LtJ	106.22 $\pm$ 0.98	-0.02
							WSB/EiJ	112.20 $\pm$ 1.25	5.96
							NZO/HILtJ	106.80 $\pm$ 1.13	0.56
							PWK/PhJ	105.30 $\pm$ 1.01	-0.95
Average percent	7	rs6381934	90.9- 96.9	39	7.3	24.4	129S1/SvJ mJ	29.55 $\pm$ 1.39	5.15

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
sleep at night							A/J	20.98 $\pm$ 1.45	-3.42
							C57BL6/J	25.24 $\pm$ 1.18	0.84
							CAST/EiJ	24.55 $\pm$ 1.13	0.15
							NOD/LtJ	22.06 $\pm$ 1.23	-2.34
							WSB/EiJ	27.27 $\pm$ 1.61	2.87
							NZO/HILtJ	27.12 $\pm$ 1.76	2.72
							PWK/PhJ	20.84 $\pm$ 1.27	-3.56
Hot plate latency	5	rs1347823 8	45.5- 61.0	6	6.7	1.64	129S1/SvJ mJ	1.57 $\pm$ 0.05	-0.07
							A/J	1.63 $\pm$ 0.05	-0.01
							C57BL6/J	1.69 $\pm$ 0.05	0.05

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Red cell distribution width	7	rs31752108	105.5 - 112.5	221	25.99	14.07	CAST/EiJ	1.64 $\pm$ 0.05	0
							NOD/LtJ	1.54 $\pm$ 0.05	-0.1
							WSB/EiJ	1.60 $\pm$ 0.05	-0.04
							NZO/HILtJ	1.59 $\pm$ 0.04	-0.05
							PWK/PhJ	1.88 $\pm$ 0.05	0.24
							129S1/SvJ mJ	13.73 $\pm$ 0.20	-0.34
							A/J	13.69 $\pm$ 0.24	-0.38
							C57BL6/J	14.85 $\pm$ 0.18	0.78
							CAST/EiJ	13.64 $\pm$ 0.17	-0.43
							NOD/LtJ	14.64 $\pm$ 0.16	0.57

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Peak activity after sleep deprivation	9	rs3376714 3	29.7 - 30.2	2	7.77	1.96	WSB/EiJ	13.20 $\pm$ 0.26	-0.87
							NZO/HILtJ	15.00 $\pm$ 0.25	0.93
							PWK/PhJ	13.39 $\pm$ 0.20	-0.68
							129S1/Svl mJ	3.38 $\pm$ 0.56	1.42
							A/J	0.46 $\pm$ 0.69	-1.5
							C57BL6/J	1.15 $\pm$ 0.70	-0.81
							CAST/EiJ	-2.04 $\pm$ 1.01	-4.00
							NOD/LtJ	1.49 $\pm$ 0.82	-0.47
							WSB/EiJ	2.19 $\pm$ 0.72	0.23
							NZO/HILtJ	1.95 $\pm$ 0.85	-0.01

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Open field locomotion in the first 3 minutes	4	rs27723618	21.1 - 30.4	16	8.23	1232.67	PWK/PhJ	3.68 $\pm$ 0.62	1.72
							129S1/Svl mJ	1276.04 $\pm$ 122.84	43.37
							A/J	1210.18 $\pm$ 92.01	-22.48
							C57BL6/J	989.28 $\pm$ 94.34	-243.39
							CAST/EiJ	1621.75 $\pm$ 90.90	389.08
							NOD/LtJ	1159.43 $\pm$ 90.90	-73.23
							WSB/EiJ	1118.83 $\pm$ 76.70	-113.84
							NZO/HILtJ	1374.05 $\pm$ 101.04	141.38

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Integrated distance from center	6	rs4949582 9	89.6 - 93.2	45	9.54	5.6	PWK/PhJ	1202.3 6 $\pm$ 73.64	-30.31
							129S1/SvJ mJ	5.68 $\pm$ 0.47	0.08
							A/J	7.50 $\pm$ 0.64	1.9
							C57BL6/J	4.86 $\pm$ 0.50	-0.74
							CAST/EiJ	5.21 $\pm$ 0.57	-0.39
							NOD/LtJ	8.52 $\pm$ 0.75	2.92
							WSB/EiJ	5.01 $\pm$ 0.43	-0.59
							NZO/HILtJ	5.98 $\pm$ 0.68	0.38
							PWK/PhJ	4.31 $\pm$ 0.59	-1.29
Periosteal circumferen	19	rs6287522	21.5-22.5	5	11.6	4.87	129S1/SvJ mJ	4.78 $\pm$ 0.07	-0.09

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
ce							A/J	4.85 $\pm$ 0.07	-0.02
							C57BL6/J	4.95 $\pm$ 0.06	0.07
							CAST/EiJ	4.78 $\pm$ 0.08	-0.09
							NOD/LtJ	4.66 $\pm$ 0.07	-0.21
							WSB/EiJ	5.16 $\pm$ 0.08	0.29
							NZO/HILtJ	5.14 $\pm$ 0.09	0.27
							PWK/PhJ	4.78 $\pm$ 0.09	-0.1
SUGGESTIVE QTLS									
Litter size	6	rs6258737	93.9 - 128.3	291	7.24	4.45	129S1/Svl mJ	4.12 $\pm$ 0.37	-0.33
							A/J	5.01 $\pm$ 0.39	0.56
							C57BL6/J	3.97 $\pm$ 0.32	-0.48

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Gonadal fat pad weight	8	rs31954766	64.5-80.8	28	6.06	0.67	CAST/EiJ	3.80 $\pm$ 0.32	-0.64
							NOD/LtJ	3.93 $\pm$ 0.34	-0.52
							WSB/EiJ	3.61 $\pm$ 0.26	-0.83
							NZO/HILtJ	2.79 $\pm$ 0.34	-1.65
							PWK/PhJ	4.66 $\pm$ 0.31	0.21
							129S1/SvImJ	0.74 $\pm$ 0.10	-0.07
							A/J	0.58 $\pm$ 0.08	0.09
							C57BL6/J	0.60 $\pm$ 0.07	0.07
							CAST/EiJ	0.82 $\pm$ 0.10	-0.15
							NOD/LtJ	0.53 $\pm$ 0.07	0.13

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean ± S.E.M	Dev. from Mean
Body weight	2	rs27338024	154.0-159.5	55	5.55	27.27	WSB/EiJ	0.47 ± 0.09	0.2
							NZO/HILtJ	0.62 ± 0.06	0.04
							PWK/PhJ	0.98 ± 0.07	-0.32
							129S1/Svl mJ	27.94 ± 0.57	-0.67
							A/J	27.39 ± 0.72	-0.13
							C57BL6/J	25.72 ± 0.62	1.55
							CAST/EiJ	33.22 ± 2.97	-5.96
							NOD/LtJ	26.30 ± 0.86	0.97
							WSB/EiJ	24.99 ± 0.74	2.28
							NZO/HILtJ	27.83 ± 0.64	-0.56

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
95 Brain weight	7	rs6263981	36.3 - 52.7	121	12.1	0.43	PWK/PhJ	28.97 $\pm$ 0.58	-1.7
							129S1/Svl mJ	0.43 $\pm$ 0.01	0
							A/J	0.42 $\pm$ 0.01	-0.01
							C57BL6/J	0.44 $\pm$ 0.01	0.01
							CAST/EiJ	0.42 $\pm$ 0.01	-0.01
							NOD/LtJ	0.44 $\pm$ 0.01	0.01
							WSB/EiJ	0.45 $\pm$ 0.01	0.02
							NZO/HILtJ	0.39 $\pm$ 0.01	-0.04
							PWK/PhJ	0.43 $\pm$ 0.01	0.01
Tail clip latency	2	rs2708518 4	11.62-12.58	5	5.33	32.57	129S1/Svl mJ	41.76 $\pm$ 2.02	9.19

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
							A/J	26.42 $\pm$ 2.77	-6.15
							C57BL6/J	31.07 $\pm$ 2.77	-1.5
							CAST/EiJ	27.61 $\pm$ 2.27	-4.97
							NOD/LtJ	37.13 $\pm$ 2.46	4.56
							WSB/EiJ	35.52 $\pm$ 2.36	2.95
							NZO/HILtJ	25.24 $\pm$ 3.18	-7.33
							PWK/PhJ	26.16 $\pm$ 2.77	-6.41
Percentage of monocytes	17	rs6311776	27.6 - 42.5	41	9.92	10.23	129S1/SvI mJ	10.07 $\pm$ 0.19	-0.16
							A/J	11.14 $\pm$ 0.22	0.9
							C57BL6/J	10.01 $\pm$ 0.30	-0.23

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Percentage of eosinophils	17	rs33325845	12.7 - 22.5	147	9.08	11.23	CAST/EiJ	10.21 $\pm$ 0.23	-0.02
							NOD/LtJ	9.60 $\pm$ 0.28	-0.63
							WSB/EiJ	9.91 $\pm$ 0.21	-0.33
							NZO/HILtJ	10.61 $\pm$ 0.26	0.38
							PWK/PhJ	10.11 $\pm$ 0.28	-0.12
							129S1/SvImJ	10.96 $\pm$ 1.13	10.96
							A/J	10.27 $\pm$ 1.16	10.27
							C57BL6/J	12.10 $\pm$ 0.85	12.1
							CAST/EiJ	14.72 $\pm$ 0.91	14.72
							NOD/LtJ	11.33 $\pm$ 1.43	11.33

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
Mean corpuscular volume	11	rs28228790	69.2 - 82.9	276	15.18	46.38	WSB/EiJ	8.86 $\pm$ 0.91	8.86
							NZO/HILtJ	10.42 $\pm$ 1.24	10.42
							PWK/PhJ	9.95 $\pm$ 1.05	9.95
							129S1/Svl mJ	48.62 $\pm$ 0.49	2.24
							A/J	46.36 $\pm$ 0.76	-0.02
							C57BL6/J	47.05 $\pm$ 0.64	0.67
							CAST/EiJ	45.44 $\pm$ 0.43	-0.93
							NOD/LtJ	47.31 $\pm$ 0.47	0.93
							WSB/EiJ	46.45 $\pm$ 0.46	0.07
							NZO/HILtJ	44.29 $\pm$ 0.76	-2.09

Table 3-2 continued.

Phenotype	Chr	Peak Marker	1.5 LOD CI (Mbp)	# Pos. Cand.	Effect Size	Mean	Strain	Mean $\pm$ S.E.M	Dev. from Mean
66 Femur length	14	rs51336560	95.8-98.9	3	8.5	15.45	PWK/PhJ	44.50 $\pm$ 0.52	-1.88
							129S1/SvJ mJ	15.37 $\pm$ 0.13	-0.08
							A/J	15.48 $\pm$ 0.12	0.03
							C57BL6/J	15.73 $\pm$ 0.14	0.28
							CAST/EiJ	15.60 $\pm$ 0.16	0.15
							NOD/LtJ	15.69 $\pm$ 0.13	0.24
							WSB/EiJ	15.55 $\pm$ 0.11	0.1
							NZO/HILtJ	14.96 $\pm$ 0.19	-0.49
							PWK/PhJ	14.92 $\pm$ 0.14	-0.53

microRNA and two genes — *Ntm* (neurotrimin) and *Snx19*. The largest interval detected for a statistically significant locus was 6.96 Mbp for red blood cell distribution width containing 232 positional candidate genes.

Periosteal circumference was mapped to a 1.5-LOD interval from 21.55 - 22.51 Mbp on Chr19 (Build 36), containing five candidate genes, including *Tmem2* and *Trpm3*. For thermal nociception, the 1.5-LOD drop interval (Chr5: 45.44 – 60.99 Mbp) around the peak marker contains just six genes, *Slit2*, *Pacrgl*, *Gpr125*, *Dhx15*, *Sod3* and *Kcnip4*. The 1.5-LOD drop interval for average minimum distance from the center of the open-field (Chr6, 89.59 – 93.20 Mbp), average percentage of sleep time over dark cycles for all baseline days (Chr7, 90.92 – 96.94 Mbp) and red cell distribution width (Chr7: 105.55-112.52 Mbp) respectively map onto gene-rich genomic regions with 45, 39 and 221 genes, the latter containing a cluster of 130 Olfr family members in addition to three compelling functional candidates, hemoglobins *Hbb-b1*, *Hbb-bh1* and *Hbb-y*.

We also mapped litter size to identify potential loci that could influence litter size thereby reflecting inbreeding depression. A suggestive QTL was found on Chr 6 for the regulation of litter size. This locus harbors several genes for which alleles have been annotated to embryonic lethality. At this locus, A/J alleles are associated with higher litter sizes, and NZO with lower litter sizes. Interestingly, these alleles have significant decrease in frequency (rs13478994, $p < .003$; rs29829339, $p < .02$), suggesting that for any surviving line with an A/J allele at this location, litter size is elevated.

Significant QTLs for some traits are the effect of alleles from single strains, while for other traits the effect is due to alleles present among multiple strains (Table 3-2). For body length, individuals with alleles originating from WSB/EiJ at the QTL were significantly different from individuals with alleles derived from other strains. The QTL for hot plate thermal nociception was driven by significant differences among individuals with alleles derived from PWK/PhJ and relative to those with other alleles except for C57BL6/J. For gonadal fat pad mass, alleles specific to PWK/PhJ were significantly different from others alleles except for CAST/EiJ and 12S1/SvImJ. For peak activity onset after sleep deprivation, wild-derived strains PWK/PhJ and CAST/EiJ differ significantly from WSB/EiJ, CAST/EiJ is significantly different from 129S1/SvImJ, PWK/PhJ is significantly different from A/J and A/J is significantly different from 129SvImJ. For other measures, a more complex pattern of strain-specific effects emerges (Table 3-2). A composite interval mapping approach enabled the detection of additional loci for several phenotypes (Table 3-3). For example, a search for additional QTLs for open field locomotor behavior, conditioned on the Chr 4 locus reveals an additional locus on Chr 9 at rs33324954, from 108 – 113 Mbp, consistent with previously reported QTL for this phenotype (90).

Discussion

The need for advanced GRPs stems from the inherent drawbacks that plague existing GRPs. These drawbacks include limited genetic variation and limited numbers of strains, which ultimately makes the task of QTL detection and

Table 3-3. Additional loci detected using composite interval mapping.

Trait	Original QTL	Second Locus	refSNP	Mbp	CI	Two locus model LOD Score
Body length	Chr 2	Chr 13	rs6216118	99.3	75.23 - 101.45	7.15
Body weight	Chr 2	Chr 8	rs31787744	66.88	60.82 - 87.93	7.28
Brain weight	Chr 7	Chr 17	rs33558716	56.63	55.67 - 64.91	9.3
		Chr 1	rs6263216	177.76	173.30 - 182.73	8.94
		Chr 13	rs29513935	31.59	30.66 - 34.12	9.23
Eosonophil %	Chr 17	Chr 4	rs27479104	118.79	103.41 - 155.32	7.24
		Chr 19	rs30748908	57.83	57.34 - 61.14	6.68
Gonadal fat pad weight	Chr 8	Chr 9	rs6387702	46.82	38.47 - 41.22	6.83
					43.20 - 50.56	
					106.92 - 107.62	
Hot plate latency	Chr 5	Chr 18	rs6203819	52.98	50.29 - 55.21	8.69
Monocytes %	Chr 17	Chr 2	rs27184231	23.43	22.07 - 25.32	9.51
					32.12 - 34.11	

Table 3-3 continued.

Trait	Original QTL	Second Locus	refSNP	Mbp	CI	Two locus model
Open field distance from center	Chr 6	Chr 3	rs30002306	47.28	35.96 - 36.25	8.71
					42.48 - 47.45	
		Chr 16	rs37822351	21.37	14.19 - 30.95	8.11
Open field locomotion	Chr 4	Chr 9	rs33324954	111.42	108.93 - 113.57	7.77
% Sleep night on average	Chr 7	Chr 1	rs51956473	146.34	136.71 - 155.81	7.14
		Chr 8	rs6304230	119.88	119.09 - 124.88	6.44
Red cell width distribution	Chr 7	Chr 2	rs6226117	178.88	171.36 - 178.97	12.38
Tail clip latency	Chr 2	Chr 5	rs32966653	56.15	53.92 - 65.37	8.42
					72.53 - 75.66	
					76.14 - 79.62	
	Chr 15	Chr 15	rs33196993	37.34	11.60 - 12.27	7.55
					12.73 - 23.44	
					25.53 - 27.34	
					30.13 - 51.60	
Periosteal circ	Chr 19	Chr 13	rs46919382	19.04	4.68 - 6.92	9.68
					18.92 - 20.95	

candidate gene identification extremely labor intensive. The availability of advanced, large and diverse panels of GRPs will help alleviate these drawbacks, while at the same time serving as an invaluable resource to the biomedical research community. The collaborative cross (CC) is one such advanced GRP for mouse genetics. Similar panels have either been or are being developed across multiple species, including *D. melanogaster* (9), *A. thaliana* (108), *Z. Mays* (29) and *C. elegans* (99, 122, 176)

The completed CC RI strain panel is expected to have increased numbers of recombination and smaller haplotype blocks resulting in a fine-grained mosaic of the eight founder strains. Simulations performed by Karl Broman (25) have supported this expectation by showing that on average, eight-way RILs have 134 recombinant haplotype blocks compared to 85 in two-way RILs. As expected, the median lengths of these haplotype blocks are smaller in eight-way RILs (8.5 cM) compared to two-way RILs (12.9 cM). Comparing the emerging CC populations (pre-CC) to the above-simulated CC population has been revealed 6% more detectable recombination breakpoints and smaller median block or segment lengths (8). Additional simulations, performed by Valdar and colleagues (213), have assessed the power and precision with which one can map QTL in the CC. In an extensive series of simulations with several assumptions, Valdar and colleagues has shown that a QTL with 5% effect size can be mapped with moderate power (0.67) and high precision (mean location error of 1.14 cM) when employing 500 RILs. QTL analysis in the emerging CC reveals that with sample

size of approximately 250 lines with only one or two mice per line, it is possible to map genetic loci with high precision and moderate to large effect sizes. Among the significant loci detected, QTL effect sizes ranged from (6.7% to 26%), while the average 1.5 LOD support interval ranged from (530 Kb to 6.96 Mb) (Table 3-3). In several instances, the emerging CC had sufficient mapping resolution to identify one or a few candidate genes. Mapping of activity after sleep deprivation revealed only three positional candidates, a microRNA and two genes, *Ntm* (neurotrimin), which encodes a neural cell adhesion molecule that plays a role in brain development (31) and *Snx19*. Allelic variation in neurotrimin could affect sleep, wake and activity by influencing developmental patterns in central nervous system structure or through dynamic effects on adult neuronal functions. Sleep deprivation maximizes the magnitude of slow-wave activity (SWA), which is directly proportional to the duration of prior periods of wakefulness (6). Increasing evidence suggests a functional role for sleep and high SWA in synaptic plasticity (86, 110). Periosteal circumference at the mid-shaft of the femur mapped to five candidates, *Gda*, *1110059E24Rik*, *Tmem2*, and *Trpm3*, a transient receptor potential channel gene. Periosteal circumference is one of many measures for skeletal size and is correlated with total body size and total skeletal size, therefore the gene underlying this QTL could be responsible for global effects on body size. Mechanistic analysis of the causative locus is the best method to interpret the specificity of these effects. Neither *1110059E24Rik* nor *Fam108b* appear to be expressed in osteoblasts or osteoclasts. The *Tmem2* gene is highly

expressed in primary osteoblasts, but the function of this gene in bone biology is unknown. The *Gda* gene appears to be moderately expressed in osteoclasts, but like *Tmem2*, the function of this gene in bone biology has not been previously determined. Expression of *Trpm3* has been demonstrated in human osteoblast-like cell lines and other genes in this family have a known role in osteoblast differentiation (4). In addition to increased recombinations, improved QTL mapping power and resolution evident in the emerging CC, the out-crossing and inbreeding process has had minimal effect on founder strain contribution. Founder strains entered the populations at a minor allele frequency of 0.125 and a genome-wide analysis in the emerging CC has revealed equal founder strain contribution at all loci. This is important since this ensures that the breeding scheme is truly randomized, balanced and that no single founder strain was being selected for or against.

Genetic analysis in the emerging CC provides valuable insights in to what is to be expected from the completed CC. The ability to have large numbers of independent, phenotypically and genotypically diverse CC RILs will facilitate high precision QTL mapping and subsequent QTG discovery. Furthermore, the high amount of genetic variation captured in the CC will make it a better suited to model complex human disorders thereby facilitating translation research.

CHAPTER IV: INTEGRATIVE FUNCTIONAL GENOMICS AND QTG IDENTIFICATION

Abstract

Despite the high heritability of addiction, relatively few underlying genes or quantitative trait genes (QTG) have been identified. Recent studies have used the integration and convergence of evidence among functional genomics studies as a successful strategy towards the efficient and rapid nomination of QTGs. GeneWeaver (www.geneweaver.org) is a web resource that facilitates integrative functional genomic analyses by providing comprehensive data sources and analytical tools. While GeneWeaver facilitates a multitude of genomic analysis, one specific application of it is the identification and prioritization of QTGs. Here, we apply this web resource towards the identification a highly likely QTG for cocaine-induced locomotor activation. Cocaine-induced locomotor activation was previously mapped in the recently expanded panel of BXD RI strains to a region on Chr 4 and Chr 15, with over 100 and over 70 positional candidate genes in their respective 1.5 LOD drop support intervals. Combinatorial analysis using GeneWeaver resulted in the identification of *Rab3b*, based on the following its co-occurrence with previously published QTL mapping studies of cocaine related behaviors, methamphetamine related climbing and cocaine and amphetamine regulated transcript (CART) abundance; convergent evidence from other substances of abuse; polymorphisms albeit upstream and intronic; and expression pattern in reward pathways. Preliminary functional evaluation of *Rab3*

complex in cocaine response was performed using mixed background *Rab3bcd* triple null mice (B6;129 *Rab3bcd*^{-/-}). *Rab3bcd*^{-/-} mice have increased cocaine-induced activation compared to *Rab3bcd*^{+/-} and C57BL/6J mice, further implicating *Rab3b* in acute cocaine response. In conclusion, integrative functional genomics, performed here within the framework of GeneWeaver, facilitates efficient *in silico* reduction of a QTL interval with numerous positional candidate genes to a handful of testable candidate QTGs.

Introduction

Advances in genotyping and phenotyping technologies have led to mapping of complex traits at an ever-increasing pace across multiple species. To date, over 4,600 published mouse Quantitative Trait Loci (QTL) exist, of which less than one percent have been successfully identified at the gene and/or mutation level. Of these 4,600 QTL, approximately 250 are alcoholism and addiction related. Among all the published QTLs, only 20 genes have been accepted as QTG (138). The rate-limiting step in the identification of QTGs is not in QTL mapping but rather in the identification of the causative variant and validation of a gene(s) as the QTG. As shown in Chapters II and III, QTL confidence intervals are influenced by a number of factors, including the mapping population, sample sizes and trait heritability. This results in a large number of positional candidate genes within the QTL confidence interval, making the task of gene identification formidable. The emergence of advanced mapping panels, such as the collaborative cross (CC) and the diversity outbred (DO), will help

achieve greater precision in QTL mapping. However, even in these advanced mapping populations, gene level resolution using QTL mapping is still not attainable for most traits, requiring the utilization of multiple approaches for the identification of causal loci and the nominations of QTGs.

The QTL to QTG approaches that have been proposed to date can be categorized as genetic and bioinformatic approaches. Genetic approaches facilitate QTG identification by fine-mapping of the original QTL by taking advantage of meiotic recombination. Strategies falling in this category include chromosome substitution strains, congenic strain pairs, interval specific congenics and fine mapping using advanced intercross lines (AIL) or heterogeneous stock (HS), recombinant inbred segregation test (RIST) and Yin-Yang crosses (50, 74, 188). Genetic approaches often take years and large numbers of resources to fine map and confirm QTLs. For example, in a BXD recombinant inbred (RI) strain panel of 100 strains, mapping resolution or the 95% confidence interval (CI) of a QTL with a 5% effect size is approximately 7 cM (74, 213). A 7 cM interval corresponds to approximately a 14 Mb genomic region. Given the average gene density of 10 genes per Mb in the mouse, one is still left with approximately 140 positional or QTL candidate genes for further validation. This problem is alleviated to a certain extent in the CC mapping population. QTL mapping results in the emerging CC population have been promising. The average 1.5 LOD support interval was 4.80 Mb with approximately 50 positional candidate genes (165). While 50 is a considerable

reduction from 140 positional candidate genes, the necessity of additional fine mapping remains. Using only genetic approaches one could easily take six or more years to fine-map and identify the QTG (74). In an attempt to accelerate QTG identification, complementary bioinformatic approaches have been proposed, including comparative genomics, combined cross analysis and sequence and expression analysis (57). These approaches help facilitate the *in silico* reduction of a QTL interval and the identification of highly likely QTGs via the use of publicly available genotypes and expression databases.

Examples of successful identification of QTGs related to addiction using genetic approaches include *Mpdz* (189) and *Kcnj9* (109). In these and all QTGs identified to date, a stepwise process was utilized in an effort to eliminate all other possibilities. The process involved fine mapping, comparing sequence and expression levels of genes across the progenitor strains, and *in vivo* confirmation. Bioinformatic approaches on the other hand, although meant to accelerate QTG identification, have yet to catch up with the large numbers of unresolved QTL. Whether one uses purely genetic approaches or combines genetic and bioinformatic approaches for the identification of QTGs, integration among diverse genomic data sets is often performed. Furthermore, translating findings from mouse to human populations is of utmost importance, so as to achieve additional progress in understanding complex human disorders.

The strategy of integrative functional genomics, complementary to genetic and bioinformatics approaches, has been successfully applied towards the

nomination of highly likely candidate genes for alcoholism, bipolar disorder and drug addiction (117). However, existing integrative functional genomics methodologies are tailored towards the needs of a specific laboratory and lack computationally efficient infrastructure in the form of databases, algorithms and user interface that facilitates genome scale integrative analyses in a seamless and scalable manner. Here, a web-based software system called GeneWeaver (www.geneweaver.org) (11) is described that facilitates convergent analyses across multiple genomic experiments, data sources, platforms and species (11). Using this system, we demonstrate the integration of empirically derived QTL mapping results with other relevant genomic data sets, with the aim of nominating highly likely QTG.

The QTL mapping results described here are for cocaine-induced locomotor activation in the recently expanded BXD RI genetic reference population (164). Drug (here, cocaine) induced locomotor activation is an animal model focused on a very specific and important characteristic of human drug abuse, i.e., initial sensitivity to drugs. Furthermore, given the heritable and complex nature of cocaine-induced locomotor activation, it is readily amenable to the approach of QTL mapping and has been mapped previously in multiple studies (20, 82, 100). To date, 13 chromosomes and multiple chromosomal regions have been identified. However, a cocaine-induced activation QTG is yet to be discovered. Here, using these previously published results in conjunction

with relevant genomic data sets within the framework of GeneWeaver (12), a putative QTG is nominated.

Once a QTG has been nominated, confirmation of the gene as the QTG necessitates *in vivo* assessment of the gene. Multiple approaches exist for the *in vivo* assessment of QTGs, including targeted gene knockouts, transgenic overexpression and rescue and viral mediated RNA interference. Of these, the targeted gene knockout approach is utilized here. The targeted gene knockout approach assesses the role of the QTG by nullifying its function in an otherwise normal animal. The detection of significant differences in the trait of interest in a comparison of null to control animals enables further validation and serves as evidence of the gene as the QTG.

Methods

QTL mapping in the BXD RI

QTL mapping for cocaine-induced locomotor activation after the second intraperitoneal (*ip*) injection of 10 mg/kg of cocaine was performed in combined male and female, male only and female only BXD RI strains using GeneNetwork (<http://www.GeneNetwork.org>) and R/QTL (33, 219). R/QTL was utilized for additional QTL model fitting and to ensure reproducibility of QTLs mapped in GeneNetwork.org. One thousand permutations were performed to obtain genome-wide suggestive and significant thresholds. Ninety-five percent

confidence intervals (CI) were obtained by extracting 1.5-LOD support interval around the peak marker for statistically suggestive and/or significant QTLs.

Integrative functional genomic analysis of the candidate genes

Positional candidate genes residing within the 1.5 LOD drop support interval of the Chr 4 and Chr 15 QTLs were deposited to GeneWeaver (www.geneweaver.org). These gene lists, called “gene sets” in GeneWeaver, were integrated with a core collection of gene sets associated with cocaine, methamphetamine and nicotine. This core collection consists of empirically derived gene sets associated with mammalian phenotype (MP) and gene ontology (GO) annotations, previously published QTLs, gene expression studies obtained from either the Neuroscience Information Framework (<http://www.neuinfo.org>) or previous publications, striatal and nucleus accumbens basal gene expression obtained from Allen Brain Atlas (<http://www.brain-map.org>), polymorphic genes among the BXD RI strain panel and lastly, BXD gene expression correlates for striatal and nucleus accumbens to locomotor response after the second *ip* injection of 10 mg/kg of cocaine using GeneNetwork (<http://www.GeneNetwork.org>). The complete list of gene sets used in this analysis is provided in Appendix Table A1.

GeneWeaver’s Phenome Map tool was utilized for the integrative analysis of each of the two candidate gene sets with the above defined core collection of gene sets (12). The result obtained using Phenome Map is a hierarchical representation of gene-to-gene set interactions across the diverse collection of

gene sets, such that gene(s) shared across a greater number gene sets are placed at the top of this map. The Phenome Map tool also enables the generation of a gene-to-gene set association matrix that can be used to calculate gene membership counts. The gene membership counts can be used to rank each list of positional candidates to obtain a ranked list (or prioritized list) of positional candidates for functional validation.

Assessment of Rab3b polymorphisms

The Variant Brower interface in GeneNetwork (33, 36, 219) was queried for the presence of SNPs among the C57BL/6J (B6) and DBA/2J (D2) strains for *Rab3b*. Once locations for SNPs were retrieved, Integrative Genomics Viewer 2.0 was used for the extraction of genomic regions for transcription factor binding site and assessment of cryptic splice sites. The Jackson Laboratory's custom C57BL/6J ("Eve") and Sanger DBA/2J bam files (105) were imported in to IGV. For the prediction of transcription factor binding sites, 1Kb upstream of *Rab3b* was used. The B6 and D2 sequences were analyzed for putative TFBS using AliBaba 2.1 (80). Putative TFBS were then scanned for the presence of the two upstream polymorphisms. For the assessment of cryptic splice sites, nucleotides ± 19 bp upstream and downstream of the SNP location (for B6 and D2) were extracted and compared against consensus splice donor, branch and acceptor sites. Figure 4-1 depicts the consensus eukaryotic sequence logos, obtained from enoLOGOS (227) used in this analysis.

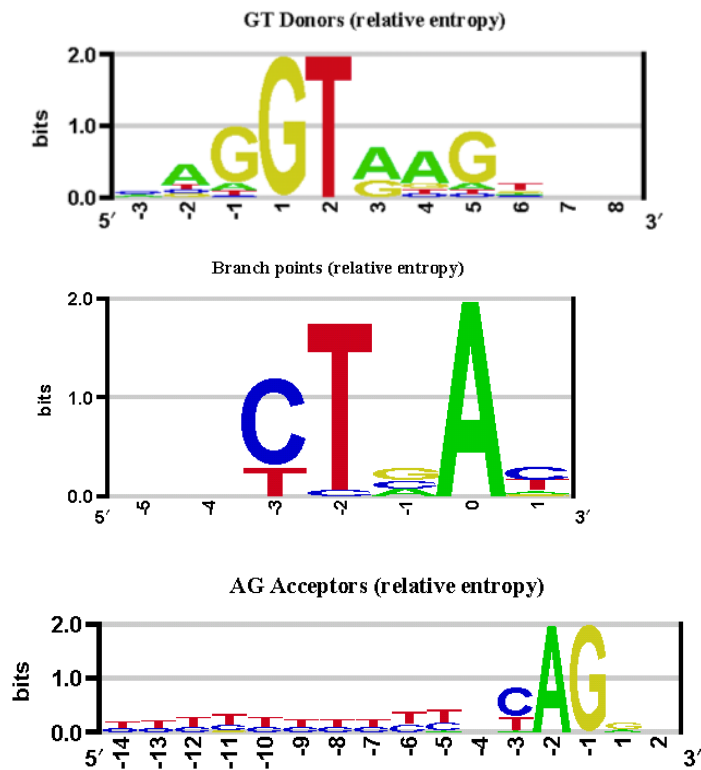


Figure 4-1. Donor, branch and acceptor consensus sequences (227).

Consensus sequences for the determination of cryptic splice sites. GT, A and AG are the most frequently seen dinucleotides at donor, branch and acceptor sequences, respectively.

Animals

Generation of mice with targeted deletions in *Rab3a*; *Rab3b*; *Rab3c* and/or *Rab3d* function has been previously described (184). *Rab3a* heterozygous and *Rab3bcd* triple null mice were obtained from The Jackson Laboratory (Strain Name: B6;129-*Rab3b*^{tm1Sud} *Rab3a*^{tm1Sud} *Rab3d*^{tm1Rja} *Rab3c*^{tm1Sud}/J ; Stock Number: 006375) and bred to produce obligatory *Rab3bcd* triple heterozygous (*Rab3bcd*^{+/-}) and triple null (*Rab3bcd*^{-/-}) offspring. Ninety-six mice belonging to three genotype (C57BL/6J, *Rab3bcd*^{+/-} and *Rab3bcd*^{-/-}) and both sexes were tested. Age and sex matched mice were randomly assigned to the following groups: saline (SAL) and cocaine (COCA). Mice were between 56 and 132 days of age at the start of behavioral testing. All mice were maintained at The Jackson Laboratory in climate- and 12 h light cycle-controlled rooms and provided acidified water and 5K52 chow (LabDiet/PMI Nutrition, St. Louis, MO) ad libitum. The Jackson Laboratory Animal Care and Use Committee approved all protocols involving mice.

Breeding and genotyping

Mutant mice, B6;129-*Rab3b*^{tm1Sud} *Rab3d*^{tm1Rja} *Rab3c*^{tm1Sud}/J carrying targeted mutations in *Rab3b*, *Rab3c* and *Rab3d* genes were acquired from the Jackson Laboratory Repository (JR#006375). Production of these mice has been described by Schluter and colleagues (184). Briefly, all genes were targeted using a neomycin resistance gene to replace exons encoding either amino acids

1-76 of *Rab3b* or exons encoding amino acids 9-84 of *Rab3c* or *Rab3d*. The constructs were independently electroporated into 129P2/OlaHsd ES cells. Chimeric mice were mated to C57BL/6J mice and the compound mutants created by intercrossing individual lines. The strain was backcrossed 2 or 3 times to C57BL/6J before being deposited in the repository. Homozygous triple null mice were received and maintained by sibling crosses. Heterozygous triple nulls were produced by crossing to C57BL/6J.

In order to isolate the *Rab3b* null allele from the *Rab3c* and *Rab3d* knockout alleles (and produce B6.129P2-*Rab3b*^{tm1sud} mice), the triple mutants were additionally backcrossed to C57BL/6J for 3 or 4 times, selecting for the *Rab3b* heterozygotes and *Rab3c* and *Rab3d* wild-type loci at each generation. Experimental *Rab3b*^{-/-}, *Rab3b*^{+/-} and *Rab3b*^{+/+} mice were generated by intercrossing *Rab3b*^{+/-} mice. Progeny were produced in the expected 1:2:1 ratio. All mice were genotyped for the presence of the wild-type or mutant *Rab3b*, *Rab3c* or *Rab3d* locus by PCR (184) on genomic DNA isolated from peripheral blood collected by retro-orbital plexus or submandibular bleeding.

Behavioral analysis

Assessment of acute and sensitized response to cocaine-induced locomotor activity

Prior to testing, mice from all three genotypes (C57BL/6J, *Rab3bcd*^{+/-} and *Rab3bcd*^{-/-}) were randomly assigned either to a saline (SAL) or cocaine (COCA)

group. Mice were brought from their home room into the test room and allowed to habituate for one hour. Mice were administered either saline or cocaine (10 mg/kg) by *ip* injection. Immediately post injection, mice were placed in an open field box and locomotor activity was recorded for a duration of 20 minutes using the EthoVision XT 8 video tracking and analysis system (Noldus Information Technology, Wageningen, The Netherlands). On Days 1 and 2 of testing all mice received saline. Day 1 measures locomotor activity in response to injection stress and novel environment, while Day 2 measures locomotor activity within a familiar environment and attenuated injection stress. On Days 3, 5, 7 and 9, mice received their assigned saline or cocaine injection *ip*.

Assessment of habituation and its effect on cocaine induced locomotor activity.

Prior to testing mice from all three genotypes (C57BL/6J, *Rab3bcd^{+/-}* and *Rab3bcd^{-/-}*) were randomly assigned to one of three groups – Group A, B or C. Group A received saline on all three consecutive days of testing. Group B received saline on Days 1 and 2 and cocaine (10 mg/kg by *ip*) on Day 3. Group C received cocaine (10 mg/kg by *ip*) on Day 1 only. Group A serves as the baseline group, while Group B facilitates the assessment of cocaine induced locomotor activation following two days of habituation. Group C enables the assessment of cocaine induced locomotor activation in a novel environment, as mice belonging to this group have not been exposed to test apparatus. Immediately post injection, mice were placed in an open field box and locomotor activity was

recorded for a duration of 20 minutes using the EthoVision XT 8 video tracking and analysis system (Noldus Information Technology, Wageningen, The Netherlands).

Tissue retrieval

All mice across the three genotypes (C57BL/6J, *Rab3bcd*^{+/-} and *Rab3bcd*^{-/-}) subjected to behavioral testing were sacrificed via decapitation 24 – 48 hrs following the last behavioral trial. In addition to these three genotypes, a naïve group of mice representing C57BL/6J (n = 8), DBA/2J (n = 8) and *Rab3bcd*^{-/-} (n = 8) were also sacrificed via decapitation in an attempt to obtain baseline levels of gene expression for *Rab3b* (C57BL/6J and DBA/2J) and *Cartpt* (C57BL/6J and *Rab3bcd*^{-/-}). Striatal regions were extracted and total striatal RNA was isolated. Striatal RNA from naïve C57BL/6J (n = 8) was used in the analysis of *Rab3b* transcript abundance and *Cartpt* expression. Striatal RNA from naïve DBA/2J (n = 8) was used in the analysis of *Rab3b* transcript lengths. Striatal RNA from naïve *Rab3bcd*^{-/-} was used in the analysis *Cartpt* expression

Statistical analysis

Statistical analyses were performed in JMP 9.0 (SAS Institute Inc., Raleigh, NC). For the assessment of intra-session habituation, for days 1 and 2 data were binned into five-minute epochs and activity change scores were calculated using the following formula: activity during the initial five minutes / [activity during the initial five minutes + activity during the final five minutes]

(119). Mean activity change scores across all three genotypes for each of the two days were calculated. These scores were compared with an activity change score of 0.5 (representing no habituation). Significantly different and lower mean activity change scores from 0.5 are representative of normal intra-session habituation. For acute response to cocaine, total distance travelled in 20 minutes (TD20) on Day 2 was subtracted from TD20 on Day 3 was analyzed using a one-way analysis of variance (ANOVA), with genotype as the only factor. Similarly, for assessment of sensitized response to cocaine, TD20 on Day 3 was subtracted from TD20 on Day 9 and analyzed using a one-way ANOVA. Following each ANOVA, planned contrast statements were utilized to test for significant differences among the levels of genotype. To assess locomotor response following repeated exposure to cocaine, a repeated measures ANOVA was performed. For this analysis day was the within-subject variable and genotype as the between-subjects variable.

qRT-PCR

Briefly, tissue samples were stored in RNAlater (Life Technologies, Carlsbad, CA) per manufacturer's instructions and homogenized in TRIzol (Life Technologies, Carlsbad, CA). Total RNA was isolated by standard TRIzol® Plus methods according to manufacturer's methods including on the column DNase digestion. The quality of the isolated RNA was assessed using an Agilent 2100 Bioanalyzer instrument and RNA 6000 Nano LabChip assay. 500ng of total RNA

was then reverse transcribed with random decamers and M-MLV RT using the Message Sensor RT Kit (Life Technologies, Carlsbad, CA).

Results

QTL mapping

Cocaine-induced locomotor activation (second *ip* injection of 10 mg/kg of cocaine) was previously mapped separately in BXD RI males and females. No QTL were detected in the males. In the females, on the other hand, suggestive QTL were detected on Chr 4 and Chr 15 respectively ($p < 0.63$, Figure 4-2). The two QTL additive model involving Chr 4 (rs13477919 @ 104.96 Mb) and Chr 15 (rs13482528 @ 38.82 Mb) accounted for 34.40 % of the trait variance. Taken independently, the Chr 4 and Chr 15 QTLs accounted for 11.71 % and 15.65 % of the trait variance respectively. No significant interaction among these loci was detected. Furthermore, at each of these loci the DBA2/J (D2) allele is the increasor allele (Figure 4-3A,B). A genome-wide pair scan was also performed, resulting in the detection of a significant Chr 4 (rs13477881 @ 103.31) x Chr 5 (rs3723202 @ 91.57) interaction (Figure 4-3C, $p < 0.001$). Inclusion of this interaction term increases the % variance accounted for by the model to 55.40%. 1.5 LOD support intervals were used to define the 95% CIs. The Chr 4 (COCIA Chr4) CI spans 101.10 (rs13477873) – 114.51 (gnf04.110.583) Mb, while the Chr 15 (COCIA Chr15) CI spans 33.43 (rs8267966) – 45.11 (rs13482547) Mb (Figure 4-4). Positional candidate genes residing in each these regions were obtained by

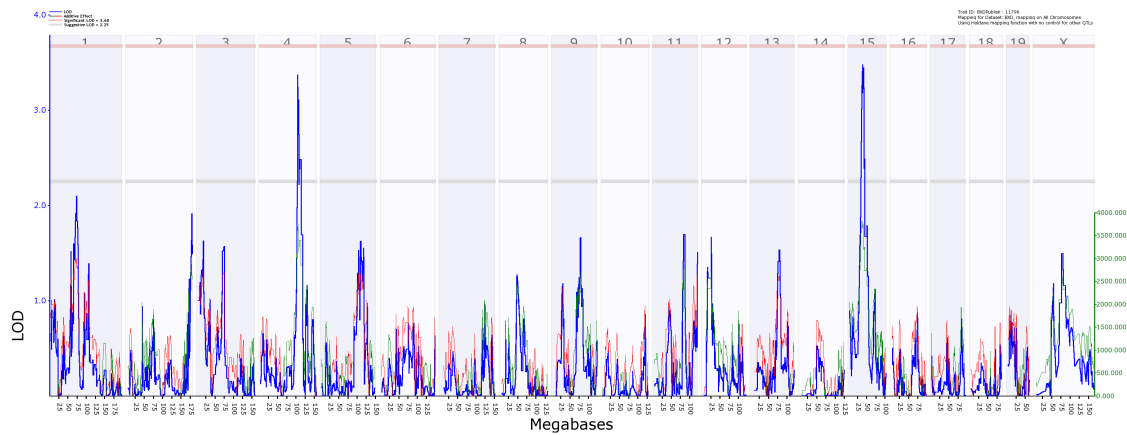


Figure 4-2. QTL Mapping of cocaine-induced locomotor activation in the expanded BXD RI panel.

QTL mapping of locomotor response following the second injection of cocaine in females resulted in the detection of suggestive QTLs on Chr 4 and Chr 15. Green lines indicate that DBA/2J is the increaser allele, while red lines indicate that C57BL/6J is the increaser allele.

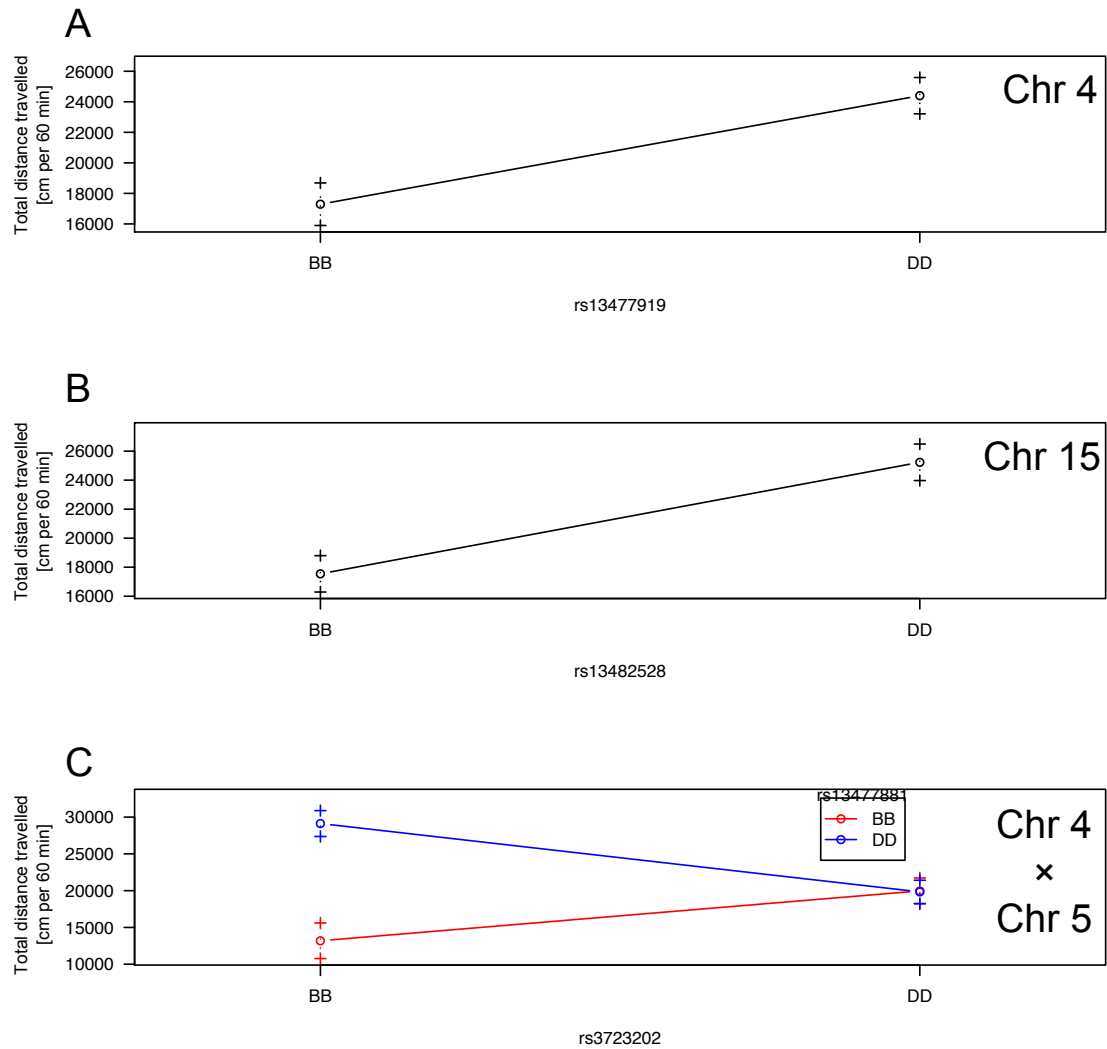


Figure 4-3. Main and interaction effect plots.

A) Effect plot for the peak marker on Chr 4 (rs13477919 @ 104.96 Mb) indicates a significant effect of genotype ($p < 0.0003$). Furthermore at this locus the D2-derived allele is the increasor allele. B) Effect plot for the peak marker on Chr 15 (rs13482528 @ 38.82 Mb) indicates a significant effect of genotype ($p < 0.0001$). C) Interaction plot for the significant Chr 4 (rs13477881 @ 103.31 Mb) x Chr 5 (rs3723202 @ 91.67 Mb) ($p < 0.0001$). There is a significant difference in the phenotypic values when the genotypes at rs3723202 and rs13482528 are BB, BB and BB, DD, respectively.

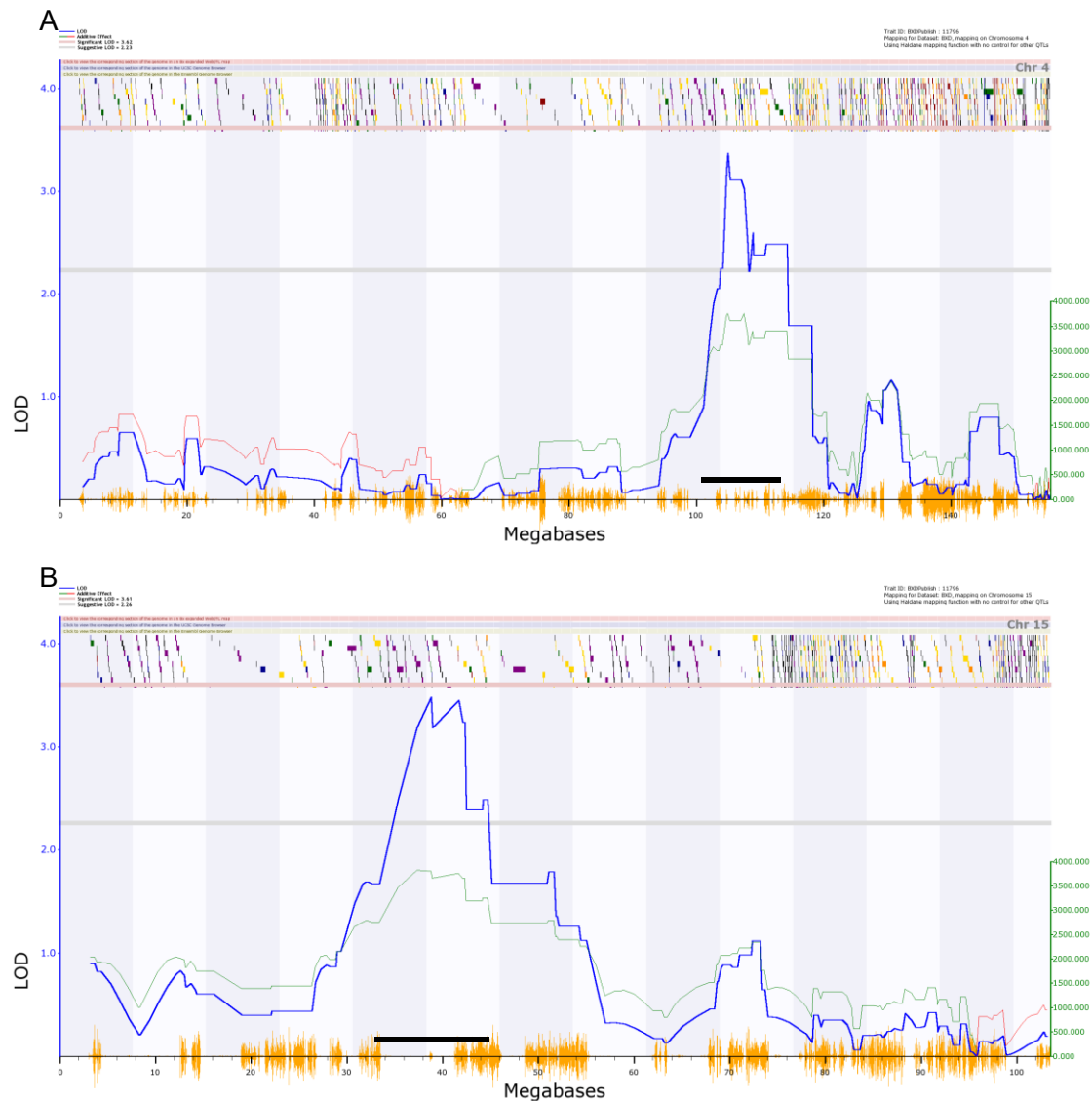


Figure 4-4. Chromosome 4 and Chromosome 15 QTL LOD support intervals.

The black horizontal lines represent the 1.5 LOD support interval on Chr 4 (A) and Chr 15 (B). The LOD support interval on Chr 4 spans 101.10 (rs13477873) – 114.51 (gnf04.110.583) Mb, while on Chr 15 it spans 33.43 (rs8267966) – 45.11 (rs13482547) Mb. 108 and 61 positional protein coding candidate genes reside in the Chr 4 and Chr 15 intervals, respectively.

searching for genes underlying the flanking marker coordinates in Mouse Genome Informatics (MGI; <http://informatics.jax.org>). Each gene list was uploaded to GeneWeaver.org (11) for integrative functional genomic analysis.

Integrative functional genomic analysis reveals *Rab3b* as a putative QTG acute cocaine response

A term search involving keywords pertaining to cocaine, methamphetamine and nicotine was performed in GeneWeaver (11) followed by a closer inspection of the search results, gave a total of 122 gene sets with an average of 72.41 genes per gene set spanning 21 publications. The Phenome Map analysis tool in GeneWeaver was used to construct an empirically derived hierarchical representation of the input gene sets for both the Chr 4 and Chr 15 candidate gene list. A ranked list of genes was obtained using the gene-to-gene set connectivity metric. *Rab3b* was the most highly ranked gene (gene-to-gene set connectivity = 7) across Chr 4 (Figure 4-5) and Chr 15. The most highly connected gene on Chr 15 in this analysis has a gene-to-gene set connectivity of four. Additional analysis of *Rab3b* reveals that *Rab3b* is polymorphic among the BXD RI panel with 46 intronic polymorphisms, specifically within introns 1 and 2 (Table 4-1). Additionally, a search for polymorphisms less than 1Kb upstream from *Rab3b* revealed the presence of two upstream polymorphisms. These upstream polymorphisms were investigated further for the possibility of serving as transcription factor binding sites (Table 4-1). Polymorphism at or near transcription factor binding sites could potentially alter transcription regulation.

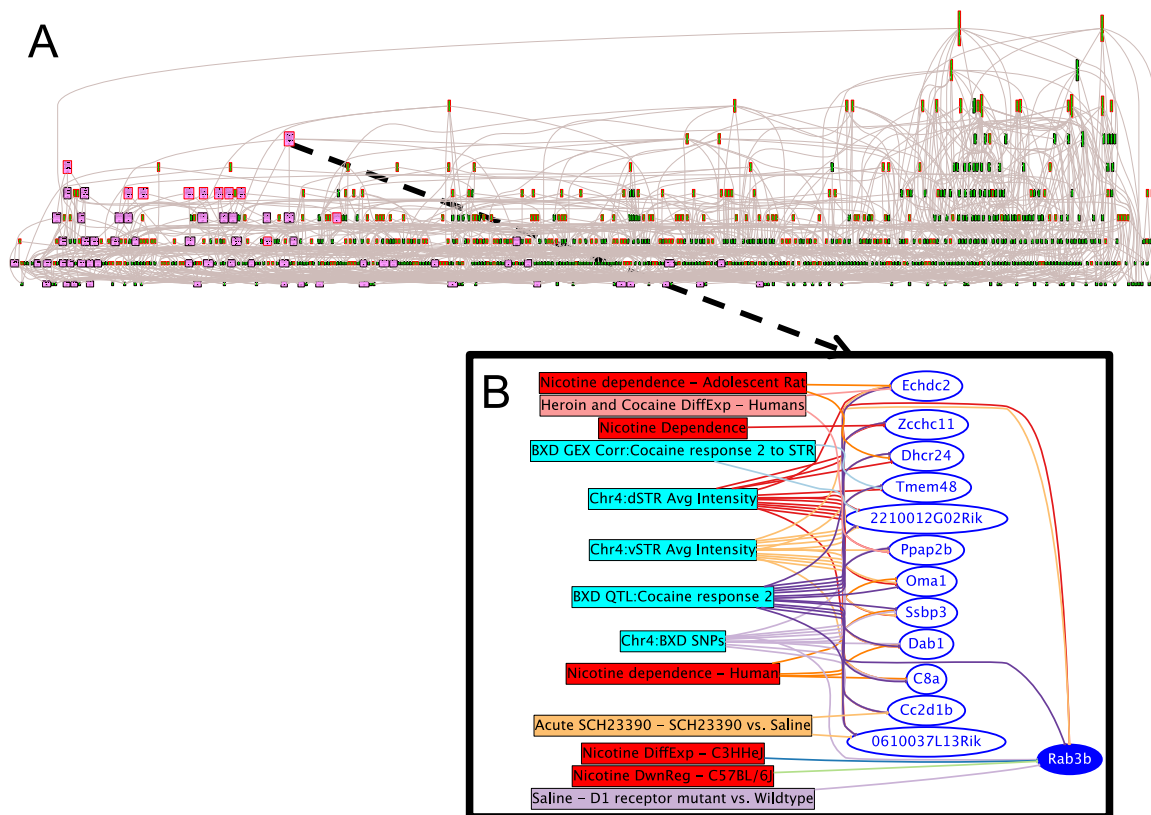


Figure 4-5. Integrative functional analysis in GeneWeaver (11).

A) Phenome map tool depicting a hierarchical representation of gene sets used in the integrative analysis. Nodes at the top are highly connected genes. Pink colored nodes represent nodes containing one or more QTL candidate genes. *Rab3b* is revealed as the most highly connected QTL candidate gene. B) GeneSet graph representing the underlying interconnections among the genes

Table 4-1 Consequence of upstream and intronic polymorphisms between C57BL/6J and DBA/2J.

No	SNP ID	Mb	Location	C57BL/6J	DBA/2J	Genomic sequence +/- 19 bp of the polymorphism	Consequence	D2 Sequence
				C57BL/6J		DBA/2J		
1	wt37-4-108551617	108.55	Upstream	A	T	ATTGT AGGTT CCTTA CACCA GCCC AGAAC GCAG GGGC G	ATTGT AGGTT CCTTA CACCT GCCCA GAACG CAGG GGCG	Predicted binding site for TF <i>Nf-1</i>
2	wt37-4-108553715	108.55	Intron	G	A	CAACA TGTAT CCAGT CTTG TTGTG TCTCA GAAG CACA	CAACA TGTAT CCAGT CTTGA TTGTG TCTCA GAAGC ACA	Splice branch site <u>TTGAT</u> I
3	wt37-4-108554864	108.55	Intron	G	A	CAACA TGTAT CCAGT CTTG TTGTG TCTCA GAAG CAGC	CAACA TGTAT CCAGT CTTGA TTGTG TCTCA GAAGC AGC	Splice branch site <u>TTGAT</u> I
4	wt37-4-108557288	108.56	Intron	C	T	CCCCA CTCCC CAGAT TTAGC GTAAC TAAAA GCAG	CCCCA CTCCC CAGAT TTAGT GTAAC TAAAA GCAG	Splice donor site <u>TTAGT</u> <u>GTAA</u>

Table 4-1 continued.

No	SNP ID	Mb	Location	C57BL/6J DBA/2J		Genomic sequence +/- 19 bp of the polymorphism	Consequence	D2 Sequencing	
						C57BL/6J DBA/2J			
						GCAT GCAT			
5	wt37-4-108568303	108.57	Intron	G	A	CAGAC TCTGG AGGG CATTG <u>G</u> GGC TTTGA TCATC TTCCA	CAGAC TCTGG AGGG <u>CATTG</u> <u>A</u> GGCT TTGAT CATCT TCCA	Splice acceptor site	<u>CATT</u> <u>CAGG</u>

Furthermore, *Rab3b* is expressed in the dorsal and ventral striatum, down-regulated in C57BL/6J and up-regulated in C3H/HeJ after exposure to nicotine (218), differentially expressed in D1 receptor mutants when compared to their wild-type controls after saline treatment (236) and is differentially expressed in the striatum in a recent RNA-Seq gene expression analysis across the C57BL/6J and DBA/2J BXD RI founder strains (21). Additionally, *Rab3b* is a candidate gene among three previously published QTLs, all of which were mapped in the BXD RI reference population, namely climbing in response methamphetamine (82), repeated movements following cocaine (100) and hypothalamic *Cart* abundance (20). In each of the three QTLs, the D2 derived alleles underlie the QTL, which is the case for our cocaine-induced locomotor activation QTL on Chr 4 (COCIA Chr4).

Remapping of previously published QTLs

Three previously published QTLs overlap COCIA Chr 4 QTL, namely, CART transcript abundance (*Crq2*) (20), methamphetamine related climbing response (82) and cocaine related behavior (*Cocrb5*, repeated movements at 5 mg/kg) (100). The advantages of utilizing previously published QTLs are: a) to achieve reduction in the QTL interval by overlapping the previously published QTLs on to the cocaine-induced activation QTL and b) to show that locomotor behaviors after exposure to drugs that affect the dopaminergic system, such as cocaine and methamphetamine, co-map to a common genomic region in multiple independent studies. In an attempt to obtain accurate peak marker locations and

confidence intervals based on the updated BXD genotypes and mouse genome builds, each of the three previously published QTLs were remapped. However, the results of this undertaking were rather surprising. None of the peak markers on Chr 4 that were originally reported were reproducible (Table 4-2). A closer inspection of the BXD genotypes used reveals that the peak loci initially reported are no longer represented in the most recent BXD genotypes, except for *Crq2*. The peak marker on Chr 4 for *Crq2* is D4Mit249. When marker regression is performed using D4Mit249 alone, the marker remains significantly associated with *Cart* abundance ($F(1,23) = 6.96, p < 0.0147$). Legacy marker data are available for *Adpf* (currently, *Plin2*) and *Msh6* in MGI. Using these data, a point biserial correlation analysis was performed (as done in the original publications). Both markers remain significantly correlated with their corresponding trait data (*Adpf*, $r = 0.67, p = 0.0008$; *Msh6*: $r = 0.51, p = 0.0107$). This suggests due to the inclusion of additional markers in current BXD genotypes, there exists newer markers that have better associations with the underlying behavioral data and hence mask the detection of the original peak loci resulting in a failure to reproduce the original QTLs.

Assessment of Rab3b polymorphisms

Forty-six intronic and two upstream polymorphisms exist between B6 and D2 (Table 4-1). The upstream polymorphisms, if within potential promoter binding sites, transcription start sites or transcription factor binding sites, could result in alterations in gene regulation. The intronic polymorphisms, on the other hand,

Table 4-2 Remapping of previously published QTLs.

	Original QTLs (Chr:cM)	Start - End (Mbp)	Marker (cM:Mbp)	Symbol	Reproducible	Updated marker-trait association (method)	Increasor allele	Allele effects at markers	
								C57BL/6J Mean $\pm$ SE	DBA/2J Mean $\pm$ SE
131	Adfp (4:38.9)	61.29 - 111.29	Adfp (40.69:82.11)	<i>Cocrb5</i> [193]	No	$r = 0.67, p = 0.0008$ (pbr)	DBA/2J	22.24 $\pm$ 5.48	52.24 $\pm$ 5.22
	D4Mit249 (4:58.2)	99.84 - 149.84	D4Mit249 (48.18:124.84)	<i>Crq2</i> [192]	No	$r = 0.51, p = 0.0107$ (pbr)	DBA/2J	23.72 $\pm$ 0.39	24.93 $\pm$ 0.24
	Ms6hm (4:40.3)	60.41 - 110.41	Ms6hm (40.3:85.41)	<i>Mrc16</i> [194]	No	$F(1,23) = 6.96, p < 0.0147$ (sma)	DBA/2J	-0.16 $\pm$ 0.33	1.08 $\pm$ 0.3

pbr: point biserial correlations, sma: single marker analysis.

could result in the creation of cryptic splice sites due to the presence of new splice acceptor, branch and donor sites. To assess the effect of the upstream polymorphisms, 1 Kbp upstream of *Rab3b* (specific to B6 and D2) were extracted for the determination of its promoter region, transcription start site and potential transcription factor binding sites. Both strain sequences have the same predicted promoter region, transcription start sites and transcription factor binding sites. However, an upstream polymorphism (wt37-4-108551617, B6: A> D2: T) exists at a predicted transcription factor binding sites for nuclear factor 1 or *Nf-1*.

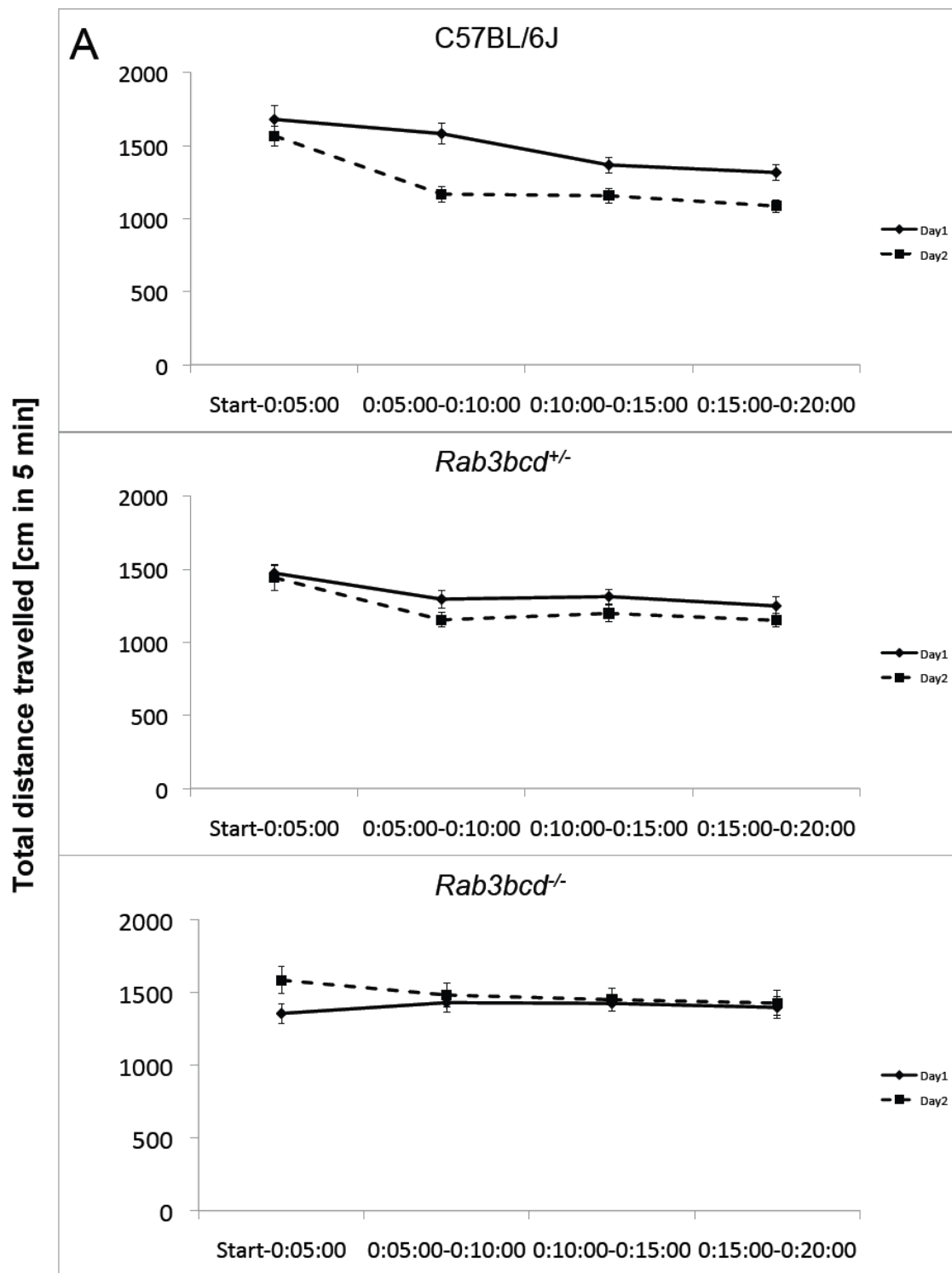
Assessment of B6 and D2 sequences around the location of the polymorphisms revealed the presence of putative splice branch sites, a splice donor site and a splice acceptor site. The two potential splice branch sites are located at 108.553715 Mb (wt37-4-108553715, B6: G> D2: A) and 108.554864 Mb (wt37-4-108554864, B6: G> D2: A), respectively. Both loci are within intron1-2 and contain the sequence TTGATT (D2), which closely resembles the consensus splice branch site sequence (TCCT(A/G)A(C/T)). The splice donor site is located at 108.557288 (wt37-4-108557288, B6: C> D2: T) within intron1-2 and contains the following sequence TTAGTGTA (D2). The splice acceptor site is located at 108.568303 (wt37-4-108568303, B6: G> D2: A) within intron2-3 and consists of the following sequence CATTGAG (D2). The donor and splice site sequence present in D2, closely resembles their corresponding consensus sequences (Figure 4-1).

Assessment of cocaine response in Rab3bcd triple null mice

Preliminary *in vivo* assessment for the role of *Rab3b* in cocaine response was performed using targeted gene disruption. *Rab3b* knockout mice were not readily available but were available by *proxy* as a triple knockout in combination with *Rab3c* and *Rab3d*. *Rab3c* and *Rab3d* are also members of the Rab3 family, as are *Rab3a* and *Rab3b*. All *Rab3* family members, except for *Rab3d* localize on synaptic vesicles. As a whole, Rab3's are involved in regulated exocytosis. Breeding is ongoing for the derivation and purification of *Rab3b* single knockout. However, the *Rab3bcd* triple knockout enables the preliminary assessment of cocaine response involving a similar set of gene products. There exists the possibility that manifestation of abnormal cocaine response could be in part due to elimination of *Rab3c* and *Rab3d*, but it should be noted that the COCIA Chr 4 QTL harbors neither of these genes.

Habituation

Assessment of inter- (Figure 4-6A) and intra-session (Figure 4-6B) habituation on Day 1 and Day 2 revealed that *Rab3bcd*^{-/-} mice fail to habituate. To assess intra-session habituation, data were binned into five-minute intervals C57BL/6J and *Rab3bcd*^{+/-} mice showed typical habituation curves, i.e. decreasing locomotor activity over time intervals. *Rab3bcd*^{-/-} mice, however, displayed typical habituations patterns for intra-session habituation only on Day 2 (Figure 4-6A). Statistically, intra-session habituation can be determined by calculating activity change scores (activity during the initial five minutes / [activity



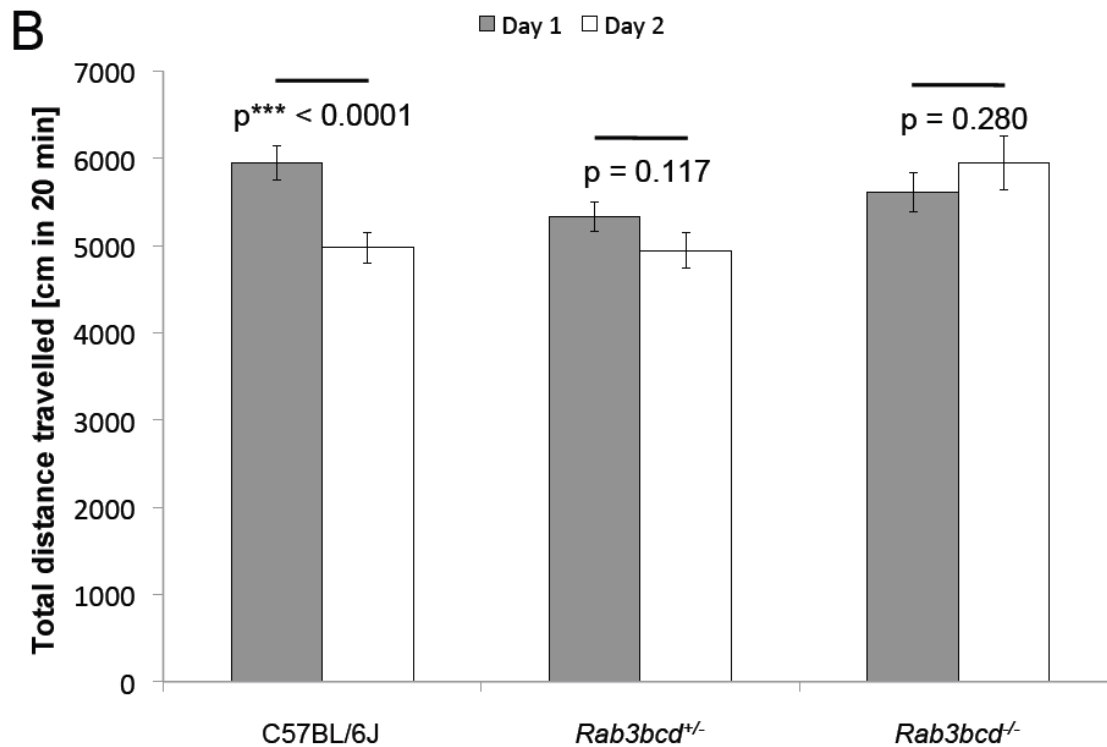


Figure 4-6. Habituation deficits in the *Rab3bcd*^{-/-}.

A) Assessment of intra-session habituation reveals that *Rab3bcd*^{-/-} mice show abnormal habituation curves. Unlike C57BL/6J and *Rab3bcd*^{+/-}, *Rab3bcd*^{-/-} mice show no decrease in distance travelled over the 20 min test period. Assessment of activity change scores revealed that C57BL/6J and *Rab3bcd*^{+/-} had normal habituation patterns, i.e. significantly different from the activity change score of 0.5 (no habituation) on Day 1 (C57BL/6J: $t = -2.95$, $df = 31$, $p = 0.006$; *Rab3bcd*^{+/-}: $t = -2.71$, $df = 31$, $p = 0.011$) and Day 2 (C57BL/6J: $t = -8.24$, $df = 31$, $p < 0.0001$; *Rab3bcd*^{+/-}: $t = -4.10$, $df = 31$, $p = 0.003$). *Rab3bcd*^{-/-} mice showed significantly different from the activity change score of 0.5 only on Day 2 (*Rab3bcd*^{-/-}: Day 1: $t = -2.89$, $df = 29$, $p = 0.775$; Day 2: $t = -2.31$, $df = 29$, $p = 0.027$). B) Compared to C57BL/6J and *Rab3bcd*^{+/-}, *Rab3bcd*^{-/-} had a marginal increase in total distance travelled on Day 2 compared to Day 1. Additionally, C57BL/6J depicted significant decreases in locomotor activity on Day 2 compared to Day 1 ($p < 0.0001$). While the *Rab3bcd*^{+/-} exhibit decreased locomotor activity on Day 2 compared to Day 1, this decrease was not significant ($p = 0.1175$). These two results suggest that *Rab3bcd*^{-/-} have inherent habituation deficits. Error bars are SE

during the initial five minutes + activity during the final five minutes]] (119).

Activity change scores equal to or nearing 0, 0.5 and 1 are representative of habituation, no habituation and increased activity compared to the initial time period, respectively. Mean activity change scores for C57BL/6J, *Rab3bcd*^{+/-} and *Rab3bcd*^{-/-} mice on Day 1 were 0.45 (SEM = 0.02), 0.46 (SEM = 0.01) and 0.50 (SEM = 0.01), respectively. On Day 2 mean activity change scores for C57BL/6J, *Rab3bcd*^{+/-} and *Rab3bcd*^{-/-} mice were 0.41 (SEM = 0.01), 0.45 (SEM = 0.01) and 0.48 (SEM = 0.01), respectively. Comparison of mean activity change scores to an activity change score of 0.5 (representing no habituation) facilitates determination of the extent of intra-session habituations among the three genotypes on days 1 and 2. On Day 1, mean activity change scores for C57BL/6J and *Rab3bcd*^{+/-} were statistically different and lower compared to 0.5 (C57BL/6J: $t = -2.95$, $df = 31$, $p = 0.006$; *Rab3bcd*^{+/-}: $t = -2.71$, $df = 31$, $p = 0.011$). Inter-session habituation revealed that while C57BL/6J and *Rab3bcd*^{+/-} mice show decreased locomotor activity on Day 2 compared to Day 1, *Rab3bcd*^{-/-} mice show a marginal increase in locomotor activity (Figure 4-6B). Specifically, C57BL/6J had significant differences in locomotor activity on Day 1 (Mean = 5946.41, SEM = 196.08) compared to Day2 (Mean = 4979.40, SEM = 175.67) ($t = -4.63$, $df = 31$, $p < 0.0001$). The decrease in locomotor activity for *Rab3bcd*^{+/-} mice on Day 1 (Mean = 5329.72, SEM = 171.41) compared to Day 2 (Mean = 4943.46, SEM = 203.39) was non-significant ($t = -1.61$, $df = 31$, $p = 0.117$). Similarly, the marginal increase in locomotor activity on Day 1 (Mean = 5709.17,

SEM = 201.50) compared to Day 2 (Mean = 5999.13, SEM = 321.62) for the *Rab3bcd*^{-/-} was also non-significant ($t = 1.1$, $df = 29$, $p = 0.280$). Furthermore, over the duration of the behavioral testing protocol, among the SAL group of mice, *Rab3bcd*^{-/-} showed consistently greater locomotor activity post saline on days 2, 3, 5, 7 and 9 compared to C57BL/6J and *Rab3bcd*^{+/-} mice (Table 4-3, Figure 4-7).

Acute response to cocaine

Analysis of acute response to cocaine (Day 3^{COCA} – Day 2^{SAL}) revealed a significant effect of Treatment $\times$ Genotype ($F(2,2) = 6.04$, $p < 0.003$) (Figure 4-8). Specific tests for significant effects of genotype for the COCA and SAL groups were performed using planned contrasts statements. Planned comparisons revealed a significant effect of genotype for the COCA groups of mice ($F(2,90) = 10.75$, $p < 0.0001$) and not the SAL group ($F(2,90) = 0.08$, $p < 0.92$). Within the COCA group, *Rab3bcd*^{-/-} mice had the greatest change in locomotor response following cocaine compared to *Rab3bcd*^{+/-} and C57BL/6J (Figure 4-8).

Response to repeated exposure to cocaine

A repeated measures analysis of variance approach was used to assess locomotor response following repeated exposure to cocaine (repeated factor being day). As assumptions of sphericity were not met, the correction factors of

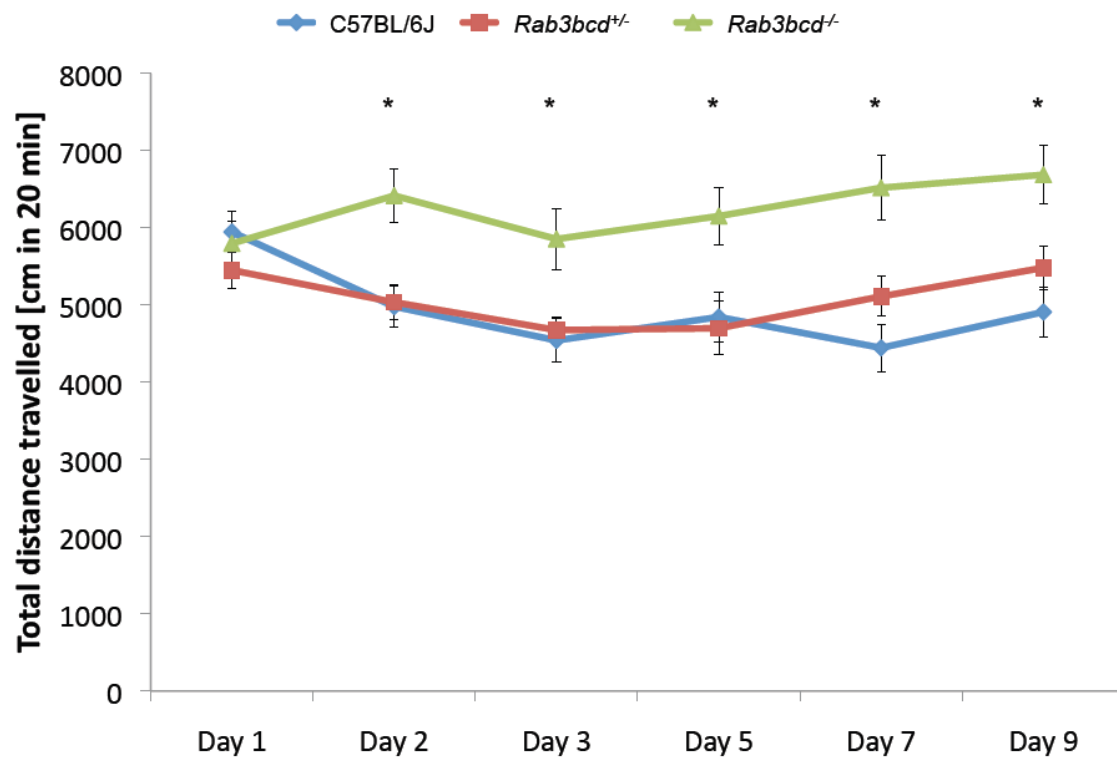


Figure 4-7. Consistent increase locomotor activity in saline-treated *Rab3bcd*^{-/-}.

Rab3bcd^{-/-} showed consistently greater locomotor activity following saline compared to C57BL/6J and *Rab3bcd*^{+/-} mice on all days except Day 1. * indicate significant differences ($p < 0.05$). Error bars are SEM.

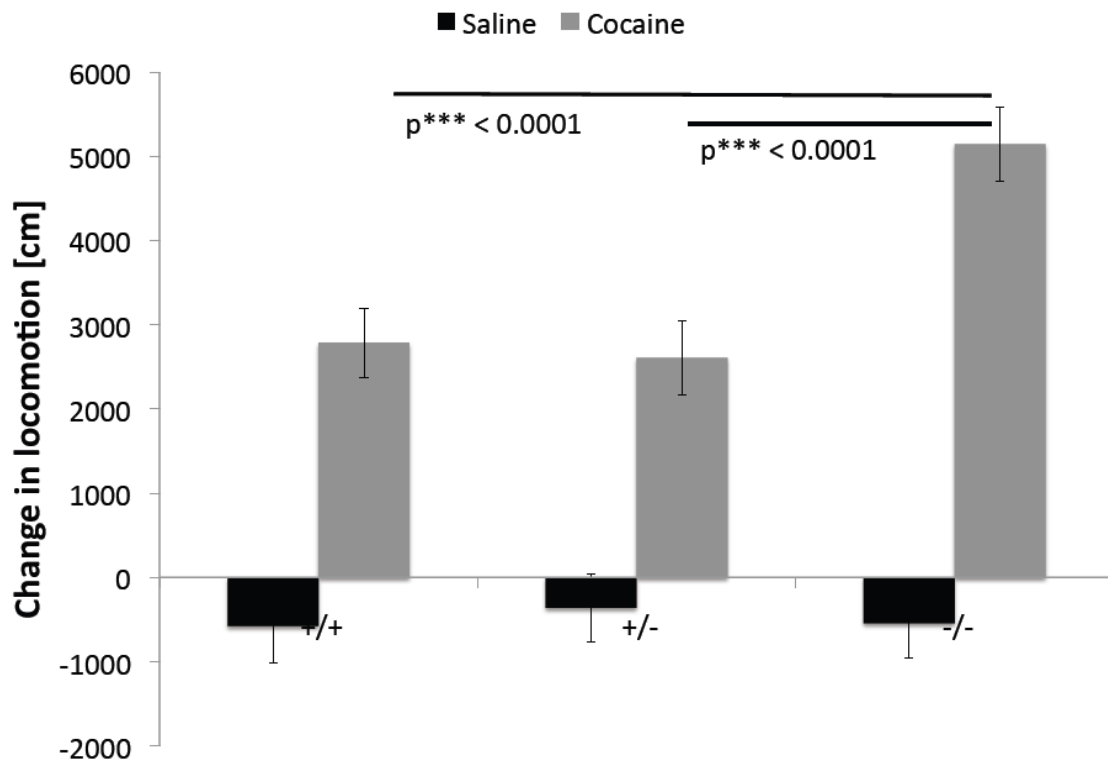


Figure 4-8. *Rab3bcd*^{-/-} and acute cocaine response.

Acute response to cocaine measured here, as the difference in total distance travelled on Day2 (post saline) from that on Day3 (post cocaine) for the Cocaine group indicates that *Rab3bcd*^{-/-} mice have a significant increase in change in locomotor activity following cocaine compared to *Rab3bcd*^{+/-} and C57BL/6J mice. No significant differences exist across the three genotypes among the mice belonging to the Saline group. Error bars are SEM.

Table 4-3. Assessment of baseline locomotor differences in *Rab3bcd*^{-/-}.

Day	Genotype1	Genotype2	Difference [Genotype1 - Genotype2, cm]	SE Difference	LCL	UCL	<i>p</i>
1	C57BL/6J	<i>Rab3bcd</i> ^{+/-}	616.69	281.76	-54.42	1287.79	0.08
	C57BL/6J	<i>Rab3bcd</i> ^{-/-}	333.07	281.76	-338.03	1004.18	0.47
	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	283.62	281.76	-387.49	954.72	0.57
2	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	1007.27	334.59	210.34	1804.19	0.01*
	<i>Rab3bcd</i> ^{-/-}	C57BL/6J	971.33	334.59	174.4	1768.25	0.01*
	C57BL/6J	<i>Rab3bcd</i> ^{+/-}	35.94	334.59	-760.99	832.87	0.99
3	<i>Rab3bcd</i> ^{-/-}	C57BL/6J	1313.46	427.37	278.46	2348.47	0.01*
	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	1179.46	413.79	177.32	2181.61	0.02*
	<i>Rab3bcd</i> ^{+/-}	C57BL/6J	134	427.37	-901.01	1169.01	0.95
5	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	1453.04	483.3	282.56	2623.52	0.01*
	<i>Rab3bcd</i> ^{-/-}	C57BL/6J	1307.62	499.15	98.76	2516.49	0.03*
	C57BL/6J	<i>Rab3bcd</i> ^{+/-}	145.41	499.15	-1063.45	1354.28	0.95
7	<i>Rab3bcd</i> ^{-/-}	C57BL/6J	2073.86	485.82	897.28	3250.44	0.001*
	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	1407.8	470.4	268.58	2547.02	0.01*
	<i>Rab3bcd</i> ^{+/-}	C57BL/6J	666.06	485.82	-510.52	1842.64	0.36

Table 4-3 continued.

Day	Genotype1	Genotype2	Difference [Genotype1 - Genotype2, cm]	SE Difference	LCL	UCL	<i>p</i>
9	<i>Rab3bcd</i> ^{-/-}	C57BL/6J	1778.71	474.66	629.17	2928.25	0.001*
	<i>Rab3bcd</i> ^{-/-}	<i>Rab3bcd</i> ^{+/-}	1205.1	459.58	92.06	2318.14	0.03*
	<i>Rab3bcd</i> ^{+/-}	C57BL/6J	573.61	474.66	-575.93	1723.15	0.45

* Indicate significant differences ($p < 0.05$)

the Greenhouse-Geisser Epsilon (G-G) and the Huynh-Feldt Epsilon (H-F) were used for the assessment of the day $\times$ genotype effect. No significant effect of day $\times$ genotype ($F(6,132) = 0.72$, G-G: $p < 0.61$, H-F: $p < 0.62$) was detected (Figure 4-9).

Sensitized response to cocaine

In addition to assessing the effects of repeated exposure to cocaine in this panel on mice, we also assessed the development of cocaine sensitization. Analysis of cocaine sensitization calculated as Day 9^{COCA} – Day 3^{COCA} locomotion revealed no significant effect of Genotype ($F(2,2) = 1.39$; $p < 0.25$).

Effect of habituation on cocaine induced locomotor activation

Assessment of the effect of habituation on cocaine induced locomotor activation led to the manifestation of a significant habituation $\times$ genotype interaction ($F(1,2) = 4.25$, $p < 0.03$). On Day 1, C57BL/6J (Mean = 16815.36, SEM = 1535.22) and *Rab3bcd*^{+/-} (Mean = 16018.17, SEM = 1535.22) mice showed increased cocaine induced locomotor activation compared to *Rab3bcd*^{-/-} (Mean = 11481.20, SEM = 1880.26) mice. Furthermore, following two days of habituation C57BL/6J (Mean = 8594.24, SEM = 1535.22) and *Rab3bcd*^{+/-} (Mean = 10876.61, SEM = 1535.22) exhibited reduced amounts of cocaine induced locomotor activation compared to Day 1. Unlike the C57BL/6J and *Rab3bcd*^{+/-}, *Rab3bcd*^{-/-} mice showed a marginal non-significant increase in cocaine induced

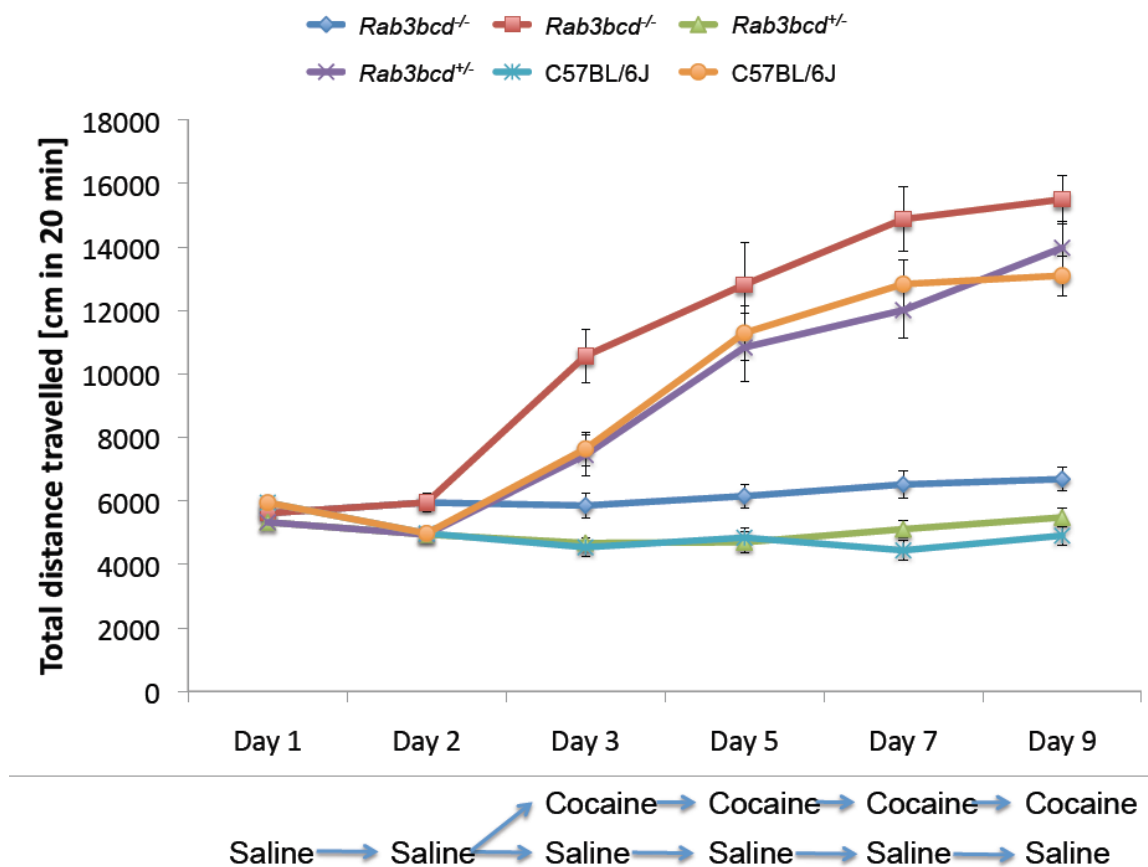


Figure 4-9. $Rab3bcd^{-/-}$ and lack of sensitized cocaine response.

On Days 1 and 2 all mice have approximately the same amount of distance travelled following saline injections. Over the course of the experiment $Rab3bcd^{-/-}$ mice have higher baseline locomotor activity compared to $Rab3bcd^{+/-}$ and C57BL/6J. Administration of cocaine (Days 3, 5, 7 and 9) results in an increase in locomotor activity for all three genotypes compared to the saline treated group with $Rab3bcd^{-/-}$ exhibiting the highest amount of locomotor activity across the three genotypes. Repeated measures ANOVA resulted in no significant increase in locomotor activity over days following administration of cocaine

locomotor activation on Day 3 (Mean = 12888.41, SEM = 1681.75) as compared to Day 1 (Mean = 11481.20, SEM = 1880.26). This result suggests that lack of habituation exhibited by the *Rab3bcd*^{-/-} may affect the magnitude of cocaine induced locomotor activation observed.

Differences in Rab3b transcript lengths between C57BL/6J and DBA/2J

There are four transcripts of *Rab3b* according to the Vertebrate Genome Annotation database (VEGA) genome browser (7, 226). Of these four, two are CCDS or consensus coding sequences of different lengths, henceforth referred to as *Rab3b*-L (OTTMUST00000018722) and *Rab3b*-S (OTTMUST00000018724) transcripts. *Rab3b*-L has an extended 5' end compared to *Rab3b*-S. As a possible means of differential regulation of *Rab3b* between the BXD progenitors, namely B6 and D2, we sought to quantify transcript abundance using two TaqMan probes, namely, Mm01258423 and Mm00772238. Mm01258423 spans Exons 1-2 and hence would uniquely target *Rab3b*-L. Mm00772238 span Exons 3–4 and targets both *Rab3b*-L and *Rab3b*-S. ΔC_t values were analyzed for differential expression across the two strains. qRT-PCR results suggest that there is no differential expression between the two strains for either of the probes (Table 4-4, *Rab3b*-L: $p < 0.24$; *Rab3b*-S: $p < 0.31$).

Table 4-4. ΔC_t values for probes targeting Rab3b-S and Rab3b-L transcripts.

Transcript ID	TaqMan Probe ID	Exons spanned	Identifier	ΔC_t – Mean $\pm$ SE	
				C57BL/6J	DBA/2J
OTTMUST00000018724	Mm00772238	Exons 3–4	<i>Rab3b</i> -S	6.29 $\pm$ 0.08	6.2 $\pm$ 0.04
OTTMUST00000018722	Mm01258423	Exons 1–2	<i>Rab3b</i> -L	6.79 $\pm$ 0.08	6.67 $\pm$ 0.05

CART transcript abundance in $Rab3bcd^{-/-}$ and C57BL/6J

A previously published QTL mapping study involving CART transcript abundance (20) overlapped our COCIA Chr4 QTL. CART or cocaine- and amphetamine-regulated transcript peptide is involved in reward and feeding-related behaviors. Studies in rats have shown that CART mRNA expression is upregulated after acute cocaine self-administration in brain regions of the mesocorticolimbic dopamine system (59). To assess the effects of *Rab3bcd* KO on *cartpt* expression, we analyzed CART expression using ΔC_t values to detect genotype (*Rab3bcd^{-/-}* and C57BL/6J) $\times$ treatment (naïve, SAL and COCA) effects (Table 4-5). The genotype $\times$ treatment effect was suggestive ($F(1,2) = 2.58$; $p < 0.08$). No significant differences were observed across genotypes for the naïve group, suggesting that baseline levels of *Cartpt* expression are not altered in *Rab3bcd* mutants. Planned contrasts were used to detect differences in CART expression following saline or cocaine treatment. Following saline and cocaine treatment *Cartpt* is differentially expressed across the two genotypes (SAL: $F(1,42) = 10.03$, $p < 0.003$; COCA: $F(1,42) = 11.10$, $p < 0.002$). Within genotype contrasts failed to detect any significant differences across the two treatments (*Rab3bcd^{-/-}*: $F(1,42) = 1.76$, $p < 0.19$; C57BL/6J: $F(1,42) = 1.84$, $p < 0.18$). These results suggest that it is unlikely that disruption of *Rab3bcd* alone induces dysregulation of *Cartpt*. However, in combination with external factors, in this case repeated handling and injection stress, an increase in *Cartpt* expression was detected (Table 4-6). There was a significant genotype

Table 4-5. *Cartpt* ΔC_t values for C57BL/6J and *Rab3bcd*^{-/-} among naïve, saline and cocaine treated mice.

Treatment	ΔC_t – Mean $\pm$ SE	
	Genotype	
	C57BL/6J	<i>Rab3bcd</i> ^{-/-}
Naïve	10.24 $\pm$ 0.19	10.12 $\pm$ 0.23
Saline	10.55 $\pm$ 0.11	9.75 $\pm$ 0.15
Cocaine	10.22 $\pm$ 0.16	9.42 $\pm$ 0.18

Gapdh was used in the calculations of ΔC_t .

Table 4-6. *Cartpt* ΔC_t values for C57BL/6J and *Rab3bcd*^{-/-} among naïve and challenged (saline and cocaine treated) mice.

Challenge	ΔC_t – Mean $\pm$ SE	
	Genotype	
	C57BL/6J	<i>Rab3bcd</i> ^{-/-}
Naïve	10.24 $\pm$ 0.19	10.12 $\pm$ 0.23
Challenge (saline or cocaine)	10.36 $\pm$ 0.11	9.58 $\pm$ 0.11

Gapdh was used in the calculations of ΔC_t .

× challenge (naïve or injection) effect when the cocaine and saline injected mice were collapsed and grouped as “challenge” ($F(1,1) = 4.69$; $p < 0.03$). Planned contrasts revealed a significant difference in *Cartpt* expression between *Rab3bcd*^{-/-} naïve and injected mice ($F(1,44) = 6.18$; $p < 0.02$). No significant difference in *Cartpt* expression was detected between C57BL/6J, naïve and challenge mice ($F(1,44) = 0.33$; $p < 0.57$).

Discussion

The highly heritable component of alcoholism and drug addiction has led to numerous studies attempting to understand and characterize the genetic basis of addiction. One method that has been widely used for the unbiased detection of genomic regions harboring polymorphic candidate genes for complex traits is

QTL mapping. Widespread utilization of this approach has resulted in the detection of large numbers of QTL harboring numerous candidate genes with only a handful of candidate genes within these QTL being successfully nominated and validated as the QTG. Most, if not all QTG have been identified using a combination of approaches such as fine mapping, sequence and expression level comparison of genes between progenitor strains or between congenics and background strains (138). While genetic methods have successfully identified QTGs, they have been extremely time- and resource-intensive, often-taking years and large numbers of mice (especially recombinant progeny required for fine-mapping a QTL interval). Complementary bioinformatic tools have been proposed as a means of accelerating QTG identification but to

date very few genes have been nominated as QTG (57, 84, 151). A strategy with tremendous potential in facilitating QTG discovery that has emerged over the last few years is integrative functional genomics. This strategy, as the name suggests, involves integration of candidate genes with empirically derived genome-scale information available via published studies that span multiple platforms, species, and curated data repositories, such as Mammalian Phenotype Ontology (MP,(194)), Gene Ontology (GO, (3)) and Kyoto Encyclopedia of Genes and Genomes (KEGG, (102)). This strategy has been successfully applied towards the analysis of bipolar disorder (117, 151). Existing integrative approaches suffer from a lack of cross-species integration, scalability and computationally efficient algorithms, in order to achieve true genome-wide integrative analysis. Cross-species integration is vital to facilitate translation of research findings from induced animal models to humans. To facilitate a theoretically comprehensive integrative analysis, i.e. no limit on the number of gene sets in the analysis, it is important that integrative approaches be scalable. By being computationally efficient, reasonable analysis run times can be achieved, making genome scale integrative analysis feasible.

Presented here is the integrative functional genomic analysis of the candidate genes for a previously mapped cocaine-induced activation QTL in BXD RI strain panel (164). By incorporating multiple independent lines of evidence, it is possible to efficiently and seamlessly query a QTL interval, in an unbiased manner, and nominate QTGs. In the current analysis, empirical information was

derived from 21 published studies, totaling 122 genomic data sets. Using these genomic data sets *Rab3b* was identified as the most highly ranked (maximum gene to gene set connectivity) positional candidate for the COCIA Chr4 QTL.

RAB proteins are proteins belonging to the family of Ras-like GTPases. These proteins are primarily involved in trafficking of secretory and endocytic vesicles of eukaryotic cells (133, 154). Sixty known RAB proteins exist, of these RAB3's are the most abundant and expressed at variable levels in the brain (85, 197, 203). The RAB3 protein family consists of four homologous isoforms, namely, RAB3A, 3B, 3C and 3D. Among the RAB3's, RAB3A, 3B and 3D co-localize on presynaptic vesicles, while RAB3D localizes on secretory vesicles on exocrine glands (72, 152, 183). RAB3's are involved in regulated exocytosis and are involved in docking, priming and fusion stages of the synaptic vesicle cycle. Specifically, *in vitro* studies have revealed that RAB3 proteins inhibit Ca^{2+} -triggered exocytosis in PC12 cells (183). Additional lines of evidence support the involvement of RAB3 in inhibition of neurotransmitter release. Overexpression of GTPase-deficient mutants of RAB3 inhibits release of neurotransmitters and hormones (92, 93, 97, 171). Overexpression of the catalytic domain of RAB3 GEP, which promotes the exchange of GDP for GTP on RAB3, reduces exocytosis (157). Decreasing the expression of *Rab3a* via antisense oligonucleotides microinjection, resulted in an increase in response to repetitive stimulations (97). *In vitro* and *in vivo* studies involving overexpression of *Rab3b* in dopamine neurons of the rat SN increases dopamine content, uptake, number

and size of synaptic vesicles, and levels of presynaptic proteins consistent with the results obtained in studies involving RAB3B (40). Within the synapse a number of putative effector proteins of RAB3 exist. These include RIMs (RIM1 α and RIM2 α), Rabphilin, NOC2, SNAP-25, synapsin, synaptogamin and CaM. Among these, it has been shown that disruption of RAB3A-CaM interaction (and not the other effector proteins) result in the lack of inhibition in exocytosis by RAB3 (46). Furthermore, overexpression of *Rab3b* also leads to an increase in CaM (190). Similar to disrupting the RAB3A-CaM interaction, one can hypothesize that disruption of the RAB3B-CaM interaction could also lead to RAB3 inhibition of exocytosis, primarily due to the redundant functionality and high degree of homology among the Rab3 proteins.

While *Rab3b*'s functional involvement in inhibiting neurotransmitter release makes it a highly attractive candidate gene for drug addiction, is it a QTG? Using the characteristics of a QTG suggested by the Complex Trait Consortium (CTC) (5) and Glazier and colleagues (78), we show that *Rab3b* satisfies multiple criteria that makes it a highly likely and valid QTG. These are:

a) *Rab3b* is polymorphic between the B6 and D2, the progenitor strains for the BXD RI genetic reference population. Two upstream and 46 intronic polymorphisms located within the first and second introns of *Rab3b* exist (Table 4-2). Detailed analysis these polymorphisms using the recent Sanger resequencing project (105) revealed important insights regarding the nature of these polymorphisms. Of the two upstream polymorphisms, one (wt37-4-

108551617, B6: A>D2: T) underlies the potential transcription factor binding site of nuclear factor 1(NF-1). This could potentially result in differential gene regulation of *Rab3b* across the B6 and D2 strains. Among the intronic polymorphisms, the D2 sequences of four SNPs closely resemble the consensus sequences for splice acceptor (wt37-4-108568303, B6: G>D2: A), branch (wt37-4-108553715, B6: G> D2: A; wt37-4-108554864, B6: G> D2: A) and donor (wt37-4-108557288, B6: C> D2: T) sites (Table 4-2). These polymorphisms could potentially create cryptic splice sites due to the presence of novel splice acceptor, branch and donor sites again resulting in differential regulation of *Rab3b* across B6 and D2 strains, b) *Rab3b* is expressed in brain regions relevant to cocaine related responses. It is expressed in the striatum, nucleus accumbens and the ventral tegmental area (www.brain-map.org). Each of these brain regions contain dopaminergic neurons and plays a central role in cocaine-mediated behaviors, c) *Rab3b* is differentially expressed following chronic nicotinic exposure across five brain regions in two inbred strains of mice, namely, C57BL/6J and C3H/HeJ (218). Specifically, *Rab3b* was down regulated in the nucleus accumbens of C57BL/6J mice, while being upregulated in the ventral tegmental area of C3H/HeJ mice. Furthermore, a recent RNA-Seq study has shown that *Rab3b* ($p < 0.02$, q-value < 0.07) is differentially expressed in the striatum of C57BL/6J and DBA/2J strains of mice (21). The differential expression of *Rab3b* across the two strains could be attributable to the presence of functional SNPS, i.e. cryptic splice sites and can be experimentally validated,

and d) The COCIA Chr4 QTL harboring *Rab3b*, has also been detected in three previous mapping studies for cart transcript abundance (20), methamphetamine related climbing response (82) and cocaine related behavior (repeated movements at 5 mg/kg) (100). It should be noted that we were not able to reproduce the results from the above-mentioned three original publications using new BXD genotypes and interval mapping. The non-reproducibility is in the location of QTLs upon remapping the original publications. Even though there is disparity in the location of the remapped QTLs compared to previously published QTLs, it must be noted that this disparity does not affect the results of the IFG analysis. *Rab3b* is still the most highly connected gene regardless of the inclusion of the previously published QTLs.

To validate *Rab3b* as a QTG, an *in vivo* gene targeting approach was used. *Rab3b* KO mice are only available as a triple KO, which also involves *Rab3c* and *Rab3d*. Among the *Rab3bcd*^{-/-} mice, two of the three RAB3 proteins that co-localize on presynaptic vesicles, namely *Rab3b* and *Rab3c*, are absent. *Rab3d* does not localize within synaptic vesicles (72, 152, 183) and may not be involved in regulated exocytosis. This could result in possible alterations in the inhibition of exocytosis by Rab3s. Given this, we would expect the *Rab3bcd*^{-/-} mice to have increased neurotransmitter release (here dopamine), which behaviorally would result in increased locomotor activity following saline and cocaine treatment. As shown in Figure 4-7, this was in fact observed; *Rab3bcd*^{-/-} mice consistently had increased locomotor activity post saline (suggesting

habituation deficits) compared to their *Rab3bcd*^{-/-} and C57BL/6J mice throughout the length of the study. Furthermore, the cocaine response observed is an acute response (Figure 4-8) rather than a sensitized response (Figure 4-9).

Interestingly, we also detected significant differences in *Cartpt* expression across naïve and treated *Rab3bcd*^{-/-} mice. This suggests a role of *Rab3bcd* in *Cartpt* variation following external stressors, in this case repeated injection and handling stress. This is of additional value as one of the previously published QTLs overlapping the COCIA Chr4 QTL was for *Cartpt* expression in the BXD RI strain panel and mapped to a region on Chr 4 that contains *Rab3b*. The similar levels of *Cartpt* after saline and cocaine exposure might be attributed to the harvesting of brain regions after 24 – 48 hr from the last cocaine injection.

Future work will involve backcrossing and purification of *Rab3b*^{-/-} mice to N5 and higher, at which stage they can be considered as incipient congenic strains followed by assessment of acute cocaine response in the *Rab3b* single nulls. While functional validation of the *Rab3b*^{-/-} mice will serve as proof of *Rab3b*'s involvement in cocaine response, it still will not unequivocally prove that it is the QTG. Conclusive proof that *Rab3b* is the QTG can be obtained via additional fine-mapping of the QTL using RISTs (50). As this approach starts with recombinants among the RI strain panel within the QTL interval, it can readily be applied to the COCIA Chr4 QTL that was mapped in the BXD RI strain panel. The COCIA Chr4 QTL interval contains twelve recombinants, namely, BXD 13, 15, 16, 28, 32, 43, 55, 67, 86, 87 90 and 100. Of these twelve, BXD 55 and BXD

86 are double recombinants. BXD 33 could also possibly be a double recombinant. BXD 15, 16, 28, 67 and 87 strains are best suited for the derivation of RISTs as they contain almost equal progenitor strain contribution in the QTL interval. There are two RIST populations, namely, RIST-F2 and RIST-BC, available for fine-mapping a QTL. Since the original QTL was detected in the BXD RI mapping population, which only allows for assessment of additive effects, the RIST-F2 approach will be most suited for fine mapping. Another approach that can be used to verify *Rab3b* as a QTG is to perform a QTL – knockout interaction test or quantitative complementation test (52, 196). In order to do so, four strains are required. Two co-isogenic strains, one containing the null allele, while the other being the wild type. These strains are called “Cross” in this approach. The other two strains are “high” and “low” phenotypic strains. These strains are referred to as “Line”. The next step involves crossing the null strain with the high and low strains, to obtain high/null and low/null F1 hybrids. Similarly, the wild-type (137) is crossed with the high and low strains to obtaining high/wt and low/wt F1 hybrids. A significant Line x Cross interaction in an ANOVA analysis is interpreted as evidence that the knockout is in fact the QTL. *Rgs2* was validated as the gene underlying the QTL using this approach (230). However, the possibility that the significant interaction is due to the gene of interest interacting with loci on other chromosomes cannot be ruled out. An approach such as reciprocal hemizygosity, first applied for the dissection of quantitative traits in yeast (126), can help resolve such issues.

While the eventual proof of a candidate gene as the QTG lies in genetic approaches, *in silico* approaches can help point future studies in the right direction through the identification and nomination of putative QTG. As shown here, using an integrative functional genomics approach within the framework of GeneWeaver (www.geneweaver.org), *Rab3b* was identified as a putative QTG for acute cocaine response. GeneWeaver facilitates the detection of putative QTGs by its extensive and well-curated data repository of functional genomic studies that span multiple species and experimental platforms. In addition to this extensive data repository, GeneWeaver's computationally robust database and efficient graph theoretical algorithmic enable integrative analysis to be performed at a truly genomic scale. These features make GeneWeaver an extremely useful and applicable resource for the acceleration of QTG identification and nomination by well-defined hypothesis driven integrative analyses.

CHAPTER V: CONCLUSION

The results presented here address the key issue for efficient nomination of Quantitative Trait Genes (QTG). Nomination and validation of a QTG is the end goal of all Quantitative Trait Locus (QTL) mapping studies. Despite the widespread utilization of QTL mapping, very few QTG have been nominated and validated. This is partly attributable to the power and resolution with which a QTL is mapped that depends heavily upon the genetic architecture of the mapping population or experimental crosses utilized. Frequently, the resolution (confidence interval; CI) of a mapped QTL is large and harbors numerous positional candidate genes, which makes the task of QTG nomination tedious. Improvements in statistical methodologies and experimental crosses have greatly improved the prospect of QTG nomination.

To address the issue of efficient QTG nomination, the methodology and results assess the current and future state for QTL to QTG discovery by:

- Assessing the power and resolution of QTL mapping in two recently developed experimental crosses, the BXD RI (*Chapter II*) and the collaborative cross (CC, *Chapter III*)
- Presenting integrative functional genomics as an alternative *in silico* approach for the successful nomination of QTGs (*Chapter IV*).

QTL mapping in the expanded BXD Recombinant Inbred and the emerging collaborative cross mouse strains

The BXD recombinant inbred (RI) strain panel has been a widely used mapping population for the study of complex traits. Prior to their recent expansion, the original panel of BXD RI strains ($n = 26$) was able to reliably detect QTLs accounting for 32% of trait variance (power = 0.90, $\alpha = 0.05$) (18). QTL accounting for such large trait variance are representative of QTL underlying Mendelian rather than complex traits. Hence, genes underlying such QTL would be Mendelian rather than quantitative trait genes. Following the expansion, the number of RI strains in the BXD panel increased to approximately 90 strains, facilitating the detection of QTL accounting for as little as 11% of trait variance (power = 0.90, $\alpha = 0.05$). Additionally, the increased number of strains makes feasible the detection of epistatic interactions. In *Chapter II*, the recently expanded BXD RI strains were utilized for genetic analysis of a multitude of complex traits. In total, 257 behavioral traits were assayed across an average of 60 BXD RI strains. The majority of the traits were amenable to QTL mapping (broad sense heritability or $h^2 > 0.30$) with better mapping resolution performance. Furthermore, QTLs accounting for approximately 16% of trait variance can be detected with 90% power ($\alpha = 0.05$) using 60 BXD RI strains. Mapping resolution, determined using the 95% CI, revealed that the mean 95% CI of significant QTL ($\text{LOD} > 3.2$ (73)) was 12.5 cM, a considerable improvement from 20-30 cM that was achieved using the initial BXD RI strain

panel. Part of this improvement is attributable to dense marker maps, but it is largely due to the increased numbers of recombination events in the expanded BXD RI strains (162). Despite these improvements, the BXD RI strain panel suffers from the unavoidable constraint of limited genetic variation. On average, two progenitor RILs capture approximately 15% of available genetic variation (genetic variation captured in the BXD RILs is about 16%), primarily due to their shared ancestry and large haplotype blocks (162, 232, 233). The limited percentage of captured genetic variability leads to “blind spots” in QTL mapping. Blind spots are regions of limited or no variation within RILs, which mask the presence and impede the detection of true QTLs (175). These limitations necessitate the generation of advanced experimental crosses, such as the CC. The CC is derived from eight inbred strains rather than two. The eight inbred strains chosen for the derivation of the CC are C57BL/6J, 129S1/SvImJ, NOD/LtJ, A/J, NZO/HiLtJ, CAST/EiJ, WSB/EiJ and PWK/PhJ (34, 42, 211). Three of the eight founder strains, namely, CAST/EiJ, PWK/PhJ and WSB/EiJ are wild-derived strains representing, *M. m. castaneus*, *M. m. musculus* and *M. m. domesticus* mouse sub-species, respectively. The eight combined inbred strains capture 89% of the available genetic variation (175). The increased genetic variation results in smaller regions of identity by descent (IBD) that consequently reduces the presence of blind spots. In addition to increased genetic variation, the CC is expected to have higher numbers of recombination events and larger numbers of independent genetically diverse RI strains, each a

fine-grained mosaic of the eight founder strains. The increased amount of genetic variation and higher number of recombinations are expected to greatly improve QTL mapping resolution in the CC, once completed. In *Chapter III*, an assessment of QTL mapping in the early breeding generations of the CC was performed. QTL mapping in this population achieved an average confidence interval (1.5-LOD support interval around the significant QTL peaks) of 3.98 Mb (approximately 2 cM) with an average of 50 positional candidate genes in these intervals (8, 165). This is a considerable reduction in QTL mapping resolution compared to the BXD RI strain panel. However, QTL mapping in the CC still does not achieve single gene resolution, necessitating additional fine mapping to further reduce the QTL interval.

In silico reduction of QTL intervals

Thus far almost all the QTGs nominated and validated have made use of genetic approaches. While successful, these approaches are considerably resource and time intensive, and do not necessarily expedite the process of QTG nomination and validation. To expedite QTG discovery, complementary bioinformatic tools have been developed. Unfortunately, current bioinformatic tools require considerable end-user involvement due to the disparate nature of empirically derived genomic data sets. This severely limits the ability with which they can be used to perform a truly cross-species and cross-platform oriented integrative genomic analysis. An approach that has emerged over the last few

years and has tremendous potential in addressing and resolving these issues is integrative functional genomics.

The appeal of integrative functional genomics lies in its ability to facilitate convergent analyses across multiple genomic experiments, platforms and species (11). This approach can help expedite QTG discovery, as candidate genes within a QTL interval can be integrated with other relevant empirically derived genomic data sets across multiple species to nominate highly relevant QTG. In *Chapter IV*, the approach of integrative functional genomics is implemented using GeneWeaver (www.geneweaver.org) that facilitates computationally efficient integrative functional genomic analyses by providing comprehensive data sources and analytical toolsets (11). Using GeneWeaver, *Rab3b* is nominated as a putative QTG for a cocaine-induced locomotor activation QTL on Chr 4 mapped in the expanded BXD RI strain panel (164). *Rab3b* meets multiple criteria that justify its classification as a QTG, including relevant polymorphism (here, intronic cryptic splice site polymorphisms), differential expression in the striatum across C57BL/6J and DBA/2J inbred strains (21) and basal expression in mesocorticolimbic regions of the brain. Additionally, functional evaluation of the *Rab3* complex in cocaine response was performed using mixed background *Rab3bcd* triple knockout (KO) mice (B6;129 *Rab3bcd*^{-/-}). This analysis revealed the role of *Rab3b* in acute cocaine response. Specifically, *Rab3bcd*^{-/-} mice have increased cocaine-induced activation compared to *Rab3bcd*^{+/-} and C57BL/6J mice in a test for acute cocaine response.

Conclusive evidence of *Rab3b* as the QTG will be obtained by: a) derivation and functional evaluation of the *Rab3b* single KO mice, b) additional fine-mapping of the cocaine-induced locomotor activation QTL on Chr 4 using recombinant inbred segregation tests (RISTs) and c) experimental validation of the potential splice branch site polymorphisms.

The nomination of *Rab3b* as the putative QTG for cocaine-induced locomotor activation speaks to the utility and efficiency of integrative functional genomics and GeneWeaver. The analysis that led to its identification consisted of 122 empirically derived genomic data sets spanning 21 publications. The efficient algorithm and database features within GeneWeaver resulted in rapid generation of a prioritized or ranked list of positional candidates within the cocaine-induced locomotor activation QTL, with *Rab3b* ranked first. Furthermore, no single study or line of evidence pointed towards *Rab3b* as the likely QTG. It was the convergence of suggestive results from these multiple independent studies that led to the nomination of *Rab3b*. This is an important point to note, since often times (and naturally) published results only contain significant findings with suggestive findings being either discarded or overlooked. In a field such as QTL mapping wherein suggestive evidence is acceptable, it is important to utilize all available information (significant or suggestive) for finding QTG underlying suggestive QTL. However, the results of such integrative analysis should be interpreted with extreme caution to ensure the validity of such findings. This

responsibility lies entirely in the hands of the researcher and must be undertaken prior to embarking upon any kind of biological validation of *in silico* findings.

The future of QTG nomination and validation

As advanced mouse genetic resources continue to become available, finding genes that underlie complex traits will undoubtedly occur at a faster pace with increased success. For example, the generation of the CC RI panel is nearing completion. As of 2011, approximately 100 CC strains were to be derived (211). The CC, as preliminary studies have revealed, will certainly improve QTL mapping resolution. Furthermore, mapping populations derived from the CC, such as the Diversity Outbred (DO), will also serve as a high-resolution mapping resource. In fact, QTL mapping studies in the DO, like the CC, have shown considerable improvements in QTL mapping resolution (125). Improved QTL mapping resolution results in fewer positional candidate genes, making the task of QTG nomination feasible. Applying the approach of integrative functional genomics to these positional candidate genes will aid in the rapid and efficient nomination highly relevant QTG. These putative QTG can then be readily validated using targeted transgenic mouse technologies. This step will be greatly expedited, in larger part due to the efforts of the Knockout Mouse Project (124) and the International Mouse Knockout Consortium (45), that are in the process of generating knockouts for all protein coding genes. These resources, in concert with high-throughput sequencing technologies such as RNA-Seq, will further aid

in the nomination of highly relevant QTG, facilitating an improved understanding of the genetic basis of complex traits.

REFERENCES

1. The Allen Brain Atlas. 2011; Available from: <http://www.brain-map.org>.
2. The Institute for Genome Research. 2011; Available from: <http://www.tigr.org>.
3. The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res* 38(Database issue): D331-335, 2010.
4. **Abed, Y.Y., et al.**, Biological reconstruction after resection of bone tumours around the knee: long-term follow-up. *J Bone Joint Surg Br* 91(10): 1366-1372, 2009.
5. **Abiola, O., et al.**, The nature and identification of quantitative trait loci: a community's view. *Nat Rev Genet* 4(11): 911-916, 2003.
6. **Achermann, P. and A.A. Borbely**, Mathematical models of sleep regulation *Front. Biosci.* 8: s683-s693 2003.
7. **Ashurst, J.L., et al.**, The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res* 33(Database issue): D459-465, 2005.
8. **Aylor, D.L., et al.**, Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Res* 21(8): 1213-1222, 2011.
9. **Ayroles, J.F., et al.**, Systems genetics of complex traits in *Drosophila melanogaster*. *Nat Genet* 41(3): 299-307, 2009.
10. **Baker, E.J., et al.**, MuTrack: a genome analysis system for large-scale mutagenesis in the mouse. *BMC Bioinformatics* 5: 11, 2004.
11. **Baker, E.J., et al.**, GeneWeaver: a web-based system for integrative functional genomics. *Nucleic Acids Res*, 2011.
12. **Baker, E.J., et al.**, Ontological Discovery Environment: a system for integrating gene-phenotype associations. *Genomics* 94(6): 377-387, 2009.
13. **Bear, B., et al.**, Small molecules targeting sodium and calcium channels for neuropathic pain. *Curr Opin Drug Discov Devel* 12(4): 543-561, 2009.
14. **Becker, J.B. and M. Hu**, Sex differences in drug abuse. *Front Neuroendocrinol* 29(1): 36-47, 2008.
15. **Beckmann, J.S. and M. Soller**, Detection of linkage between marker loci affecting quantitative traits in crosses between segregating populations. *Theoretical and Applied Genetics* 76: 228-236, 1988.
16. **Belknap, J.K.**, Effect of within-strain sample size on QTL detection and mapping using recombinant inbred mouse strains. *Behav Genet* 28(1): 29-38, 1998.
17. **Belknap, J.K. and A.L. Atkins**, The replicability of QTLs for murine alcohol preference drinking behavior across eight independent studies. *Mamm Genome* 12(12): 893-899, 2001.
18. **Belknap, J.K., et al.**, Type I and type II error rates for quantitative trait loci (QTL) mapping studies using recombinant inbred mouse strains. *Behav Genet* 26(2): 149-160, 1996.

19. **Bennett, B., et al.**, Genetic dissection of quantitative trait loci specifying sedative/hypnotic sensitivity to ethanol: mapping with interval-specific congenic recombinant lines. *Alcohol Clin Exp Res* 26(11): 1615-1624, 2002.
20. **Boone, E.M., et al.**, Genetic regulation of hypothalamic cocaine and amphetamine-regulated transcript (CART) in BxD inbred mice. *Brain Res* 1194: 1-7, 2008.
21. **Bottomly, D., et al.**, Evaluating gene expression in C57BL/6J and DBA/2J mouse striatum using RNA-Seq and microarrays. *PLoS One* 6(3): e17820, 2011.
22. **Boyartchuk, V.L., et al.**, Multigenic control of *Listeria monocytogenes* susceptibility in mice. *Nat Genet* 27(3): 259-260, 2001.
23. **Brazma, A., et al.**, ArrayExpress--a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 31(1): 68-71, 2003.
24. **Brigman, J.L., et al.**, Genetic relationship between anxiety-related and fear-related behaviors in BXD recombinant inbred mice. *Behav Pharmacol* 20(2): 204-209, 2009.
25. **Broman, K.W.**, The genomes of recombinant inbred lines. *Genetics* 169(2): 1133-1146, 2005.
26. **Broman, K.W. and S. Sen**, A Guide to QTL Mapping with R/qtl. Statistics for Biology and Health 2009: Springer.
27. **Broman, K.W., et al.**, R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19(7): 889-890, 2003.
28. **Brown, A.C., et al.**, ExQuest, a novel method for displaying quantitative gene expression from ESTs. *Genomics* 83(3): 528-539, 2004.
29. **Buckler, E.S., et al.**, The genetic architecture of maize flowering time. *Science* 325(5941): 714-718, 2009.
30. **Burgess-Herbert, S.L., et al.**, Practical applications of the bioinformatics toolbox for narrowing quantitative trait loci. *Genetics* 180(4): 2227-2235, 2008.
31. **Chen, S., et al.**, Neurotrimin expression during cerebellar development suggests roles in axon fasciculation and synaptogenesis. *J Neurocytol* 30(11): 927-937, 2001.
32. **Chesler, E.J., et al.**, Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat Genet* 37(3): 233-242, 2005.
33. **Chesler, E.J., et al.**, WebQTL: rapid exploratory analysis of gene expression and genetic networks for brain and behavior. *Nat Neurosci* 7(5): 485-486, 2004.
34. **Chesler, E.J., et al.**, The Collaborative Cross at Oak Ridge National Laboratory: developing a powerful resource for systems genetics. *Mamm Genome* 19(6): 382-389, 2008.
35. **Chesler, E.J., S.L. Rodriguez-Zas, and J.S. Mogil**, In silico mapping of mouse quantitative trait loci. *Science* 294(5551): 2423, 2001.

36. **Chesler, E.J., et al.**, Genetic correlates of gene expression in recombinant inbred strains: a relational model system to explore neurobehavioral phenotypes. *Neuroinformatics* 1(4): 343-357, 2003.
37. **Chesler, E.J., et al.**, Influences of laboratory environment on behavior. *Nat Neurosci* 5(11): 1101-1102, 2002.
38. **Chesler, E.J., et al.**, Identification and ranking of genetic and laboratory environment factors influencing a behavioral trait, thermal nociception, via computational analysis of a large data archive. *Neurosci Biobehav Rev* 26(8): 907-923, 2002.
39. **Cheverud, J.M., et al.**, Healing quantitative trait loci in a combined cross analysis using related mouse strain crosses. *Heredity (Edinb)*, 2011.
40. **Chung, C.Y., et al.**, Functional enhancement and protection of dopaminergic terminals by RAB3B overexpression. *Proc Natl Acad Sci U S A* 106(52): 22474-22479, 2009.
41. **Churchill, G.** The Diversity Outbred. 2011 [cited 2012 1/05/]; Available from: <http://jaxmice.jax.org/jaxnotes/514/514b.html>.
42. **Churchill, G.A., et al.**, The Collaborative Cross, a community resource for the genetic analysis of complex traits. *Nat Genet* 36(11): 1133-1137, 2004.
43. **Churchill, G.A. and R.W. Doerge**, Empirical threshold values for quantitative trait mapping. *Genetics* 138(3): 963-971, 1994.
44. **Cicila, G.T., et al.**, High-resolution mapping of the blood pressure QTL on chromosome 7 using Dahl rat congenic strains. *Genomics* 72(1): 51-60, 2001.
45. **Collins, F.S., J. Rossant, and W. Wurst**, A mouse for all reasons. *Cell* 128(1): 9-13, 2007.
46. **Coppola, T., et al.**, Disruption of Rab3-calmodulin interaction, but not other effector interactions, prevents Rab3 inhibition of exocytosis. *EMBO J* 18(21): 5885-5891, 1999.
47. **Cox, A., et al.**, Combining QTL data for HDL cholesterol levels from two different species leads to smaller confidence intervals. *Heredity (Edinb)* 105(5): 426-432, 2010.
48. **Crabbe, J.C., et al.**, Polygenic and single-gene determination of responses to ethanol in BXD/Ty recombinant inbred mouse strains. *Neurobehav Toxicol Teratol* 5(2): 181-187, 1983.
49. **Darvasi, A.**, In silico mapping of mouse quantitative trait loci. *Science* 294(5551): 2423, 2001.
50. **Darvasi, A.**, Experimental strategies for the genetic dissection of complex traits in animal models. *Nat Genet* 18(1): 19-24, 1998.
51. **Darvasi, A. and M. Soller**, Advanced intercross lines, an experimental population for fine genetic mapping. *Genetics* 141(3): 1199-1207, 1995.
52. **De Luca, M., et al.**, Dopa decarboxylase (Ddc) affects variation in *Drosophila* longevity. *Nat Genet* 34(4): 429-433, 2003.

53. **Demarest, K., et al.**, Further characterization and high-resolution mapping of quantitative trait loci for ethanol-induced locomotor activity. *Behav Genet* 31(1): 79-91, 2001.
54. **Demarest, K., et al.**, Identification of an acute ethanol response quantitative trait locus on mouse chromosome 2. *J Neurosci* 19(2): 549-561, 1999.
55. **Denny, P., et al.**, Mapping of the IDDM locus Idd3 to a 0.35-cM interval containing the interleukin-2 gene. *Diabetes* 46(4): 695-700, 1997.
56. **DiPetrillo, K., et al.**, Genetic analysis of blood pressure in C3H/HeJ and SWR/J mice. *Physiol Genomics* 17(2): 215-220, 2004.
57. **DiPetrillo, K., et al.**, Bioinformatics toolbox for narrowing rodent quantitative trait loci. *Trends Genet* 21(12): 683-692, 2005.
58. **Doerge, R.W. and G.A. Churchill**, Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142(1): 285-294, 1996.
59. **Douglass, J. and S. Daoud**, Characterization of the human cDNA and genomic DNA encoding CART: a cocaine- and amphetamine-regulated transcript. *Gene* 169(2): 241-245, 1996.
60. **Dukhanina, O.I., et al.**, Linkage map and congenic strains to localize blood pressure QTL on rat chromosome 10. *Mamm Genome* 8(4): 229-235, 1997.
61. **Dupuis, J. and D. Siegmund**, Statistical methods for mapping quantitative trait loci from a dense set of markers. *Genetics* 151(1): 373-386, 1999.
62. **Durrant, C., et al.**, Collaborative Cross mice and their power to map host susceptibility to *Aspergillus fumigatus* infection. *Genome Res* 21(8): 1239-1248, 2011.
63. **Dussor, G.O., et al.**, Cholinergic modulation of nociceptive responses in vivo and neuropeptide release in vitro at the level of the primary sensory neuron. *Pain* 107(1-2): 22-32, 2004.
64. **Edgar, R., M. Domrachev, and A.E. Lash**, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30(1): 207-210, 2002.
65. **Einat, H.**, Different behaviors and different strains: potential new ways to model bipolar disorder. *Neurosci Biobehav Rev* 31(6): 850-857, 2007.
66. **Eisener-Dorman, A.F., et al.**, Quantitative trait locus and haplotype mapping in closely related inbred strains identifies a locus for open field behavior. *Mamm Genome* 21(5-6): 231-246, 2010.
67. **Fawcett, G.L., et al.**, Fine-mapping of obesity-related quantitative trait loci in an F9/10 advanced intercross line. *Obesity (Silver Spring)* 18(7): 1383-1392, 2010.
68. **Fehr, C., et al.**, Congenic mapping of alcohol and pentobarbital withdrawal liability loci to a <1 centimorgan interval of murine chromosome 4: identification of Mpdz as a candidate gene. *J Neurosci* 22(9): 3730-3738, 2002.

69. **Feng, G.H., et al.**, Mouse pale ear (ep) is homologous to human Hermansky-Pudlak syndrome and contains a rare 'AT-AC' intron. *Hum Mol Genet* 6(5): 793-797, 1997.
70. **Fernandez, J.R., et al.**, Sex-exclusive quantitative trait loci influences in alcohol-related phenotypes. *Am J Med Genet* 88(6): 647-652, 1999.
71. **Ferraro, T.N., et al.**, Fine mapping of a seizure susceptibility locus on mouse Chromosome 1: nomination of Kcnj10 as a causative gene. *Mamm Genome* 15(4): 239-251, 2004.
72. **Fischer von Mollard, G., et al.**, rab3 is a small GTP-binding protein exclusively localized to synaptic vesicles. *Proc Natl Acad Sci U S A* 87(5): 1988-1992, 1990.
73. **Flint, J.**, Analysis of quantitative trait loci that influence animal behavior. *J Neurobiol* 54(1): 46-77, 2003.
74. **Flint, J., et al.**, Strategies for mapping and cloning quantitative trait genes in rodents. *Nat Rev Genet* 6(4): 271-286, 2005.
75. **Fuller, J.L.**, Measurement of Alcohol Preference in Genetic Experiments. *J Comp Physiol Psychol* 57: 85-88, 1964.
76. **Gagliani, S.M., et al.**, The genetic control of neocortex volume and covariation with neocortical gene expression in mice. *BMC Neurosci* 10: 44, 2009.
77. **Glass, G.V.**, Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher* 5(10): 3, 1976.
78. **Glazier, A.M., J.H. Nadeau, and T.J. Aitman**, Finding genes that underlie complex traits. *Science* 298(5602): 2345-2349, 2002.
79. **Goldowitz, D., et al.**, Large-scale mutagenesis of the mouse to understand the genetic bases of nervous system structure and function. *Brain Res Mol Brain Res* 132(2): 105-115, 2004.
80. **Grabe, N.** AliBaba2.1. 2012; Available from: <http://www.gene-regulation.com/pub/programs/alibaba2/index.html>.
81. **Gribkoff, V.K.**, The therapeutic potential of neuronal K V 7 (KCNQ) channel modulators: an update. *Expert Opin Ther Targets* 12(5): 565-581, 2008.
82. **Grisel, J.E., et al.**, Quantitative trait loci affecting methamphetamine responses in BXD recombinant inbred mouse strains. *J Neurosci* 17(2): 745-754, 1997.
83. **Grupe, A., et al.**, In silico mapping of complex disease-related traits in mice. *Science* 292(5523): 1915-1918, 2001.
84. **Guo, A.Y., et al.**, ERGR: An ethanol-related gene resource. *Nucleic Acids Res* 37(Database issue): D840-845, 2009.
85. **Gurkan, C., et al.**, Large-scale profiling of Rab GTPase trafficking networks: the membrome. *Mol Biol Cell* 16(8): 3847-3864, 2005.
86. **Hanlon, E.C., et al.**, Effects of skilled training on sleep slow wave activity and cortical gene expression in the rat. *Sleep* 32(6): 719-729, 2009.

87. **Henderson, N.D., et al.**, QTL analysis of multiple behavioral measures of anxiety in mice. *Behav Genet* 34(3): 267-293, 2004.
88. **Hinrichs, D., et al.**, Analysis of inbreeding depression in the first litter size of mice in a long-term selection experiment with respect to the age of the inbreeding. *Heredity* 99(1): 81-88, 2007.
89. **Hirsch, J.**, Behavior-Genetic Analysis 1967, New York, NY: McGraw-Hill.
90. **Hitzemann, R., et al.**, Multiple cross mapping (MCM) markedly improves the localization of a QTL for ethanol-induced activation. *Genes Brain Behav* 1(4): 214-222, 2002.
91. **Holt, M., T. Meuwissen, and O. Vangen**, Long-term responses, changes in genetic variances and inbreeding depression from 122 generations of selection on increased litter size in mice. *J Anim Breed Genet* 122(3): 199-209, 2005.
92. **Holz, R.W., et al.**, Evidence for the involvement of Rab3A in Ca(2+)-dependent exocytosis from adrenal chromaffin cells. *J Biol Chem* 269(14): 10229-10234, 1994.
93. **Iezzi, M., et al.**, Subcellular distribution and function of Rab3A, B, C, and D isoforms in insulin-secreting cells. *Mol Endocrinol* 13(2): 202-212, 1999.
94. **Iraqi, F.**, Fine mapping of quantitative trait loci using advanced intercross lines of mice and positional cloning of the corresponding genes. *Exp Lung Res* 26(8): 641-649, 2000.
95. **Iraqi, F.A., G. Churchill, and R. Mott**, The Collaborative Cross, developing a resource for mammalian systems genetics: a status report of the Wellcome Trust cohort. *Mamm Genome* 19(6): 379-381, 2008.
96. **Jagodic, M. and T. Olsson**, Combined-cross analysis of genome-wide linkage scans for experimental autoimmune encephalomyelitis in rat. *Genomics* 88(6): 737-744, 2006.
97. **Johannes, L., et al.**, The GTPase Rab3a negatively controls calcium-dependent exocytosis in neuroendocrine cells. *EMBO J* 13(9): 2029-2037, 1994.
98. **Johnson, R.A. and D.W. Wichern**, Applied Multivariate Statistical Analysis. 1998, Upper Saddle River, NJ.: Prentice Hall.
99. **Johnson, T.E. and W.B. Wood**, Genetic analysis of life-span in *Caenorhabditis elegans*. *Proc Natl Acad Sci U S A* 79(21): 6603-6607, 1982.
100. **Jones, B.C., et al.**, Quantitative-trait loci analysis of cocaine-related behaviours and neurochemistry. *Pharmacogenetics* 9(5): 607-617, 1999.
101. **Kanehisa, M. and S. Goto**, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28(1): 27-30, 2000.
102. **Kanehisa, M., et al.**, KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40(Database issue): D109-114, 2012.
103. **Kang, H.M., et al.**, Efficient control of population structure in model organism association mapping. *Genetics* 178(3): 1709-1723, 2008.

104. **Kao, C.H. and M.H. Zeng**, A study on the mapping of quantitative trait loci in advanced populations derived from two inbred lines. *Genet Res (Camb)* 91(2): 85-99, 2009.
105. **Keane, T.M., et al.**, Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature* 477(7364): 289-294, 2011.
106. **Kempermann, G., et al.** Hippocampal gene networks in BXD mice. . in *5th Annual Meeting of the Complex Trait Consortium*. . 2006. Chapel Hill, North Carolina.
107. **Korstanje, R. and B. Paigen**, From QTL to gene: the harvest begins. *Nat Genet* 31(3): 235-236, 2002.
108. **Kover, P.X., et al.**, A Multiparent Advanced Generation Inter-Cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* 5(7): e1000551, 2009.
109. **Kozell, L.B., et al.**, Mapping a barbiturate withdrawal locus to a 0.44 Mb interval and analysis of a novel null mutant identify a role for Kcnj9 (GIRK3) in withdrawal from pentobarbital, zolpidem, and ethanol. *J Neurosci* 29(37): 11662-11673, 2009.
110. **Krueger, J.M., et al.**, Sleep as a fundamental property of neuronal assemblies. *Nat Rev Neurosci*. 9(12): 910-919, 2008.
111. **Kurian, S.M., et al.**, Identification of blood biomarkers for psychosis using convergent functional genomics. *Mol Psychiatry* 16(1): 37-58, 2011.
112. **LaCroix-Fralish, M.L., et al.**, Patterns of pain: meta-analysis of microarray studies of pain. *Pain* 152(8): 1888-1898, 2011.
113. **Lander, E. and L. Kruglyak**, Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 11(3): 241-247, 1995.
114. **Lander, E.S. and D. Botstein**, Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121(1): 185-199, 1989.
115. **Lariviere, W.R., et al.**, Heritability of nociception. III. Genetic relationships among commonly used assays of nociception and hypersensitivity. *Pain* 97(1-2): 75-86, 2002.
116. **Le-Niculescu, H., et al.**, Convergent functional genomics of genome-wide association data for bipolar disorder: comprehensive identification of candidate genes, pathways and mechanisms. *Am J Med Genet B Neuropsychiatr Genet* 150B(2): 155-181, 2009.
117. **Le-Niculescu, H., S.D. Patel, and A.B. Niculescu**, Convergent integration of animal model and human studies of bipolar disorder (manic-depressive illness). *Curr Opin Pharmacol* 10(5): 594-600, 2010.
118. **Lein, E.S., et al.**, Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445(7124): 168-176, 2007.
119. **Leussis, M.P. and V.J. Bolivar**, Habituation in rodents: a review of behavior, neurobiology, and genetics. *Neurosci Biobehav Rev* 30(7): 1045-1064, 2006.

120. **Li, N., et al.**, Involvement of CGRP and CGRPI receptor in nociception in the basolateral nucleus of amygdala of rats. *Neurosci Lett* 443(3): 184-187, 2008.
121. **Li, R., et al.**, Combining data from multiple inbred line crosses improves the power and resolution of quantitative trait loci mapping. *Genetics* 169(3): 1699-1709, 2005.
122. **Li, Y., et al.**, Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet* 2(12): e222, 2006.
123. **Liu, P., et al.**, Candidate lung tumor susceptibility genes identified through whole-genome association analyses in inbred mice. *Nat Genet* 38(8): 888-895, 2006.
124. **Lloyd, K.C.**, A knockout mouse resource for the biomedical research community. *Ann N Y Acad Sci* 1245(1): 24-26, 2011.
125. **Logan, R.W., et al.**, High-precision QTL mapping of complex behavioral traits in the Diversity Outbred (DO) mouse population. *Unpublished*, 2012.
126. **Long, A.D., et al.**, Genetic interactions between naturally occurring alleles at quantitative trait loci and mutant alleles at candidate loci affecting bristle number in *Drosophila melanogaster*. *Genetics* 144(4): 1497-1510, 1996.
127. **Lynch, M. and B. Walsh**, Genetics and Analysis of Quantitative Traits 1998, Sunderland, MA: Sinauer Associates.
128. **Machado-Vieira, R., F. Kapczinski, and J.C. Soares**, Perspectives for the development of animal models of bipolar disorder. *Prog Neuropsychopharmacol Biol Psychiatry* 28(2): 209-224, 2004.
129. **Mackay, T.F.**, The genetic architecture of quantitative traits. *Annu Rev Genet* 35: 303-339, 2001.
130. **Mackay, T.F., E.A. Stone, and J.F. Ayroles**, The genetics of quantitative traits: challenges and prospects. *Nat Rev Genet* 10(8): 565-577, 2009.
131. **Manenti, G., et al.**, Outbred CD-1 mice carry the susceptibility allele at the pulmonary adenoma susceptibility 1 (Pas1) locus. *Carcinogenesis* 24(6): 1143-1148, 2003.
132. **Manly, K.F., R.H. Cudmore, Jr., and J.M. Meer**, Map Manager QTX, cross-platform software for genetic mapping. *Mamm Genome* 12(12): 930-932, 2001.
133. **Martinez, O. and B. Goud**, Rab proteins. *Biochim Biophys Acta* 1404(1-2): 101-112, 1998.
134. **Matin, A., et al.**, Susceptibility to testicular germ-cell tumours in a 129.MOLF-Chr 19 chromosome substitution strain. *Nat Genet* 23(2): 237-240, 1999.
135. **McCall, R.D. and D. Frierson, Jr.**, Evidence that two loci predominantly determine the difference in susceptibility to the high pressure neurologic syndrome type I seizure in mice. *Genetics* 99(2): 285-307, 1981.
136. **McClearn, G.E., J.R. Wilson, and W. Meredith**, The use of isogenic and heterogenic mouse stocks in behavioral research. Contributions to

- behavior-genetic analysis. The mouse as a prototype ed. G. Lindzey and D.D. Thiessen 1970: New York: Appleton-Century-Crofts.
137. **Mignery, G.A., et al.**, Structure and expression of the rat inositol 1,4,5-trisphosphate receptor. *J Biol Chem* 265(21): 12679-12685, 1990.
 138. **Milner, L.C. and K.J. Buck**, Identifying quantitative trait loci (QTLs) and genes (QTGs) for alcohol-related phenotypes in mice. *Int Rev Neurobiol* 91: 173-204, 2010.
 139. **Mitra, R., et al.**, Social stress-related behavior affects hippocampal cell proliferation in mice. *Physiol Behav* 89(2): 123-127, 2006.
 140. **Mogil, J.S., et al.**, Identification of a sex-specific quantitative trait locus mediating nonopioid stress-induced analgesia in female mice. *J Neurosci* 17(20): 7995-8002, 1997.
 141. **Moore, K.J.**, Insight into the microphthalmia gene. *Trends Genet* 11(11): 442-448, 1995.
 142. **Morahan, G., L. Balmer, and D. Monley**, Establishment of "The Gene Mine": a resource for rapid identification of complex trait genes. *Mamm Genome* 19(6): 390-393, 2008.
 143. **Moreau, Y., et al.**, Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet* 19(10): 570-577, 2003.
 144. **Mott, R. and J. Flint**, Simultaneous detection and fine mapping of quantitative trait loci in mice using heterogeneous stocks. *Genetics* 160(4): 1609-1618, 2002.
 145. **Mott, R., et al.**, A method for fine mapping quantitative trait loci in outbred animal stocks. *Proc Natl Acad Sci U S A* 97(23): 12649-12654, 2000.
 146. **Mulligan, M.K., et al.**, Toward understanding the genetics of alcohol drinking through transcriptome meta-analysis. *Proc Natl Acad Sci U S A* 103(16): 6368-6373, 2006.
 147. **Mulligan, M.K., et al.**, Molecular profiles of drinking alcohol to intoxication in C57BL/6J mice. *Alcohol Clin Exp Res* 35(4): 659-670, 2011.
 148. **Nadeau, J.H. and W.N. Frankel**, The roads from phenotypic variation to gene discovery: mutagenesis versus QTLs. *Nat Genet* 25(4): 381-384, 2000.
 149. **Nadeau, J.H., et al.**, Analysing complex genetic traits with chromosome substitution strains. *Nat Genet* 24(3): 221-225, 2000.
 150. **Niculescu, A.B., 3rd, et al.**, Identifying a series of candidate genes for mania and psychosis: a convergent functional genomics approach. *Physiol Genomics* 4(1): 83-91, 2000.
 151. **Niculescu, A.B. and H. Le-Niculescu**, Convergent Functional Genomics: what we have learned and can learn about genes, pathways, and mechanisms. *Neuropsychopharmacology* 35(1): 355-356, 2010.
 152. **Nonet, M.L., et al.**, *Caenorhabditis elegans* rab-3 mutant synapses exhibit impaired function and are partially depleted of vesicles. *J Neurosci* 17(21): 8061-8073, 1997.

153. **Norgard, E.A., et al.**, Replication of long-bone length QTL in the F9-F10 LG,SM advanced intercross. *Mamm Genome* 20(4): 224-235, 2009.
154. **Novick, P. and M. Zerial**, The diversity of Rab proteins in vesicle transport. *Curr Opin Cell Biol* 9(4): 496-504, 1997.
155. **O'Brien, S.J., et al.**, The promise of comparative genomics in mammals. *Science* 286(5439): 458-462, 479-481, 1999.
156. **Ogorevc, J., P. Dovc, and T. Kunej**, Comparative genomics approach to identify candidate genetic loci for male fertility. *Reprod Domest Anim* 46(2): 229-239, 2011.
157. **Oishi, H., et al.**, Localization of the Rab3 small G protein regulators in nerve terminals and their involvement in Ca²⁺-dependent exocytosis. *J Biol Chem* 273(51): 34580-34585, 1998.
158. **Park, Y.G., et al.**, Multiple cross and inbred strain haplotype mapping of complex-trait candidate genes. *Genome Res* 13(1): 118-121, 2003.
159. **Patel, S.D., et al.**, Coming to grips with complex disorders: genetic risk prediction in bipolar disorder using panels of genes identified through convergent functional genomics. *Am J Med Genet B Neuropsychiatr Genet* 153B(4): 850-877, 2010.
160. **Peirce, J.L., et al.**, Genome Reshuffling for Advanced Intercross Permutation (GRAIP): simulation and permutation for advanced intercross population analysis. *PLoS One* 3(4): e1977, 2008.
161. **Peirce, J.L., et al.**, How replicable are mRNA expression QTL? *Mamm Genome* 17(6): 643-656, 2006.
162. **Peirce, J.L., et al.**, A new set of BXD recombinant inbred lines from advanced intercross populations in mice. *BMC Genet* 5: 7, 2004.
163. **Pennacchio, L.A.**, Insights from human/mouse genome comparisons. *Mamm Genome* 14(7): 429-436, 2003.
164. **Philip, V.M., et al.**, High-throughput behavioral phenotyping in the expanded panel of BXD recombinant inbred strains. *Genes Brain Behav* 9(2): 129-159, 2010.
165. **Philip, V.M., et al.**, Genetic analysis in the Collaborative Cross breeding population. *Genome Res* 21(8): 1223-1238, 2011.
166. **Phillips, T.J., et al.**, Localization of genes affecting alcohol drinking in mice. *Alcohol Clin Exp Res* 18(4): 931-941, 1994.
167. **Pletcher, M.T., et al.**, Use of a dense single nucleotide polymorphism map for in silico mapping in the mouse. *PLoS Biol* 2(12): e393, 2004.
168. **Plomin, R., et al.**, Use of recombinant inbred strains to detect quantitative trait loci associated with behavior. *Behav Genet* 21(2): 99-116, 1991.
169. **Podolin, P.L., et al.**, Localization of two insulin-dependent diabetes (Idd) genes to the Idd10 region on mouse chromosome 3. *Mamm Genome* 9(4): 283-286, 1998.
170. **Podolin, P.L., et al.**, Congenic mapping of the insulin-dependent diabetes (Idd) gene, Idd10, localizes two genes mediating the Idd10 effect and eliminates the candidate Fcgr1. *J Immunol* 159(4): 1835-1843, 1997.

171. **Regazzi, R., et al.**, Expression, localization and functional role of small GTPases of the Rab3 family in insulin-secreting cells. *J Cell Sci* 109 (Pt 9): 2265-2273, 1996.
172. **Rhodes, D.R., et al.**, Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res* 62(15): 4427-4433, 2002.
173. **Rhodes, D.R., et al.**, Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 101(25): 9309-9314, 2004.
174. **Riordan, J.R., et al.**, Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. *Science* 245(4922): 1066-1073, 1989.
175. **Roberts, A., et al.**, The polymorphism architecture of mouse genetic resources elucidated using genome-wide resequencing data: implications for QTL discovery and systems genetics. *Mamm Genome* 18(6-7): 473-481, 2007.
176. **Rockman, M.V. and L. Kruglyak**, Recombinational landscape and population genomics of *Caenorhabditis elegans*. *PLoS Genet* 5(3): e1000419, 2009.
177. **Rodgers, D.A. and C.G. Mc**, Mouse strain differences in preference for various concentrations of alcohol. *Q J Stud Alcohol* 23: 26-33, 1962.
178. **Rosen, G.D., et al.**, Genetic modulation of striatal volume by loci on Chrs 6 and 17 in BXD recombinant inbred mice. *Genes Brain Behav* 8(3): 296-308, 2009.
179. **Saccone, S.F., et al.**, Supplementing high-density SNP microarrays for additional coverage of disease-related genes: addiction as a paradigm. *PLoS One* 4(4): e5225, 2009.
180. **Samocha, K.E., et al.**, Fine mapping of QTL for prepulse inhibition in LG/J and SM/J mice using F(2) and advanced intercross lines. *Genes Brain Behav* 9(7): 759-767, 2010.
181. **Sax, K.**, The Association of Size Differences with Seed-Coat Pattern and Pigmentation in PHASEOLUS VULGARIS. *Genetics* 8(6): 552-560, 1923.
182. **Schafer, J. and K. Strimmer**, A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat Appl Genet Mol Biol* 4: Article32, 2005.
183. **Schluter, O.M., et al.**, Localization versus function of Rab3 proteins. Evidence for a common regulatory role in controlling fusion. *J Biol Chem* 277(43): 40919-40929, 2002.
184. **Schluter, O.M., et al.**, A complete genetic analysis of neuronal Rab3 function. *J Neurosci* 24(29): 6629-6637, 2004.
185. **Schmidt, C., et al.**, A meta-analysis of QTL for diabetes-related traits in rodents. *Physiol Genomics* 34(1): 42-53, 2008.

186. **Seyfried, T.N. and G.H. Glaser**, Genetic linkage between the AH locus and a major gene that inhibits susceptibility to audiogenic seizures in mice. *Genetics* 99(1): 117-126, 1981.
187. **Shifman, S., et al.**, A high-resolution single nucleotide polymorphism genetic map of the mouse genome. *PLoS Biol* 4(12): e395, 2006.
188. **Shifman, S. and A. Darvasi**, Mouse inbred strain sequence information and yin-yang crosses for quantitative trait locus fine mapping. *Genetics* 169(2): 849-854, 2005.
189. **Shirley, R.L., et al.**, Mpdz is a quantitative trait gene for drug withdrawal seizures. *Nat Neurosci* 7(7): 699-700, 2004.
190. **Sidhu, R.S. and R.P. Bhullar**, Rab3B in human platelet is membrane bound and interacts with Ca(2+)/calmodulin. *Biochem Biophys Res Commun* 289(5): 1039-1043, 2001.
191. **Silver, L.M.**, Mouse Genetics: Concept and Applications 1995: Oxford University Press.
192. **Singer, J.B., et al.**, Genetic dissection of complex traits with chromosome substitution strains of mice. *Science* 304(5669): 445-448, 2004.
193. **Singer, J.B., et al.**, Mapping quantitative trait loci for anxiety in chromosome substitution strains of mice. *Genetics* 169(2): 855-862, 2005.
194. **Smith, C.L. and J.T. Eppig**, The mammalian phenotype ontology: enabling robust annotation and comparative analysis. *Wiley Interdiscip Rev Syst Biol Med* 1(3): 390-399, 2009.
195. **Smith, J.D., et al.**, In silico quantitative trait locus map for atherosclerosis susceptibility in apolipoprotein E-deficient mice. *Arterioscler Thromb Vasc Biol* 23(1): 117-122, 2003.
196. **Steinmetz, L.M., et al.**, Dissecting the architecture of a quantitative trait locus in yeast. *Nature* 416(6878): 326-330, 2002.
197. **Stenmark, H. and V.M. Olkkonen**, The Rab GTPase family. *Genome Biol* 2(5): REVIEWS3007, 2001.
198. **Storey, J.D.**, A direct approach to false discovery rates. *Journal of the Royal Statistical Society: Series B* 64(3): 479 - 498, 2002.
199. **Stridh, P., et al.**, Fine-mapping resolves Eae23 into two QTLs and implicates ZEB1 as a candidate gene regulating experimental neuroinflammation in rat. *PLoS One* 5(9): e12716, 2010.
200. **Su, A.I., et al.**, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* 101(16): 6062-6067, 2004.
201. **Su, Z., et al.**, Farp2 and Stk25 are candidate genes for the HDL cholesterol locus on mouse chromosome 1. *Arterioscler Thromb Vasc Biol* 29(1): 107-113, 2009.
202. **Swiger, L.A., et al.**, The variance of intraclass correlation involving groups with one observation. *Biometrics* 20: 818 - 826, 1964.
203. **Takai, Y., T. Sasaki, and T. Matozaki**, Small GTP-binding proteins. *Physiol Rev* 81(1): 153-208, 2001.

204. **Talbot, C.J., et al.**, High-resolution mapping of quantitative trait loci in outbred mice. *Nat Genet* 21(3): 305-308, 1999.
205. **Talbot, C.J., et al.**, Fine scale mapping of a genetic locus for conditioned fear. *Mamm Genome* 14(4): 223-230, 2003.
206. **Tang, P.L., et al.**, Genome-wide haplotype association mapping in mice identifies a genetic variant in CER1 associated with BMD and fracture in southern Chinese women. *J Bone Miner Res* 24(6): 1013-1021, 2009.
207. **Tartaglia, L.A., et al.**, Identification and expression cloning of a leptin receptor, OB-R. *Cell* 83(7): 1263-1271, 1995.
208. **Taylor, B.A., H.G. Bedigian, and H. Meier**, Genetic studies of the Fv-1 locus of mice: linkage with Gpd-1 in recombinant inbred lines. *J Virol* 23(1): 106-109, 1977.
209. **Taylor, B.A., et al.**, Genotyping new BXD recombinant inbred mouse strains and comparison of BXD and consensus maps. *Mamm Genome* 10(4): 335-348, 1999.
210. **Threadgill, D.W.**, Meeting report for the 4th annual Complex Trait Consortium meeting: from QTLs to systems genetics. *Mamm Genome* 17(1): 2-4, 2006.
211. **Threadgill, D.W., et al.**, The collaborative cross: a recombinant inbred mouse population for the systems genetic era. *ILAR J* 52(1): 24-31, 2011.
212. **Trullas, R. and P. Skolnick**, Differences in fear motivated behaviors among inbred mouse strains. *Psychopharmacology (Berl)* 111(3): 323-331, 1993.
213. **Valdar, W., J. Flint, and R. Mott**, Simulating the collaborative cross: power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics* 172(3): 1783-1797, 2006.
214. **Valdar, W., et al.**, Mapping in structured populations by resample model averaging. *Genetics* 182(4): 1263-1277, 2009.
215. **Valdar, W., et al.**, Genetic and environmental effects on complex traits in mice. *Genetics* 174(2): 959-984, 2006.
216. **Wahlsten, D., et al.**, Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proc Natl Acad Sci U S A* 103(44): 16364-16369, 2006.
217. **Walter, N.A., et al.**, High throughput sequencing in mice: a platform comparison identifies a preponderance of cryptic SNPs. *BMC Genomics* 10: 379, 2009.
218. **Wang, J., et al.**, Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment. *Genes Brain Behav* 7(1): 78-87, 2008.
219. **Wang, J., R.W. Williams, and K.F. Manly**, WebQTL: web-based complex trait analysis. *Neuroinformatics* 1(4): 299-308, 2003.

220. **Wang, M., et al.**, Fine mapping and identification of candidate pulmonary adenoma susceptibility 1 genes using advanced intercross lines. *Cancer Res* 63(12): 3317-3324, 2003.
221. **Wang, X., et al.**, Using advanced intercross lines for high-resolution mapping of HDL cholesterol quantitative trait loci. *Genome Res* 13(7): 1654-1664, 2003.
222. **Waterston, R.H., et al.**, Initial sequencing and comparative analysis of the mouse genome. *Nature* 420(6915): 520-562, 2002.
223. **Weller, J.I.**, Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics* 42(3): 627-640, 1986.
224. **Wergedal, J.E., et al.**, Mapping genetic loci that regulate lipid levels in a NZB/B1NJxRF/J intercross and a combined intercross involving NZB/B1NJ, RF/J, MRL/MpJ, and SJL/J mouse strains. *J Lipid Res* 48(8): 1724-1734, 2007.
225. **Whitney, G., G.E. McClearn, and J.C. DeFries**, Heritability of alcohol preference in laboratory mice and rats. *J Hered* 61(4): 165-169, 1970.
226. **Wilming, L.G., et al.**, The vertebrate genome annotation (Vega) database. *Nucleic Acids Res* 36(Database issue): D753-760, 2008.
227. **Workman, C.T., et al.**, enoLOGOS: a versatile web tool for energy normalized sequence logos. *Nucleic Acids Res* 33(Web Server issue): W389-392, 2005.
228. **Wuschke, S., et al.**, A meta-analysis of quantitative trait loci associated with body weight and adiposity in mice. *Int J Obes (Lond)* 31(5): 829-841, 2007.
229. **Yalcin, B., et al.**, Unexpected complexity in the haplotypes of commonly used inbred strains of laboratory mice. *Proc Natl Acad Sci U S A* 101(26): 9734-9739, 2004.
230. **Yalcin, B., et al.**, Genetic dissection of a behavioral quantitative trait locus shows that Rgs2 modulates anxiety in mice. *Nat Genet* 36(11): 1197-1202, 2004.
231. **Yalcin, B., et al.**, Sequence-based characterization of structural variation in the mouse genome. *Nature* 477(7364): 326-329, 2011.
232. **Yang, H., et al.**, On the subspecific origin of the laboratory mouse. *Nat Genet* 39(9): 1100-1107, 2007.
233. **Yang, H., et al.**, Subspecific origin and haplotype diversity in the laboratory mouse. *Nat Genet* 43(7): 648-655, 2011.
234. **Yu, X., et al.**, Fine mapping of collagen-induced arthritis quantitative trait loci in an advanced intercross line. *J Immunol* 177(10): 7042-7049, 2006.
235. **Zhang, B., S. Kirov, and J. Snoddy**, WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33(Web Server issue): W741-748, 2005.

- 236. **Zhang, D., et al.**, Repeated cocaine administration induces gene expression changes through the dopamine D1 receptors. *Neuropsychopharmacology* 30(8): 1443-1454, 2005.
- 237. **Zhang, Q.Y., et al.**, Interval mapping and congenic strains for a blood pressure QTL on rat chromosome 13. *Mamm Genome* 8(9): 636-641, 1997.
- 238. **Zhang, S., et al.**, Fine mapping of a major locus on chromosome 10 for exploratory and fear-like behavior in mice. *Mamm Genome* 16(5): 306-318, 2005.

APPENDIX

Table A1. List of gene sets used in the integrative functional genomic analysis using GeneWeaver (11).

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS1132	Genetic influences on impulsivity, risk taking, stress responsivity and vulnerability to drug abuse and addiction.	16251987	NA
GS1229	Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse.	16710320	Differentially expressed among both cocaine and heroin abusers
GS1231	Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse.	16710320	Upregulated in both cocaine and heroin abusers
GS1232	Distinctive profiles of gene expression in the human nucleus accumbens associated with cocaine and heroin abuse.	16710320	Upregulated in cocaine abusers but downregulated in heroin abusers
GS1256	Gene Expression in Human Hippocampus from Cocaine Abusers Identifies Genes which Regulate Extracellular Matrix Remodeling. Candidate genes, pathways and mechanisms for bipolar	18000554	NA
GS3649	(manic-depressive) and related disorders: an expanded convergent functional genomics approach.	15314610	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS14885	Nicotine causes age-dependent changes in gene expression in the adolescent female rat brain.	17234382	NA
GS14888	Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment.	17504244	Differentially expressed genes modulated by nicotine in five combined brain regions (Amyg, HP, NA, PFC and VTA) for C3H/HeJ mice
GS14890	Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment.	17504244	Downregulated genes in five combined brain regions, Amyg, HP, NA, PFC and VTA, in C57BL/6J mice
GS14891	Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment.	17504244	Upregulated genes in five combined brain regions, Amyg, HP, NA, PFC and VTA, in C57BL/6J mice
GS14892	Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment.	17504244	Downregulated genes in five brain regions, NA, PFC, and VTA, in C3H/HeJ mice
GS14893	Strain- and region-specific gene expression profiles in mouse brain in response to chronic nicotine treatment.	17504244	Upregulated genes in three brain regions, NA, PFC, and VTA, in C3H/HeJ mice
GS14905	Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs.	17407593	Genes implicated in nicotine dependence

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS14906	Molecular genetics of nicotine dependence and abstinence: whole genome association using 520,000 SNPs.	17407593	Genes implicated in smoking cessation
GS14907	Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers.	19009022	Nicotine Dependence Genes found in Genome-wide Association Studies (GWAS)
GS14908	Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers.	19009022	Nicotine Abstinence Genes found in Genome-wide Association Studies (GWAS)
GS14909	Gene profiling the response to repeated cocaine self-administration in dorsal striatum: a focus on circadian genes.	18452895	NA
GS15464	Region-specific transcriptional response to chronic nicotine in rat brain.	11478936	NA
GS18839	Nicotine modulates the expression of a diverse set of genes in the neuronal SH-SY5Y cell line.	12588870	Downregulated Nicotine-dependent genes in human neuroblastoma (SH-SY5Y) cells
GS18841	Nicotine modulates the expression of a diverse set of genes in the neuronal SH-SY5Y cell line.	12588870	Upregulated Nicotine-dependent genes in human neuroblastoma (SH-SY5Y) cells
GS26287	Novel genes identified in a high-density genome wide association study for nicotine dependence.	17158188	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS26304	Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs.	17135278	Genes with SNPs associated with nicotine dependence
GS26305	Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs.	17135278	Nicotine dependence Candidate Genes combined from GWAS and SNP association methods
GS26324	Gene Expression in Human Hippocampus from Cocaine Abusers Identifies Genes which Regulate Extracellular Matrix Remodeling	18000554	NA
GS46978	A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations.	18000554	Nicotine Dependence in Australian DNA-Pools
GS46980	A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations.	18000554	Comorbid Alcohol and Nicotine Dependence in Australian DNA-Pools
GS46981	A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations.	18000554	Meta-Analysis of Nicotine Dependence in Australian and Dutch Samples
GS46983	A genomewide association study of nicotine and alcohol dependence in Australian and Dutch populations.	18000554	Nicotine Dependence in Dutch Sample

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS75633	The dissection of transcriptional modules regulated by various drugs of abuse in the mouse striatum.	20459597	Genes regulated by methamphetamine in mouse striatum, 1h
GS75635	The dissection of transcriptional modules regulated by various drugs of abuse in the mouse striatum.	20459597	Genes regulated by methamphetamine in mouse striatum, 2h
GS75636	The dissection of transcriptional modules regulated by various drugs of abuse in the mouse striatum.	20459597	Genes regulated by methamphetamine in mouse striatum, 4h
GS75637	The dissection of transcriptional modules regulated by various drugs of abuse in the mouse striatum.	20459597	Genes regulated by methamphetamine in mouse striatum, 8h
GS86511	Methamphetamine induces dopamine D1 receptor-dependent endoplasmic reticulum stress-related molecular events in the rat striatum. Distinctive profiles of gene expression in the human	19564919	Table 1: METH administration causes changes in ER stress-related genes
GS86571	nucleus accumbens associated with cocaine and heroin abuse.	16710320	NA
GS86697	Essential role of the histone methyltransferase G9a in cocaine-induced plasticity. Distinctive profiles of gene expression in the human	20056891	NA
GS86937	nucleus accumbens associated with cocaine and heroin abuse.	16710320	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS87050	Repeated cocaine administration induces gene expression changes through the dopamine D1 receptors.	15770241	395 genes are differentially down- (-) or up-(+) regulated in the CPu in D1 receptor mutant mice compared to wild-type mice after repeated saline injections
GS87302	Gene expression profile of the nucleus accumbens of human cocaine abusers: evidence for dysregulation of myelin.	15009677	NA
GS87306	Repeated cocaine administration induces gene expression changes through the dopamine D1 receptors.	15770241	109 genes are differentially down- (-) or up-(+) regulated in the CPu in D1 receptor mutant mice compared to wild-type mice after repeated cocaine administration Table S2: Full list of significantly regulated transcripts in postmortem aPFC from cocaine, phencyclidine and / or cannabis abusers. (Whole Table) [DRG]
GS87365	Transcriptional changes common to human cocaine, cannabis and phencyclidine abuse.	17205118	
GS87516	Methamphetamine induces dopamine D1 receptor-dependent endoplasmic reticulum stress-related molecular events in the rat striatum.	19564919	Table S3: SCH23390-sensitive METH-responsive genes in the rat striatum (173 genes)

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS87527	Methamphetamine induces dopamine D1 receptor-dependent endoplasmic reticulum stress-related molecular events in the rat striatum.	19564919	Table S4: SCH23390-resistant METH-responsive genes in the rat striatum (372 genes)
GS93572	GO:0050780 dopamine receptor binding	NA	
GS93573	GO:0050780 dopamine receptor binding	NA	
GS94395	GO:0060160 negative regulation of dopamine receptor signaling pathway	NA	
GS94413	GO:0007191 activation of adenylate cyclase activity by dopamine receptor signaling pathway	NA	
GS94417	GO:0007195 inhibition of adenylate cyclase activity by dopamine receptor signaling pathway	NA	
GS94576	GO:0001963 synaptic transmission, dopaminergic	NA	
GS95079	GO:0031752 D5 dopamine receptor binding	NA	
GS95081	GO:0031750 D3 dopamine receptor binding	NA	
GS95082	GO:0031751 D4 dopamine receptor binding	NA	
GS95241	GO:0042417 dopamine metabolic process	NA	
GS95389	GO:0005330 dopamine:sodium symporter activity	NA	
GS95559	GO:0060161 positive regulation of dopamine receptor signaling pathway	NA	

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS95590	GO:0019811 cocaine binding	NA	NA
GS95732	GO:0035240 dopamine binding	NA	NA
GS95842	GO:0007269 neurotransmitter secretion	NA	NA
GS96672	GO:0042220 response to cocaine	NA	NA
GS96863	GO:0042053 regulation of dopamine metabolic process	NA	NA
GS97110	GO:0042416 dopamine biosynthetic process	NA	NA
GS97790	GO:0033602 negative regulation of dopamine secretion	NA	NA
GS97887	GO:0004500 dopamine beta-monooxygenase activity	NA	NA
GS97959	GO:0001593 dopamine D4 receptor activity	NA	NA
GS97960	GO:0004952 dopamine receptor activity	NA	NA
GS97961	GO:0001591 dopamine receptor activity, coupled viaGi/Go	NA	NA
GS97962	GO:0001590 dopamine D1 receptor activity	NA	NA
GS98438	GO:0001670 dopamine D2 receptor activity	NA	NA
GS99511	GO:0032227 negative regulation of synaptic transmission, dopaminergic	NA	NA
GS99512	GO:0032226 positive regulation of synaptic transmission, dopaminergic	NA	NA
GS99513	GO:0032225 regulation of synaptic transmission, dopaminergic	NA	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS99535	GO:0051583 dopamine uptake	NA	NA
GS99536	GO:0051585 negative regulation of dopamine uptake	NA	NA
GS99537	GO:0051584 regulation of dopamine uptake	NA	NA
GS99538	GO:0051586 positive regulation of dopamine uptake	NA	NA
GS99966	GO:0060159 regulation of dopamine receptor signaling pathway	NA	NA
GS100542	GO:0006585 dopamine biosynthetic process from tyrosine	NA	NA
GS100663	GO:0007212 dopamine receptor signaling pathway	NA	NA
GS101218	GO:0060158 activation of phospholipase C activity by dopamine receptor signaling pathway	NA	NA
GS101469	GO:0005329 dopamine transmembrane transporter activity	NA	NA
GS101610	GO:0031749 D2 dopamine receptor binding	NA	NA
GS101611	GO:0031748 D1 dopamine receptor binding	NA	NA
GS101683	GO:0015872 dopamine transport	NA	NA
GS102100	GO:0045963 negative regulation of dopamine metabolic process	NA	NA
GS102436	GO:0042420 dopamine catabolic process	NA	NA
GS103206	GO:0045964 positive regulation of dopamine metabolic process	NA	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS103507	GO:0048148 behavioral response to cocaine	NA	NA
GS103863	GO:0001588 dopamine receptor activity, coupled viaGs	NA	NA
GS103955	GO:0014059 regulation of dopamine secretion	NA	NA
GS103980	GO:0001589 dopamine D5 receptor activity	NA	NA
GS104211	GO:0033603 positive regulation of dopamine secretion	NA	NA
GS104228	GO:0033076 isoquinoline alkaloid metabolic process	NA	NA
GS105153	GO:0071542 dopaminergic neuron differentiation	NA	NA
GS105758	GO:0001592 dopamine D3 receptor activity	NA	NA
GS107749	MP:0002556 abnormal cocaine consumption	NA	NA
GS108180	MP:0004077 abnormal striatum morphology	NA	NA
GS108435	MP:0009754 enhanced behavioral response to cocaine	NA	NA
GS108445	MP:0009758 impaired behavioral response to cocaine	NA	NA
GS109244	MP:0001905 abnormal dopamine level	NA	NA
GS109499	MP:0009943 abnormal olfactory bulb periglomerular cell morphology	NA	NA
GS110717	MP:0001906 increased dopamine level	NA	NA
GS111278	MP:0005643 decreased dopamine level	NA	NA
GS111849	MP:0001988 cocaine preference	NA	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS113111	MP:0010149 abnormal synaptic dopamine release	NA	NA
GS113769	MP:0003243 abnormal dopaminergic neuron morphology	NA	NA
GS113770	MP:0003244 loss of dopaminergic neurons	NA	NA
GS114711	Discordant Nicotine Twins Significant S-score Genes	NA	NA
GS114712	Concordant Nicotine Twins Significant S-score Genes	NA	NA
GS121068	Genome-wide association for nicotine dependence and smoking cessation success in NIH research volunteers.	19009022	NA
GS122445	Cocaine interacting genes (MeSH:D003042) in CTD	NA	NA
GS122980	Methamphetamine interacting genes (MeSH:D008694) in CTD	NA	NA
GS123349	Dopamine Antagonists interacting genes (MeSH:D018492) in CTD	NA	NA
GS124877	arachidonyl dopamine interacting genes (MeSH:C474941) in CTD	NA	NA
GS126894	dopamine quinone interacting genes (MeSH:C104705) in CTD	NA	NA
GS126983	Dopamine interacting genes (MeSH:D004298) in CTD	NA	NA
GS128000	ABA Average Intensity for Striatum dorsal region Chr 4	NA	NA
GS128002	ABA Average Intensity for Striatum ventral region Chr 4	NA	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS128083	ABA Average Intensity for Striatum dorsal region Chr 15	NA	NA
GS123084	ABA Average Intensity for Striatum ventral region Chr 15	NA	NA
G128007	BXD SNP Count Chr4	NA	NA
GS128085	BXD SNP Count Chr 15	NA	NA
GS128080	RhoA, encoding a RhoGTPase, is associated with smoking initiation. Positional candidate genes for Cocaine Sensitization - TOTAL	17284169	NA
GS128185	locomotion (cm in 1 hr) [Cocaine] in Females for Suggestive BXD QTL on Chr4 from 102 to 112 Mb	19958391	NA
GS128186	Gene expression correlations for Nucleus accumbens to cocaine induced locomotor activation following the first (10 mg/kg) exposure to cocaine	19958391	NA
GS128187	Gene expression correlations for Striatum to cocaine induced locomotor activation following the first (10 mg/kg) exposure to cocaine	19958391	NA
GS128188	Gene expression correlations for Nucleus accumbens to cocaine induced locomotor activation following the second (10 mg/kg) exposure to cocaine	19958391	NA

Table A1 continued.

Gene set ID	Gene set Name	Pubmed ID	Additional information
GS128189	Gene expression correlations for Striatum to cocaine induced locomotor activation following the second (10 mg/kg) exposure to cocaine	19958391	NA
GS128205	Chromosomal loci that influence oral nicotine consumption in C57BL/6J x C3H/HeJ F2 intercross mice.	17010155	NA
GS33776	Positional candidate genes for Cocaine Sensitization - TOTAL locomotion (cm in 1 hr) [Cocaine] in Females for Suggestive BXD QTL on Chr15 from 30 to 55 Mb	19958391	NA

VITA

Vivek M. Philip received his Master's in Computer Science from Annasaheb Magar Mahavidyalaya affiliated with the University of Pune, Maharashtra, India in 2003. Upon graduation, he obtained an additional Master's degree in Computer from the University of Louisiana at Lafayette, Lafayette, LA in 2006. His Master's thesis work entailed protein structure validation. Immediately upon completion of his Master's degree, he began his doctoral studies in the Genome Science and Technology graduate program at the University of Tennessee, Knoxville, TN.