COMPUTATIONAL METHODS FOR STUDYING PARENT-OF-ORIGIN EFFECTS VIA RECIPROCAL MOUSE CROSSES

Daniel G. Oreper

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum of Bioinformatics and Computational Biology

Chapel Hill
2018

Approved by:

Daniel Pomp

Leonard McMillan

Michael Love

William Valdar

Fernando Pardo-Manuel de Villena

ABSTRACT

Daniel G. Oreper: Computational methods for studying parent-of-origin effects via reciprocal mouse crosses
(Under the direction of William Valdar)

Imprinted genes have been linked with diseases ranging from cancer, to metabolic syndromes, to psychiatric illness. For psychiatric illness in particular, numerous lines of evidence, both from human and mouse studies, suggest imprinted genes affect behavior along with brain development and function. Nonetheless, the effect of imprinted genes on most complex traits is not well characterized. Moreover, the architecture of environment-by-imprinting effects is even less well-understood.

The lack of characterization is likely due to the general difficulty of observing "parent-of-origin effects" (POEs), which typically arise in mammals from maternal effects—or from imprinting. To study POE/environment-by-POE, we can employ a relatively neglected but maximally powerful POE-detection system: the reciprocal cross (RX). Towards this end, we develop and apply computational methods for designing and analyzing RX experiments. Here, these techniques are applied in the context of RXs of inbred lines of mice, with a focus on behavior—but these techniques could be similarly employed in any model organism subject to POE, and on any complex trait.

The first set of methods focuses on the analysis of expression and behavioral data from RXs of a single pair of classical inbred mouse strains, with offspring exposed in utero to various diets. In this analysis, we detected dozens of POE/diet-by-POE on gene expression, a handful of similar effects on behavior, and a possible connection between POE on expression and behavior. Motivated by these results, we engaged in a similar but larger study, the CC-POE, in which we RXd *multiple* pairs of inbred lines drawn from the Collaborative Cross (CC)—a panel of multiparental recombinant inbred mouse strains. To aid in the CC-POE design, we developed a novel method for selecting an optimal set of reciprocal crosses: the Reciprocal Cross Explorer. Finally, with the goal of analyzing CC-POE

data, we develop a resource for variant imputation in the CC: the Inbred Strain Variant Database (available online at `https://isvdb.unc.edu`). Taken together, methods developed in this dissertation represent progress towards a new way of studying POEs via RXs.

PREFACE

Chapter 2 has been adapted with permission from a submission to G3. Chapter 3 is adapted from a manuscript in preparation for submission. Chapter 4 has been adapted with permission from a submission to G3.

ACKNOWLEDGEMENTS

There are many people who have made my PhD journey possible.

I want to first thank my mentor, William Valdar, not only for developing my scientific ability, but for his guidance, perspective, example, and care for my development as a human being.

For their excellent sense for humor, their feedback, and their friendship, I thank my lab, members past and present: Yuying Xie, Alan Lenarcic, Jeremy Sabourin, Zhaojun Zhang, Robert Corty, Greg Keele, Paul Maurizio, Yunjung Kim, Wes Crouse, Kathie Sun, and especially Yanwei Cai, who was instrumental in the ISVdb project.

I thank our collaborators in the Pardo–Manuel de Villena and Tarantino labs—especially Fernando Pardo–Manuel de Villena, Darla Miller, Leeanna Hyacinth, Rachel McMullan, Lisa Tarantino, and Sarah Schoenrock—for envisioning and working together with us on the experiments that made my computational efforts possible, and for their encouragement. Thank you to Andrew Morgan for imparting some of his tremendous understanding and love of research.

I thank my committee—Leonard McMillan, Daniel Pomp, Michael Love, as well as William Valdar and Fernando Pardo–Manuel de Villena— for refining my ideas, for providing their insight, and for helping me remember that a PhD is foremost about education.

I thank UNC, BBSP, and BCB, the organizations that have created a program conducive to learning, intellectual risk-taking, and collaboration. I thank Sausyty Hermreck, John Cornett and Cara Marlow, the administrators who have patiently shepherded me through the PhD process.

Thank you to ITS—especially Mike Barker and Jeff Roach—for their help in making the ISVdb a publicly available resource, from generously providing high performance disk space, to implementation guidance. I also want to thank ITS for the powerful computing cluster that they have developed and maintained. I thank the UNC functional Genomics Core and Mike Vernon, the UNC high throughput sequencing facility, and the UNC Systems Genetics Core for their support. Thank you to Sara Selitsky and Matt Kanke for their guidance in miRNA-related analysis.

I thank my sources of funding: the grants awarded by the National Institute of Mental Health, the National Human Genome Research Institute Center of Excellence, the PhRMA Foundation Pre Doctoral Fellowship, and the National Institute of Health T32 training grant that BCB chose to award to me.

I thank my girlfriend Cindy, for proudly cheering me on throughout weeks of writing, for her ready smile, and for her love.

And finally, I thank my loving family, who, from start to finish, have listened to my frustrations, celebrated my successes, and have always believed in me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

129	129S1Sv/ImJ
AJ	A/J
ASR	Accoustic Startle Stimulus
B6	C57BL/B6J
CAST	CAST/EiJ
CC	Collaborative Cross
CC-POE	The Collaborative Cross Parent-of-origin Effect Experiment
CORT	Corticosterone
CTP	Combined Tail Probability
FDR	False Discovery Rate
FST	Forced Swim Test
FWER	Family-wise Error Rate
GMCP	Generalized Maximum Coverage Problem
HMM	Hidden Markov Model
IBD	Identical-by-descent
ISVdb	Inbred Strain Variant Database
JAGS	Just Another Gibbs Sampler
LMM	Linear Mixed Model
lncRNA	Long Noncoding RNA
MCMC	Markov Chain Monte Carlo
MDD	Methyl Donor Deficient
ME	Methyl Enriched
MegaMUGA	Mega Mouse Universal Genotyping Array
MEXCLP	Maximum Expected Covering Location Problem
MPP	Multiparenal Population
MRCA	Most Recent Common Ancestor
NIH	National Institute of Health
NIMH	National Institute of Mental Health

NOD	NOD/ShiLtJ
NZO	NZO/HILtJ
OF	Open Field
PD	Protein Deficient
PND	Postnatal Day
POE	Genetic-background-by-parent-of-origin Effect
PPI	Prepulse Inhibition
PWK	PWK/PhJ
qPCR	Quantitative Polymerase Chain Reaction
RA	Response Amplitude
Rexplorer	Reciprocal Cross Explorer
RF1	Reciprocal F1 hybrid
RIA	Radioimmuno Assay
RX	Reciprocal Cross
scPLS	Single Cell Partial Least Squares
SIH	Stress-induced Hyberthermia
SSVA	Supervised Surrogate Variable Analysis
Std	Standard Diet
SV	Surrogate Variable
SVA	Surrogate Variable Analysis
UNC	University of North Carolina
VDD	Vitamin D Deficient
WSB	WSB/EiJ

CHAPTER 1
Introduction

Imprinted genes have been estimated to play a role in as many as 100 diseases (Ubeda and Wilkins, 2008), having been at least tentatively linked with maladies ranging from cancer, to metabolic syndromes, to psychiatric illness (Kalish et al., 2014). Diseases more definitively known to be caused in part by imprinted gene mutations and/or defective imprinting include Beckwith-Wiedemann, Russell–Silver, Prader–Willi and Angelman Syndromes, as well as Albright hereditary osteodystrophy, and transient neonatal diabetes (Robertson, 2005; Kalish et al., 2014). These are all complex diseases, though in some cases they can be caused by a single gene mutation—such as deletion of *UBE3A*, which causes Angelman Syndrome. But even for Angelman syndrome, 10% of cases cannot be explained by any mutation. More broadly, there is still a gap in understanding of imprinting-related diseases, and in general, the effect of imprinted genes on complex traits is not well-characterized.

The lack of characterization is likely due to the difficulty of directly observing the "parent-of-origin effects" (POEs) that imprinted genes exert on complex traits. For imprinted genes, each inherited allele's expression changes depending on its parent-of-origin, and traits affected by imprinted alleles are in turn subject to parent-of-origin effects (Lawson et al., 2013). As a result, identifying imprinting-driven POE on complex traits requires that reciprocal heterozygotes for a given imprinted locus exist in the population under study; for example, assuming an "A" and "B" allele exist at some imprinted locus, "AB" organisms (maternal A) need to be compared to "BA" organisms.

The outbred populations typically used for studying POE—be they human or model organism—can generate the requisite reciprocal heterozygotes. But these populations are not ideal, in part because POE in these outbred populations can be confounded with genetic differences at every other locus. An alternate, but rarely used population for studying POE, is one consisting of reciprocal F1 hybrids (RF1s), each generated by a reciprocal cross (RX); in a RX of inbred strains S1 and S2, any

resulting female S1xS2 and S2xS1 RF1s are (almost) genetically identical, differing only in allelic parent-of-origin. Consequently, by comparing the S1xS2 and S2xS1 subpopulations, POE can be detected without confounding, and with maximal power.

Reciprocal crosses of model organisms are the focus of this dissertation: I describe the development of computational methods for employing reciprocal crosses to study POEs on complex traits. These techniques are also applied in this work, specifically to crosses of inbred lines of mice, and with an additional focus on POEs on behavior. But these same techniques could be employed in any model organism subject to parent-of-origin effects, and on any complex trait. The first set of techniques, described in chapter 2, focuses on the analysis and integration of multiple modes of data, in the context of a reciprocal cross of a single pair of strains. The second set of techniques, described in chapter 3, focuses on experimental design—specifically the optimal selection of multiple genetically distinct reciprocal crosses from a panel of candidate crosses. Third, in chapter 4, I describe an online resource for variant imputation, which was designed to help map POE in an experiment using multiple reciprocal crosses. Before delving more deeply into these three areas, the remainder of the introduction more fully elucidates the biology of imprinting and POE, the mouse resources we use, and the motivation behind the development of each of the three areas.

1.1 Imprinted genes

Imprinted genes are subject to an epigenetic process whereby either either the maternally or paternally inherited allele (depending on the gene), is (at least partially) silenced (Crowley et al., 2015; Bartolomei and Ferguson-Smith, 2011) relative to the other allele.

This asymmetry is believed to largely (though not necessarily exclusively) result from differential methylation of gametic DNA, with certain regions of DNA being methylated in egg cells, and other regions methylated in sperm cells. After fertilization, methylation at these gametic differentially methylated regions (DMRs) is then maintained in somatic cells during cell division. Gametic DMRs are also associated with nearby "somatic DMR", regions affected by parental-chromosome-specific methylation that is only acquired *after* fertilization.

Both somatic and gametic DMRs are associated with the silencing of alleles in the vicinity of the methylated region in one parental chromosome, but not the other. The silenced alleles are not

necessarily on the methylated chromosome; in fact, for imprinted gene clusters controlled by gametic DMR—which account for the majority of known imprinted genes—methylation seems to silence a long noncoding RNA (lncRNA) while at the same time *activates* other genes in the vicinity of the methylation mark. The exact mechanisms underlying this effect are uncertain, but two of the major hypotheses are that either: i) methylation interferes with the formation of an insulator, allowing an enhancer to activate expression of nearby alleles on the methylated chromosome (and preventing that enhancer from instead activating the lncRNA); or ii) methylation silences a lncRNA, preventing the lncRNA from in turn silencing nearby alleles on the methylated chromosome. (Barlow and Bartolomei, 2014).

Among the obstacles to studying imprinted gene effects is the fact that imprinting can be developmental stage specific, and tissue specific—even to the extent that different regions of the brain exhibit different patterns of imprinting (Koerner et al., 2009; Prickett and Oakey, 2012). Consequently, an effect measured at the wrong time or in the wrong tissue may not be observed. On the other hand, imprinted gene effects present certain opportunities:

1. There are only $\sim$150 mouse genes typically identified as imprinted (Blake et al., 2010) (although in much of this document we end up using a slightly larger set that includes imprinted genes identified in Crowley et al. (2015)), so effects following a parent-of-origin dependent pattern may be more readily mapped back to these genes.

2. Given that maintenance of imprinting depends on availability of methyl donors (Crider et al., 2012), imprinted genes may present an ideal path for understanding the interaction of genetics and environmental exposures—in particular dietary methyl donors—on development.

1.2 Parent-of-origin effects and an introduction to the reciprocal cross

Hager et al. (2008) used "parent-of-origin-dependent effect" to describe any genetic effect that causes phenotypic differences in reciprocal heterozygotes. Similarly, Lawson et al. (2013) described parent-of-origin effects as the phenomenon in which an allele's effect changes depending on whether it is maternally or paternally inherited. Accordingly, imprinted genes can be described as exerting POEs; for imprinted genes, each allele's expression depends on its parent-of-origin, and traits affected by imprinted alleles are then in turn subject to parent-of-origin-effects.

Parent-of-origin-effect-causing mechanisms such as imprinting exist at a locus whether or not they are observed. But if a parent-of-origin effect is to be observed, it requires that genetic variation exist at the causal locus. This is perhaps most clearly demonstrated in the context of the reciprocal cross (RX). Suppose inbred mouse strains B6 and NOD were reciprocally crossed (RXd). Such a RX generates two reciprocal F1 hybrid (RF1) populations, B6xNOD and NODxB6.

Referring to Figure 1.1, suppose that some locus is silenced due to maternal imprinting. Only the gray version of the allele is expressed in B6xNOD, whereas only the blue version is expressed in NODxB6. Consequently, if the two alleles differ in their effect on some phenotype, a POE on the phenotype will be observable in the form of a phenotypic difference between the B6xNOD and NODxB6 populations. By contrast, suppose B6 and NOD bear the same gray allele (or two different alleles identical in their effect on the phenotype) at the locus controlling phenotype: in this case, since in both RF1 subpopulations the gray version of the allele is expressed, imprinting has no apparent effect, and so a parent-of-origin effect can not be observed. Thus, when we claim that a parent-or-origin effect has been observed, we are actually claiming that a parent-of-origin effect *interacting with genetic background* has been observed. As a shorthand, we will primarily refer to background-dependent parent-of-origin effects in the rest of this document simply as POEs.

Such imprinting-by-genetic interaction POEs have been described in multiple contexts (Georges et al., 2003; Vrana et al., 2000; Wolf et al., 2014; Schultz et al., 2015), although they have not always been named as such, and typically have been detected using populations other than RF1s.

1.2.1 Maternal factors and POEs

In mammals, maternal factors affecting offspring can include include maternal behavior (Peripato and Cheverud, 2002), oocyte composition (Tong et al., 2000), and *in utero* environment (Cowley et al., 1989; Kirkpatrick and Lande, 1989). To the extent that such factors depend on maternal genetics at some locus, RF1s that are reciprocal heterozygotes at that same locus will differ. This is the primary weakness of the RX: it cannot distinguish between POEs driven by imprinted genes and those driven by maternal factors.

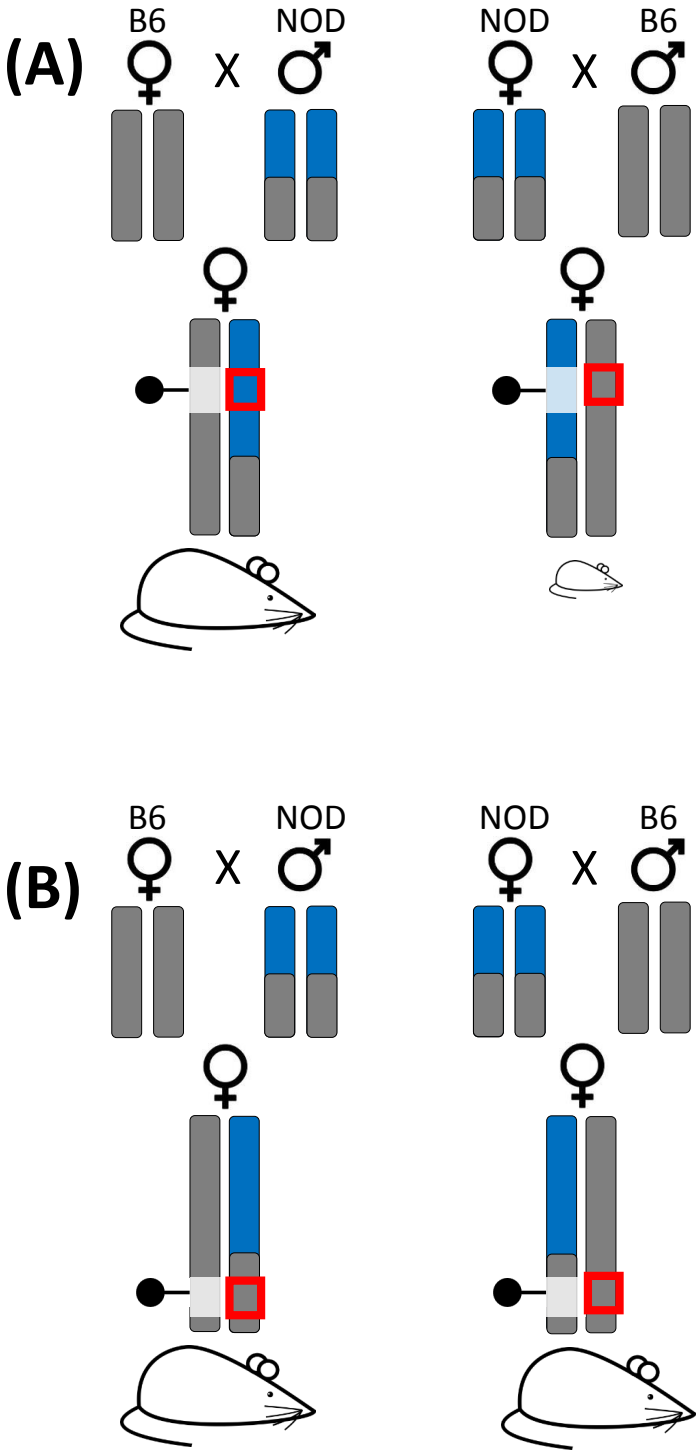


Figure 1.1: A reciprocal cross investigating POE on size. A) A reciprocal cross in which the locus causal to POE is reciprocally heterozygous. Imprinting-based silencing of the maternally inherited chromosome results in the blue allele being expressed in B6xNOD, whereas the gray allele is expressed in NODxB6. Consequently, the two RF1s differ in the size of animal, and a genetic-background-by-POE can be detected. B) A reciprocal cross in which the locus causal to POE is homozygous in the RF1s. Despite the fact that imprinting exists, silencing has no effect because in both populations the gray allele is expressed. Consequently, the two RF1s are identical in average size, and no genetic-background-by-POE can be detected.

1.2.2 On the power of the RX to study POE

Returning to Figure 1.1A, we note that the comparison is limited to only include RF1s that were female. In this case, both RF1 subpopulations would be almost (save for mitochondria) genetically identical, differing only in allelic parent-of-origin. Accordingly, any phenotypic differences can be attributed with high likelihood to POEs. Not only does the RX generate reciprocal heterozygotes at many loci—which are necessary for detection of POEs—but the RX does so in a manner such that POEs are not confounded with other genetic effects. It is this property that makes the RX maximally powerful for studying POE.

1.3 Other approaches and populations for studying POEs

Most of the imprinting-driven POEs identified (or tentatively identified) so far have been found by observing large effects caused by uniparental disomy, imprinted gene knockouts, and overexpression assays (Wolf et al., 2008; Cleaton et al., 2014; Dent and Isles, 2014). Such studies, while effective, are laborious, expensive, and may not be able to identify more subtle effects because of the disruptive effect of a large genetic perturbation (Wolf et al., 2008).

An alternate and more cost effective approach is employ association or QTL mapping, whose results can then be used prioritize target validation. The typical populations used in these association studies are outbred—e.g., F2, backcross populations, and heterogeneous stock populations (Lawson et al., 2013). Such populations, just like RF1s, can generate the requisite reciprocal heterozygotes per locus needed to study POE. The advantages of such outbred populations over RF1s are that: 1) POE can be detected simultaneously with non-POE genetic effects; and 2) POE arising from imprinting vs maternal effects can be disambiguated—a significant difference between reciprocal heterozygote (at some locus) offspring from heterozygote (at that same locus) mothers can be ascribed to imprinting rather than to a maternal effect (Hager et al., 2008).

However, outbred populations have disadvantages as well: i) Due to the fact that every animal is genetically distinct in an outbred population (unlike in RF1 populations), alternate parent-of-origin states can never be observed in the exact same genetic background, and this confounding limits the power of outbred populations to estimate POE; ii) in many of the outbred populations, especially F2s, determining each allele's parent-of-origin at heterozygous loci is challenging (i.e., AB and BA cannot

be distinguished) and needs to be imputed using additional information (Wolf et al., 2008)—whereas in RF1s, if the parental genomes are known, the allelic parent-of-origin in the offspring is known with total certainty; and iii) the irreplicability of outbred animals makes it impossible to perfectly recreate genetic state for a validation study, or for a studying investigating some treatment effect, whereas RF1s are (almost) perfectly reproducible.

1.4 POE on behavior

Much of this dissertation describes analysis of POE on behavior in mice. Here, we provide motivation, describing evidence for the utility and tractability of studying behavioral POE, drawing from both human and animal studies.

1.4.1 POE on psychiatric illness

The lifetime prevalence of mental illness among Americans has been estimated to be $\sim$47% (Insel, 2008), and effective treatment options are limited (Sachs et al., 2007; Naber and Lambert, 2009; Sultzer et al., 2008). The necessary insights for devising new treatments may require a better understanding of POE: multiple psychiatric diseases exhibit patterns of transmission consistent with POEs (Davies et al., 2001; Isles and Wilkinson, 2000). In particular, the canonical demonstration of imprinted gene POEs on psychiatric disease is given by Prader–Willi and Angelman syndromes, both of which can be caused by improper imprinting of the cluster of imprinted genes in the 15q11-13 region (Dykens et al., 2011). Copy number variants in 15q11-13 have also been associated with ASD (Dykens et al., 2011) and schizophrenia (McNamara and Isles, 2013). Outside of 15q11-13, a mutation in *LRRTM1* has been associated paternally with schizophrenia (Francks et al., 2007; Linhoff et al., 2009).

Despite these assorted POE findings in humans—as well as other findings more generally suggesting the heritability of psychiatric illness (Lee and Avramopoulos, 2014)—the precise genetic and epigenetic mechanisms underlying inherited susceptibility are generally not well understood. Studying behaviors that model psychiatric illness in an experimentally tractable organism such as mouse, may provide a potential avenue to gaining this understanding.

1.4.2 Mouse models of POE on behavior

Mouse has the virtue of rapid gestation and development and is versatile as a model for behavioral genetics and environmental perturbation. Moreover, imprinting is functionally consistent between mice and humans (Bartolomei and Ferguson-Smith, 2011).

So far, studies using mouse models have found that imprinting affects brain development, function, and behavior. Many imprinted genes are active (some exclusively) in the brain (Prickett and Oakey, 2012), especially during embryogenesis (Wilkinson et al., 2007). Among the functions characterized so far: *Igf2* and *Igf2r* affect brain size and organization (Wilkinson et al., 2007), and $G_S\alpha$ affects control of nutritional resources (Bartolomei and Ferguson-Smith, 2011). Affecting behavior: *Peg1* and *Peg3* affect maternal nesting, pup-gathering and pup-grooming; *Gnasxl* deletions prevent mice from suckling properly; *Nesp* affects exploratory behavior; and *Grb10* affects social dominance (Dent and Isles, 2014).

Adding to the appeal of mouse, a handful of mouse studies have identified POE on behavior specifically using the RX. Affected behaviors include various measures of emotionality, as well as urinary odor preferences (Putterman, 1998; Isles et al., 2001; Calatayud and Belzung, 2001; Calatayud et al., 2004).

1.4.3 Diet-by-POE on behavior

Behavioral phenotypes may also be particularly well-suited for studying the interaction of POE with developmental diet. Rodent and human studies have demonstrated that certain perinatal diets affect both imprinting and behavior: for example, perinatal protein deficiency (PD) and vitamin D deficiency (VDD) both induce methylation changes (Vucetic et al., 2010; Lillycrop et al., 2007; Kesby et al., 2010, 2012) and alter behaviors that model schizophrenia (Burne et al., 2004a,b, 2006; Palmer et al., 2008; Franzek et al., 2008; Kesby et al., 2006, 2010; Burne et al., 2006; Harms et al., 2008, 2012; Turner et al., 2013; Vucetic et al., 2010). Similarly, other perinatal diets that imply a deficiency in methyl donors have been linked to reduced methylation in the brain (Davison et al., 2009; Niculescu et al., 2006; Konycheva et al., 2011), increased anxiety-like behaviors (Ferguson et al., 2005; Konycheva et al., 2011), and changes in learning and memory (Konycheva et al., 2011; Berrocal-Zaragoza et al., 2014). In general, epigenetic effects have repeatedly been shown

to be sensitive to maternal diet during the prenatal period: classically in agouti mouse experiments (Waterland and Jirtle, 2003); and observationally in studies of human physiology, mental health, and gene expression during the Dutch Hunger Winter (Heijmans et al., 2008; Tobi et al., 2009).

1.5 The Collaborative Cross as a platform for studying POE on behavior

One RX can detect POE, but in principle, multiple RX could *map* POE. Chapter 3 and chapter 4 are largely devoted to developing tools towards this end. In particular, these chapters focus on employing RX of inbred mouse lines drawn from the Collaborative Cross (CC) reference population. Before describing the overarching effort to map POE, we describe the properties of the CC first.

The Collaborative Cross (CC) is a large panel of recombinant inbred mouse lines derived from a genetically diverse set of eight inbred founder strains: A/J (AJ), C57BL/6J (B6), 129S1Sv/ImJ (129), NOD/ShiLtJ (NOD), NZO/HlLtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB). These eight founder strains were first outcrossed for three generations to produce mice with contributions from all eight founder strains. These outcrosses were initiated, with different founder orderings, in over 1000 independent breeding funnels (Shorter et al., 2017). Mice within each funnel were subsequently inbred for multiple generations until two or more animals were identified by MegaMUGA genotyping collectively as having over 90% of the genome fixed (i.e., homozygous and consistent for a founder haplotype).

These animals, hereafter termed the most recent common ancestors (MRCAs), were then chosen to become the obligate ancestors of all subsequent generations and bred to produce a distinct CC strain. The set of MRCAs from all strains composes the CC's obligate ancestors, that is, the set of individuals that together circumscribes the initial genetic material that can be passed on to subsequent CC mice.As a result of this breeding scheme, the inbred CC strain genomes are random and independent mosaics of the eight founder haplotypes Collaborative Cross Consortium (2012); Srivastava et al. (2017) (Figure 1.2; more details are available at `http://csbio.unc.edu/CCstatus/index.py?run`).

This combination of independent genomes and high genetic diversity, along with the reproducibility of inbred strains, has made the CC a unique resource in mammalian genetics, and early studies on the CC have begun to exploit these features (Aylor et al., 2011; Ferris et al., 2013; Phillippi et al., 2014; Rasmussen et al., 2014; Mosedale et al., 2017; Green et al., 2017; Gralinski et al., 2017).

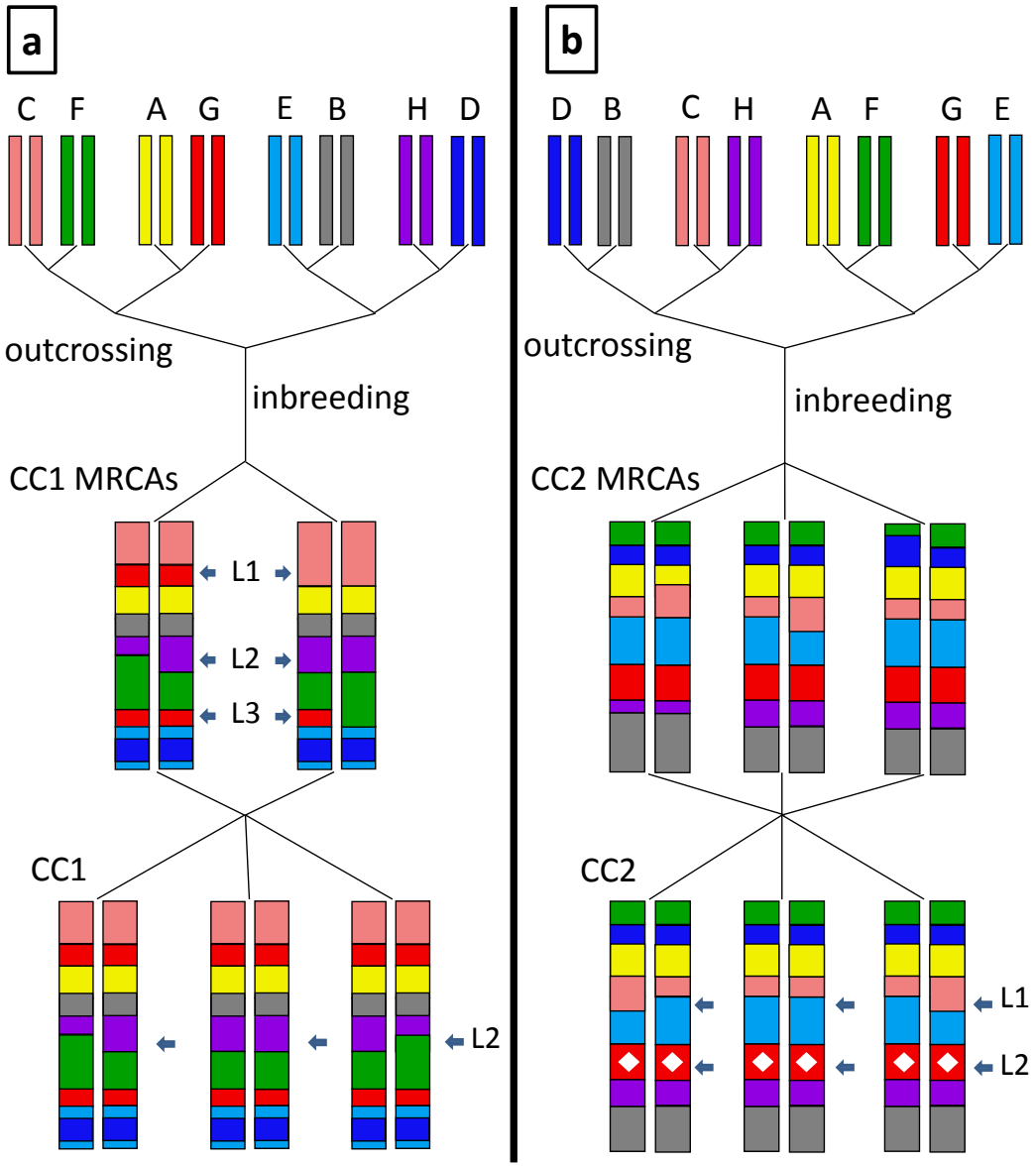


Figure 1.2: Breeding process for two CC strains. Both funnels begin by outcrossing the same eight founders, but the initial outcrossing order differs, resulting in completely independent populations per funnel. Animals are outcrossed for three generations, then inbred until genotyping reveals at least two animals with over 90% consistent homozygosity by haploytpe. These homozygous animals (a.k.a., the MRCAs) are chosen to become the obligate ancestors for the CC strains; all subsequent generations of a CC strain descend from a subset of the MRCAs. In (**a**), arrows show CC1 MRCA regions of inconsistent homozygosity (L1) and residual heterozygosity (L2, L3). After further inbreeding, only L2 continues to segregate. In (**b**), the CC2 MRCA set includes three animals rather than two. After further inbreeding, only L1 continues to segregate, but a de-novo mutation has become fixed at L2.

1.6 Multiple RXs: the CC-POE study

RXs of even a single pair of strains can powerfully detect POE. But RXs of multiple pairs of strains could be used to perform a sort of *mapping* of POE. In more detail, suppose RXs of multiple pairs of strains were performed. Each of the resulting pairs of RF1s would have their own heterozygosity and homozygosity mosaic. Given that a POE cannot be observed without reciprocal heterozygosity at some locus, if the same POE is observed in multiple sets of genetically distinct RF1s, it suggests that the POE arose from the loci that are reciprocally heterozygous in every genetically distinct RF1; each additional genetically distinct RX that detects POE progressively narrows the space of candidate POE. If we further assume that POEs we observe are caused by imprinting rather than maternal effects (admittedly a strong assumption) we can further narrow the space of candidate loci.

This type of approach, along with the goal of mapping behavioral POE and diet-by-POE, motivated the CC-POE study, a focus of part of this dissertation. The overarching CC-POE experiment is illustrated in Fig 1.3, and described below at a high level.

We selected RXs from the panel of CC lines, using Rexplorer (chapter 3), an experimental design tool. Nine genetically distinct RXs were generated, with resulting RF1s exposed in utero to one of four of diets. Males were unused, but of the female RF1s, one subset was behaviorally phenotyped and the other subset was concurrently expression phenotyped (RNA-seq of whole brain tissue). To our knowledge, the only previous effort to use a panel of RXs (technically, reciprocal *backcrosses*) to perform POE mapping was presented in Gonzalo et al. (2007), a study employing over 200 inbred strains of corn.

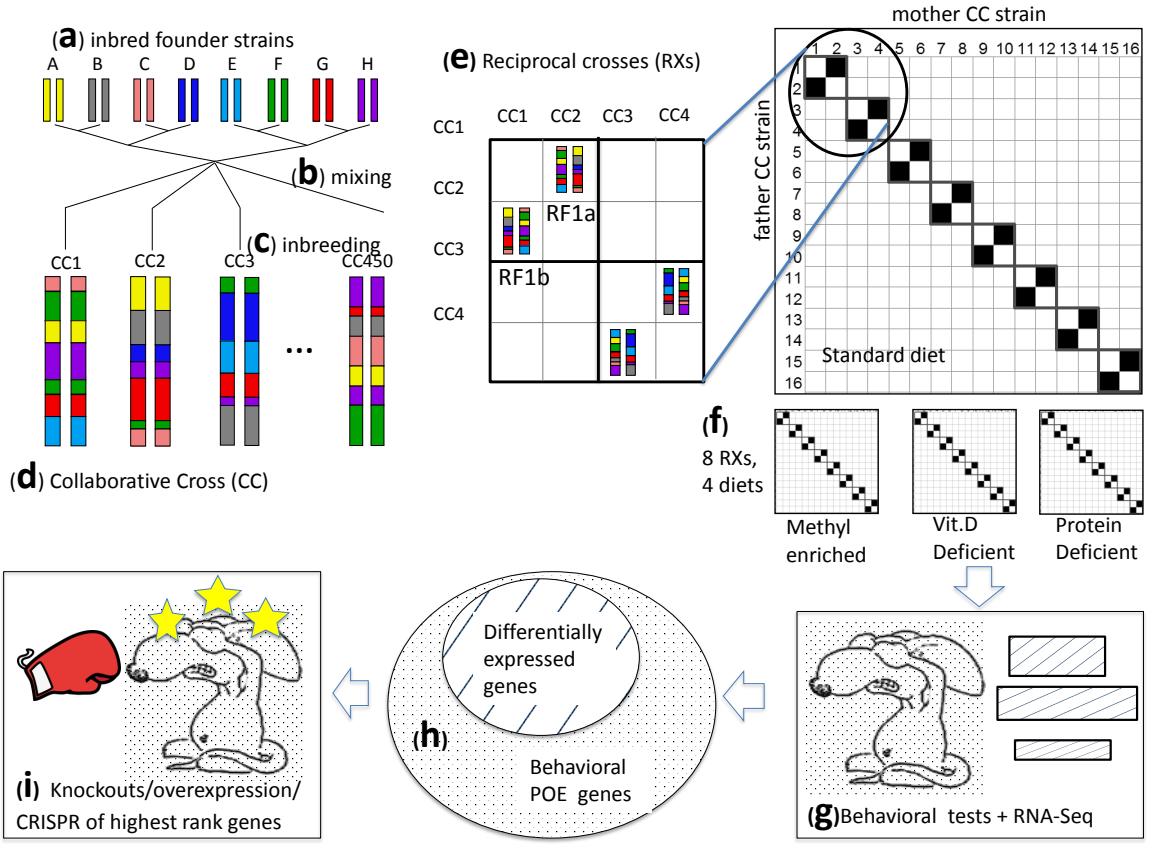


Figure 1.3: A Collaborative Cross (CC) experiment to map POE using RXs. The CC is an existing population; it is not being developed in this project: a) The 8 inbred founder lines of the CC (including NOD and B6) b) were outbred in a funnel breeding scheme, mixing their genomes c) then inbred for multiple generations d) resulting in the CC, a panel of recombinant inbred lines. e) A RX is composed of genetically identical RF1a and RF1b populations that differ in allelic parent-of-origin. f) Using 16 of the CC lines, we planned to generate 8 distinct RXs, on 4 diets (though we ended up with 9 RXs) g) One subset of RF1s (including both direction) was behaviorally tested, and the other was RNA-seqd. h) A behavioral and expression data model will rank imprinted genes as causal to POE. i) The top genes will be further biologically validated.

1.7 Chapter progression

The introduction (this chapter) provides background and motivation useful for understanding the three primary efforts presented in this dissertation.

In the first effort, we develop and apply techniques for analyzing data from a RX of a single pair of classical inbred lines, to detect POEs and diet-by-POEs on behavior and gene expression. Both types of data are integrated into a unified Bayesian mediation model. Multiple testing correction methods, including a bespoke permutation testing procedure are employed. These techniques lay the groundwork for the future analysis of the much larger CC-POE experiment involving multiple RXs of CC lines.

The second effort occurs in the context of the CC-POE, an experiment which is motivated (in part) by the desire to map POE detected in the first effort; towards this end, the CC-POE employs RXs of multiple pairs of CC lines. In this context, we describe the development and application of an experimental design method, the "Reciprocal Cross Explorer" (Rexplorer), for choosing the best RXs from a panel of inbred line for the purposes of POE-mapping. An operations research approach is taken to ensuring that POE is detected while simultaneously ensuring mapping resolution.

In the third effort, motivated by a desire to improve the fidelity of Rexplorer, as well by the need to analyze the CC-POE data, we develop a variant imputation resource for the CC population and for RXs of CC lines. This resource, the "Inbred strain variant database" (ISVdb) is publicly available online, and can be useful to any researcher performing design or analysis of CC experiments—and not just for POE studies.

This dissertation concludes with key findings and potential future efforts.

CHAPTER 2

Reciprocal F1 hybrids of two inbred mouse strains reveal parent-of-origin and perinatal diet effects on behavior and expression. [1]

2.1 Introduction

It is well established that susceptibility to psychiatric disorders arises from a combination of genetics and environmental exposures (Lee and Avramopoulos, 2014). Less well-studied is the phenomenon that this susceptibility seems to vary depending on whether certain harmful alleles were carried by the mother— or by the father (Davies et al., 2001; Isles and Wilkinson, 2000). That is, it is unclear to what extent the heritable component of disease risk is driven by parent-of-origin effects (POEs). Especially poorly understood is the extent to which POEs depend upon environmental context during development, and therefore how alternate environmental exposures could modulate a POEs impact on disease risk. A better understanding of POEs and their environmental modifiers could lead to improved interpretation of existing studies, to more effective experimental design, and even to novel public health interventions. Nonetheless, rigorous estimation of POEs in humans is difficult, especially on complex traits; even in animals it requires specialized experimental design attuned to POE biology.

Hager et al. (2008) used "parent-of-origin-dependent effect" to describe any genetic effect that causes phenotypic differences in reciprocal heterozygotes. Similarly, here we use "POE" as a shorthand for any effect driven by the interaction of genetic background with either maternal factors (e.g., maternal behavior, etc.), or imprinting, an epigenetic process in which either the maternally or paternally inherited allele of certain genes is at least partially silenced (Crowley et al., 2015; Bartolomei and Ferguson-Smith, 2011).

[1] This chapter has been adapted from a manuscript submitted to G3. The citation will be as follows: Oreper D., Schoenrock S., McMullan R. C., Ervin R., Farrington J., et al., 2018 Reciprocal F1 hybrids of two inbred mouse strains reveal parent-of-origin and perinatal diet effects on behavior and expression. G3: Genes, Genomes, Genetics.

Imprinting-driven POEs may be particularly relevant to psychiatric disease given the numerous lines of evidence suggesting imprinted genes affect behavior as well as brain development and function, drawn from both human and animal studies Imprinted genes may present an ideal path for understanding the interaction of genetics and environmental exposures — especially diet — on development: not only can imprinting be developmental-stage (and tissue)-specific (Koerner et al., 2009; Prickett and Oakey, 2012), but it is also believed to largely result from differential allelic methylation, and thus to require dietary methyl donors (Crider et al., 2012). For previous human and animal studies that have demonstrated or suggested the importance of POE or diet-by-POE on psychiatric illness and behavior, see section 1.4.

2.1.1 Reciprocal F1 hybrids (RF1s) for investigating POE and its environmental modifiers

The points above motivate an experiment to directly determine the extent of POEs on psychiatric disease across multiple perinatal dietary exposures in a simple, controlled, and replicable system — something only possible in an animal model. An ideal population is provided by (female) reciprocal F1 hybrids (RF1s) of inbred strains: in female RF1s, genetic background is constant (save for mitochondria), and only the direction of inheritance varies, allowing POEs to be measured directly. RF1s have been used to identify POEs on behavior in a handful of studies so far (Putterman, 1998; Isles et al., 2001; Calatayud and Belzung, 2001; Calatayud et al., 2004). Here we exploit the replicability of RF1s further to study the unconfounded effect of an environmental modifier on POE, varying diet while genetic background stays constant. To our knowledge, this approach has only been followed in Schoenrock et al. (2017), our recent related study in which we reciprocally crossed Collaborative Cross strains.

Here we examine, under four different *in utero* dietary exposures, behavior and expression in RF1s of the inbred mouse strains C57BL/6J (B6) and NOD/ShiLtJ (NOD). B6 and NOD inbred strains were selected because: 1) B6 is the reference genome and is the best characterized strain with respect to behavior; 2) B6 and NOD are among the founder strains for the Collaborative Cross, a population that is an area of focus for our labs; 3) B6 and NOD were both readily available, and B6-NOD crosses generate large litters, facilitating replication; 4) NOD is genetically similar enough to B6, that standard B6-expression microarrays were appropriate for NOD alleles as well (Oreper

et al., 2017c), but different enough that a substantial number of POEs on gene expression could still be revealed by B6-NOD RF1s.

Our replication of the RF1s under four different *in utero* dietary exposures serves several purposes, namely to: 1) increase the likelihood of observing POE, as POE may be diet-specific; 2) estimate the extent to which POE generalizes across alternate perinatal dietary exposures; and 3) estimate the perinatal diet effect itself.

Our study, the first to examine the connection between POE on expression and POE controlled behavior, demonstrates: 1) the presence of POEs on behavior and gene expression, many of which are robust to differences in perinatal diet; 2) a possible explanatory pathway connecting imprinting, to gene expression, to behavior; and 3) the usefulness of our approach as a template for further animal model studies of POE and developmental exposures on complex traits.

2.2 Experimental Materials and Methods

2.2.1 Mice

C57BL/6J (B6) and NOD/ShiLtJ (NOD) mice originated from a colony maintained by Gary Churchill at Jackson Laboratory, and were transferred in 2008 to the FPMdV lab at UNC (this originating colony also produced the G1 breeders of the CC; see Srivastava et al. 2017). Six-week old B6 females (3-8 dams/diet) and NOD females (3-5 dams/diet) were transferred from the FPMdV lab to the Tarantino lab at UNC and acclimated for one week. At 7 weeks of age, dams were placed on one of 5 different diets. At 12 weeks, dams were mated with males of the opposite strain to produce either B6xNOD or NODxB6 F1 hybrids (dam strain listed first; Figure 2.1B). Pregnant dams remained on their experimental diet until litters were weaned, ensuring that offspring were exposed to the diet throughout the entire perinatal period. At postnatal day (PND) 21, F1 hybrids were weaned onto a standard laboratory chow (Pico rodent chow 20; Purina, St. Louis, MO, USA) (Figure 2.1A). F1 hybrids were bred in one vivarium, but then transferred to a separate behavioral testing vivarium, and were then acclimated to this testing vivarium for at least one week before testing began. Mice were housed in a specific pathogen free facility on a 12-hour light/dark cycle with lights on at 7 A.M. All procedures and animal care were approved by the UNC Institutional

Animal Care and Use Committee and followed the guidelines set forth by the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals.

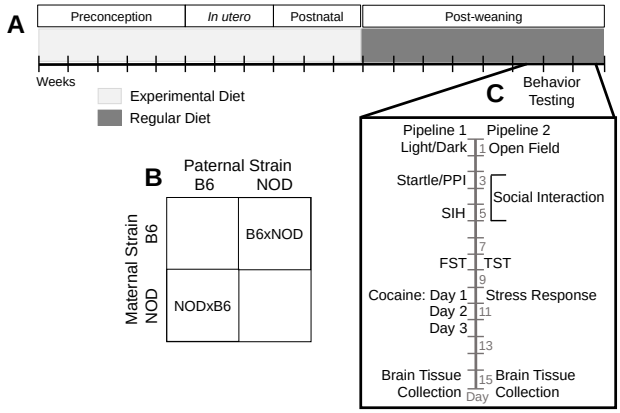


Figure 2.1: Experimental design to assess POE, perinatal diet, and diet-by-POE on behavior and gene expression in reciprocal F1 hybrids (RF1s). Female NOD/ShiLtJ (NOD) and C57BL/6J (B6) mice were placed on one of 4 experimental diets (protein deficient, vitamin D deficient, methyl enriched, standard) at 7 weeks of age (A). After 5 weeks, NOD females were mated to B6 males and B6 females to NOD males forming NODxB6 and B6xNOD RF1 hybrids, respectively (B). Dams remained on their experimental diet throughout gestation and the postnatal period. At PND 21, female F1 hybrids were weaned and placed onto a regular laboratory diet (A). Upon reaching adulthood at PND 60, F1 hybrids were tested in one of two behavioral pipelines. After behavioral testing, mice were euthanized, and their brain tissue collected for gene expression analysis via microarray and qPCR (C).

2.2.2 Experimental Diets

The following diets, purchased from Dyets Inc. (Bethlehem, PA), were administered: vitamin D deficient (VDD; #119266), protein deficient (PD; 7.5% casein; #102787), methyl donor deficient (MDD; #518892), methyl donor enriched (ME; #518893) and control (Std; #AIN-93G). The PD and VDD diets were nutritionally matched to the Std diet while the MDD was matched to the ME diet; Table A.1 specifies each diet's nutrient composition. Food and water were available *ad libitum* throughout the experiment.

2.2.3 Behavior Assays

To ensure a standardized genetic background that included the sex chromosomes, the only tested F1 hybrids were female. Mice were 61.7 days old 2.6 standard deviations at the onset of testing. All behavioral testing was performed during the light part of the light/dark cycle between

8:00 A.M. and 12:00 P.M. Mice were placed into one of two behavioral pipelines (Figure 2.1C) to assess anxiety- and depressive-like behavior, stress response, sensorimotor gating or response to a psychostimulant: Pipeline 1— light/dark assay, startle/prepulse inhibition (PPI), stress-induced hyperthermia (SIH), forced swim test (FST) and cocaine response (N = 91); Pipeline 2— open field (OF), social interaction test, tail suspension and restraint stress (N = 87). In total, 34 behavioral measures were collected, with 22 in pipeline 1 and 12 in pipeline 2 (Table 2.1). For each pipeline, mice were tested in 3 batches, over 3 months. Offspring from both RF1 directions, as well as from at least 2 diet exposures, were included in each batch, to avoid confounding. For each diet and RF1 direction, we tested litters from at least 2 dams (N = 4 1.4; see Table A.2 for dam and offspring counts). One mouse in the NODxB6 ME group was euthanized due to injury on the day of social interaction testing; there is no data for this mouse for social interaction or for any subsequent test. There is no restraint stress data for another 4 mice (1 NODxB6 ME, 2 B6xNOD Std, 1 B6xNOD VDD), due to either death in the restrainer or insufficient serum collected for radioimmuno assay (RIA) analysis of corticosterone (CORT) levels.

2.2.3.1 Open Field (OF)

Mice were placed in the OF arena for 10 minutes. The OF apparatus (ENV-515-16, Med Associates, St. Albans, VT, USA) was a 43.2x43.2x33 cm arena, consisting of a white Plexiglas floor and clear Plexiglas walls with infrared detection beams at 2.54 cm. intervals on the x, y, and z axes that automatically tracked mouse position and activity throughout each experimental session. The apparatus was in a sound-attenuating chamber (73.5x59x59 cm) fitted with two overhead light fixtures containing 28-V lamps. Mice were placed in the OF arena for 10 minutes. The OF apparatus (ENV-515-16, Med Associates, St. Albans, VT, USA) was a 43.2x43.2x33 cm arena, consisting of a white Plexiglas floor and clear Plexiglas walls with infrared detection beams at 2.54 cm. intervals on the x, y, and z axes that automatically tracked mouse position and activity throughout each experimental session. The apparatus was in a sound-attenuating chamber (73.5x59x59 cm) fitted with two overhead light fixtures containing 28-V lamps. Mice were scored for total distance traveled (cm), average velocity (cm/s), number of vertical movements (rearing), and percent time spent in the center of the arena (a 22.86 cm^2 central part of the arena). These measurements were recorded in 5 bins of 2-minute width, and were scored in post-session analyses using Activity Monitor 5.1 software

(Med Associates). The testing apparatus was cleaned with a 0.25 % bleach solution between test subjects.

2.2.3.2 Social Interaction

Social approach was measured in a 3-chamber social interaction apparatus during a 20-minute test (described fully in Moy et al. (2007)). Briefly, the first 10 minutes was a habituation period in which the test mouse was given free access to all 3 chambers. The total number of transitions between all chambers during this 10 min period was measured. During the second 10 minutes, the test mouse was given the choice between a chamber containing a circular mesh enclosure that held a stranger mouse (B6), and a chamber containing an empty mesh enclosure. The amount of time the test mouse spent in the chamber with the stranger mouse was recorded and is reported as "percent stranger time", a measure of social preference.

2.2.3.3 Tail Suspension

Mice were suspended by a piece of laboratory tape wrapped around the tail and hung from a hook at the top of a 24.13 cm x 17.78 cm x 17.78 cm white acrylic enclosure. Mice were videotaped for the entire 4-minute session, and videotapes were analyzed for immobility during the last 2 minutes using the Actimetrics Freeze Frame analysis program (Actimetrics, Wilmette, IL). Percent immobility during the last two minutes is reported as a measure of depressive-like behavior (Miller et al., 2010).

2.2.3.4 Restraint Stress

Restraint was used to elicit a stress response that was then quantified by measurement of CORT levels in the serum. A retroorbital blood sample was taken immediately prior to placing the mice into a Broome-Style restraint tube (Plas Labs, Inc., Lansing, MI, USA) for 10 minutes. Immediately upon removal from the restrainer, a second retroorbital eye bleed was performed. Whole blood was centrifuged to isolate serum, and then the CORT levels were measured by competitive RIA per the manufacturers protocol (MP Biomedicals, Santa Ana, CA, USA).

2.2.3.5 Light/Dark

The open field arena described above was converted to a light dark apparatus by placement of an opaque polycarbonate black box that occupied one third of the arena space, thus allowing the mouse to choose between the light or dark part of the apparatus. Mice were placed in the lighted area immediately adjacent to and facing the entry to the dark enclosure and remained in the apparatus for 10 minutes. The amount of time (sec), distance moved (cm) and number of transitions between the dark and light zones was scored in 5-minute bins in post-session analyses using Activity Monitor 5.1 software (Med Associates). The testing apparatus was cleaned with a 0.25 % bleach solution between test subjects.

2.2.3.6 Startle and prepulse inhibition (PPI)

Acoustic startle and PPI of the startle response were both measured using the San Diego Instruments SR-Lab system (San Diego, CA), and following the protocol in Moy et al. (2012). Mice were placed in a plexiglas cylinder located in a sound-attenuating chamber that included a ceiling light, fan, and a loudspeaker that produced the acoustic stimuli (bursts of white noise). Background sound levels (70 dB) and calibration of the acoustic stimuli were confirmed with a digital sound level meter. Each test session consisted of 42 trials, presented following a 5-min habituation period. There were 7 types of trials: no-stimulus trials, trials with a 120 dB acoustic startle stimulus (a.k.a., ASR), and 5 trials in which a 20 ms prepulse stimulus (74, 78, 82, 86, or 90 dB) was presented 100 ms before the onset of the 120 dB startle stimulus. The different trial types were presented in 6 blocks of 7, in randomized order within each block, with an average intertrial interval of 15 sec (range: 10 to 20 s). Measures were taken of the startle response amplitude (RA) for each trial, defined as the peak response recorded from the onset of startle stimulus to the end of the 65-msec sampling. The PPI for each prepulse sound level was calculated as:

$$\text{PPI} = 100 - \left[\frac{\text{RA with prepulse \& startle stimulus}}{\text{RA with only startle stimulus}}\right] \times 100$$

2.2.3.7 Stress-induced hyperthermia (SIH)

Each tested mouse was individually removed from its home cage, and then its body temperature (T1) was measured. Specifically, a lubricated digital thermometer probe was inserted 1-1.5 cm into the rectum for approximately 10 seconds. The mouse was then returned to its home cage, and 10 minutes later the temperature measurement was repeated (T2). The difference in body temperature, $\Delta T = T2 - T1$, was used as a measure of anxiety-like behavior. Basal temperature was measured for all mice within a single cage in under a minute, to avoid increases in body temperature due to anticipatory stress.

2.2.3.8 Forced swim test (FST)

Mice were placed in a glass-polycarbonate cylinder (46cm tall X 21cm in diameter) filled with water (25-28 C) to a depth of 15 cm for 6 minutes. The duration of immobility during the last 4 minutes of the test period was scored using Ethovision 7.0 automated tracking software (Noldus, Leesburg, VA). Immobility was defined as no movements other than those needed to stay afloat. Mice were monitored continuously, and removed if they were unable to keep their nose or heads above water for more than 30 seconds. Percent immobility was reported as a measure of depressive-like behavior.

2.2.3.9 Cocaine-induced locomotor activation

Cocaine-induced locomotor activity was measured over a 3-day test protocol in the OF arena described above. On days 1 and 2, mice were given an intraperitoneal injection of saline before being placed into the OF arena for 30 minutes, and then returned to their home cage. The day 3 protocol was nearly identical, but instead of saline, mice were injected with 20 mg/kg cocaine (Cocaine HCl; Sigma-Aldrich, St. Louis, MO). The total distance traveled was calculated for each day, and cocaine-induced locomotor activation was calculated by subtracting the distance on day 2 from day 3.

2.2.3.10 Body Weight

Adult body weight was recorded for mice in pipeline 1 prior to startle/PPI and cocaine administration.

2.2.4 Gene Expression

To identify genes subject to POE and/or perinatal-diet effect, whole-brain expression was measured by microarray, and key expression results were later validated with qPCR.

2.2.4.1 Tissue extraction

Three days after completion of behavioral testing, mice were euthanized, cerebellar tissue was removed, and the brain was split midsagitally into left and right hemispheres. Brain tissue was flash frozen in liquid nitrogen. Right brain hemispheres were pulverized using a BioPulverizer unit (BioSpec Products, Bartlesville, OK). Pulverization batches were designed to prevent contamination between mice from different crosses or diets.

2.2.4.2 RNA extraction

Total RNA was extracted from 25 mg of powdered brain hemisphere tissue using an automated bead-based capture technology (Maxwell 16 Tissue LEV Total RNA Purification Kit, AS1220; Promega, Madison, WI). Purified mRNA was evaluated for quality and quantity by Nanodrop Spectrophotometer (Thermo Scientific).

2.2.4.3 Microarray expression measurement

Of the 178 behaviorally-phenotyped, female B6xNOD and NODxB6 F1s, 96 females were selected for microarray measurement of gene expression. The choice of 96 mice was balanced to include both directions of reciprocal cross offspring, all 4 diets, as well as both behavioral test pipelines, while simultaneously maximizing the number of represented litters. Gene expression was measured using the Affymetrix Mouse Gene 1.1 ST Array. All samples were processed by the Functional Genomics Core at UNC.

2.2.4.4 qPCR expression measurement

Commercially available Taqman qPCR assays for *Carmil1* (Life Technologies, Mm01158156_m1) and *Meg3* (Life Technologies, Mm00522599_m1) were used to estimate gene expression levels. For each sample, mRNA was retro-transcribed to cDNA using 200ng of starting RNA (SuperScript III First-Strand Synthesis System, 18080051; ThermoFisher Scientific, Waltham, MA) following the manufacturers protocol. The amplification curve was calibrated using an *Rfng* (Life Technologies, Mm00485703_m1) reference assay. All assays were performed following the manufacturers protocol on an ABI StepOne Plus Real-Time PCR System (Life Technologies, Carlsbad, CA), and in duplicate; each sample was assayed on 2 of 3 available plates. Samples were plated such that breeding batch, which explained much of the microarray expression variance, was partially confounded with qPCR plate. Cycle thresholds were determined using ABI CopyCaller v2.0 software on default settings. All available brain samples were assayed, regardless of hemisphere.

2.3 Computation and statistical models

2.3.1 Statistical Analysis of Behavior

Diet effects, POE, and diet-by-POE were evaluated using a mosgtly similar linear mixed model (LMM) for every behavior. Specifically, each behavioral phenotype was transformed to ensure residual normality (see In depth: subsection 2.4.2), and then modeled by an LMM that: 1) controlled for batch and any test-specific nuisance factors; 2) controlled for population structure by modeling dam as a random effect; and 3) modeled diet, PO, and diet-by-PO using categorical fixed effects. See subsection 2.4.1 for more details.

Every behavioral LMM was fit in R (R Core Team, 2016) using lme4 (Bates et al., 2015) and p-values calculated by a type I (i.e., sequential) sum of squares ANOVA using Satterthwaite's approximation using lmerTest (Kuznetsova et al., 2015). To account for multiple testing, the p-values were pooled over all behaviors in each pipeline, but separated per effect type (diet effects, POE, diet-by-POE); then, each pipeline/effect type combination was subject to a Benjamini-Hochberg false discovery rate correction, generating q-values (Benjamini and Hochberg, 1995).

Most test-specific nuisance factors were modeled as fixed effects, including: 1) the swimming chamber for the forced swim test; 2) the testing order for the stress induced hyperthermia and restraint stress tests; and 3) the box holding the stranger mouse for the sociability test. In repeated measures models of the startle/PPI phenotypes, random effects were used for pup and chamber (subsection 2.4.1)

For ASR data, the modeled outcome was the raw ASR divided by the mouse body weight. For the PPI at each prepulse intensity, the modeled outcome was the average PPI response divided by the weight-adjusted ASR value— a weight-and-ASR-adjusted PPI.

2.3.2 Microarray Preprocessing

Microarray probe alignments to the GRCm38.75 C57BL/B6J reference genome (the reference we use throughout) were used to infer probe binding locations (In depth: subsection 2.4.7). Using these locations, along with Affymetrix Power Tools (APT) 1.18 software (Affymetrix, 2017), probes and probesets at biased/uninformative binding locations (In depth: subsection 2.4.8) were masked. Masking reduced the original set of 28,440 non-control probesets to only 20,099 probesets (representing 19,224 unique genes, including X chromosome genes). For each remaining probeset, RMA (Irizarry et al., 2003) was applied to the non-masked probes to compute a probeset expression score. Each probeset's position was defined as the binding location of its first non-masked probe. The expression of one mouse was inadvertently measured twice; these probeset measurements were pairwise averaged.

2.3.3 Statistical Analysis of Gene Expression

Data from 95 microarray-assayed mice and 20,099 probesets was used to test diet effects, POE, and diet-by-POE on gene expression as follows. For each probeset: 1) fixed nuisance effects were regressed out of the expression score to generate adjusted expression values (see below); 2) the adjusted expression was transformed to ensure residual normality; 3) the resulting values were tested for diet, POE and diet-by-POE using an LMM that accounted for dam (using the R package nlme Pinheiro et al. 2016).

The p-value distribution for each effect type appeared to be inflated. To correct for the inflation, p-values were adjusted by a genomic-control-like procedure (Dadd et al., 2009) whereby, for all

p-values within an effect-type, an inflation factor was estimated and then divided out (In depth: subsection 2.4.5). Then, to control for multiple testing, we used two complementary approaches: Benjamini-Hochberg false discovery rate (FDR; Benjamini and Hochberg 1995), applied separately per effect type; and family-wise error rate (FWER) control, using a bespoke permutation procedure that makes minimal parametric assumptions while accounting for between-probeset correlations (see In depth: subsection 2.4.6).

2.3.3.1 Adjusted expression, SSVA estimated nuisance factors

Prior to testing for diet effects, POE and diet-by-POE, expression values for each probeset were first adjusted by regressing out nuisance effects; this was done to facilitate permutation-based threshold calculation (see In depth: subsection 2.4.6). Nuisance effects were estimated by fitting a simple linear model (to the original expression) that accounted for nuisance factors only — batch, pipeline, and a set of estimated unobserved factors. These unobserved factors were themselves estimated using a modified form of Supervised Surrogate Variable Analysis (Leek 2014), which we adapted to accommodate random effects (see In depth: subsection 2.4.4).

2.3.4 Analysis of imprinting status

Each microarray probeset was classified as measuring imprinted gene expression, if its probe sequences either: 1) hybridized to the sequence of an imprinted gene identified in Mousebook (Blake et al., 2010) or in Crowley et al. (2015); or 2) hybridized within 100bp of these known imprinted genes. All together, 241 probesets were classified as measuring imprinted regions, corresponding to 182 unique imprinted genes. Each probeset was also categorized as to whether it revealed a significant (q-value < 0.05) POE on expression of the probed region. Fishers exact test was used to calculate p-values for the association between imprinting status and significant expression POE.

2.3.5 Analysis of qPCR validation data

An apparent POE on microarray expression of *Carmil1* and a diet-by-POE on *Meg3* were validated by analysis of their respective qPCR data as follows. Each gene's qPCR relative-cycle-threshhold (relative to *Rfng*, In depth: subsection 2.4.9) was transformed for residual normality, and then modeled by an LMM that accounted for pipeline, the interaction of breeding batch with qPCR

plate (as a random efffect), and dam (random effect), as well as the diet, POE, and diet-by-POE effects. LMMs were fit using lme4 (Bates et al., 2015), with p-values computed using lmerTest (Kuznetsova et al., 2015). qPCR data analysis was repeated in three sets of mice: 1) 85 mice assayed by both microarray and qPCR; 2) 30 mice newly assayed by qPCR alone; and 3) all 115 qPCR'd mice.

2.3.6 Mediation Analysis

POEs were observed upon several behaviors, as well as upon the expression of the non-imprinted gene, *Carmil1*. To identify (potentially imprinted) genes exerting POE on these outcomes, we applied a genomewide mediation analysis. That is, for each outcome above, and for each potential mediator gene, we tested whether the gene's expression mediated POE on the outcome (details in In depth: subsection 2.4.10). For completeness, and to generate percentile-based significance thresholds, we tested every gene as a candidate mediator whether or not we observed POE on the candidate in mediation-free analysis.

This test was performed using a model (see Figure 2.9 notation) in which the outcome was the sum of: 1) outcome-specific nuisance effects (which also affect the candidate mediator gene); 2) a diet-specific *direct* effect of parent-of-origin (c'_d), and 3) a diet-specific *indirect* effect of parent-of-origin, that is mediated by way of POE on the candidate mediator gene's expression ($a_d b$). Candidate mediator genes with a significant *average* indirect effect ($ab = \overline{a_d} b$) on POE were identified as true mediators. Candidate mediator genes for which the indirect and direct effect had opposite signs were further classified as suppressors.

We note that in this model, diet does *not* modulate the effect of mediator expression on outcome; the indirect effect is diet-specific only insofar as diet affects mediator expression.

2.3.6.1 Mediation analyzed using a Bayesian approach

Most simple mediation analyses are handled using frequentist methods. However, our mediation model required that we estimate an indirect effect across multiple diets, all while accounting for the random effect for dam. For this type of complexity, a Markov Chain Monte Carlo (MCMC)-based Bayesian approach was ideal, providing the necessary flexibility to easily provide point and interval estimates of the indirect effect, all without the need to derive an analytic form (Yuan and

MacKinnon, 2009; Wang and Preacher, 2015). Our mediation model, described in more detail in In depth: subsection 2.4.10, was implemented in JAGS [Just Another Gibbs Sampler; Plummer (2003, 2016)]. Posterior medians and credible intervals for direct and indirect effects were estimated from Gibbs samples. To obtain a measure of "mediation significance", we estimated the indirect effect's "Combined Tail Probability" (CTP): the minimum of the sample-based, upper and lower tail probabilities of the indirect effect, where we deemed CTP $\leq .05$ significant (as used in, *e.g.*, Schoenrock et al. 2016).

2.3.6.2 Mediation of *Carmil1* expression

Mediation modeling of the *Carmil1* expression outcome was restricted to data from mice in which expression was measured. Batch, pipeline, and dam (a random effect), were modeled as nuisance effects acting on both the mediator gene and on *Carmil1*.

2.3.6.3 Mediation of behavior

All behavior outcomes were tested for gene mediation of POE, whether or not expression-free analysis had revealed POE on that outcome. Modeling was restricted to data from mice in which expression and behavior were both measured. Dam, batch, and behavior-specific covariates were modeled as nuisance effects on both mediator and outcome. Pipeline was *not* modeled, as each behavior was only measured in one pipeline. For PPI outcomes, groups of measurements from the same mouse/prepulse intensity were averaged together into a single value.

2.3.6.4 Aggregate mediation of behavior

To quantify each gene's aggregate level of mediation over *all* behaviors, we defined a statistic inspired by the Fisher combined p-value (Fisher, 1925): the "Combined Tail Probability" (CTP; In depth: subsection 2.4.10). Aggregate levels of mediation were also assessed by counting how often a given mediator was among the 3 most significant mediators for any behavior.

2.3.7 Reporting significant genes vs. probesets

The number of genes we report as significantly affected by some factor (*e.g.*, diet) is generally not equal to the number of significantly affected probeset measurements. The mismatch arises

because some genes (*e.g.*, *Snord 115*) are assayed by more than one probeset, and some probesets simultaneously assay more than one gene (*e.g.*, overlapping genes). For each significantly affected multi-gene probeset, we propagate significance to all of its assayed genes.

2.3.8 Test for miRNA regulation of significantly affected genes

To evaluate the validity of the diet-by-POE on *Mir341*, we tested whether the set of other genes showing diet-by-POE (by FDR) was enriched for *Mir341*'s predicted targets of regulation. Specifically, we used miRHub (Baran-Gale et al., 2013), allowing it to consider all miRNA targets predicted by TargetScan (Agarwal et al., 2015), regardless of whether those targets were conserved in another species.

2.3.9 Segregating variant determination

Variants segregating between NOD and B6 with $> .95$ probability were identified using ISVdb (Oreper et al., 2017c).

2.3.10 Computational resources

Computation was performed on Longleaf, a slurm based cluster at UNC. Up to 400 jobs were run at a time in parallel. Computation completed in approximately 6 days.

2.4 Computational Methods: in depth

2.4.1 Behavior Models

The LMM used to model behavioral phenotypes (excluding the startle/PPI phenotypes) was as follows. The behavioral outcome y_{mi} of mouse mi was modeled as

$$f(y_{mi}) \;=\; \text{intcov}_{mi} \;+\; \text{diet}_{d[m]} \;+\; \text{POE}_{s[m]} \;+\; \text{diet.by.POE}_{(sd)[m]} \;+\; \text{dam}_{m} \;+\; \varepsilon_{mi}\,, \quad (2.1)$$

where mi denotes the ith mouse of mother m; $d[m]$ denotes mother m's diet, where $d = 1, \ldots, 4$, corresponding to diets Std, ME, VDD and PD; $s[m]$ denotes the mother's strain, where $s = 1, 2$ corresponds to B6 and NOD respectively; $(sd)[m]$ denotes the mother's diet and strain combination.

Modeled effects consisted of: intcov$_{mi}$, a fixed intercept and a set of (behavior-specific) fixed effect covariates; diet$_d$, a fixed effect of diet d; POE$_s$, a fixed effect of POE (technically, strain-by-POE); diet.by.POE$_{sd}$ a fixed effect of diet-by-POE; and dam$_m$, a random effect of dam. The function $f()$ is a transformation chosen to ensure the residuals ε_{mi} are approximately normal (see In depth: subsection 2.4.2).

2.4.1.1 Startle/PPI Models

For every prepulse intensity, 6 measurements of the average startle response were taken per mouse (all in the same chamber). The startle/PPI LMMs therefore accounted for repeated measures. Letting $y_{mi,j}$ be mouse mi's jth measurement, we modeled:

$$f(y_{mi,j}) = \text{intcov}_{mi} + \text{diet}_{d[m]} + \text{POE}_{s[m]} + \text{diet.by.POE}_{b[m]} + \\ \text{chamber}_{h[mi]} + \text{dam}_m + \text{pup}_{mi} + \varepsilon_{mi,j} \quad (2.2)$$

where chamber$_{h[mi]}$ is a random effect of chamber, and pup$_{mi}$ is the random effect of mouse mi.

2.4.2 Variable transformation procedure

A transformation procedure was applied to both the expression and the behavior phenotypes to ensure residual normality. For a given LMM requiring a transformation of the outcome y, the procedure was as follows. Center and scale y to mean 0 and standard deviation 1 to give z. Apply a shifted Box-Cox transformation (Sakia, 1992; Box and Cox, 1964), restricted to the ladder of powers $\lambda \in \{-3, -2, -1, -.5, 0, .5, 1, 2, 3\}$ to give in each case values $z^{(\lambda)}$. For each transformation $z^{(\lambda)}$, the LMM is fitted, and residual normality is evaluated using the Shapiro-Wilk W statistic (Shapiro and Wilk, 1965); denote the optimal λ as $\hat{\lambda}$. If $\hat{\lambda} \in \{0, .5, 1, 2\}$, then use $z^{(\hat{\lambda})}$; if $\hat{\lambda} \in \{-2, -1, -.5\}$, then additionally negate the value, in order to ensure the monotonicity of the transformation and thereby improve interpretability of effect estimates; if the $\hat{\lambda} \in \{-3, 3\}$, then discard the transformation and instead apply a rank inverse normal transform (Van der Waerden, 1952). Rescale the selected transformed variable to mean 0 and standard deviation 1.

2.4.3 Microarray expression models

Expression was first adjusted by regressing out nuisance factors, and then the *adjusted* expression was modeled to test diet, POE, and diet-by-POE. This two-step process was employed to facilitate permutation testing later on.

2.4.3.1 Generation of the adjusted expression outcome

Letting $y_{mi,j}$ be the average expression of probes in probeset j for mouse mi, we obtained adjusted expression values as residuals $\hat{\varepsilon}_{mi,j}$ from the linear model:

$$f(y_{mi,j}) = \text{intcov}_{mi,j} + \text{SV}_{mi,j} + \varepsilon_{mi,j}\,, \tag{2.3}$$

where the covariates in $\text{intcov}_{mi,j}$ were the nuisance effects of pipeline and behavioral batch. The $\text{SV}_{mi,j}$ term modeled fixed effects for 7 "surrogate variables" (SVs), which represented aggregate effects of unobserved confounding on the microarray (see In depth: subsection 2.4.4). Specifically, $\text{SV}_{mi,j} = \sum_{k=1}^{7} \beta_{k,j} v_{mi,k}$, where $v_{mi,k}$ is mouse mi's value for the kth SV, and $\beta_{k,j}$ is the fixed effect of that SV on the expression of probeset j. (Estimation of the SVs themselves is described in In depth: subsection 2.4.4)

2.4.3.2 Model of adjusted expression outcome

For each probeset j, adjusted expression (a.k.a., the residuals from Eq 2.3) was then analyzed using the LMM,

$$f^{(\varepsilon)}(\hat{\varepsilon}_{mi,j}) \;=\; \mu_j \;+\; \text{diet}_{d[m],j} \;+\; \text{POE}_{s[m],j} \;+\; \text{diet.by.POE}_{(sd)[m],j} \;+\; \text{dam}_{m,j} \;+\; \epsilon_{mi,j}\,, \tag{2.4}$$

where μ_j and $\epsilon_{mi,j}$ are the intercept and residual error, $f^{(\varepsilon)}$ is a transformation that may be different from f in Eq 2.3, and other terms are defined as in Eq 2.1.

2.4.4 Surrogate variable estimation allowing for random effects

Gene expression measurements by microarray are typically affected by many unobserved factors, some of which can have a large confounding effect on transcript levels across many genes. One way

to control for such unobserved factors is to first model their aggregate effects as linear combinations of "surrogate variables" (SVs; Leek and Storey 2007), and then include these SVs as predictors in subsequent modeling, and/or regress these effects out (as in In depth: subsection 2.4.3).

Here we mostly— deviating somewhat to accommodate random effects and variable transformation— follow the Supervised Surrogate Variable Analysis (SSVA) approach of Leek (2014), which defines the SVs using negative control probes; success of this approach requires that unobserved confounding effects arise from technical rather than biological variation. As a further aside, we note that our approach is also largely equivalent to the "remove unwanted variation with negative control genes" (RUVg) strategy (Risso et al., 2014), applied to microarray data.

In our implementation of SSVA, we first estimate a standardized matrix of the aggregate effects that arise from unobserved factors, E. For each negative control probe $c = 1, \ldots, C$, we fitted the LMM

$$f(y_{mi,c}) \;=\; \text{intcov}_{mi,c} + \text{diet}_{d[m],c} + \text{POE}_{s[m],c} + \text{diet.by.POE}_{(sd)[m],c} + \text{dam}_{m,c} + \varepsilon_{mi,c}\,,$$

where terms are defined as in Eq 2.1 and Eq 2.4, and where the estimated residuals, $\hat{\varepsilon}_{mi,c}$, were standardized and stored in n-vector $\mathbf{e}_c$. These steps were repeated for all C negative control probes to give the $n \times C$ matrix $\mathbf{E}$.

Let the SVD of E be denoted as $\mathbf{U\Sigma V}'$. Under this parameterization, the space of aggregate unobserved factor effects on the control probes is (by construction) spanned by the n columns of $\mathbf{U}$. Since a model for main probes that included all n columns as surrogate variables would be unidentifiable, the first $K = 7$ columns of $\mathbf{U}$ were chosen as an approximating subset of surrogate variables. $K = 7$ was chosen by following the strategy described in Sun et al. (2012) for K-selection in SVA with random effects: a plot of the squared eigenvalues from $\mathbf{\Sigma}$ was examined, and it revealed an inflection point at 7 eigenvalues.

Of note, the original implementation of SSVA did not regress any effects out of control probes, under the assumption that these probes should be unaffected; in contrast, we regress these effects out before computing eignevectors. We justify this by noting that if in fact the treatments of interest somehow did affect the control probes, we would not want these treatment effects to be incorporated

into the surrogate variables. And if the control probes truly are unaffected by any of the observed experimental factors, then there should be no harm in residualizing out these size-zero effects.

2.4.5 Bias-adjustment for gene expression p-values

For some effect types that were tested in the gene expression model of Eq 2.4, the distribution of nominal p-values across all transcripts was consistent with those p-values being downwardly biased. To remove this bias, which would otherwise invalidate our use of FDR, we applied an empirical adjustment similar to the genomic control procedure of Devlin et al. (2001) (see also Dadd et al. 2009). Let p_j be the p-value associated with a given effect type (diet, POE, or diet-by-POE) on the jth probeset, let $F(x)$ be the cumulative distribution function for the χ^2_1 density, and define $x_j = F^{-1}(p_j)$ and $\mathbf{x} = (x_1, \ldots, x_m)$. Under unbiasedness, p-values associated with testing for given effect should, under the null, have a uniform distribution, $p_j \sim \text{Unif}(0, 1)$, such that $x_j \sim \chi^2_1$. Assuming most results are in fact null, in the dataset as a whole we would expect $\text{median}(\mathbf{x}) \simeq F^{-1}(0.5)$. However, if significances were systematically inflated, the null x_j's would appear as if from a scaled χ^2_1 such that $x_j/\lambda \sim \chi^2_1$ with inflation factor $\lambda > 1$. Therefore, we correct for this systematic inflation by first estimating the inflation factor as $\hat{\lambda} = \text{median}(\mathbf{x})/F^{-1}(0.5)$ and then calculating bias-adjusted p-values as $\tilde{p}_j = F(x_j/\hat{\lambda})$.

2.4.6 Permutation-based FWER thresholds for gene expression p-values

For gene expression, empirical p-value thresholds that controlled for the family-wise error rate (FWER) across all probesets were determined by permutation. A separate FWER threshold was computed per effect of interest (diet, parent-of-origin, and diet-by-parent-of-origin). Below, we describe the permutations that were generated, the statistic that was collected per permutation, and how this was translated into a significance threshold.

2.4.6.1 Structure of permutation

For every permutation-tested effect type, we generated a separate set of $W = 401$ permutations (including the identity permutation), $w = 1 \ldots, W$. Litters were taken as exchangeable units; diet/strain labels were permuted amongst the dams, and all pups of a given dam were assigned their

dam's diet/strain label. Permuting *labels*, rather than outcomes, enabled us to allow for varying litter sizes between dams.

For the main effects we employed a form of restricted permutation (Anderson and Braak, 2003; Good, 2005); i.e., for parent-of-origin effects, we randomly permuted the strain labels ($\mathbf{s}$ in Eq 2.4) between dams that had been *exposed to the same diet*, whereas for diet effects, we randomly permuted diet labels ($\mathbf{d}$) between dams *of the same strain*.

For the interaction effect of diet-by-POE, we employed a form of unrestricted permutation (Anderson and Braak, 2003; Good, 2005) of the interaction labels. In particular, we permuted the interaction labels $\mathbf{g}$ between dams. However, the $\mathbf{s}$ and $\mathbf{d}$ labels were held constant even as the interaction labels $\mathbf{g}$ were permuted.

2.4.6.2 Permutation statistic and threshold computation

For each permutation w and probeset $j = 1, \ldots, J$ we fitted the expression LMM of Eq 2.4. Note that the modeled outcome in this equation is adjusted gene expression from which *all nuisance covariates have already been regressed*; following Gail et al. (1988), this residualization was performed to facilitate exchangeability for the effects of interest. For every permutation, the fitting of 2.4 included recalculation of the transformation $f^{(\varepsilon)}$. Furthermore, for every permutation, we bias-adjusted (through genomic control, In depth: subsection 2.4.5) the p-values, $\tilde{\mathbf{p}} = \tilde{p}_1^{(w)}, \ldots, \tilde{p}_J^{(w)}$ and recorded the minimum, $p_{\text{min}}^{(w)}$.

The set of W such minimum p-values from all permutations was then used to estimate the FWER $\alpha = 0.05$ threshold via modeling of a generalized extreme value (GEV) distribution after Dudbridge and Koeleman (2004); Manly (2006). Specifically, a GEV was fitted to $T_w = -\log[p_{\text{min}}^{(w)}]$ for $w = 1, \ldots, W$ using R package evir (Pfaff and McNeil, 2012), and the fitted GEV was used to estimate the upper 5% quantile, $T_{\alpha=.05}$. $T_{\alpha=.05}$ was then translated back into a threshold on the p-value scale as $p_{\alpha=.05} = e^{-T_{\alpha=.05}}$. Note that, as a conservative measure, the GEV fit included the identity permutation.

2.4.7 Probe alignments and estimated probeset positions

Probe alignments were downloaded from the Ensembl 38.75 funcgen database (Yates et al., 2016). Notably, this database contained alignments for MoGene1.0 ST probes, rather than for the

MoGene1.1 ST probes that we used in our experiment. To address this mismatch, we imputed 1.1 alignments by using the fact that every 1.1 probe is identical to at least one 1.0 probe in sequence (though not in probe id); we formed correspondences from each 1.1 probe to its identical-sequence 1.0 probe alignment. Since most probes aligned to multiple positions, we estimated per probe and per probeset, the "intended" target position, defining this self-referentially as the position that minimizes the sum of distances between probes in the same probeset.

2.4.8 Criteria for masking biased and uninformative probes/probesets

APT masking was used to eliminate four types of probes: 1) probes aligning to ≥ 100 locations; 2) probes aligning outside of annotated exons; 3) probes whose "interior" (basepairs 3-21) aligned to regions in which NOD possesses a variant relative to B6, i.e., probes with a binding affinity difference between strains (Dannemann et al., 2009), where NOD variants were extracted from the Inbred Strain Variant Database (Oreper et al., 2017c); or 4) redundant probes mapping to the same position. Following probe masking, probe*sets* were eliminated if they contained <4 non-masked probes, or if every remaining non-masked probe measured <32 units of expression across all samples.

2.4.9 qPCR analysis

2.4.9.1 qPCR model

Letting $y'_{mi,j}$ be the qPCR relative cycle threshold for a targetted gene (*Meg3* or *Carmil1*), we modeled:

$$f(y'_{mi,j}) = \text{intcov}_{mi} + \text{diet}_{d[m]} + \text{POE}_{s[m]} + \text{diet.by.POE}_{b[m]} + \text{dam}_{m} + \text{batch.plate}_{a[mi]} + \varepsilon_{mi,j}\ ,$$

where intcov includes the intercept and behavioral pipeline, batch.plate$_a$ is a random effect of the combination a of breeding batch and qPCR plate, and the other terms are akin to those defined in the microarray model (In depth: subsection 2.4.3).

2.4.9.2 qPCR normalization

The raw value measured by qPCR is a target gene's cycle threshold. To allow comparison between qPCR batches, which can vary in replication efficiency, the cycle threshold for a target gene must be normalized by some reference gene that is unaffected by biological factors. As such, rather than modeling the cycle threshold, we model the relative cycle threshold, defined as $\Delta Ct = Ct_{\text{target}} - Ct_{\text{reference}}$. The ΔCt relative cycle threshold represents the relative gene expression level of the target gene on the log scale (Didion et al., 2015). The larger ΔCt is, the less the target gene expression.

We chose *Rfng* as the reference gene, because microarray data suggested negligible effects of diet, POE and diet-by-POE on *Rfng* expression. Specifically, each candidate reference gene was assigned a score equal to the minimum of the p-values for POE, diet-by-POE, and diet effects on the candidate reference's microarray-measured expression. *Rfng* had the largest such score.

2.4.10 Bayesian mediation model

Mediation analysis is typically posed as the estimation of the model in Figure 2.2: An intervention or predictor variable X affects an outcome Y either directly or/and through an observed mediator outcome M. In our case, X is reciprocal direction (*i.e.*, parent-of-origin, coded as the maternal strain), M is the expression of a mediator gene, and Y is the outcome of primary interest, either expression of *Carmil1* or a behavioral phenotype. By common convention, the effect of X on M, *i.e.*, the POE on M, is denoted a, which in our case is a_d to allow different effects under each diet d, and the effect of M on Y is denoted b. The product $a_d b$ is then the expression-mediated effect of parent-of-origin on Y, conditional on the diet d, and our primary quantity of interest is this value averaged over diets, $ab = \overline{a_d} b$. The direct effect of X on Y after accounting for mediation by M is denoted c', which in our case is analogously diet-specific and denoted here as c'_d with average direct effect $c' = \overline{c'_d}$. (Not explicitly calculated here but used elsewhere is c, which would be the effect of X on Y if mediation were unmodeled.) When ab and c' have opposite signs, mediation by way of M acts to suppress the overall parent-of-origin effect on the outcome Y.

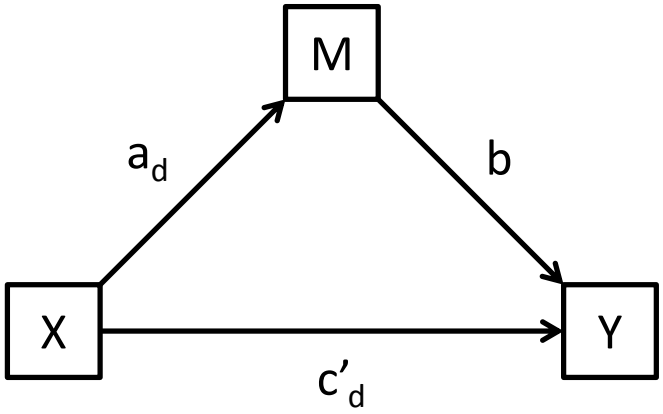


Figure 2.2: Multilevel mediation model in which the levels are diets. X represents the maternal-strain treatment, Y is the outcome (behavior or expression), and M is the mediating gene expression factor. a_d is the (diet-specific) effect of the treatment on the mediator, c'_d is the (diet-specific effect) direct effect of the treatment on the outcome, and b is the (diet-independent) effect of the mediator value on the outcome.

2.4.10.1 Linked LMMs

Our mediation model for the effect of a gene-expression-mediator z on an outcome y is specified via two linked LMMs as:

$$f(y_{mi}) = \text{intcov}_{mi} + \text{diet}_{d[m]} + \text{POE}_{s[m]} + \text{diet.by.POE}_{(sd)[m]} + \underbrace{b \cdot f^{(z)}(z_{mi})}_{\text{effect of } z \text{ on } y} + \text{dam}_m + \varepsilon_{mi}, \tag{2.5}$$

where $f^{(z)}$ denotes a transformation that may be different from f, b is the effect of mediator z on y, and the combined contribution of POE and diet.by.POE provides the direct effect c'_d. Meanwhile, mediator z is simultaneously modeled as

$$f^{(z)}(z_{mi}) \;=\; \text{intcov}^{(z)}_{mi} \;+\; \text{diet}^{(z)}_{d[m]} \;+\; \text{POE}^{(z)}_{s[m]} \;+\; \text{diet.by.POE}^{(z)}_{(sd)[m]} \;+\; \text{dam}^{(z)}_{m} \;+\; \varepsilon^{(z)}_{mi}\,, \tag{2.6}$$

where, for example, the notation $\text{intcov}^{(z)}_{mi}$ means the same regression input as intcov_{mi} but with regression coefficients specific to mediator z rather than outcome y, and the combined contribution of $\text{POE}^{(z)}$ and $\text{diet.by.POE}^{(z)}$ provides the effect a_d. Specifically, the correspondence of Eq 2.5 and

Eq 2.6 to the more general mediation analysis is as follows:

$$
\begin{aligned}
X_{mi} &= I(s[m] = \text{NOD}) && \text{(\textit{i.e.}, 1 if NOD maternal strain, 0 o/w)} \\
M_{mi} &= f^{(z)}(z_{mi}) \\
Y_{mi} &= f(y_{mi}) \\
b &= b && \text{(effect of } z \text{ on } y\text{)} \\
a_d &= \text{POE}^{(z)}_{\text{NOD}} + \text{diet.by.POE}^{(z)}_{\text{NOD},\,d} \\
c'_d &= \text{POE}_{\text{NOD}} + \text{diet.by.POE}_{\text{NOD},\,d} \\
a &= \overline{a_d} && \text{(all-diets-average POE on } z\text{)} \\
ab &= \overline{a_d} b && \text{(all-diets-average mediated POE on } y\text{)} \\
c' &= \overline{c'_d} && \text{(all-diets-average direct POE on } y\text{)}
\end{aligned}
$$

where POE_{NOD} is the effect of switching from an NOD mother to a B6 mother, and $\text{diet.by.POE}_{\text{NOD},\,d}$ is the additional effect of this for diet d.

2.4.10.2 Transformations, expression adjustment, priors, and MCMC sampling.

Prior to fitting the Bayesian mediation model, all candidate mediators and outcomes were transformed using the same process as described earlier; *i.e.*, transforms were chosen to ensure normality using the frequentist, mediation-free models (subsection 2.4.2, C). Additionally, akin to the mediation-free microarray analysis, surrogate variable effects (In depth: subsection 2.4.4) were regressed out of every gene's expression prior to mediation modeling. However, unlike the mediation-free analysis of expression, batch and pipeline were *not* regressed out, and were included as nuisance effects on mediator and outcome in the mediation model. Priors were specified as follows, noting that M and Y by construction have means of 0 and standard deviation 1: fixed effects (*i.e.*, all effects except dam) were given priors of $\text{N}(0, 5^2)$; and the random effect of dam was modeled as drawn from $\text{N}(0, \tau^2)$ with $\tau^2 \sim \text{Unif}(0, 25)$. Model fitting proceded by running a single MCMC chain for 16,000 timesteps, of which the first 3,200 were discarded (*i.e.*, as burn-in), and the last 12,800 were retained for estimation.

2.4.10.3 Combined Tail Probability: a statistic to quantify aggregate mediation

To quantify the extent to which a given gene's expression mediated POE on multiple outcomes, we use a statistic inspired by the Fisher combined p-value that we refer to as the "Combined Tail Probability" (CTP). The CTP is the probability that a value drawn from χ^2_{2K} is at least as extreme as the statistic $T = -2\sum_k^K ln(p_k)$, where K is the number of outcomes tested for mediation, and p_k is the CTP for the mediator's indirect effect on outcome k. Although the implicit distributional assumption is not strictly justified, the CTP associated with T provides a statistic for evaluating which mediators are strongest in aggregate.

2.4.11 Data Availability

Data and supplemental results files are stored on Zenodo at https://doi.org/10.5281/zenodo.1168578 (Oreper et al., 2018). File S1 contains detailed descriptions of all supplemental files. File S2 contains chromosome sizes. File S3 contains exon data. File S4 contains Snord data. Files S5, S6, and S7, contain imprinted genes from Crowley et al. (2015), Mousebook, and the union thereof, respectively. File S8 contains NOD variants. File S9 contains covariates for RF1s. Files S10 and S11 contain Affymetrix library files for the Exon 1.1 ST and 1.0 ST microarrays, respectively. File S12 contains 1.0 ST probe binding locations. File S13 contains raw (CEL) microarray-measured expression for RF1s. File S14 contains a summary of microarray expression— the output from APT-summarize, but with default args. File S15 contains pulverized brain data, pre-qPCR validation. File S16 contains qPCR data. File S17 contains behavior models for mediation analysis. File S18 contains bodyweights. File S19 contains cocaine responses. File S20 contains FST data. File S21 contains light/dark data. File S22 contains OF data. File S23 contains restraint stress data. File S24 contains SIH data. File S25 contains sociability data. File S26 contains startle/PPI data. File S27 contains tail suspension data. File S28, S29, and S30 contain POE, diet, and diet-by-POE expression modeling results, respectively. File S31 and S32 contain mediation analysis results for the *Carmil1* and behavior outcomes, respectively. Code to generate results is available at https://github.com/danoreper/mnp2018.git.

2.5 Results

2.5.1 Overview and key results

NOD and B6 mice were reciprocally crossed, with F1 hybrids exposed perinatally to Std, VDD, ME, MD, and PD diets (the MD diet was eventually dropped due to a near total lack of reproductive/weaning productivity; Table A.2). Following weaning, the female F1 hybrids were tested in one of two different pipelines, each of which consisted of a different set of behavioral tests (Figure A.1). Following behavioral testing, whole brain gene expression was measured via microarray. Analysis and validation lead to the following key results (detailed in subsequent subsections):

- Parent-of-origin affected 7 behaviors, including multiple locomotor behaviors and SIH behavior.
- Perinatal diet affected body weight and PPI behavior.
- Diet-by-POE acted on OF percent center time.
- Diet, POE, and diet-by-POE significantly (by FWER) acted on expression of 37, 15, and 16 genes respectively.
- The significance of diet's effect on expression was primarily driven by ME.
- Notable POE were observed on *Snord 115*, *Airn*, and most significantly on *Carmil1*, a non-imprinted gene.
- The *Carmil1* POE was qPCR-validated in two sets of mice: the microarrayed mice, and a new set of mice.
- Genes affected by POE are enriched for imprinting.
- POE on *Carmil1* seems to be mediated (specifically, suppressed) by the expression of the imprinted gene *Airn*;
- *Carmil1*, and *Snord 115*, and especially *Airn* seem to mediate POE on multiple behaviors. These, along with other identified mediators of behavioral POE, tend to be suppressors.

2.5.2 Effects on behavior

At a nominal level, POE, diet, and diet-by-POE acted significantly upon 7, 4, and 2 behaviors, respectively. Post-FDR correction, POE, diet, and diet-by-POE acted upon 3, 0, and 0 behaviors, respectively. Table 2.1 shows per-variable p-values, whereas Table A.5 shows tukey p-values for variable level contrasts.

2.5.2.1 POE acts upon several locomotor behaviors, as well as SIH and PPI outcomes

Across several assays and both pipelines, a significant POE was observed on 5 different assessments of locomotor behavior. In all 5 assesments, NODxB6 mice moved more than B6xNOD mice. In pipeline 1, in the Light/Dark test, a POE was observed on both total distance and distance moved on the dark side of the arena (p=0.0493, q=0.181; p=0.0187, q=0.103 respectively), but not on light side distance (p=0.273; Figure 2.3A). Also in pipeline 1, in the cocaine response assay, a POE was observed on total OF distance, on both the baseline and the habituation day (Day 1, p=0.000671, q=.00975; Day 2, p=0.00221, q=0.0162 respectively) (Figure 2.3B). In pipeline 2, in a separate set of OF-assessed mice, a POE was observed upon total-distance moved (p=0.013, q=0.156; (Figure 2.3B).

POE was also observed on post-stress temperature in the SIH assay, with B6xNOD mice having higher temperatures (p=0.000887, q=0.00975; Figure 2.4). A smaller, non-significant effect in the same direction was also seen for both basal temperature (SIH-T1) and change in temperature (SIH-delta), consistent with a small difference in basal temperature being magnified after stress. A significant POE was also observed on PPI at 82 decibels, with B6xNOD mice exhibiting a higher percent PPI than NODxB6 (p=0.0307 and q=0.00274; Figure A.2A). A similar effect (to that at 82 decibels) was observed at 86 decibels, but it was not significant (Figure A.2A).

2.5.2.2 Diet has nominally significant effects on body weight and PPI

At a nominal level, perinatal diet significantly affected body weight (p=0.00541, q=.0595; Figure 2.5), with mice exposed to ME diet weighing less than mice exposed to Std and VDD diets (Tukey post-hoc p=0.0228 and p=0.0402). Diet also significantly affected measures of sensorimotor gating: in particular, PPI at 82 decibels (p=0.00274, q=0.0595; Figure A.2B). At 78 decibels, PD had a non-significant (p=.0714, q =.524), but similar effect (Figure A.2B). At both 78 and 82 decibels,

Pipeline	Test	Phenotype	Covariates	p-value			FDR adjusted p-value		
				POE	Diet	DietxPOE	POE	Diet	DietxPOE
1	Light/Dark	Total Distance	Batch, Dam	**0.0493***	0.481	0.99	0.181	0.814	0.99
		Distance Dark		**0.0187***	0.646	0.985	0.103	0.836	0.99
		Distance Light		0.273	0.247	0.905	0.43	0.68	0.99
		% Time Dark		0.373	0.175	0.392	0.547	0.561	0.92
		% Time Light		0.226	0.129	0.341	0.414	0.561	0.92
		Total Transitions		0.0772.	0.904	0.61	0.243	0.904	0.92
	Startle/Prepulse Inhibition	AS50 Average	Batch, Chamber, Dam	0.399	0.617	0.0904.	0.548	0.836	0.731
		AS50 Latency		0.935	0.149	0.432	0.98	0.561	0.92
		Average PPI 74	Batch, Chamber, Dam, Pup	0.217	0.481	0.565	0.414	0.814	0.92
		Average PPI 78		0.22	0.0714.	0.636	0.414	0.524	0.92
		Average PPI 82		**0.0307***	**0.00274****	0.445	0.135	0.0595.	0.92
		Average PPI 86		0.123	0.179	0.669	0.301	0.561	0.92
		Average PPI 90		0.988	0.62	0.0997.	0.988	0.836	0.731
	Stress-Induced Hyperthermia	SIH-T1	Batch, Test Order, Dam	0.273	0.828	0.61	0.43	0.904	0.92
		SIH-T2		**0.000887*****	0.628	0.0624.	**0.00975****	0.836	0.731
		SIH-Delta		0.648	0.879	0.473	0.839	0.904	0.92
	Forced Swim	% Immobility	Batch, Arena, Dam	0.111	0.317	0.531	0.301	0.776	0.92
	Cocaine Response	Day1 Distance	Batch, Dam	**0.000671*****	0.43	0.332	**0.00975****	0.814	0.92
		Day2 Distance		**0.00221****	0.47	0.325	**0.0162***	0.814	0.92
		Day3 Distance		0.782	0.692	0.876	0.906	0.846	0.99
		Day3-Day2 Distance		0.73	0.771	0.897	0.892	0.893	0.99
	Body Weight	Body Weight	Batch, Dam	0.908	**0.00541****	0.913	0.98	0.0595.	0.99
2	Open Field	Distance Moved	Batch, Dam	**0.013***	0.647	0.555	0.156	0.647	0.832
		% Center Time		0.319	0.234	**0.0144***	0.638	0.592	0.172
		Average Velocity		0.428	0.128	0.145	0.638	0.511	0.435
		Jump Counts		0.788	0.312	0.223	0.788	0.592	0.447
		Vertical Counts		0.0763.	0.103	0.932	0.318	0.511	0.932
		Boli Count		0.466	0.113	0.301	0.638	0.511	0.517
	Social Interaction	% Time Stranger	Batch, Stranger Box, Dam	0.425	0.493	0.182	0.638	0.592	0.438
		Transitions		0.705	0.633	0.72	0.769	0.647	0.864
	Tail Suspension	% Immobility	Batch, Dam	0.536	0.305	0.652	0.643	0.592	0.864
	Restraint Stress	Basal CORT	Batch, Test Order, Dam	0.478	0.475	0.923	0.638	0.592	0.932
		10 min CORT		0.0796.	0.372	0.0735.	0.318	0.592	0.388
		Δ CORT		0.113	0.412	0.097.	0.338	0.592	0.388

Table 2.1: POE, perinatal diet effect, and diet-by-POE on behavioral phenotypes. For each phenotype, the table shows the modeled variables, along with the p-values of interest, and their corresponding q-values (FDR), which account for multiple testing within a behavioral pipeline. Significant values are **bolded**, and *, **, and ***, indicate significance levels of *0.05, **0.01, ***0.001 respectively. POE = parent of origin effect; PPI = prepulse inhibition; CORT = corticosterone; SIH-T1 = basal temperature; SIH-T2 = post-stress temperature; SIH-delta = (T2-T1)

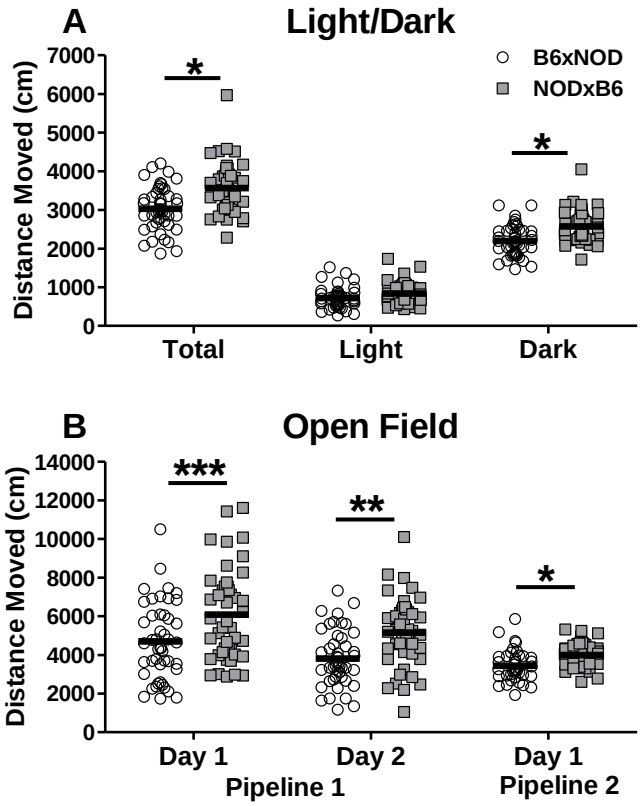


Figure 2.3: POEs on locomotor behavior are consistent across behavioral tests and pipelines. (A) Light side, dark side, and total distance moved in the light/dark arena for individual B6xNOD (n=46) and NODxB6 (n=45) mice (bars indicate mean). (B) OF distance moved for B6xNOD (n=46) and NODxB6 (n=45) mice, in Pipeline 1 on Day 1 and 2 of a 30 min cocaine response test when the mice received an ip saline injection; Distance moved in Pipeline 2 in a separate 10 min OF test (B6xNOD:n=39, NODxB6:n=48). For all assays, NODxB6 mice move significantly more than B6xNOD mice. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

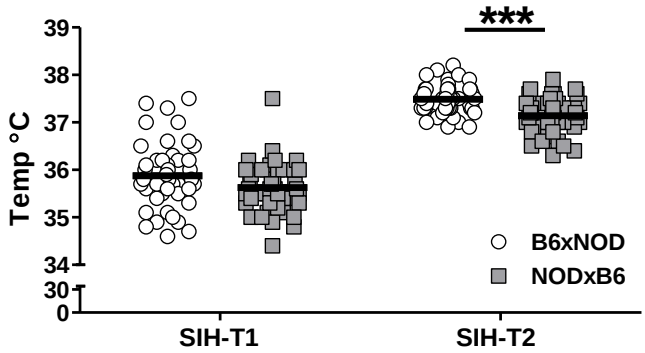


Figure 2.4: POEs on baseline (SIH-T1) and post-stress induced temperature (SIH-T2) in the stress induced hyperthermia test. Data are for individual B6xNOD (n=46) and NODxB6 (n=45) mice (bars indicate mean). For SIH-T2 B6xNOD mice have higher temperature than NODxB6 mice. A similar, though non-significant pattern seems to occur in the the SIH-T1 data

PPI seemed greatest for PD mice compared to other diets, although individual contrasts were not significant (Figure A.2B, Table A.5)

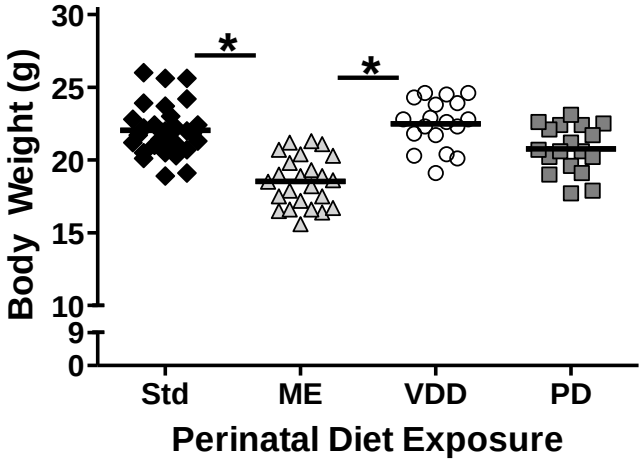


Figure 2.5: Effect of perinatal diet exposure on body weight in adulthood. Body weight of individual mice (bars indicate mean) exposed to either standard (Std, n=31), methyl enriched (ME, n=24), protein deficient (PD, n=18) or vitamin D deficient (VDD, n=18) diet during the perinatal period. Perinatal diet significantly affected body weight (p=0.00541). *indicates a significant difference between ME from Std and VDD mice ($p < 0.05$)

2.5.2.3 Diet interacts with parent-of-origin to alter percent center time

A nominally significant diet-by-POE was observed on percent center time in the OF test (p=0.0144, q=0.172; Figure 2.6). In this test, NODxB6 mice exposed to VDD and PD diets spent more time in the center of the arena than diet-matching B6xNOD mice, but no such difference was seen for ME or Std diets. Similar but non-significant effects were seen on OF locomotor activity (Figure A.3).

2.5.3 Effects on whole-brain gene expression

Gene expression at each microarray probeset was tested for POE, diet effects, and diet-by-POE. Significance was assessed in two ways: using the false discovery rate (FDR), and using a more conservative, permutation-based family wise error rate (FWER) threshold. The FDR (q-value = 0.05) and FWER (adjusted p-value = 0.05) thresholds were nearly identical for POE, were similar for diet-by-POE, but were over two orders of magnitude different for diet, with FWER more conservative.

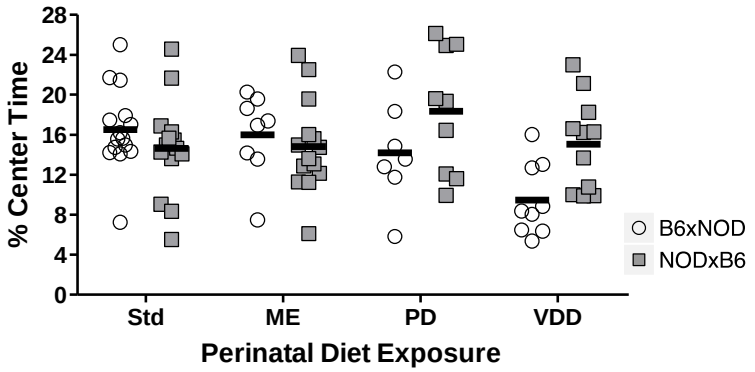


Figure 2.6: Perinatal diet-by-POE on percent center time in the 10 min OF test, for B6xNOD and NODxB6 mice exposed to Std (n=15,14), ME (n=8,14), PD (n=7,9) or VDD (n=9,11) diets; although no individual contrast is significant, diet-by-POE (p=0.0144) is significant overall.

2.5.3.1 POE detected on 15 genes, 9 imprinted.

POE was FWER-significant for 15 genes (Table A.9; Figure 2.7), a significant subset of which (nine) were imprinted ($p < 2.2x10^{-16}$). Across the 16 genes, greater expression was not associated with either cross direction (seven more expressed in NODxB6; ten more expressed in B6xNOD). Both patterns were seen in imprinted genes *Snord 113* and *Snord 115*, depending on the subregion (Table A.9). Significant POEs clustered on (Figure 2.7A) chromosome 7 in the vicinity of the imprinted *Snord 115/116* family, and on chromosome 12 near the imprinted *Snord 113* family.

2.5.3.2 POE on non-imprinted *Carmil1* validated by qPCR.

The most significant POE was on *Carmil1* ($-\log_{10}(p) = 13.8$). This POE was consistent across diets (Figure 2.7B), and was validated by qPCR. qPCR was performed on 115 mice, 85 of which had already been assayed by microarray. POE on *Carmil1* was significant whether considering qPCR data from all 115 (p=6.3e-7), only the 85 (p=4.4e-07), or the qPCR-only 30 (p=9.7e-11) (see Table A.4).

2.5.3.3 According to FWER, diet affects 37 (solely non-imprinted) genes

The most significantly affected was *Cnot2* ($-\log_{10}(p) = 7.4$). For 35 of the 37 genes (Table A.10, Figure A.5), significance was driven by the ME diet: across the 4 diets, these 35 genes were either most or least expressed in ME mice (See the "ME group rank" field in Table A.10; Figure A.5). By the less stringent FDR threshold, diet significantly affected 958 genes. This included even Y

Effect Type	Significance threshhold		# Significantly affected		
	Type	-log$_{10}$(thresh)	Probesets	Genes	Imprinted genes
POE	FWER	5.08	20	15	9
Diet	FWER	4.97	33	37	0
Diet x POE	FWER	4.68	17	16	1
POE	FDR	4.25	26	19	10
Diet	FDR	2.61	983	958	12
Diet x POE	FDR	3.43	149	154	7

Table 2.2: Microarray-measured effects on expression. For each effect type/significance threshold type, the table specifies the significance threshold value, as well as the number of probesets, genes, and imprinted genes whose expression was significantly affected. Note that: i) some probesets measure multiple genes, and some genes are measured by multiple probesets; ii) the FDR and FWER thresholds for diet differ greatly; iii) imprinting is enriched among genes subject to POE, and iv) by FWER, diet does not affect any imprinted gene, whereas one imprinted gene is subject to diet-by-POE

chromosome genes (Table A.10), suggesting, since we only use females, that the FWER threshold is more appropriate.

2.5.3.4 According to FWER, diet-by-POE affects 16 genes, with only *Mir341* imprinted

Not only was *Mir341* the only significantly affected imprinted gene, but it was also the most significantly affected ($-\log_{10}$(p) =6.5; Table A.11). However, despite *Mir341* being expected to regulate hundreds of genes (Targetscan) the 149 (*FDR selected*-genes significantly subject to diet-by-POE were not enriched for *Mir341*'s predicted regulatory targets (p=.999; using miRHub; Baran-Gale et al. (2013)). Following *Mir341*, the imprinted gene *Meg3* was the next most significant imprinted gene, subject diet-by-POE (but only by FDR; $-\log_{10}$(p) =4.4; Figure 2.8; Figure A.4A). However, this weakly significant effect on *Meg3* was not reproduced in qPCR validation (Table A.4).

2.5.4 Mediation of POE by way of gene expression

2.5.4.1 POE on the gene expression of non-imprinted gene *Carmil1* may be mediated by *Airn*

The microarray and qPCR-based evidence for POE on *Carmil1* expression raised the question: given that *Carmil1* is not known to be imprinted, might *Carmil1* expression be regulated (*i.e.*, mediated) by some imprinted genes expression?

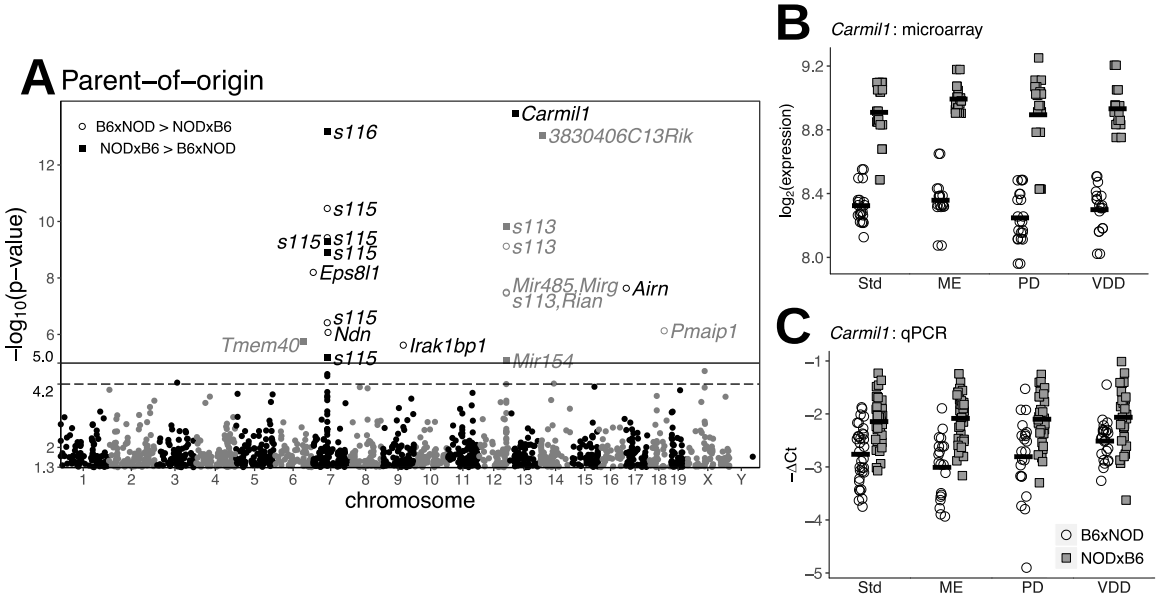


Figure 2.7: POEs on *Carmil1* gene expression. (A) Manhattan-like plot of p-values of POE on microarray-based gene expression; each point corresponds to a probeset's genomic location, coupled with the p-value of POE on expression at that location. Probesets with a nominal p-value $> .05$ are not shown. The dashed and solid lines represent the FDR and FWER thresholds, respectively. Probesets above the FWER threshold are labeled with the gene(s) that they interrogate. The *S113, S115,* and *S116* labels are shorthand for *Snord 113, Snord 115, Snord 116* respectively. Labeled points are shaped according to whether expression was greater in B6xNOD or NODxB6. The most significant POE is on *Carmil1*. (B) Raw microarray expression data for *Carmil1*; circles and squares represent expression for B6xNOD and NODxB6 hybrids, respectively. POE on expression is evident under all dietary exposures. (C) qPCR validation data for *Carmil1*, showing the same significant pattern of POE in all dietary exposures, confirming the microarray findings. In any qPCR assay, increased expression *reduces* ΔCt; consequently, we use the y-axis to depict $-\Delta$Ct, ensuring that an increased y-value represents increased expression in both (B) and (C).

We first attempted to answer this question through a ChIPBase-driven analysis (Yang et al., 2013) of predicted and recorded transcription factor binding sites. We found that the protein product of *Wt1*, an imprinted gene, might bind upstream of *Carmil1*— suggesting that the POE on *Carmil1* might be mediated by *Wt1*. However, we deemed this hypothesis unlikely given that, in our data, *Wt1* expression levels were unaffected by POE (p=.267).

This focused bioinformatic analysis having failed to clearly identify a mediator, we applied a genome-wide analysis: for every microarray-measured gene, we tested whether its expression mediated the POE on *Carmil1* expression. The model used to test for mediation is shown in Figure 2.9.

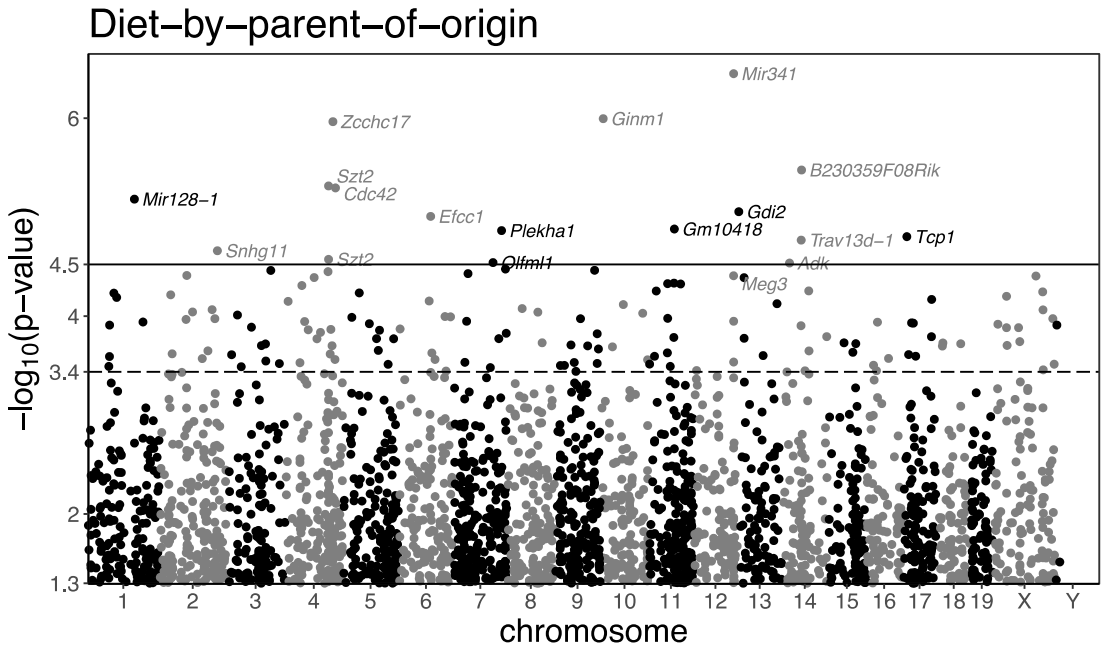


Figure 2.8: Manhattan-like plot of P-values of diet-by-POE effects on gene expression. Plotting format is similar to that used in Figure 2.7A. The dashed line represents the FDR threshold, and the solid line represents the FWER threshold. *Mir341* expression is the most significantly affected by diet-by-POE. Note that *Meg3*, an imprinted gene just below the FWER threshhold, is also labelled.

The expression of 8 different genes was found to significantly (Combined Tail Probability, CTP$<$.05) mediate POE on *Carmil1* expression. For 7 of these 8 genes, their mediation (*i.e.*, indirect) effect acted against the direct effect (Figure 2.9); rather than explaining POE, expression of these 7 genes actually suppressed the overall POE on *Carmil1*. *3830406C13Rik*, a non-imprinted protein coding gene of unknown function (Yue et al., 2014), was the most significant (CTP=.00289) overall mediator of POE on *Carmil1*. *Airn* was the most significant (CTP=.0134) mediator that was imprinted; specifically, *Airn* acted to suppress POE on *Carmil1* (Table A.6; Figure 2.10).

2.5.4.2 POE on behavior may be mediated by *Carmil1* and *Airn*.

We repeated a similar genome-wide POE-mediation analysis for every behavioral outcome (including behaviors without significant POE in mediation-free analysis). A significant (CTP$<$.05) gene mediator of POE was observed for 10 of the 34 modeled behavioral outcomes. POE on some outcomes was mediated by more than one gene, and some genes mediated POE on more than one outcome. Although 16 different significant mediator-outcome pairs were observed, there were only 6

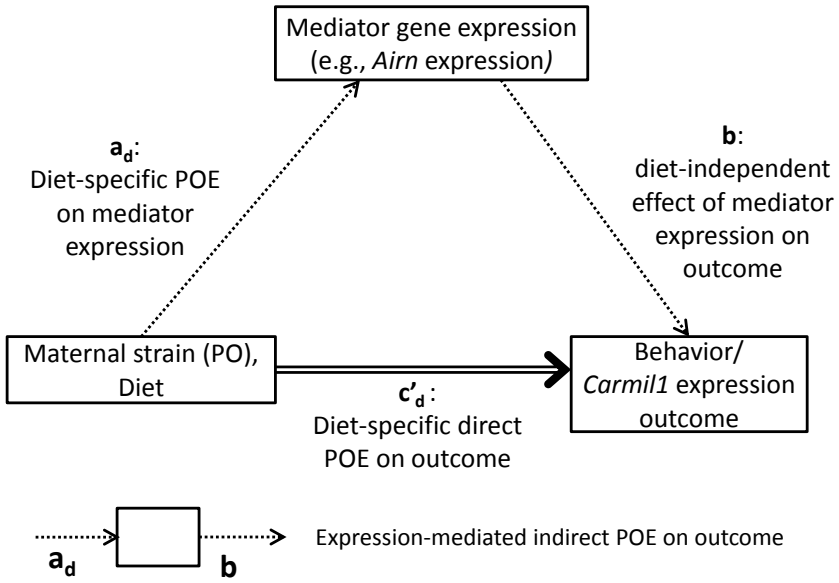


Figure 2.9: Model of gene-expression mediation of POE on the outcome, which is either behavior or *Carmil1* expression. Parent-of-origin, encoded as the maternal strain, in conjunction with diet, acts both directly upon the outcome, with effect size c'_d, and indirectly upon the outcome, with effect size $a_d b$. This indirect effect is composed of the diet specific POE on some mediator's expression (a_d) and the diet-independent effect of the mediator's expression on the outcome (b). Not shown in this figure for clarity, but present in the actual model, are nuisance effects of dam, pipeline, batch, and behavior specific covariates, that all can affect both mediator expression and the outcome. Mediation is determined by testing whether the average indirect effect ($ab = \overline{a_d} b$) is significant according to its Combined Tail Probability (CTP).

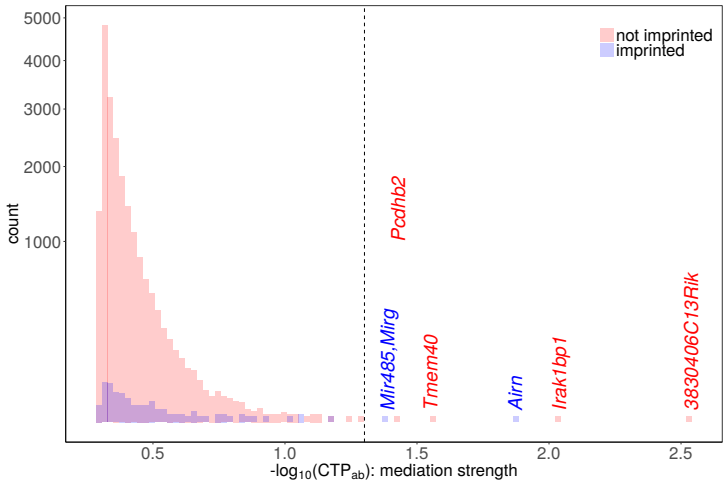


Figure 2.10: Histograms of the $-\log_{10}$ Combined Tail Probabilities (CTPs) for candidate gene mediators of POE on *Carmil1* expression. The red and blue histograms correspond to CTPs for non-imprinted and imprinted candidate mediators, respectively. Mediators whose mediation effect has a CTP$<$.05 (the dashed line threshold) are labeled. Notably, the imprinted gene *Airn* is one of the top 3 mediators of POE on *Carmil1*.

distinct genes/gene families significantly mediating any behavioral outcome: *Snord 113*, *Snord 115*, *Snord 116*, *3830406C13Rik*, *Rian*, *Carmil1*. In 15 of the 16 significant mediator-outcome pairings, the gene expression mediator suppressed POE; *i.e.*, 15 of 16 gene mediators acted in the opposite direction of the direct POE on behavior (Table A.7).

To determine each genes mediation of POE on behavior in the aggregate, we combined the CTPs for a given gene, over all behaviors, into a single metric: the "Combined Tail Probability" (CTP; In depth: subsection 2.4.10). By this metric, 21 probesets, corresponding to 17 distinct genes/gene families mediated POE on behavior in the aggregate at CTP$<$.05. Even though *Airn* was not a significant mediator for any individual behavior (see above), it was the most significant mediator in the aggregate (CTP=5.09e-05). *Airn* was followed closely by (a subregion of) *Snord 115* (CTP=.000408) and *Carmil1* (CTP=.000518). See Table A.8.

To gain further insight into aggregate mediation, for each outcome we determined the 3 most significant POE gene-mediators. Each gene was then scored according to the number of behaviors for which it was one of the 3 top mediators. According to this metric, *Airn* was the most notable mediator, acting as one of the 3 most-significant POE-mediators for 12 behavioral outcomes, while *Carmil1* was a top-3 mediator for 8 outcomes. The enrichment for *Airn*, *Carmil1*, and *Snord 115* in the sets of top-3 mediators is also readily apparent in Figure 2.11: for each behavior, genes with a significant CTP are labelled, as are *Carmil1* and *Airn* if they were among the top 3 mediators; mediation CTPs for *Airn* and *Carmil1* are often extreme.

2.6 Discussion

Our study identifies POEs on behavior, POEs on gene expression, and shows that many of these — with notable exceptions — are robust to differences in perinatal diet. We also provide evidence for a possible explanatory pathway connecting imprinting to gene expression to behavior, and are the first study to do this.

But beyond its specific results, our study also serves to advance a general protocol based on reciprocal F1s for studying POE and perinatal environment effects on a complex trait. The RF1 holds genetic background constant while varing parent of origin, making it the most powerful design for detecting POE. To investigate the interaction of developmental-environment with POE, we further

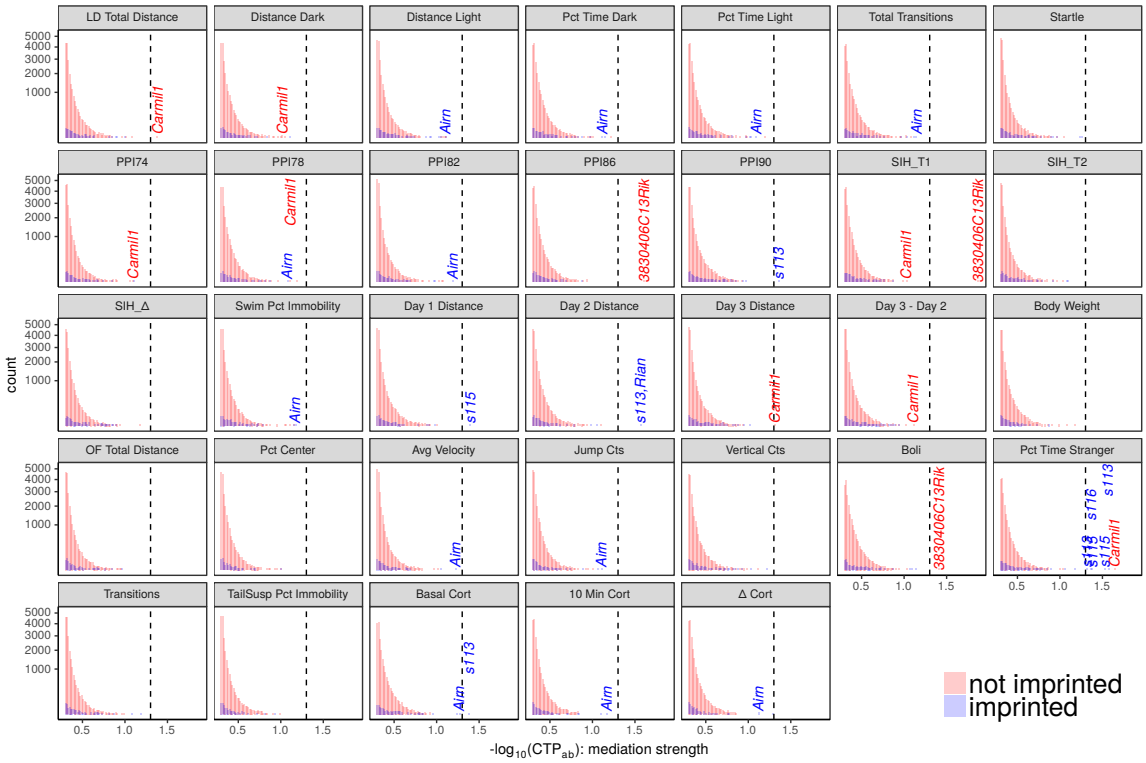


Figure 2.11: Histograms of the $-\log_{10}$ Combined Tail Probabilities (CTPs; akin to p-values) for candidate gene mediators of POE on various behaviors. Each panel corresponds to a behavioral outcome that may be mediated by gene expression. The red and blue histograms correspond to CTPs for non-imprinted and imprinted candidate mediators, respectively. Mediators having a CTP$<$.05 (threshold denoted by the dashed line) are labelled, as are *Airn* and *Carmil1* when they are one of the top-3 mediators. These two genes show up repeatedly (along with *Snord 113/115*) as one of the top mediators per behavior, especially when mediator/behavior pairs with a non-significant mediation p-value are also considered.

varied *in utero* nutrition, using four different diets: diet was a relatively easy variable to control, and ample evidence suggested its importance in POE. By repeating the behavioral and expression assays under multiple dietary conditions, we: 1) enabled detection of environment-by-POE, 2) hedged our bets, as an effect that would be unobservable in one environment might be amplified in another, and 3) enabled detection of POE that generalizes across environments.

In the remainder, we discuss the range of mechanisms that might explain POE as discoverable by our approach; our specific results on POE, diet, and diet-by-POE; and lastly, we reflect on the use of replicable vs non-replicable populations for POE discovery and investigation.

2.6.1 Coding-POE vs eQTL-POE, and POE observability

The two examined groups of female RF1s, NODxB6 and B6xNOD, were (aside from mitochondria) genetically identical. Consequently, differences in phenotype between these two groups could with high probability be attributed to POE. But for such observable POE to exist, imprinting/maternal factors must have interacted with a locus differing in sequence between parents (Figure 2.12A). This difference driving the POE could have been in a coding region, making it a "coding-POE", and/or in a regulatory region, making it an "eQTL-POE".

In coding-POE, the expressed allele's coding sequence differs between the two cross directions. Consequently, the RF1 populations are equal in total expression, but allele-specific expression (ASE) differs (Figure 2.12B). Although the microarrays in our study cannot quantify ASE, ASE differences can still manifest as an observable POE on an emergent phenotype such as behavior, or as POE on total expression of a downstream gene.

By contrast, eQTL-POE could arise by way of non-coding cis-eQTLs that alter total expression of an imprinted gene. For example, an eQTL-POE could arise from differences in promoter attractiveness, (Figure 2.12C). Or, perhaps more interestingly, eQTL-POE could arise by way of genetic background-dependent loss of imprinting (Vrana, 2007; Duselis et al., 2005; Wolf et al., 2014) (Figure 2.12D). In our study, all directly-observed POE on expression are necessarily instances of eQTL-POE, because we did not employ assays capable of measuring ASE.

eQTL-POE and coding-POE both require a genetic difference between parents in some imprinted or maternally-affected gene. However, any gene can exhibit POE— provided it is regulated by

the imprinted/maternal-effect gene. This trans effect can occur either by way of coding-POE (Figure 2.12E) or eQTL-POE (Figure 2.12F).

Both types of POE may be undetected by our study. As mentioned above, coding-POE is unobservable by our microarrays. Additionally, by measuring expression once, $\sim$8 weeks after birth, we may have failed to observe POE during transient, developmental-stage-specific imprintin. And by measuring whole-brain gene expression, we may have occluded POE arising from imprinting that is specific to subregions of the brain (Koerner et al., 2009; Prickett and Oakey, 2012).

2.6.2 POE on expression

All 9 imprinted genes that were subject to POE contain non-coding variants that differ between NOD and B6, a finding consistent with cis-driven, eQTL-POE (Figure 2.12C,D). However, six of the genes subject to POE were non-imprinted, including *Carmil1*. POE on such genes may be driven by maternal effects, or perhaps by trans-acting imprinted regulators (as in Figure 2.12F).

2.6.3 Mediation of POE on *Carmil1*

To determine potential imprinted regulators of *Carmil1* expression, we applied mediation analysis, identifying *Airn*. Unexpectedly however, *Airn* exerted its mediation effect in the opposite direction of the overall POE on *Carmil1* (ab and c' have opposite signs in Table A.6), suggesting that *Airn* suppresses POE on *Carmil1* in trans. All but one of the other significant mediators also acted as POE suppressors. The lack of explanatory mediation in the same direction as the overall POE may be due to the many unobservable forms of POE on expression: genes that fail to exhibit POE in their own expression cannot be statistically significant mediators of POE on another genes expression. Alternatively, *Airn* and the other imprinted mediator may be suppressing unobserved maternal effects on *Carmil1*.

2.6.4 POE on behavior and its mediation by gene expression

Five behaviors were significantly affected by POE, four of which were locomotor behaviors. The enrichment for POE could in part be due to increased power: locomotor activity has been found to be among the most stable of behaviors across laboratories and time (Crabbe et al., 1999; Wahlsten et al., 2006), resulting in more power to observe group differences. Also, however, given that locomotor

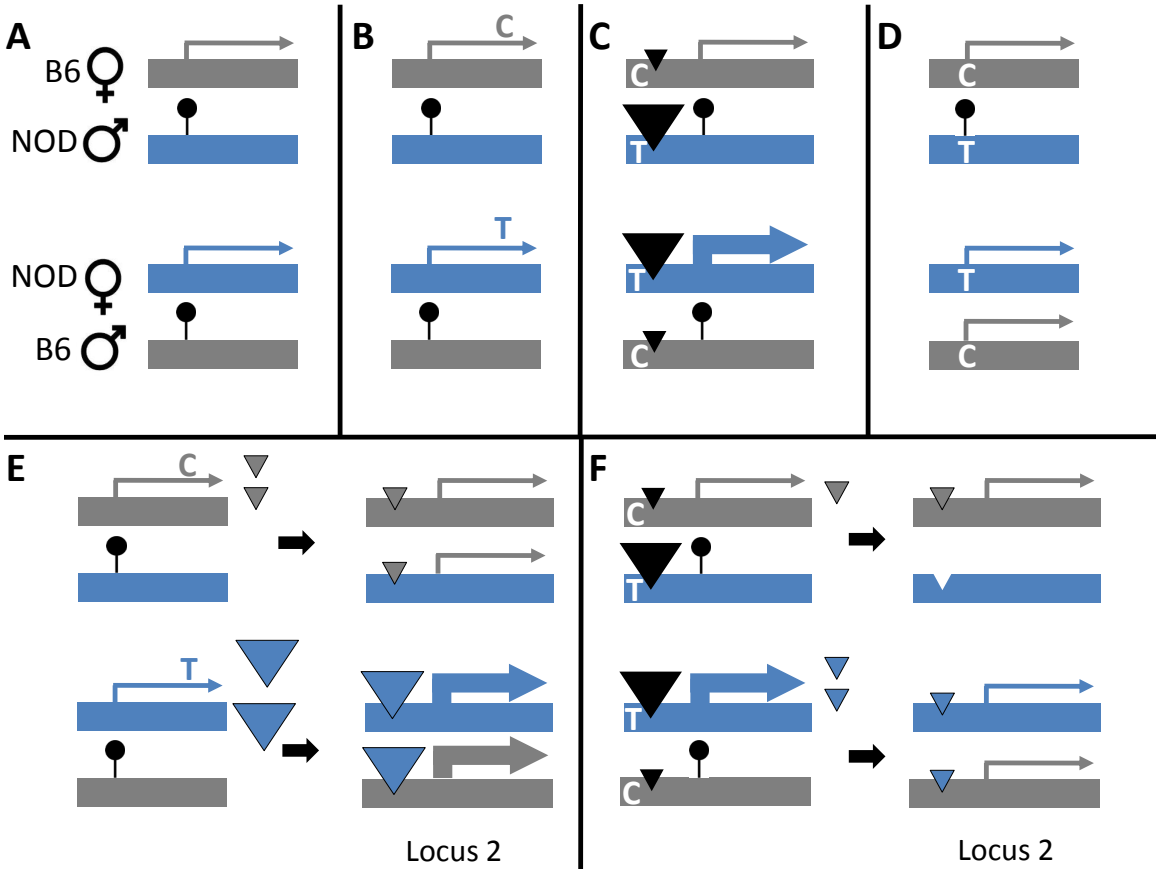


Figure 2.12: Examples of cis/trans coding-POE and eQTL-POE, in RF1s. Examples depict an imprinted gene which is fully active when maternally inherited, but fully silenced when paternally inherited. Similar examples could be constructed for maternal effects. (A) A lack of observable POE in spite of imprinting: B6 and NOD are identical in sequence, so whichever allele is silenced, the resulting expression product is the same in both RF1 directions. (B) Coding-POE: B6 and NOD differ in coding sequence causing allele-specific expression differences between RF1 directions (unobservable by microarray). (C) eQTL POE: the NOD promoter attracts a more effective transcription factor (TF), so NODxB6, in which the NOD allele is expressed, yields more expression. (D) eQTL-POE driven by background-dependent imprinting: imprinting is lost in B6, so NODxB6, in which imprint-silencing affects neither allele, yields more expression. (E) Trans coding-POE upon locus 2: locus 2 is identical between NOD and B6, but it is regulated by an imprinted TF whose NOD version is more efficient; so in NODxB6, in which the NOD TF is expressed, locus 2 expression is increased. (F) Trans eQTL-POE upon locus 2: locus 2 is regulated by an imprinted TF whose NOD version has a stronger promoter; so in NODxB6, in which the NOD allele for the TF is expressed, increased availability of the TF increases microarray-observable locus 2 expression.

activity has been used to measure rodent emotionality (Hall, 1934) and predict addiction-related behavior (Piazza et al., 1989), our POE results on locomotor activity may suggest that POEs are in fact important determinants of emotionality and/or addiction.

For the 5 POE-significant behaviors (and the other non-POE-significant behaviors too), POE must have been driven by some gene subject to POE; to identify such genes, we applied mediation analysis, finding 17 genes that mediate behavioral POE. However, for 16 of the 17 genes, the estimated mediation effect was to suppress POE; i.e, these genes did not explain the overall POE on behavior. We posit that explanatory POE on gene expression may simply have been unobservable, for the reasons described earlier.

2.6.5 *Airn* and *Carmil1* as mediators of POE

The most commonly shared mediators of behavior were *Carmil1* and *Airn*, with *Airn* also being the top mediator of POE on *Carmil1*.

Airn's mediation of POE is likely trans-acting. *Airn* is an imprinted, paternally-expressed, long non-coding RNA (lncRNA), which to our knowledge has not been found to affect any complex trait directly. Rather, *Airn* is known to control imprinting of three nearby maternally-expressed genes: *Slc22a2*, *Slc22a3*, and *Igf2r* (Cleaton et al., 2014). But none of the three genes were at all significant mediators of POE on any outcome of interest in our dataset. So, akin to other lncRNAs and imprinted genes found to affect distal gene expression (Vance and Ponting, 2014; Gabory et al., 2009), we posit that *Airn* may be exerting POE on behavior by affecting distal genes, such as *Carmil1* or *Snord 115* (as in Figure 2.12E). Our study is underpowered to directly examine this two-step mediation hypothesis.

Carmil1 may provide a link between cytoskeleton dynamics and cell migration, and behavioral change. *Carmil1* has a known cellular role in: 1) interacting with Capping Protein, which regulates actin elongation; and 2) activating the small GTPase *Rac1*, an important regulator of cytoskeletal dynamics (Gonzalez-Billault et al., 2012). Such actin cytoskeleton dynamics, critical for cytokinesis and cell migration (Rottner et al., 2017), are important throughout the lifespan for neurodevelopment and neural plasticity (Menon and Gupton, 2016; Gordon-Weeks and Fournier, 2014). In *C. elegans*, neuronal cell and axon growth cone migration has been shown to be negatively regulated by *CRML-1*,

the homolog of *Carmil1* (Vanderzalm et al., 2009). Our study, in a mammal, is the first to find a direct association between variation in *Carmil1* expression and behavior.

2.6.6 Caveats to mediation analysis of POE on *Carmil1* and behavior

We note that our analysis was applied one candidate mediator at a time; thus, any significant mediators may simply be co-expressed with the true mediator gene(s). We also note that for both mediation analyses (*Carmil1*/behavior outcome) we assumed a direction of causality in which some imprinted gene mediates POE on the outcome; although this might seem intuitive, it cannot be verified, and the "outcome" might actually mediate the imprinted gene.

Our directionality assumption is particularly uncertain in the behavioral analysis: expression in the brain was, out of necessity, measured after behavior; consequently, stressful behavioral assays could have altered expression. In future studies, we intend to address this weakness by a matching-based imputation: behavior-unperturbed expression will be imputed in behaviorally-assayed mice using expression data from mice that were unexposed but are genetically identical and otherwise perfectly matched (cf. related matching-based designs in Crowley et al. 2014)

2.6.7 Diet effects

Our data revealed significant diet effects on gene expression, with significance primarily driven by extremely low/high expression under the ME diet. This may be unsurprising given the direct role of methyl donors on DNA methylation and, consequently, on the regulation of gene expression. Future perinatal-diet studies may benefit from a ladder of methyl enrichment values.

Notably, diet significantly altered the expression of only 37 genes according to the strict FWER threshold, but 958 according to the FDR threshold. Some of these additional hits are likely false-positives (e.g. Y chromosome genes). But it is conceivable that diet did in fact cause a systemic, diet-buffering change to the overall network of gene expression levels (MacNeil and Walhout, 2011). Indeed, our FDR numbers seem consistent with earlier work: in the few examples (to our knowledge) of FDR corrected results from previous rodent studies of perinatal diet effects on gene expression (Mortensen et al., 2009; Altobelli et al., 2013; Barnett et al., 2015), 500-1000 genes were differentially expressed.

Although diet significantly altered the expression of numerous genes, the only complex phenotypes affected were body weight and PPI, and those effects were barely significant. The lack of significant diet effects on behavior, even in the presence of expression changes, is surprising but not entirely unexpected. Among other possibilities, the diet effects on behavior may be too small to overcome a sample size that was split among four different diets. Moreover, we measured a limited set of behaviors that may not have been altered by diet-driven gene expression changes.

2.6.8 Diet-by-parent-of-origin effects

Although our study perturbed nutrients involved in imprinting, the only imprinted gene subject to diet-by-POE and passing FWER was *Mir341*. The next most significant imprinted gene, which passed FDR but not FWER, was *Meg3*. However, these results are both uncertain: the genes predicted to be regulated by *Mir341* do not seem to manifest diet-by-POE effects in our data; and *Meg3*'s diet-by-POE was observed in microarray data but failed to replicate in qPCR data (Figure A.4).

In addition to inevitable lower power for testing interaction effects, the relative lack of observed diet-by-POE on imprinted genes may also be in part due to the aforementioned transience and/or tissue specificity of some imprinted genes (Ivanova et al., 2012), or because our diets are insufficiently extreme to elicit diet-by-POE. Insufficiently extreme diets may also contribute to lack of diet-by-POE on most of our behaviors (save for percent center time).

As for the 16 non-imprinted genes subject to diet-by-POE (by FWER), these may be regulated by imprinted genes that are subject to the aforementioned unobservable diet-by-POE (Figure 2.12E,F). Or, perhaps more likely, the 16 imprinted genes are controlled by maternal effects.

2.6.9 Studying POE in replicable vs non-replicable (outbred) populations

A number of previous studies of POE on complex traits have used outbred populations, such as F2, backcross, or heterogeneous stocks (Lawson et al., 2013). The advantages of such outbred populations over the RF1 are that: 1) POE can be detected simultaneously with non-POE genetic effects; and 2) POE arising from imprinting vs maternal effects can be disambiguated— a significant difference between reciprocal heterozygote (at some locus) offspring from heterozygote (at that locus) mothers can be ascribed to imprinting rather than to a maternal effect (Hager et al., 2008).

However, outbred populations have disadvantages as well: due to the fact that every animal is genetically distinct in an outbred population, alternate parent-of-origin states can never be observed in the exact same genetic background; this confounding limits the power of outbred populations to estimate POE. By contrast, in the RF1, individuals of alternate parent-of-origin state can always be perfectly matched in the same genetic background (save for the mitochondrial genome), allowing unconfounded and unbiased estimates of the causal POE. Moreover, whereas the irreplicability of outbred animals makes it impossible to perfectly recreate genetic state for a validation study, *e.g.*, a future study evaluating the effect of a knockout of *Carmil1* on behavioral POE, this is readily available for the RF1.

Only a handful of other studies have used an RF1 strategy to study POE on complex mammalian traits such as behavior. But none of these studies (save for a recent one of our own in Schoenrock et al. (2017)) simultaneously varied environment. Nor have other RF1 studies simultaneously measured gene expression. We are the first to demonstrate that combining the RF1 design, environmental perturbation, and observation of gene expression, provides a powerful paradigm for studying environment-by-POE on a complex mammalian trait.

CHAPTER 3

Rexplorer: optimal reciprocal cross selection for mapping parent-of-origin effects

3.1 Introduction

Imprinted genes have been estimated to play a role in as many as 100 diseases (Ubeda and Wilkins, 2008), having been linked with maladies ranging from cancer, to metabolic syndromes, to psychiatric illness (Kalish et al., 2014). Psychiatric illness may be a particularly important manifestation of imprinted gene mutations, as numerous lines of evidence from mouse and human studies suggest imprinted genes affect behavior as well as brain development and function (see section 1.4). Nonetheless, despite such evidence for the importance of imprinted genes, their effect on most complex traits is not well characterized.

This lack of characterization is likely due to the difficulty of directly observing the "parent-of-origin effects" (POEs) that imprinted genes exert on complex traits. In more detail, imprinted genes are subject to an epigenetic process whereby either the maternally or paternally inherited allele (depending on the gene), is (at least partially) silenced (Crowley et al., 2015; Bartolomei and Ferguson-Smith, 2011). That is, for imprinted genes, each allele's expression depends on its parent-of-origin, and traits affected by imprinted alleles are in turn subject to parent-of-origin effects (Lawson et al., 2013). As a result, identifying imprinting-driven POE on complex traits requires that reciprocal heterozygotes for a given imprinted locus exist in the population under study; for example, assuming an "A" and "B" allele exist at some imprinted locus, "AB" organisms (maternal A) need to be compared with "BA" organisms.

The requisite reciprocal heterozygotes can be generated in the outbred populations typically used for studying POE. But these populations are not ideal, in part because POE in can be confounded with genetic differences at every other locus. An alternate, but relatively unused population, is one consisting of reciprocal F1 hybrids (RF1s), each generated by a reciprocal cross (RX); in a RX of inbred strains S1 and S2, any resulting female S1xS2 and S2xS1 RF1s are (almost) genetically

identical, differing only in allelic parent-of-origin. Consequently, by comparing the S1xS2 and S2xS1 subpopulations, POE can be detected without confounding, and with maximal power.

3.1.1 Rationale for the development of Rexplorer

Motivated by the existing evidence for POE on behavior, as well as by the power of RXs, we engaged in a POE pilot study employing RXs of C57BL/6J (B6) with NOD/ShiLtJ (NOD) mice. Behavior and gene expression data from the B6xNOD and NODxB6 RF1s suggested the presence of POEs (Oreper et al., 2018). These results then motivated a follow-up study, which we refer to here as the CC-POE study. The CC-POE's design largely mirrored that of the pilot, but rather than reciprocally crossing mice from a single pair of strains, we intended to reciprocally cross *eight* pairs of parental lines (Schoenrock et al., 2017) and then measure behavior and expression in the eight resulting sets of RF1s (Figure 3.1).

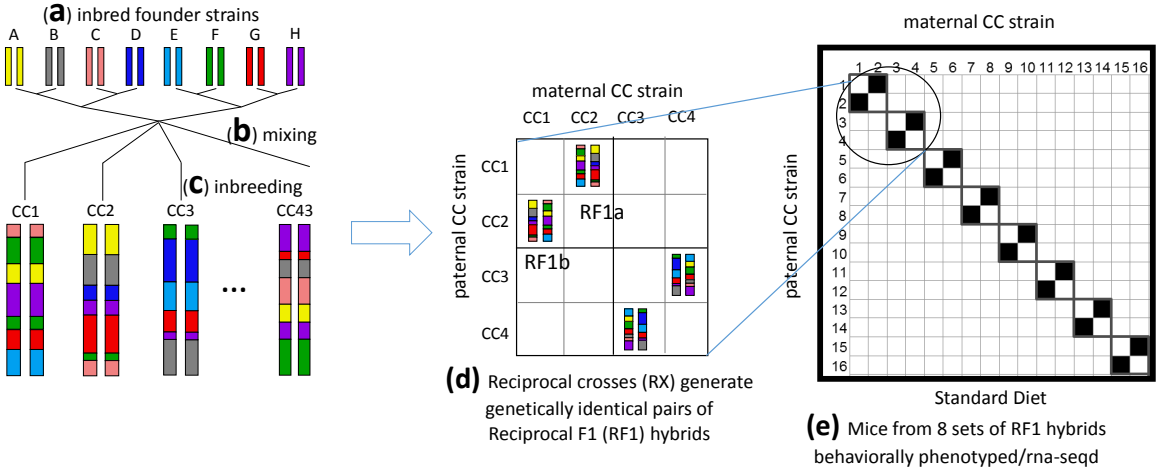


Figure 3.1: A Collaborative Cross (CC) experiment to map POE using reciprocal crosses (RXs). The CC is an existing population; it is not being developed in this project: a) The 8 inbred founder lines of the CC (including NOD and B6) b) were outcrossed in different founder orderings for 3 generations and c) then inbred for multiple generations in a funnel breeding scheme resulting in the CC, a panel of recombinant inbred lines. d) 8 pairs of CC lines were chosen from among 43 available lines, and reciprocally crossed (RXd), generating 8 genetically identical pairs of reciprocal F1 hybrids (RF1s), that differ in allelic parent-of-origin. e) These RF1 mice were either behaviorally tested or whole-brain RNA-seqd. For completeness—although this aspect is not relevant to Rexplorer—we note that the whole experiment was repeated under 4 different perinatal diets.

In part, the intent of generating 8 additional RF1 populations was to coarsely map POEs on behavior back onto imprinted loci: if well-chosen, each additional RX would be able to progressively

narrow the search space of imprinted loci potentially causal to POE. Toward this end, parental lines for the CC-POE RXs were drawn from the Collaborative Cross (CC) reference population, an existing panel of recombinant inbred lines of mice whose genomes are independent, high in genetic diversity (Shorter et al., 2017; Oreper et al., 2017c), and, critically, are derived in part from NOD and B6 (along with 6 other founder lines). But the more difficult decision was: which CC lines in particular should be crossed?

At the time of the experiment, 43 CC lines were available, allowing us to potentially generate over 900 genetically distinct RF1 populations. Generating every such RF1 was clearly impractical given breeding, housing, and other costs. On the other hand, a set of 8 RXs carelessly chosen could have resulted in RF1s that lacked mapping resolution, or that even failed to detect POE at all. To avoid these small-population pitfalls, we selected the RXs using the Reciprocal Cross Explorer (Rexplorer), a method we have developed to select RXs from a panel of candidate inbred strain parents, for the purpose of studying POE.

3.1.2 Relationship of Rexplorer to existing methods

The experimental design problem that Rexplorer seeks to address—of selecting RXs from a panel of inbred lines—could be considered as partially related to that of selecting animals for breeding: both problems require computationally exploring a large space of potential crosses, and then choosing those that optimize properties of the resulting population. For example, breeding selection methods used in livestock and for agriculture involving genomic estimation of breeding value (Meuwissen et al., 2001) employ genetic information towards optimizing some phenotype (e.g., milk yield). Other breeding algorithms may seek to also optimize genetic properties of the bred population such as genetic diversity, and/or to limit inbreeding (Kemper et al., 2012). And some breeding methods go so far as to optimize a *mate* selection index (Kinghorn, 2000, 2011), which measures the goodness of a set of *particular pairings* of animals.

But although breeding selection bears similarities to the Rexplorer problem, Rexplorer does not seeks to optimize for a particular phenotype; indeed, Rexplorer selects crosses without any knowledge of the parental phenotypes, relying solely on genetic information. Furthermore, in the Rexplorer context of crossing *inbred* lines, and unlike in breeding selection, the genetic state of the potential RF1 offspring from candidate breeding pairs is known with near-total certainty; it

is as though these offspring already exist. As such, Rexplorer attempts to solve a problem that is in fact quite closely related to that of selective phenotyping, a process (first proposed in Darvasi 1998), in which a subset of genotyped organisms are prioritized for costly phenotyping, in order to most efficiently map quantitative trait loci. Rexplorer performs a similar function—using *imputed* offspring genetic state—to prioritize a subset of organisms for costly *generation*, which are to be used for mapping POE.

Previous approaches to selective phenotyping include those described in: Jin et al. (2004) and Huang et al. (2013a), in which greedy and clustering algorithms were employed to maximize the phenotyped organisms' genetic diversity; Jannink (2005), in which a greedy algorithm maximized the number of mapping intervals for which at least one phenotyped animal's genome contained a recombination event, to maximize mapping resolution; and in Vision et al. (2000) and Xu et al. (2005), in which greedy and optimal algorithms were employed to minimize the sum of squares of "bin lengths", defined as the lengths of contiguous intervals in which no recombination occurs within any of the phenotyped population—by minimizing bin lengths, mapping resolution is optimized.

Although selective phenotyping approaches are motivating, Rexplorer has been designed specifically to optimize the experimental design of RF1 populations for studying POE. We present simulations comparing its performance to naive selection, and also employ Rexplorer in the CC-POE study. Our efforts provide a method and demonstration of experimental design for studying POE in any model organism.

3.2 Methods

Given a candidate set of RXs, Rexplorer seeks to select a subset of RXs that is optimal for studying POE arising from imprinted genes. Towards this end, Rexplorer chooses RXs that maximize a metric on the imprinted loci. Per locus contributions to the metric are weighted according to the types of heterozygosity in the RF1 populations. Maximization of the metric is formulated as equivalent to a generalized maximum coverage problem, which is solved using integer programming.

3.2.1 Optimized metric

Rexplorer optimizes a metric which is intended to reward sets of crosses that, regardless of the mapping algorithm, should generate a population conducive to mapping POE, and in a manner which hedges over all candidate loci as being potentially causal to POE. Towards this end, the Rexplorer metric rewards crosses with high potential "exploration" and "discrimination", concepts that we describe below (illustrated in Fig 3.2).

exploration A POE driven by some locus can only be detected by observing individuals that are reciprocally heterozygous at that locus. Accordingly, a RX whose RF1 offspring are heterozygous at a locus can "explore" that locus for POE. The more loci a RX can explore, the greater the chance it will detect POE.

discrimination If a locus causal to POE exists, a RX that generates RF1s whose genomes are mostly heterozygous is likely to detect the POE. However, such a RX will fail to provide mapping resolution: a detected POE could have come from any heterozygous locus. POE detection can only enable an observer to discriminate between a pair of loci as potentially causal, if one locus is heterozygous and the other homozygous. Accordingly, we describe a RX as being able to "discriminate" between a pair of loci. The more locus pairs a RX discriminates between, the finer the RXs mapping resolution.

Combining these two concepts, the overall metric optimized by Rexplorer is a weighted sum of exploration and discrimination.

Exploration and discrimination were described in the context of a single RX, but both can be adapted to apply to a set of RXs as well: for a set of RXs, the explored loci are those which are explored by at least one RX in the set, whereas the discriminated pairs of loci are those discriminated by at least one RX.

3.2.1.1 Weights for exploration and discrimination and "best-possible weight" RX scoring

Not all types of exploration are necessarily equivalent. Suppose that at some locus, alleles A and B only differ by a single base pair, whereas alleles C and D differ by a multi-base insertion: a comparison of CD vs DC animals is more likely to reveal POE than one of AB vs BA animals.

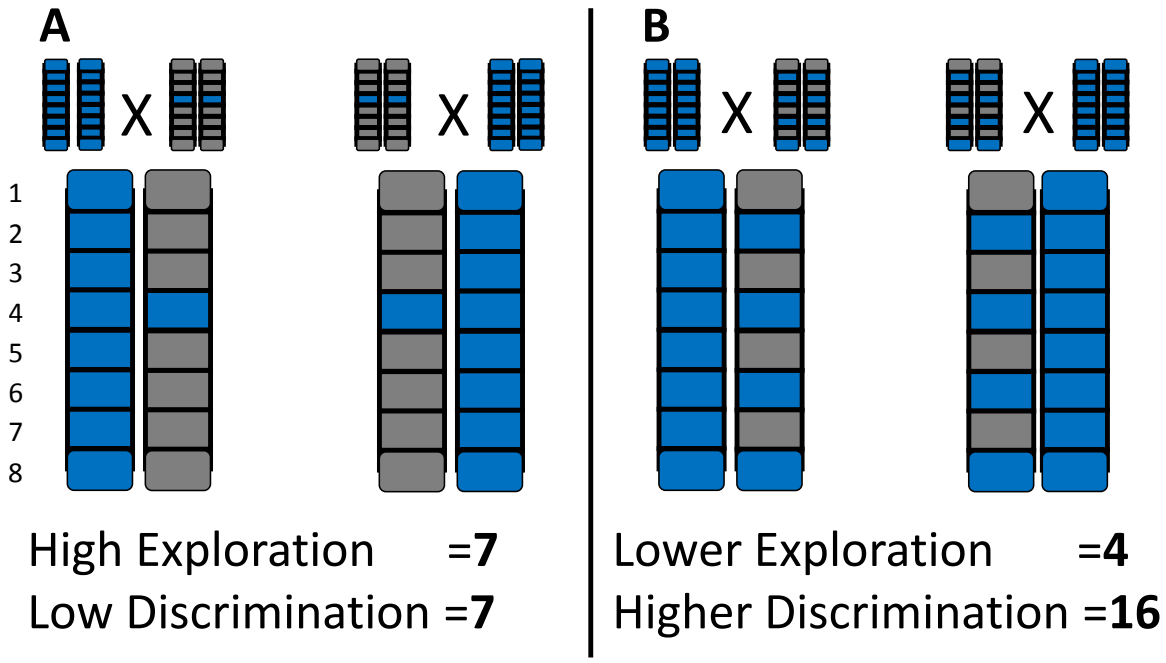


Figure 3.2: Exploration and discrimination metrics in the context of a single RX. A) A RX with high exploration. The resulting RF1s are heterozygous at every locus but #4, and so if a POE exists, this RX is highly likely to reveal that POE in a comparison between the phenotypes of the two RF1 directions. Exploration can be quantified by counting the number of heterozygous loci: in this case, 7. On the other hand, this RX has low discrimination; even if a POE were detected by this RX, it would be unclear which locus caused the POE. Discrimination can be quantified by counting all the pairs of loci in which one locus is homozygous, and the other is heterozygous. Discriminated pairs include (1,4), (2,4), (3,4), (4,5), (4,6), (4,7), and (4,8), adding up to a discrimination score of 7. B)) A different RX, with lower exploration—only 4 loci are reciprocally heterozygous—but higher discrimination—a total of 16 unique locus pairs include one homozygous and one heterozygous locus.

Accordingly, we can assign RX exploration by CD/DC heterozygosity a higher weight. Similarly, not all discrimination is equivalent: for a locus pair, if one locus is homozygous, but the other locus is a barely heterozygous AB/BA locus, the two loci are nearly indiscriminable—and so, RX discrimination in which one of the two loci is CD/DC should be assigned a higher weight.

In addition to RX-by-locus specific weighting, some loci may be believed a priori to be more likely than others to be causal to POE—independent of the RX. Exploration and discrimination weights can both be adjusted accordingly, on a per locus basis.

We decided to account for these weights in the Rexplorer metric as follows: a locus contributes to the Rexplorer score according to the best possible selected RX weight at that locus; if a locus can be explored by only one selected RX, that RX's locus-specific weight is accrued, but if exploration is possible by more than one selected RX, the maximum locus-specific weight over these RX is accrued instead. Discrimination weights for locus-pairs are accounted for similarly: the best possible weight at each locus-pair is used. An example of such "best-possible weight" scoring is illustrated in Figure 3.3.

	L1	L2	L3	L4	L5	L6
RX1	0	1	0	0	1	1
RX2	0	0	1	0	10	1
RX3	0	0	0	10	0	0
Score	0	1	1	10	10	1

Figure 3.3: Best possible weight scoring for exploration by 3 RXs. The first three rows of this table encode a RX, the columns encode loci, and each cell (of the first three rows) contains the weight associated with exploring a particular locus by a particular RX. For example, RX3 can explore locus 4 and accrue 10 points worth of exploration. The last row shows the score at each locus assuming all 3 RX are used. At each locus, the best possible weight is accrued; if two RXs can explore the same locus (i.e., locus 5), the largest weight is accrued. The total exploration score is 23.

Best-possible weight scoring leads naturally to a maximum coverage formulation, described further in subsection 3.2.2: "Weighted maximum coverage formulation".

3.2.2 Weighted maximum coverage formulation

To choose RXs that maximize both exploration and discrimination, the Rexplorer method formulates RX choice as a modified version of the **generalized maximum coverage problem (GMCP)** (Cohen and Katzir, 2008). To provide intuition, an example of the GMCP is as follows: suppose a business can profit from a set of tasks, each of which can be completed once. The business, subject to a fixed budget, can hire from a pool of candidates, each of whom can complete (i.e., cover) a different subset of the tasks, but with differing ability. And some tasks yield more profit than others. The business wants to hire the subset of candidates that would accrue maximal profit. By analogy, every possible RX is like a candidate that can be hired. Each RX has differing ability, depending on its heterozygosity, to explore some loci (i.e., complete certain tasks), but cannot explore some loci at all. Similarly, each RX can discriminate between some pairs of loci, but not others. Some loci are more important (i.e., weighted) than others. If more than one RX can explore a given locus/discriminate between loci, we assume results from the best RX will be used. On a budget allowing k RXs, we would like to maximize a weighted sum of discrimination and exploration (profit). Once RX selection is formulated as a GMCP, several existing integer programming solvers can be used to determine a solution (Hutter et al., 2010); Rexplorer uses Gurobi (Gurobi Optimization, Inc., 2015). Although GMCP is NP-hard in general (Cohen and Katzir, 2008), in the context of the CC-POE, the problem size allows for an exact solution.

The specific GMCP formulation Rexplorer uses is presented below, with variable definitions in Table 3.1.

The maximized weighted sum is

$$\max \quad \theta \sum_{i \in I} \alpha_{h,i} Y_{h,i} + (1-\theta) \sum_{(i,i') \in I^2} \beta_{h,i,i'} Z_{h,i,i'} \,, \tag{3.1}$$

and this maximization is performed subject to the constraints:

$$X_j \in \{0,1\} \qquad j \in J \tag{3.2}$$

$$Y_{h,i} \in \{0,1\} \qquad i \in I \tag{3.3}$$

$$Z_{h,i,i'} \in \{0,1\} \quad (i,i') \in I^2, h \in H_i \cup H_{i'} \tag{3.4}$$

$$\sum_{j \in J} X_j = k \qquad (k \text{ RXs must be selected from the candidate RXs}) \tag{3.5}$$

$$\sum_{j \in J_{h,i}} X_j \geq Y_{h,i} \qquad i \in I \qquad (\text{If a RX in } J_{h,i} \text{ is selected, it can explore locus } i, \text{ accruing value } \alpha_{h,i}) \tag{3.6}$$

$$\sum_{h \in H_i} Y_{h,i} \leq 1 \qquad i \in I \qquad (\text{Exploration is rewarded at most once per locus; redundant/weaker selected RXs accrue no value}) \tag{3.7}$$

$$\sum_{j \in J_{h,i,i'}} X_j \geq Z_{h,i,i'} \quad (i,i') \in I^2, h \in H_i \cup H_{i'} \qquad (\text{If a RX in } J_{h,i,i'} \text{ is selected, it can discriminate between locus } i \text{ and } i', \text{ accruing value } \beta_{h,i,i'}) \tag{3.8}$$

$$\sum_{h \in H_i \cup H_{i'}} Z_{h,i,i'} \leq 1 \quad (i,i') \in I^2 \qquad (\text{Discrim. is rewarded at most once per locus-pair; redundant/weaker selected RXs accrue no value}) \tag{3.9}$$

$$\sum_{j \in J_s} X_j \leq 1 \qquad s \in S \qquad (\text{No two selected RXs can share a parental strain}) \tag{3.10}$$

Where terms are defined as follows:

I	The set of loci for which we are choosing an optimal RX set.
J	The set of all candidate RXs, from which RXs are selected.
J_s	The set of all candidate RXs, in which strain s is a parent.
H_i	The set of all possible types of heterozygosities at locus i, across all candidate RXs
$J_{h,i}$	The set of all candidate RXs that are heterozygous of type h at locus i.
k	The number of RXs to select from the candidate set of RXs.
X_j	1 iff candidate RX j is selected, 0 otherwise
$Y_{h,i}$	1 iff the ith locus is assigned to be explored by a RX which at locus i has heterozygosity type h, 0 otherwise. Since at most 1 RX can be assigned to explore a locus, this is not the same as a RX being *capable* of exploring a locus.
$\alpha_{h,i}$	A weight reflecting the value of exploring the ith locus with heterozygosity type h, which also takes into account prior belief that locus i is causal to POE.
I^2	The set of all unique locus pairs—with (i, i') and (i', i) considered to be equivalent.
$J_{h,i,i'}$	The set of all candidate RXs that that are homozygous at locus i, but heterozygous with type h at locus i', or vice versa.
$Z_{h,i,i'}$	1 iff locus pair (i, i') is assigned to be discriminated by a RX which at locus i or i' has heterozygosity type h. 0 otherwise.
$\beta_{h,i,i'}$	A weight reflecting the value of discriminating between the ith and i'th loci using heterozygosity type h. Also accounts for prior belief that locus i or i' is causal to POE.
θ	Weight denoting the importance of exploration vs. discrimination.

Table 3.1: Terms in the maximum cover formulation

For emphasis, we reiterate that the maximized metric is computed solely using genetic information and no phenotype information.

3.2.2.1 Departures from the typical GMCP

The Rexplorer formulation departs from the typical GMCP formulation in three ways:

1. Rather than encoding the two types of elements to be covered simply as elements, we specify them separately: exploration elements, and discrimination pair elements. This admittedly complicates the formulation, but it does makes more explicit exactly how Rexplorer defines coverage of exploration vs. discrimination. The explicit discrimination terms are accounted for in the bottom half of Table 3.1.

2. Ordinarily, weights in the GMCP can be different for every candidate-set/element combination; however in the Rexplorer context, there are only a handful of different types of heterozygosities per locus, and we consider exploration by two different RXs with the same heterozygosity to be equivalent—so only as many weights as there are heterozygosities per locus are needed. Accordingly, we did *not* need a separate variable per RX-locus combination (i.e., $Y_{j,i}$), and

could instead account for maximum exploration coverage using $Y_{h,i}$ variables. Similar reasoning holds for discrimination coverage. By capitalizing on the small number of possible heterozygosity types relative to the number of RXs, we reduce the GMCP problem size by a factor of at least 20.

3. We include a RX-to-RX constraint encoding that no RX cannot be selected if one of its parental strains has already been used in another selected RX—see Equation 3.2.2. The intent is to preserve genetic independence between RXs.

3.2.3 CC-POE-specific inputs to Rexplorer

Although Rexplorer accepts a general set of inputs, in order to actually apply Rexplorer to CC-POE design, we needed to determine problem-specific inputs. In particular, we needed to determine:

1. An input set of candidate loci over which to optimize (I in subsection 3.2.2)

2. A specification of which RXs would be heterozygous (and with which alleles) at each locus (i.e., $J_{h,i}$ in subsection 3.2.2),

3. A set of heterozygosity-and-locus-specific exploration and discrimination weights (i.e., $\alpha_{h,i}$, $\beta_{h,i}$) in subsection 3.2.2).

Before describing the CC-POE-specific inputs we chose, we note that our subjective choices here (especially with respect to (i) and (iii)) were based in large part upon previous studies that had RXd CC *founder* strains and revealed POE; we reasoned that POEs revealed in RXs of CC founders were likely to also be revealed in RXs of CC lines. Specifically, we drew upon results from Crowley et al. (2015), in which RXs between NZO, PWK, and CAST revealed POE on whole-brain gene expression, and from Oreper et al. (2018), in which RXs of NOD and B6 revealed POE on whole-brain expression, as well as on behavior.

3.2.4 Candidate loci

For the choice of candidate loci, we elected to use a set of imprinted genes (Oreper et al., 2018) composed of the union of the genes identified as imprinted in Mousebook (Blake et al., 2010) and the genes identified as imprinted in Crowley et al. (2015).

3.2.5 Heterozygosity determination per RX based on identity-by-descent per locus

In the context of CC-POE design, we defined heterozygosity according to non-identity-by-descent: RF1s generated by a RX are defined as heterozygous at locus i if the two parental CC lines of that RX are, at that locus, non-identical-by-descent (non-IBD). We note that identity by descent was determined with respect to some ancestor of the CC founders; consequently, at some loci, even two different CC *founders* could themselves be IBD with one another. IBD determination per CC line and locus was extracted from the online resource developed in Wang et al. (2012), which in turn was based in large part upon IBD analysis of the CC founders in Yang et al. (2011).

3.2.6 Exploration and discrimination weights

In our exploration weight scheme (shown in Table 3.2), we upweighted imprinted genes found to be brain-expressed and/or subject to a POE in Crowley et al. (2015). We also upweighted heterozygosity types that mirrored the NODxB6 pilot RXs. Specifically, (on a per-locus basis) we assigned non-zero weight to 4 types of "heterozygosity-by-descent": i) "NOD-B6", in which one parental strain is IBD with NOD, the other parental strain is IBD with B6, and NOD is non-IBD with B6; ii) "NOD-*", where one parental strain is IBD with NOD, and the other strain is neither IBD with NOD nor IBD with B6; iii) "B6-*"—akin to (ii), but with one parental strain IBD with B6; and iv) "*-*", in which both parental strains are neither IBD to B6 nor to NOD, and the two parental strains are also non-IBD to each other.

For the choice of locus-pair discrimination weights (shown in Table 3.3), we used a simpler scheme which only took into account heterozygosity type. Locus pairs in which one locus was NOD-B6 heterozygous were given the largest weight. Pairs in which one locus was either NOD-*, or was B6-*, were given lower weight, and all other locus pairs were given 0 weight.

	Brain-expressed & strain-by-POE	Brain-expressed & no-strain-by-POE	No brain expression
NOD-B6	40	32	4
B6-*	36	28	0
NOD-*	36	28	0
-	12	8	0

Table 3.2: Exploration weights for the CC-POE study. The rows encode heterozygosity types, whereas the columns describe locus characteristics (as measured in (Crowley et al., 2015)) that are independent of heterozygosity. Each cell contains the corresponding weight for that type of locus and heterozygosity. For example, cell (B6-*, "Brain Expressed & no strain-by-POE") specifies that an exploration weight of $\alpha_{h,i} = 28$ should be applied to loci (i) that are brain-expressed but showed no POE, and that are explored by a RX whose heterozygosity type (h) is B6-* (where * denotes any founder that is non-IBD to B6 or to NOD).

NOD-B6	4
B6-*	2
NOD-*	2

Table 3.3: Discrimination weights for CC-POE study. Each contains the weight $\beta_{h,i,i'}$ for a pair of loci in which locus is homozygous and the other locus is some type of heterozygous. For example, cell NOD-B6 specifies that a discrimination weight of $\beta_{h,i,i'} = 4$ should be applied to those locus pairs (i, i') that are discriminated by a RX whose heterozygosity type (h) is NOD-B6 (at either i or i').

3.3 Results

To select CC lines to be RXd for the CC-POE, we applied Rexplorer. It chose the following 8 pairs of parental lines to be reciprocally crossed: (CC001/Unc, CC011/Unc), (CC041/TauUnc, CC051/TauUnc), (CC004/TauUnc, CC017/Unc), (CC023/GeniUnc, CC047/Unc), (CC028/GeniUnc, CC038/GeniUnc), (CC006/TauUnc, CC026/GeniUnc), (CC003/Unc, CC014/Unc), and (CC035/Unc, CC062/Unc).

3.3.1 Comparison of Rexplorer results to alternate methods

Having computed this "Rexplorer solution", in order to gain an understanding of its relative usefulness we compared its exploration and discrimination scores to those of the following alternate methods:

1. “Random”, in which RXs are selected at random, subject only to the constraints that: i) no selected RX could share a parental line with another, and ii) a total of 8 RX were to be selected. Random effectively models a naive experimental design.

2. A “Select-all” algorithm, in which every available RX is selected, ignoring any and all constraints. This effectively provides an upper bound on any algorithm’s performance.

Although exploration and discrimination are metrics that are internal to Rexplorer—and thus Rexplorer would be expected to succeed in optimizing them—an examination of these metrics across different approaches provides a starting point for comparison. This comparison is shown in Figure 3.4.

Rexplorer outperforms Random across both metrics, and is nearly as valuable as the best possible Select-all solution.

3.3.2 Breeding and analysis of data from the Rexplorer crosses

Having examined the relative value of the Rexplorer solution, we engaged in a breeding program to actually generate the RF1s for each Rexplorer-selected RX. At this point, it was discovered that the pairing of (CC028/GeniUnc, CC038/GeniUnc) could not produce offspring in one of its reciprocal directions—(CC038/GeniUnc x CC028/GeniUnc)—and so would not be useful in detecting POE. To compensate for the loss of this RX, and to hedge against future reproductive failures, 2 additional RXs were added to the CC-POE study, primarily based on their observed high fertility (rather than on their Rexplorer exploration/discrimination score). Both of the compensatory RXs successfully generated RF1 offspring.

An initial set of analyses of behavioral data from the 9 viable RXs revealed POE on multiple phenotypes, with 8 of 9 RXs revealing at least one POE (Schoenrock et al., 2017)— suggesting that the Rexplorer solution successfully prioritized locus exploration for POE. With regard to Rexplorer’s ability to ensure mapping resolution, we have not yet completed the eventual coarse-mapping of observed POEs in this dataset.

3.4 Discussion & future work

In this work, we developed Rexplorer, a method for selecting an ideal set of RXs for studying POE,and also demonstrated its viability and usefulness by applying Rexplorer to an ongoing ex-

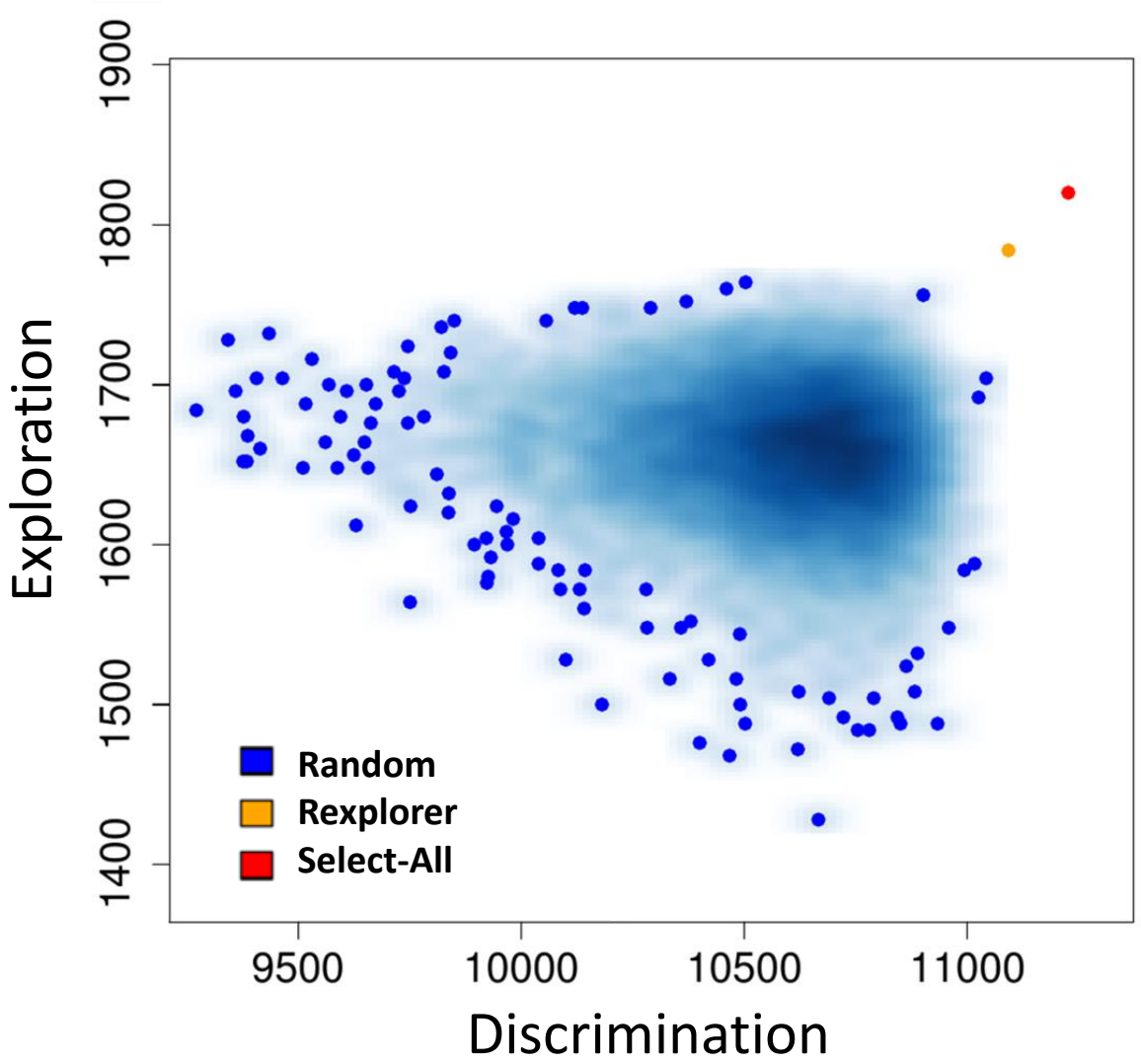


Figure 3.4: A comparison of the Rexplorer selected RXs for the CC-POE to other approaches' selected RXs. Each point corresponds to the exploration and discrimination of a set of RXs. Each point-color encodes the approach that generated that RX set, where approaches shown include: i) Select-all, an unlimited budget solution (red); random RX set selection under an 8 RX budget (blue); and iii) the Rexplorer solution under our budget (orange). A random solution was generated 1000 times, and most of these solutions are only represented as part of a cloud, rather than as individual points.

periment, the CC-POE. In a comparison to other approaches, the Rexplorer solution dominated all random solutions, and was nearly as good as a solution composed of every possible RX, under an infinite budget (a.k.a, select-all). And the Rexplorer selected crosses were generated, resulted in published work, and the detection of multiple POE.

Despite Rexplorer's success in the CC-POE, there are several potential areas for improvement. Perhaps most pressing is the need for Rexplorer to generate RX sets that account for the possibility of breeding failure: although in the CC-POE only a single RX failed, Dobzhansky-Muller incompatibility theory (Dobzhansky, 1936; Muller, 1942) suggests crosses of inbred lines in general will be prone to failure, and existing CC cross-breeding data (personal communications with Darla Miller) suggests cross-breeding of CC lines fails at a rate of 20%. At present, Rexplorer, being a maximum-cover approach, prioritizes RX sets that lack redundancy; this does not, however, provide any particular resiliency to breeding failures.

One way to account for potential breeding failure would be to reformulate the RX selection problem not as a maximum cover, but rather as a Maximum Expected Covering Location Problem (MEXCLP). Under this formulation, each RX would be assumed to have a constant probability p of failure, and RXs would be chosen to maximize *expected* exploration/discrimination (Daskin, 1983). Other more involved formulations, such as the Adjusted MEXCLP, could be drawn upon to allow for different probabilities of failure per RX (Batta et al., 1989).

A second potential area for improvement is in the assignment of exploration/discrimination weights. For applying Rexplorer to the CC-POE study, we used heterozygosity-by-descent per-locus to determine weights. However, our weighting scheme did not take into account the size, the number of know variants, or the variant types within each locus. One way to address this issue would be to incorporate CC variant info from the ISVdb (Oreper et al., 2017c) into the weighting scheme per locus. We do note, however, that even if variant information is included, future iterations of Rexplorer are likely to continue to optimize over loci—contiguous discrete regions—rather than over individual variants; preliminary results suggest optimization at the variant scale is unsolvable in a reasonable amount of time.

Last, an ideal implementation of Rexplorer would readily allow for more loci and candidate RXs, and still generate a solution. Although solution generation for the CC-POE did not tax Rexplorer's current integer programming implementation—and nor did a problem with ten times as many loci—

even larger instances of this NP-hard problem do become unsolvable. As such, it may prove to be necessary to instead use some form of randomized greedy algorithm, an approach found to perform well in a different selective phenotyping context (Vision et al., 2000).

CHAPTER 4

Inbred Strain Variant Database (ISVdb): A repository for probabilistically informed sequence differences among the Collaborative Cross strains and their founders [1]

4.1 Introduction

The Collaborative Cross (CC) is a large panel of recombinant inbred mouse strains derived from a genetically diverse set of eight inbred founder strains—A/J (AJ), C57BL/6J (B6), 129S1Sv/ImJ (129), NOD/ShiLtJ (NOD), NZO/HlLtJ (NZO), CAST/EiJ (CAST), PWK/PhJ (PWK), and WSB/EiJ (WSB)—which were outcrossed and then inbred in a multi-funnel breeding scheme. Within each funnel, mice were inbred until two or more animals were identified by MegaMUGA genotyping collectively as having over 90% of the genome fixed (i.e., homozygous and consistent for a founder haplotype). These animals, hereafter termed the most recent common ancestors (MRCAs), were then chosen to become the obligate ancestors of all subsequent generations and bred to produce a distinct CC strain.

As a result of their breeding scheme, the inbred CC strain genomes are random and independent mosaics of the eight founder haplotypes (Collaborative Cross Consortium (2012); Srivastava et al. (2017); more details in section 1.5, Figure 1.2, and in `http://csbio.unc.edu/CCstatus/index.py?run`).

These properties, along with the reproducibility of inbred strains have made the CC a unique resource in mammalian genetics. But to make optimal use of the CC strains, it is desirable to have an accurate catalog of the genetic differences between them—specifically, the positions and other characteristics of all known CC strain genetic variants—and to be able to predict which variants will be polymorphic in future hypothetical crosses of CC strains and of CC strains with other laboratory strains.

[1] This chapter has been adapted from a manuscript published in G3. The citation is as follows: Oreper, D., Cai, Y., Tarantino, L. M., Villena, F. P.-M. d., and Valdar, W. (2017). Inbred Strain Variant Database (ISVdb): A Repository for Probabilistically Informed Sequence Differences Among the Collaborative Cross Strains and Their Founders. G3: Genes, Genomes, Genetics, 7(6):16231630.

Such information could be determined directly by sequencing, but CC sequencing, soon to be released for a single male per strain Srivastava et al. 2017, is not yet easily accessible. Furthermore, sequencing from a single animal will not resolve uncertainty arising from residual heterozygosity, since two animals from the same strain could easily differ at residually heterozygous loci. More generally, whole-genome sequencing in most organisms is expensive and inconvenient.

A commonly used alternative is haplotype-based variant imputation, whereby comparatively sparse and cheap genotyping data is combined with more complete information about allelic state in (even extremely distant) relatives to infer allelic state at ungenotyped positions in the target sample. This typically involves inferring shared haplotype blocks and, by assuming that individuals sharing a haplotype block also share the corresponding allelic state, using this to impute genotype (Li et al., 2009; Marchini and Howie, 2010). A broad array of such imputation methods have been developed for use in humans (eg, Hawley and Kidd 1995; Scheet and Stephens 2006; Marchini et al. 2007; Browning and Browning 2009; Howie et al. 2009; Li et al. 2010) and livestock (eg, VanRaden et al. 2011; Hickey et al. 2012; Sargolzaei et al. 2014); these typically either start by inferring haplotypes from the genotype data or by approximating the pool of extant haplotypes via a large reference panel, and then use those haplotypes as a means to impute variant genotypes, which are assumed to be the primary interest.

In multiparental populations (MPP) of model organisms, where the founder haplotypes are usually known, it is more common for primary interest to focus on reconstructing the haplotype mosaic itself, eg, for the purpose of linkage disequilibrium mapping (Mott et al., 2000; Liu et al., 2010; King et al., 2012; Fu et al., 2012; Gatti et al., 2014; Verbyla et al., 2014; Zheng et al., 2015). In such cases, haplotype-based imputation of specific variants may proceed as a second, refinement step to inform fine-mapping and candidate prioritization (Yalcin et al., 2005; Tian et al., 2011). In any analyses using imputed variants, it important to note that the haplotype based variant imputation is inherently probabilistic. A failure to account for variant imputation uncertainty can negatively affect the robustness of downstream decisions (eg, overconfidence in a functional assignment), and also can produce misleading estimation of association significance and/or variant effects (eg, Marchini et al. 2007; Guan and Stephens 2008; Kutalik et al. 2011; Zheng et al. 2011; Zhang et al. 2014).

Haplotype-based variant imputation lends itself particularly well to MPPs because the haplotype blocks and the variants within are drawn from a known and relatively limited number of founders that

can be (more) affordably deeply sequenced and genotyped. This in turn reduces variant imputation uncertainty. For MPP RI strains, in particular, once an animal's variants are imputed, the need for even sparse genotyping is largely obviated in its inbred descendants—they are effectively genotyped as well. Haplotypes can, and have been, similarly imputed for the entire CC population based on the CC MRCAs. In particular, a Hidden Markov Model (HMM) based method (Fu et al., 2012) has previously been applied to MegaMUGA genotyping of the CC MRCA animals coupled with MegaMUGA genotyping from founder animals, (Welsh et al., 2012; Morgan et al., 2015; Srivastava et al., 2017) to impute a probabilistic estimate of each CC strain's haplotype mosaic. Sequencing of the founders by the Sanger Institute has provided a catalog of the sequence variants within the founder haplotypes (Keane et al., 2011) as well as their predicted functional consequences (McLaren et al., 2010), allowing for CC variant imputation.

However, although probabilistic imputed descriptions of CC haplotypes are already available, the final step of imputing probabilistic CC variant state using these haplotypes is currently left up to the researcher. This imputation step can be time-consuming, especially genome-wide, and it typically requires the researcher to develop their own ad-hoc, non-trivial scripts to parse and process input files. We seek to ease this burden by creating the Inbred Strain Variant Database (ISVdb). This database computes and stores imputed probabilistic CC variant information once, and then provides efficient, uniform and convenient access through a publicly accessible webtool.

4.2 ISVdb stored data and functionality

For all Sanger Institute sequencing variant positions that are exonic (or 100 bp upstream or downstream), polymorphic between CC founders, and correspond to SNP/indels, the ISVdb provides conveniently accessible information on the following:

- unphased genotypes for CC strains/founders, including the functional consequences per genotype, per transcript.
- unphased haplotype pairs (hereafter, "diplotypes") for CC strains, derived probabilistically from MegaMUGA genotyping;
- unphased genotypes for hypothetical F1 crosses amongst and between CC and founder strains;

- unphased diplotypes for hypothetical F1 crosses amongst and between CC and founder strains.

All information in the ISVdb is associated with a probability, to reflect the uncertainty of inference (discussed in the next section). The ISVdb interface is designed to be practical, oriented towards concrete tasks. For example, the ISVdb could be used to answer the following questions:

- Given microarray measurements of CC expression, which probes ought to be masked from analysis to minimize the effect of differential hybridization due to variants within the corresponding probed regions?
- Where should PCR primers be designed to bind so as to avoid differential hybridization, while still amplifying informative regions?
- What are the alleles per variant per CC strain in a given region, in order to perform association mapping?
- Given a pair of CC strains, or set of pairs of CC strains, where would the resulting F1 offspring be heterozygous? Which CC strains could be crossed with one another, or against a founder strain, to ensure that a certain region is heterozyogous in the offspring?
- Which CC strains contain a stop-gain codon in a particular gene?
- What is the ratio of missense to synonymous mutations on a particular chromosome?
- Which regions are fixed across all CC strains? Which regions are still segregating in a subset of CC strains?
- Which regions are most uncertain, either in haplotype or in genotype, across CC strains?

4.3 ISVdb preserves uncertainty

The primary purpose of the ISVdb is to provide genome-wide, inferred CC genotypes. But this inference depends on several processes and measurements that are themselves imprecise: 1) sequencing of founder strains was imperfect so some founder variant calls are ambiguous or incorrect; 2) uncertainty in the sparse, genotyping-based estimates of CC diplotypes; 3) the CC strains themselves are still segregating in some regions.

Properly representing and accounting for such sources of uncertainty is essential to avoid inaccuracies in downstream inference. Consider the effect of diplotype uncertainty on predicting the functional consequences of the alleles in an F1 hybrid from two CC mice.

Suppose one of the CC parents had a 40:30:30 probability of AJ/AJ vs 129/129 vs NOD/NOD diplotype, and where AJ carries a synonymous mutation, whereas 129 and NOD both carry a stop-gain mutation. Assuming the most likely diplotype, AJ/AJ, would imply a synonymous mutation in the F1, even though there is a greater (60%) probability of a stop-gain mutation.

The closest similar resource to the ISVdb, the CC "pseudogenomes" set, (Morgan and Welsh, 2015; Huang et al., 2013b) was designed primarily for sequence alignment: it employs most-likely point estimates of genotype and assumes all alleles are fixed. Therefore, a key secondary goal of the ISVdb is to provide a resource that retains all of the aforementioned uncertainty in CC and F1 genotype inference. The ISVdb achieves this by storing multiple records per variant, in which each record includes a probability of that variant state.

4.4 Materials and Methods

4.4.1 Inputs for Database Construction

All inputs are based on the GRCm38 mouse reference assembly. Probabilistic estimates of CC unphased diplotypes were computed as of 2016-03-24. These diplotype estimates were derived from a hidden Markov model (HMM) applied to MegaMUGA microarray measurements (Morgan and Welsh, 2015).

Founder variants were determined using the Sanger Institute, Mouse Genomes Project, mouse variant VCF files, REL-1410 (from 2014-10), corresponding to Ensembl release 75 of GRCm38 (Keane et al., 2011). These VCF files included SNPs and indels (1-100bp). Exon boundaries were drawn from Ensembl release 75 as well (Ramasamy et al., 2013).

4.4.2 CC genotype and diplotype inference: derivation

By tracing the potential transmission path of alleles from founders to CC strains, an expression can be derived for the probabilistic distribution of the unphased CC genotype, entirely in terms of known quantities. That is, we can derive an expression for the unphased CC genotype at each founder

variant position, in terms of known unphased founder genotype probabilities from sequencing, and known unphased MegaMUGA haplotyping at markers. Along the way, we can also derive the probability distributions of the CC diplotypes. The equations below are specific to a given variant; they need to be recalculated for every founder variant and CC strain combination.

First, note that the unphased CC genotype probability distribution can be defined in terms of that of the phased CC genotype:

$$p(UG_c = \{a, a'\}) = p(G_c = (a, a') + p(G_c = (a', a)) \,, \tag{4.1}$$

where UG_c is the unphased genotype of CC strain c, $\{a, a'\}$ is an unphased genotype with (possibly identical) alleles a and a', G_c is the phased genotype of CC strain c, and (a, a') is a phased genotype such that a is inherited maternally, and a' is inherited paternally.

The probability of the (typically unobservable) phased CC genotype can be expressed in terms of phased founder diplotype as

$$p(G_c{=}(a, a')){=}\sum_{(h,h')\in\mathcal{H}\times\mathcal{H}} p(G_c = (a, a') \mid D_c = (h, h')) \cdot p(D_c = (h, h')) \,, \tag{4.2}$$

where D_c is the phased diplotype of c at the variant, (h, h') describes the phased diplotype composed of (possibly identical) haplotypes h and h', which transmit alleles a and a' respectively, and $\mathcal{H}$ is the set of all haplotypes.

Assuming maternal and paternal alleles are transmitted independently, the first term in the product of the right hand side of (4.2) can be expressed as:

$$p(G_c = (a, a') \mid D_c = (h, h')) = p(T_h = a) \cdot p(T_{h'} = a') \,, \tag{4.3}$$

where T_h is the allele transmitted from parental haplotype h. T_h in turn depends on the *founder* h genotype, which is uncertain due to potential sequencing error, and on the number of a alleles in that founder genotype, which could be 0 (homozygous for the minor allele), 2 (homozygous for the major allele), or even 1 (heterozygous), as some loci are not fully inbred in the founders. Assuming

transmission of either copy is equally likely,

$$p(T_h = a) = \sum_{g \in \mathcal{G}} \left(\frac{1}{2} \text{number of } a \text{ alleles in } g \right) \cdot p(UG_h = g)\,, \tag{4.4}$$

where UG_h is the unphased genotype for the founder of haplotype h, $\mathcal{G}$ = the set of all possible genotypes, and g is a genotype.

The second term in the product within Figure (4.2), $p(D_c = (h, h'))$, also needs to be derived in terms of known values: we do not know the probability of a given diplotype at any arbitrary variant position. Rather, we *do* know diplotype probabilities of the markers to the left and right of each variant position. As such, we can linearly interpolate between the two markers:

$$p(D_c = (h, h')) = \frac{w_l \cdot p(D_{(c,l)} = (h, h')) + w_r \cdot p(D_{(c,r)} = (h, h'))}{w_l + w_r} \tag{4.5}$$

where w_l and w_r are the distances from the variant position to the left-nearest and right-nearest haplotyping markers, respectively. $D_{(c,l)}$ and $D_{(c,r)}$ are the phased diplotypes of the cc strain at the left-nearest and right-nearest haplotyping markers, respectively.

Eq. (4.5) is expressed in terms of phased diplotypes, but our observed MegaMUGA marker diplotype probabilities are unphased. If we assume that both phasings are equally likely; then for the left marker (similarly for the right),

$$p(D_{(c,l)}) = p(D_{(c,l)}) = \frac{1}{2} p(UD_{(c,l)}) \tag{4.6}$$

where UD is the unphased diplotype. The probabilities of each unphased diplotype are known from microarray assays of the CC strains.

This concludes the derivation: using (4.1)-(4.6), the unphased CC genotype distribution, $p(UG_c)$, can be fully expressed from known unphased founder genotype distribution, $p(UG_h)$ and known unphased CC marker diplotype distributions, $p(UD_{(c,m)})$.

4.4.3 Genotype and Diplotype inference for simulated F1 offspring

Once CC genotype/diplotype probabilities have been inferred, unphased genotype/diplotype probabilities for F1 offspring between CC strains and/or founder strains can be inferred as well. As in the previous derivation we can begin by expressing unphased genotype in terms of phased genotype:

$$p(UG_{12} = \{a, a'\}) = p(G_{12} = (a, a')) + p(G_{12} = (a', a)) \tag{4.7}$$

Where UG_{12} is the unphased genotype of the offspring of inbred strain 1, and inbred strain 2. G_{12} is the phased genotype.

Each phased genotype can then be rewritten in terms of the alleles transmitted from its parent strains. Assuming transmission of an allele from strain 1 is independent of transmission of an allele from strain 2,

$$p(G_{12} = (a, a')) = \frac{p(T_1 = a) \cdot p(T_2 = a') + p(T_1 = a') \cdot p(T_2 = a)}{2} \tag{4.8}$$

Combining (4.7), (4.8) and (4.4), the unphased offspring genotype distribution can be expressed in terms of known quantities. Diplotype probabilities for F1 offspring can be derived nearly identically.

4.4.4 Functional consequence inference

Functional consequences per variant, and per transcript, have been predicted in the *founder* strains by the variant effect predictor (McLaren et al., 2010). We assume that any CC strain inheriting a founder's haplotype, inherits the same transcripts and functional consequences as were in the founder in that haplotype region. This assumption that the genetic background will not affect functional consequence is not necessarily true: if a recombination event were to occur within a gene (joining the sequence of two different founders into the same gene in a CC strain), and upstream of some variant, that variant might no longer have the same effect. Nonetheless, our assumption is mostly reasonable: given the relatively small number of recombinations per CC line – on average 135 (Srivastava et al., 2017) – the number of mid-gene recombinations is necessarily small, and an even smaller number of these recombinations will actually have an effect on downstream variants. What might in fact be problematic to functional prediction within the CC, more so than within-gene

recombination, is the original process by which functional consequences were predicted in the founder strains: predictions were made a single variant at a time, without accounting for other, potentially compensatory variants within the same gene.

Of note, functional consequences in the ILVDB are represented as uncertain; i.e, the probability of a given CC genotype is also applied to the functional consequence of that genotype.

4.4.5 Database and GUI implementation

Scripts to parse VCF and haplotype files, perform genotype and diplotype inference, and store the resulting processed information in a MariaDB database, were implemented using a combination of Python, VCFtools (Danecek et al., 2011), and R (R Core Team, 2016). The ISVdb online GUI was implemented using the Python Flask library. The GUI was deployed online on the Carolina Cloud Apps managed platform, provided by UNC Information Technology Services.

Two sets of tables are stored within the MariaDB database: 1) An almost completely normalized set of tables with minimal redundancy, and 2) A set of pre-joined, non-normalized tables derived from the normalized table that are designed to allow the GUI (which provides for typical ISVdb use cases) to return results efficiently, using a minimum number of joins. This second set of tables was an intentional trade-off of space for time.

A few additional database optimizations of note were necessary, especially to rapidly generate probabilistic variant states for F1 crosses. In particular, the ISVdb.v1.1 "database" is actually implemented (using R code) as a collection of smaller databases, in which each smaller database represents a single chromosome. Where applicable, most tables were indexed by variant and strain. Tables sizes were reduced by dropping those variant diplotypes (and corresponding genotypes) whose probability was less than $.001$. Consequently, probability distributions at some variants may not sum exactly to 1.

4.4.6 Data Availability

The ISVdb version corresponding to this publication is ISVdb.v1.1, uploaded on March 15, 2017. A frozen snapshot of ISVdb.v1.1, including the v1.1 website interface and the variant information it provides, will be maintained at `http://isvdb.unc.edu/archive`. Subsequent ISVdb versions will be housed in the archive as well. The frozen ISVdb.v1.1 contents, and the inputs used

to generate those contents, are also permanently stored on Zenodo at https://doi.org/10.5281/zenodo.399474 (Oreper et al., 2017b). All ISVdb.v1.1 results can be recreated from File S8 and File S9, which are briefly described below, and in further detail in File S1. As subsequent versions of ISVdb are developed, their contents will also be archived (at another location) on Zenodo.

For ISVdb.v1.1: File S1 contains detailed descriptions of all supplemental files. File S2 contains marker diplotype data for CC strains. File S3 contains non-mitochondrial snps for founders. File S4 contains indels for founders. File S5 contains mitochondrial snps for founders. File S6 contains the custom format specification for these VCF files. File S7 contains BL37.75 exons (along with other genomic features)

File S8 contains an ISVdb.v1.1 dump of the imputed CC diplotypes per strain and per chromosome, in CSV format. File S9 contains an ISVdb.v1.1 dump of the imputed CC genotypes per strain and per chromosome, in CSV format.

Code used to generate the ISVdb and its GUI is available at https://github.com/danoreper/ISVdb.git (Oreper et al., 2017a).

4.5 Results and Discussion

The ISVdb houses and provides GUI access to imputed probabilistic genotype and diplotype data, for all 8 founders and all (as of 2016-03-24) 72 CC strains. CC allelic state is imputed at all exonic (+/- 100bp) founder SNPs and indels (which can be as long 100bp), but not at founder large structural variant positions (Morgan et al., 2017).

According to the ISVdb estimates of allelic state across strains, the genotype in most variants is known with high certainty (File S10, Figure S1). Variants with uncertain genotype appear widely and evenly distributed across the genome (File S10, Figures S2:S103). Residual heterozygosity, a key driver of uncertainty, is estimated to affect 3.1% of exonic (+/- 100bp) variants overall, but can vary dramatically between strains and chromosomes; e.g., the proportion of variants affected by residual heterozygosity ranges from 0 in CC003 on chr 2, to .38 in CC056 on chr 8 (Table S1). Note that heterozygous variants are defined as those with at least a 25% chance of continuing to segregate.

Approximately 72.6% of (the polymorphic) CC strain variants are identical to the B6 mouse reference genome (Table S1). Intronic (8.8%), downstream (4.3%), non-coding transcript (3.4%) and

upstream variants (2.9%) differing with respect to B6 are the next most common mutations, while alleles expected to have a large effect, such as stop-gain (.003%) or stop-loss (.001%) mutations, are extremely rare (Table S2).

4.5.1 Database accessibility/usability: ISVdb GUI

The intended interface to ISVdb data is through the publicly accessible ISVdb GUI, hosted at `https://isvdb.unc.edu`; the GUI allows what we believe to be the most common types of queries.

The ISVdb GUI can be broken up into roughly 3 panels:

- A **Primary query** panel that allows the user to: i) query the ISVdb for inbred strain genotypes, diplotypes, and F1 genotypes and diplotypes ii) specify strains of interest; and to iii) specify the genomic region(s) of interest—by basepair window, by genes, or by (internal ISVdb) variant IDs.

- A **Secondary restriction** panel that allows the user to limit results: i) to the max probability estimate of variant state; ii) and/or only to that variant state which is more likely than a user-specified probability threshold; iii) to variants of a particular zygosity; iv) for genotype and genotype cross queries, to variants having particular functional consequences.

- An **Output** panel that allows the user to: i) submit the primary query with secondary restrictions; ii) save the full results by opening a download URL in a browser, and; iii) examine the (first 1000) results in a sortable and searchable table displayed online.

Additionally, the GUI provides: i) a link to complete archived versions of the ISVdb, and ii) a link to CSV dump files of genotype and diplotype, per strain, per chromosome.

4.5.2 GUI-based genotype query

The ISVdb is most typically used to determine the genotype of a set of CC strains in some region. When the ISVdb GUI is queried for genotype, it produces a table with the following columns:

- variant_id: an internal ISVdb variant ID per variant

- chrom: chromosome of variant
- pos: variant start position, in mm10 coordinates
- strain: the inbred strain– either a CC or a founder strain.

The unphased genotype (allele_1 and allele_2 are arbitrary):

- allele_1: the sequence of one allele at the variant.
- allele_2: the sequence of the other allele at the variant.
- prob: the probability that at this variant, this is the actual genotype. Note that at any given variant position there is a distribution of possible genotypes; as such there will be multiple rows representing each variant position, each with its own probability.
- is_max: Whether this is the maximum likelihood genotype at this variant.
- gene_id: The Ensembl ID of a gene enclosing the variant. There may be multiple overlapping genes enclosing a variant, resulting in a separate row per gene for the same variant.
- transcript_id: The Ensembl ID of a transcript enclosing a variant. A single variant will usually be enclosed by multiple transcripts, each of which is affected differently by the variant; i.e., there will different consequences per transcript, at the same variant, necessitating a separate row per transcript for the same variant.
- consequence_1: the predicted consequence of allele_1 on the transcript, with respect to B6. If allele_1 is the B6 allele, the consequence is ”reference”. *Note that a consequence is always with respect to some transcript.*
- consequence_2: the predicted consequence of allele_2 on the transcript.

4.5.3 Example workflow for a genotype query

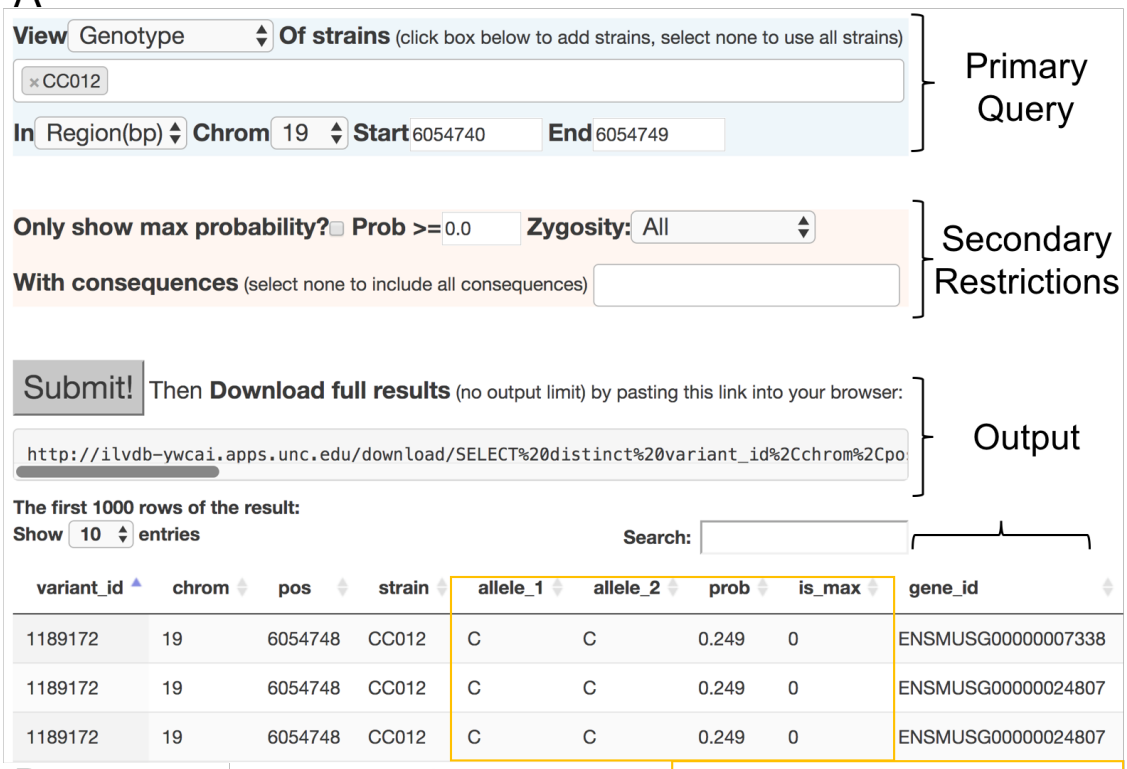
We provide an example of a genotype query to illustrate a partial ISVdb workflow, and also to demonstrate how uncertainty is represented in the ISVdb by storing multiple rows for a single variant in a single strain. Details are provided in the Figure 4.1 caption.

4.5.4 Genotype queries are similar to other ISVdb queries

Genotype queries are just one of the four types of queries enabled by the ISVdb GUI. The remaining types of ISVdb queries are almost identical to a genotype query, and all of them represent uncertainty in the same manner as a genotype query. Rather than describing them exhaustively, we will emphasize how each differs from a genotype query.

- **Genotype cross query:** rather than accepting a list of strains, this query only accepts two strains, simulates their F1 offspring, and returns data nearly identical in structure to genotype query data– except for a second strain per record.
- **Diplotype query:** there is no notion of a functional consequence with regard to a diplotype, and thus, diplotype queries return neither functional consequences, nor transcript IDs, which are closely tied to functional consequence in the ISVdb. Additionally, instead of records with (allele1,allele2) genotype, a diplotype query returns records with (haplotype1,haplotype2) diplotypes.
- **Diplotype cross query:** just like the genotype cross query, this query accepts two strains as input, and simulates their F1. It returns diplotype data which is nearly identical in structure to that from a diplotype query.

A



B

gene_id	transcript_id	consequence_1	consequence_2
ENSMUSG00000007338	ENSMUST00000007482	3_prime_UTR_variant	3_prime_UTR_variant
ENSMUSG00000024807	ENSMUST00000138532	downstream_gene_variant	downstream_gene_variant
ENSMUSG00000024807	ENSMUST00000156550	downstream_gene_variant	downstream_gene_variant

Figure 4.1: A) Example workflow of the ISVdb online GUI. A user has queried the genotype of CC0012, on chromosome 19, from 6054740:6054749. The "Primary Query" panel also allows additional strains, and/or specification of the region by genes instead. The user is interested in all zygosity variants, of all consequences, and all probabilities, and thus has applied no secondary restriction. After the user clicked "Submit!", a URL to download the resulting table was generated, as well as an online version of the table. The first three rows of the table are shown here: noticeably they all represent the *same variant in the same strain*. The difference between the rows is highlighted in the yellow box: the genotype per row and its associated probability. Collectively, the three rows represent that there is a 25% probability of a C/C genotype, 25% of T/T, and 50% of T/C at this variant in CC0012. B) The remaining wrapped columns of output from part A (A was too wide). Note that each genotype has a different consequence, accentuating that only accounting for the most likely genotype would cause a non-negligible loss of information. Also, note that this figure was pieced together from a screen capture to fit on a single page.

4.5.5 Incorporation of CC Sequencing and up-to-date CC genotyping

The basis for the current version of ISVdb variant calls is integration of MegaMUGA genotyping data from the MRCAs of each CC strain with the sequencing data from a single mouse per founder strain (Keane et al., 2011; Fu et al., 2012; Srivastava et al., 2017) Consequently, the ISVdb currently has important limitations regarding completeness, heterozygosity, and imputation precision. With regard to completeness, in the generations since the MRCAs, additional mutations have accumulated and, thanks to the extremely small effective population size within each CC strain, rapidly become fixed. Sequencing results suggest that the number of variants now segregating in the CC has increased by 2% since the MRCA generation. (Srivastava et al., 2017). Similarly, many of the regions harboring residual heterozygosity in the MRCA animals have by now become fixed. Regarding the precision of the imputation itself, the ability to construct the underlying haplotype mosaic is limited by, among other things, the resolution of the MegaMUGA genotyping array (or of any array) and in particular how well that array can mitigate the inherent difficulties arising with inferring haplotype state at recombination breakpoints and regions of identity by descent between founders. Our imputed variants thus reflect an incomplete and uncertain view of the current generation's CC genomes.

To address these limitations and gain a deeper understanding of the CC population, finer resolution data from a more recent breeding generation has been collected: the genome of a single male per CC strain has been sequenced (Srivastava et al., 2017), and this will not only identify the de novo CC mutations but also reduce the uncertainty in CC strain genotypes at known variants. Nonetheless, since sequencing is limited to a single animal per strain, it does not by itself provide a definitive answer genome-wide, largely due to residual heterozygosity. In the near future, a set of 3 other males per strain from the sequencing generation will be genotyped on the MUGA platform (PMV personal communication, 2017).

The ISVdb will progressively incorporate these new sets of data into its variant representation according. Towards this end, outstanding tasks include improved representation of the following:

- **Whole genome**: The published version of ISVdb focuses on exon data, but we have imputed variants across the whole genome, and all that remains is to make them publicly available.
- **Denovo variants**: Inclusion of sequenced de-novo variants as new records in the ISVdb.

- **Known MUGA variants**: Improved imputation of known MUGA variant by incorporating sequencing data; ISVdb probabilities will be some sort of weighted sum of the data sources.
- **Founder genotypes** CC sequencing data can indirectly provide us with ultra-deep founder sequencing, which can be used for a form of ancestral variant imputation.
- **Structural Variants**: Structural variant data (>100bp) is available and can be incoporated.

As the reference genome and its exon annotations changes, the ISVdb will be updated as well.

In addition to updating the ISVdb with more accurate state, incorporation of the newest high resolution data will allow a useful comparison between generations. In particular, detection of loci that have become fixed since the MRCA generation will open a new line of inquiry as to the possible selective advantage of the newly fixed alleles.

In summary, we have developed a database that stores imputed probabilistic variant state for CC strains and founder strains, and can rapidly generate probabilistic variants states for F1 populations. Imputed state includes alleles as well as predicted functional consequences of those alleles. This resource is a useful complement to sequencing data, and is easily accessible at `https://isvdb.unc.edu`.

CHAPTER 5
Conclusion and future efforts

In this conclusion, for each work chapter we will: i) summarize a subset of the major developments; ii) reconsider the implications of some of the results—with much of this portion being speculative; and iii) propose potential future efforts.

5.1 Reciprocal F1 Hybrids of NOD and B6

In chapter 2, we used RXs of a single pair of classical inbred strains, NOD and B6, exposing offspring to one of four diets, in order to detect POE, diet effects, and diet-by-POE on behavior and gene expression. Among the notable pieces of work in this effort, we developed a bespoke surrogate variable estimation procedure and permutation testing procedures, as well as a Bayesian mediation approach for integrating expression and behavior data; to our knowledge this is the first application of Bayesian mediation analysis in this context. The paper describing our efforts has been submitted. But some questions remain, and we discuss them below.

5.1.1 Discussion, with speculation

POEs on total expression were detected on genes throughout the genome, and although detections were enriched for imprinted genes, some genes subject to POE were non-imprinted. These non-imprinted genes are probably either regulated by an imprinted gene or subject to maternal effects. Although mediation analysis can suggest imprinted regulators, the true mechanism of POE on these non-imprinted genes cannot be known with certainty in our study design. If RNA-seq data were available, examination of ASE could increase our certainty: for a non-imprinted gene to be regulated by an imprinted gene, the ASE patterns of the two genes should match. Similarly, the POEs we detected on behavior (primarily locomotor-related) may be driven by imprinting or maternal effects, but in this context, ASE would be less helpful.

In general, POEs were quite significant in the contexts of both expression and behavior. By contrast, it seemed to be much easier to detect significant diet effects and diet-by-POE on expression than on behavior: diet effects and diet-by-POEs on behavior were not significant after FDR correction, whereas diet effects and diet-by-POEs were readily detected on the total expression of dozens of genes. Beyond the typical signal-to-noise-ratio limitations in studying behavior, possible reasons for the absence of diet/diet-by-POE on behavior include the following:

1. It was possible to significance test (almost) every gene. By contrast, although we tried to select a representative set of behavioral assays, there may be a near infinite set of behaviors that could be tested, and so it may be easy to overlook the behaviors affected by diet.

2. Individual genes may have been significantly more or less expressed depending on diet, but mouse behavior, which can be thought of as an emergent property of all gene expression taken together, may have evolved to be robust in the presence of most environmental perturbation; as such, the diets may have been insufficiently extreme to overcome compensatory pathways.

3. Given that over 20,000 genes were tested for significant effects, surrogate variable estimates of unobserved batch effects could be constructed by borrowing information between genes; and these surrogate variable effects (these results were not shown) often explained a large percentage of expression variance. By contrast, in a study akin to ours with far fewer (only 34) behaviors over which to borrow information, we could neither estimate nor control for unobserved batch effects on behavior.

Perhaps our most interesting biological finding was the suppression of *Carmil1* expression by *Airn*. To our knowledge, our study provides the first report of *Airn*—a lncRNA whose transcription is believed to control an immediate-vicinity imprinted gene cluster (Latos et al., 2012)—somehow affecting a gene distal to the cluster. If data from the CC-POE turn out to be consistent with this finding, a more targeted experiment focusing on the relationship between *Airn* and *Carmil1* may be worthwhile.

5.1.1.1 Changes to the existing study

Were I to redesign and/or perform further analysis on this or a similar study, I would consider the following:

1. Given the difficulties in achieving enough power to observe diet and diet-by-POE on behavior, I would restrict a new experiment to consider only two diets: standard and methyl enriched (given that methyl enrichment drove the majority of diet effects on expression). Alternately, a ladder of methyl enrichment values may be appropriate. Using 4 entirely different diets may have hedged our bets in the current study, but at the cost of an overly great reduction in power.

2. In the current set of analyses, we employed a statistical model in which each outcome was a value (behavior or expression) from a single animal. Given our experience in a similar study (Schoenrock et al. (2016), not discussed in this dissertation), we may have had more power to detect POE had we modeled *differences* between matching mice. That is, for every NODxB6 mouse on some diet, we would identify a matching B6xNOD mouse on the same diet, and then model the difference in phenotype between these matched mice as a function of diet and other covariates. No two mice could ever be matched perfectly, but the differences (e.g., dam, batch, etc.), could, with non-significant effort, be accounted for in the model.

3. We only employed permutation-based FWER testing in the context of testing gene expression. But a similar procedure could be derived for behavioral testing. Essentially, behavior-specific covariates would need to be regressed out of each behavior prior to permuting diet and/or parent-of-origin labels. As an aside, this point also emphasizes our difficulties with observing significant diet effects and diet-by-POE on behavior; none of these effects were significant following multiple testing correction even using FDR rather than a more conservative FWER.

4. Our mediation analysis searched for imprinted mediators of POE on *Carmil1*, the non-imprinted gene most significantly affected by POE. It would be interesting to see whether likely imprinted mediators could be identified for the other non-imprinted genes significantly affected by POE.

5. Inspection of the microarray images derived from CEL files revealed patches of irregularities, at different locations in each microarray. We decided that these were minor effects not worth

accounting for, but an ideal analysis would automatically search for these irregularities, and mask out the probes within such patches. With training data, image processing tools could be developed toward this end, and be enormously helpful for many other researchers; microarrays are still used extensively instead of RNA-seq in many contexts.

6. As mentioned earlier, surrogate variable estimation of unobserved batch effects had a large impact on detection of effects on total gene expression. To estimate these effects in our study, we adapted SSVA, a method which estimates surrogate variables based on negative control probe expression. By contrast, SVA is a method that estimates unobserved batch effects based on the main probes, and so we believed regressing out SVA-based surrogate variable would result in the elimination of true positives; it was for this reason we chose SSVA. However, recent developments in surrogate variable estimation include "Single cell partial least squares" (scPLS; Chen and Zhou 2017), which estimates surrogate variables by simultaneously modeling *both* spike-in controls, and target gene expression; similarly, scPLS could be used to estimate surrogate variables by simultaneously modeling *both* negative control probes and primary probes. This type of method may better balance avoidance of false positives and false negatives.

5.2 Rexplorer

In chapter 3, we developed a novel method for selecting optimal RXs from a panel of inbred lines for the study of POE. We then applied this method to the CC panel in the CC-POE project. The method took an operations research approach toward simultaneously ensuring POE would be detected while maintaining mapping resolution. We then used the method to select 8 RXs to be used in the CC-POE study, and showed that our choice of crosses was better than any set of crosses picked at random. Although partially similar ideas have been presented in the context of selective phenotyping, we formulated our approach to specifically focus on the study of POE. We are in the process of adapting the dissertation chapter corresponding to Rexplorer for paper submission.

5.2.1 Desired improvements

We were a bit fortunate in our application of Rexplorer: only one of its RXs selected for the CC-POE failed, even though we might have expected on average two RX to fail. Preliminary simulation-based efforts (not shown) to evaluate the robustness of our choice of RXs suggested our solution would remain fairly good at detecting POE and maintaining mapping resolution even if 1 or 2 crosses failed— but in no way did we succeed in searching for an optimally robust solution; our initial attempts at this were too computationally intensive. As such, the most pressing improvement is to incorporate a measure of resiliency into our formulation. One way to do this would be integrate our GMCP formulation with the MEXCLP formulation— which would allow for a (constant) probability of failure per RX to be taken into account in the selection of RXs. If a linear programming formulation of this sort proves computationally unfeasible, a randomized-greedy algorithm may be a nearly optimal alternative.

5.3 ISVdb

In chapter 4, with the intention of analyzing CC data from the CC-POE, we developed a probabilistically informed database of imputed variants in CC lines. The paper describing this effort has been published. We have also provided a publicly available website for querying variants, at `http:isvdb.unc.edu`. In addition to users external to UNC, our database is currently being used by other Valdar lab members in non-CC-POE-related merge-analyses of CC data.

5.3.1 Desired Improvements

In the conclusion of the ISVdb paper, we listed several desired (and intended) improvements. I restate a few of the key desired improvements here, and also summarize progress.

1. The initial implementation of the ISVdb used an SQL database to store its data. Due to the enormous disk-space requirements of this implementation, we limited ISVdb contents to only include variants within exons. In the months following publication, we have reimplemented the underlying representation of ISVdb to be a collection of compressed flat files. By virtue of existing and extremely fast decompression algorithms, along with the speed of the data.table

"fread" function for reading files into memory (Dowle and Srinivasan, 2017), our new implementation is nearly as fast, but uses two orders of magnitude less disk space. Accordingly, we now represent the entire genome in the ISVdb. Although we use this whole genome database within the Valdar lab, we are still working on making this data publicly available.

2. At the same time that the ISVdb was being developed, finer resolution data from a more recent breeding generation was collected: the genome of a single male per CC strain was sequenced (Srivastava et al., 2017). The de novo CC mutations and newly fixed alleles from this data need to be incorporates into the ISVdb. Nonetheless, since sequencing is limited to a single animal per strain, it cannot by itself provide a definitive answer genome-wide, largely due to residual heterozygosity.

3. Related to the de novo mutations detected by sequencing, the ISVdb does not currently represent structural variants (greater than 100 bp). These may become critical in future CC analyses.

5.3.2 Desired use: variant modeling

The ISVdb is a useful resource for the design or analysis of any study employing the CC. However, the primary motivation for its development was to map POE in CC-POE data. Below, we propose one way of incorporating ISVdb information into models of CC-POE study expression and behavior. Both of these models will share the same form, described in progressively more detail below.

By the experimental design of the CC-POE, each RX generates well-matched pairs of individuals that are genetically identical, and environmentally similar (same diet and cage), but that *differ in allelic parent-of-origin*. Thus (leaving out any interaction effects with diet for now) we can can model the phenotypic difference (akin to [(Gonzalo et al., 2007)]) within reciprocally-matched pairs as a sum of an overall RX POE, and a variant specific POE:

$$Y_{ijk} - Y_{ikj} \equiv \Delta Y_{ijk} = POE_{jk} + \theta_{jkv} + \epsilon_{ijk}, \;\; \epsilon_{ijk} \sim \mathrm{N}(0, \sigma^2) \tag{5.1}$$

In the ith matched pair, Y_{ijk} is the phenotype of an individual with maternal strain j, paternal strain k; Y_{ikj} is the phenotype of its matched pair; and ΔY_{ijk} is the modeled difference between the two phenotypes. POE_{jk} represents the cumulative POE (both imprinting and maternal effects) from a RX of lines j and k, excluding the POE contribution—θ_{jkv}—from variant v. ϵ_{ijk} is unmodeled noise.

At this point, ISVdb variant information can be incorporated. Each θ_{jkv} corresponds to the POE caused by a difference between the maternal and paternal alleles, which we denote $allele_{jv}$ and $allele_{kv}$, respectively. We could encode each $allele_{*v}$ at a variant as an integer that depended on the ISVdb predicted allele and predicted functional consequence; alleles with very different consequences such as the reference allele vs. a stop mutation should have a large difference in their $allele$ coefficient, to reflect the greater likelihood of such differences driving POE. Using this sort of reasoning to encode allele effects, we could specify the model:

$$\Delta Y_{ijk} = POE_{jk} + (allele_{jv} - allele_{kv})\theta_v + \epsilon_{ijk} \tag{5.2}$$

Where θ_v encodes the effect that variant v has on POE, assuming a "one-unit" difference between the alleles of parent j and k. In the simplest case, assuming a biallelic model with the two alleles encoded as 0 and 1, the design matrix for this LM would be 0 for mouse pairs that were homozygous at variant v, and 1 or -1 for pairs that were heterozygous at v (depending on the direction of the cross); the need for heterozygosity to reveal POE is encoded in this model.

Significant θ_v would correspond to variants with a higher likelihood of being causal to POE. Genes containing those significant variants would be prioritized for further study.

This model could be extended in a straightforward way to include interaction effects with diet, although we note that there would be no main effect of diet; animals within each matched pair are both on the same diet, so such an effect would be zeroed out.

We also note that ΔY_{ijk} will vary depending on the matchings that are chosen: there were typically 4 mice on the same diet and generated by the same RX in each cage, with two mice in one direction of the cross, and two in the other, yielding two different ways that animals could be matched in that cage. Accordingly, there are on the order of $2^{\#Cages}$ possible vectors ΔY_{ijk} that

could be modeled. To account for the uncertainty associated with modeling different sets of matched pairs, we could employ a multiple imputation technique as described in Schoenrock et al. (2016).

5.4 Overall conclusion

In this work, we developed methods for analysis of POE and diet-by-POE using RXs of a single pair of inbred strains. Employing these methods, our results motivated a follow up study, the CC-POE, involving multiple diets and RXs of multiple pairs of inbred CC lines. We selected the RXs for the CC-POE by developing a new experimental design tool. To aid with analysis of data from the CC-POE, we developed a database of imputed variants in CC lines. Looking at this dissertation more broadly, we have built a set of tools, and used these tools to demonstrate new ways of studying POE via reciprocal crosses.

APPENDIX A

Supplemental materials for "Reciprocal F1 Hybrids of two inbred mouse strains reveal parent-of-origin and perinatal diet effects on behavior and expression"

A.1 Figures

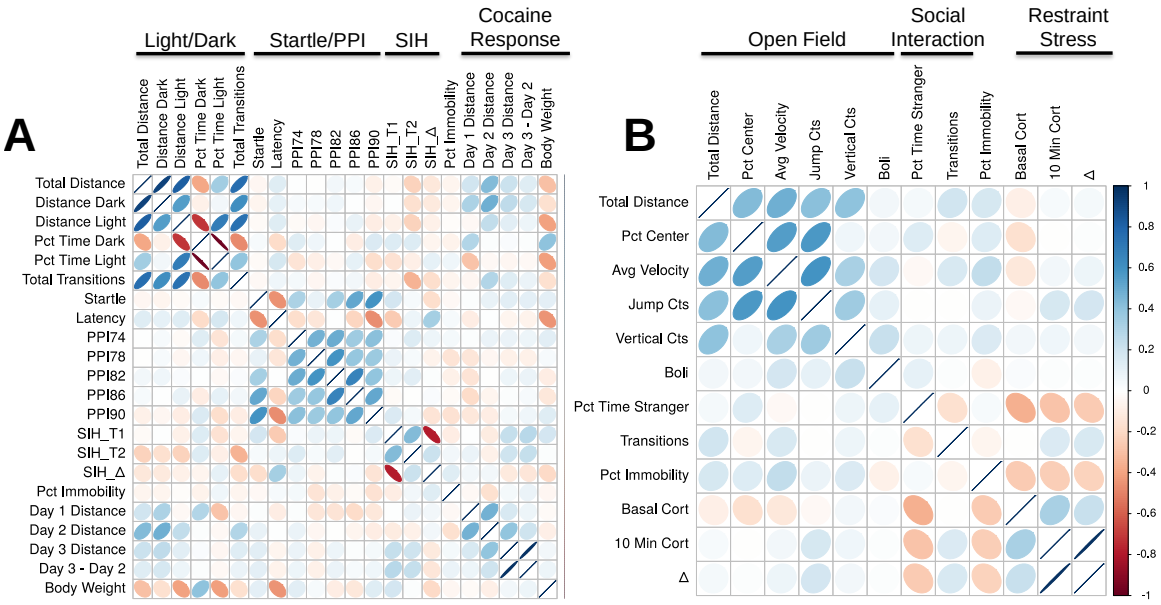


Figure A.1: Correlation of behavioral phenotypes. A) Pipeline 1 behaviors. B) Pipeline 2 behaviors

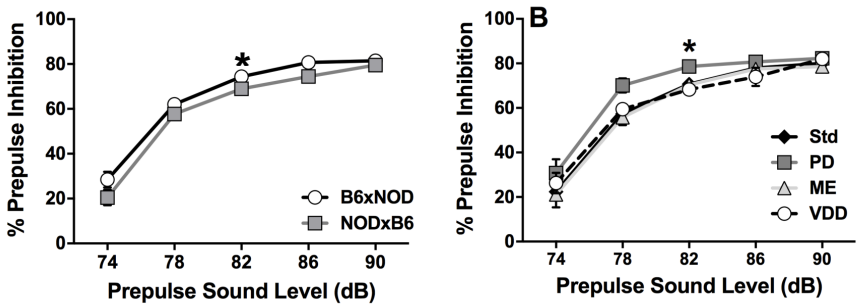


Figure A.2: POE and perinatal diet effect on sensorimotor gating. (A) POE on sensorimotor gating; mean (bars indicate SEM) percent prepulse inhibition for B6xNOD (n=46) and NODxB6 (n=45) mice. At 82 dB, a significant POE was observed on PPI (p=0.0307). (B) Diet effect on senorimotor gating; mean (bars indicate SEM) percent prepulse inhibition for standard (Std, n=31), vitamin D deficient (VDD, n=18), methyl enriched (ME, n=24) and protein deficient (PD, n=18) groups. At 82 dB, diet exposure significantly affected prepulse inhibition (p=0.00274).

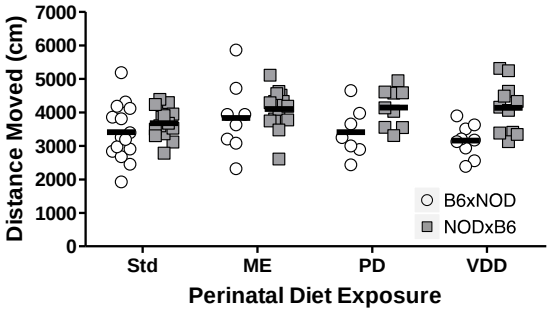


Figure A.3: Non-significant but suggestive diet-by-POE on distance moved in a 10min open field test.) Data are individual B6xNOD and NODxB6 animals exposed to Std (N=15,14), ME (N=8,14), PD (N=7,9) or VDD (N=9,11) diet. The pattern across diets follows that for percent center time (Figure 2.6).

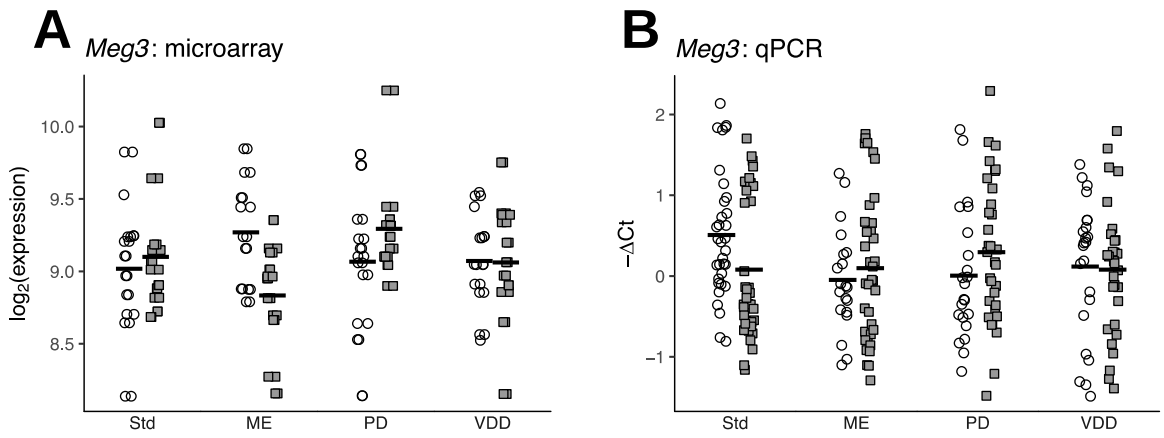


Figure A.4: Perinatal diet-by-POE on Meg3 gene expression levels. Each point corresponds to *Meg3* expression of an individual B6xNOD or NODxB6 mouse exposed to standard (Std), vitamin d deficient (VDD), methyl enriched (ME), or protein deficient (PD) diet, as measured by A) microarray, and B) Taqman qPCR analysis. *Meg3* expression was significantly subject to diet-by-POE in the microarray analysis but not in qPCR validation.

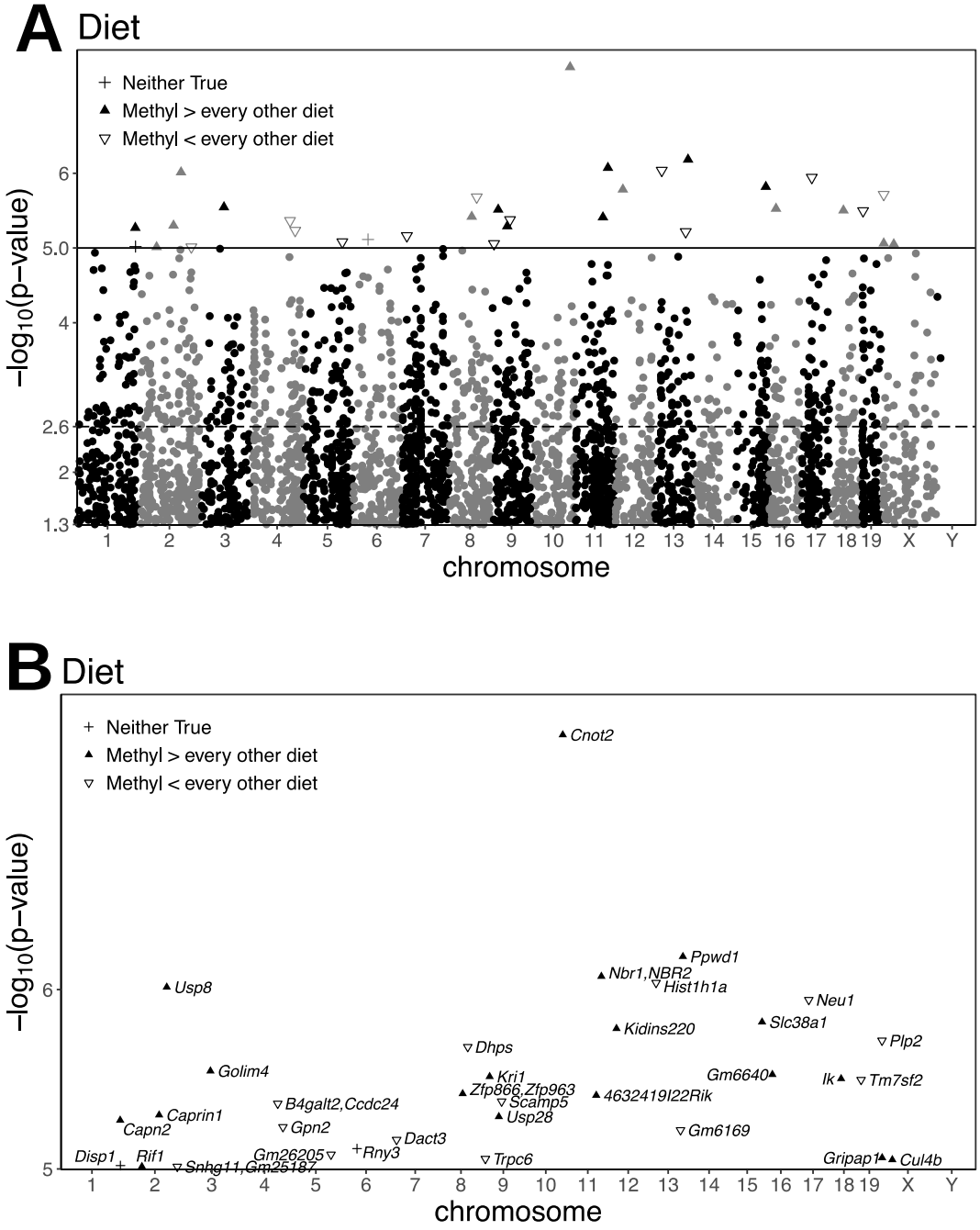


Figure A.5: P-values of diet effects on gene expression. (A) Manhattan-like plot of p-values of diet effect on microarray-measured expression; each point corresponds to a probeset location and the p-value of the diet effect on that probeset's expression. The dashed line represents the FDR threshold, and the solid line represents the FWER threshold. Probesets above the FWER threshold are marked by a shape, which depends on whether ME exposed mice have (on average) the highest expression relative to mice on the other diets (up arrow), the lowest expression relative to mice on the other diets (empty down arrow), or somewhere in middle of the 4 diets (plus sign); expression on the methyl diet is almost always at one or the other extreme. (B) Zoomed in view of just the genes significantly affected by diet. *Cnot2* is the most significantly affected

A.2 Tables

	AIN-93G Standard (Std;110700)		Protein Deficient (PD;102787)		Vitamin D Deficient (VDD;119266)		Methyl Enriched (ME;518893)		Methyl Deficient (MDD; 518892)	
Ingredient	**g/kg**	**kcal/kg**	**g/kg**	**kcal/kg**	**g/kg**	**kcal/kg**	**g/kg**	**kcal/kg**	**g/kg**	**kcal/kg**
Casein	200	716	75	269	200*	716	-	-	-	-
L-Cystine	3	12	0.9	3.6	3	12	-	-	-	-
Sucrose	100	400	100	400	100	400	382.19	1528.76	396.57	1586.28
Cornstarch	397.486	1430.9496	481.196	1732.3056	397.486	1430.9496	100	360	100	360
Dyetrose	132	501.6	160	608	132	501.6	100+	363	100+	363
Soybean Oil	70	630	70	630	70	630	50++	450	50++	450
t-Butyl hydroquinone	0.014	0	0.014	0	0.014	0	-	-	-	-
Cellulose	50	0	50	0	50	0	50	0	50	0
Choline Bitartrate	2.5	0	2.5	0	2.5	0	14.48	0	0	0
Mineral Mix #210025	35	30.8	35@ (#213266)	30.8	35	30.8	-	-	-	-
Vitamin Mix # 310025	10	38.7	10	38.7	10** (#319255)	38.7	10	38.7@@ (#300050)	10	38.7@@ ,# (#317754)
Calcium Phosphate Dibasic	-	-	10.97	0	-	-	-	-	-	-
Calcium Carbonate	-	-	4.42	0	-	-	-	-	-	-
Primex	-	-	-	-	-	-	100	900	100	900
Salt Mix #215001 (no Fe Added)	-	-	-	-	-	-	35	16.45	35	16.45
Sodium Bicarbonate	-	-	-	-	-	-	4.3	0	4.3	0
Ferric Citrate, U.S.P.	-	-	-	-	-	-	0.33	0	0.33	0
Succinyl Sulfathiazole	-	-	-	-	-	-	10	0	10	0
L-AA	-	-	-	-	-	-	143.7	574.8	143.8	575.2

Table A.1: Nutritional content of experimental diets. The PD and VDD diets were nutritionally matched to the Std diet and the ME was matched to the MDD. Red indicates the main nutritional component that was changed in each diet (pelleted). The product number associated with each diet, vitamin mix, and mineral mix are provided (Dyets, Inc; Bethlehem, PA). @Ca and P free, @@Vitamin K1/Dextrose mix free w/ addition of menadione sodium bisulfite; #folic acid free; *vitamin free, ** no vitamin D, +dextrin instead of dyetrose, ++corn oil instead of soybean oil, #folic acid free

Pipeline	Strain	Diet	# of Dams	# F1 Female Offspring	# Tested/Batch		
					1	2	3
1	B6xNOD	Standard	8	20	5	6	9
		Methyl Enriched	3	9	9	0	0
		Protein Deficient	6	9	4	5	0
		Vitamin D Deficient	3	8	0	4	4
	NODxB6	Standard	4	11	7	4	0
		Methyl Enriched	4	15	6	9	0
		Protein Deficient	3	9	2	3	4
		Vitamin D Deficient	3	10	5	5	0
2	B6xNOD	Standard	5	15	5	6	4
		Methyl Enriched	3	8	4	4	0
		Protein Deficient	3	7	7	0	0
		Vitamin D Deficient	3	9	0	3	6
	NODxB6	Standard	5	14	4	10	0
		Methyl Enriched	3	14	9	5	0
		Protein Deficient	4	9	4	5	0
		Vitamin D Deficient	4	11	4	7	0

Table A.2: Diets, number of dams, and female F1 hybrids per diet, broken down by various categories. The number of female F1 hybrids tested and the number of dams that produced those females in each behavior pipeline is broken down by reciprocal direction and perinatal diet exposure. The last column shows the breakdown of groups tested by behavior batch within each pipeline.

Dam Strain	Diet Exposure	# Dams Mated	# Litters	% Pregnant	# Pups Born	# Pups PND21	% Survival PND21	Mean Litter Size	# F Pups PND21	% F Pups	# F Tested
C57BL/6J (B6)	Standard	11	11	100	73	71	97.3	6.6 ± 2.4	35	49.30	35
	Methyl Enriched	10	9	90	65	47	72.3	7.2 ± 1.9	21	44.68	17
	Methyl Donor Deficient	12	2	16.7	18	0	0	9	-	-	-
	Protein Deficient	12	12	100	52	44	84.6	4.3 ± 2.1	21	47.73	16
	Vitamin D Deficient	10	10	100	54	52	96.3	6 ± 3	22	42.31	17
NOD/ShiLtJ (NOD)	Standard	7	7	100	63	61	96.8	9 ± 1.1	25	40.98	25
	Methyl Enriched	8	7	87.5	63	52	82.5	9 ± 1.3	29	55.77	29
	Methyl Donor Deficient	9	3	33.3	6	0	0	2	-	-	-
	Protein Deficient	10	9	90	65	55	84.6	6.9 ± 1.7	20	36.36	18
	Vitamin D Deficient	8	6	75	51	47	92.2	8.5 ± 2.1	21	44.68	21

Table A.3: Effect of perinatal diet and strain on breeding fitness. PND 21 = postnatal day 21 (time of weaning); F = females

Effect	Dataset	# Pups	p value
POE on *Carmil1*	full	115	6.3e-07
POE on *Carmil1*	old	85	4.4e-07
POE on *Carmil1*	new	30	9.7e-11
DietxPOE on *Meg3*	full	115	0.39
DietxPOE on *Meg3*	old	85	0.72
DietxPOE on *Meg3*	new	30	0.48

Table A.4: qPCR-based analysis of POE on *Carmil1* expression and Diet-by-POE on *Meg3* expression. Results shown for 3 datasets: i) mice that were both microassayed and qPCR'd ("old" data); ii) mice that were never microassayed but were qPCR'd ("new"); and iii) the union of the first 2 datasets ("full"). Parent-of-origin significantly affects qPCR-measured expression of *Carmil1* in all 3 datasets, whereas diet-by-parent-of-origin does not significantly affect qPCR-measured expression of *Meg3* in any dataset.

Table A.5: Tukey contrast p-values. For each phenotype, the table shows the modeled variables, along with the Tukey p-values for contrasts between all pairs of diets, and also between all pairs of diet-by-parent-of-origin effects. NOD:NODxB6 indicates an interaction between diet D with descent from maternal NOD. Significant values are **bolded**, and *, **, and ***, indicate significance levels of *0.05, **0.01, ***0.001 respectively. POE = parent of origin effect; PPI = prepulse inhibition; CORT = corticosterone; SIH-T1 = basal temperature; SIH-T2 = post-stress temperature; SIH-delta = (T2-T1)

Mediator Gene	Imprinted	Mediation Effect		Direct Effect		Suppressor
		ab	CTP	c'	CTP	
3830406C13Rik	FALSE	-0.465	0.00289	2.27	<7.81e-05	TRUE
Irak1bp1	FALSE	0.158	0.00914	1.62	<7.81e-05	FALSE
Airn_10441787	TRUE	-0.166	0.0134	1.97	<7.81e-05	TRUE
Tmem40	FALSE	-0.115	0.0281	1.91	<7.81e-05	TRUE
Pcdhb2	FALSE	-0.123	0.0372	1.92	<7.81e-05	TRUE
Mir485,Mirg	TRUE	-0.116	0.0409	1.93	<7.81e-05	TRUE

Table A.6: Coefficients and Combined Tail Probabilities (CTPs; akin to a p-value) for the significant gene mediators of *Carmil1* expression. The average mediation coefficient, a.k.a., the indirect effect, is ab, and is grouped together with its corresponding CTP. The direct effect, c', is also grouped together with its associated CTP. The mediation effect is generally *suppressing* the direct effect of POE; for 7 of 8 significant mediators, ab is opposite in sign to c'. *Airn* (as measured at probeset 10441787) is the most significant imprinted gene mediating the expression of *Carmil1*.

Behavior	Mediator Gene	Imprinted	Mediation Effect		Direct Effect		Suppressor
			ab	CTP	c'	CTP	
LD Total Distance	*Carmil1*	FALSE	-1.39	0.0408	2.15	0.0393	TRUE
Boli	*3830406C13Rik*	FALSE	-0.857	0.0386	0.867	0.171	TRUE
Day 1 Distance	*s115_10563911*	TRUE	0.348	0.0398	0.304	0.339	FALSE
Day 2 Distance	*s113,Rian*	TRUE	0.66	0.0271	-0.0562	0.478	TRUE
Day 3 Distance	*Carmil1*	FALSE	-1.5	0.0486	1.62	0.0912	TRUE
Basal Cort	*s113_10398354*	TRUE	0.734	0.0415	-1.24	0.12	TRUE
SIH_T1	*3830406C13Rik*	FALSE	-1.43	0.0134	1.01	0.208	TRUE
Pct Time Stranger	*Carmil1*	FALSE	1	0.0227	-0.658	0.297	TRUE
Pct Time Stranger	*s113_10398354*	TRUE	0.709	0.0272	-0.526	0.317	TRUE
Pct Time Stranger	*s115_10563949*	TRUE	0.665	0.0296	-0.432	0.366	TRUE
Pct Time Stranger	*s115_10563959*	TRUE	0.594	0.0415	-0.381	0.377	TRUE
Pct Time Stranger	*s116_10564209*	TRUE	-0.489	0.043	0.824	0.22	TRUE
Pct Time Stranger	*s113_10398370*	TRUE	0.567	0.0477	-0.231	0.427	TRUE
PPI86	*3830406C13Rik*	FALSE	1.45	0.0245	-1.86	0.0422	TRUE
PPI90	*s113_10398370*	TRUE	0.633	0.0435	-0.495	0.319	TRUE

Table A.7: Coefficients and Combined Tail Probabilities (CTPs; akin to a p-value) for the significant gene mediators of behavior. Mediation of POE was tested using each probeset's expression as a mediator, against each behavior as an outcome; the behavior-probeset pairs in this table are the nominally significant associations. The mediator probeset is named according to the gene that is probed, followed by the specific probeset ID that was found to be significant if more than one probeset interrogates that gene. As was the case of mediation of expression, the mediation effect, a.k.a., the indirect effect, is ab, and is grouped together with its corresponding CTP. The direct effect, c', is also grouped together with its associated CTP. As the coefficients are on a transformed scale, they are not especially informative, but they do demonstrate that for 17 of the 18 significant mediator-behavior pairings, the direct and indirect effect act opposite one another; i.e., when ab has the opposite sign of c', mediation is suppressing the direct effect. We note that *Carmil1* and *Airn* both appear as significant mediators of behavior.

Mediator Gene	Imprinted	# Behavior POEs		CTP
		strongly (top-3) mediated	suppressed	
Airn_10441787	TRUE	12	27	5.09e-05
Carmil1	FALSE	8	28	0.000518
s115_10563949	TRUE	7	23	0.000408
s115_10563911	TRUE	6	24	0.000857
1700063H04Rik	FALSE	4	18	0.0234
s113,Rian	TRUE	4	25	0.0152
3830406C13Rik	FALSE	4	24	0.00366
s113_10398370	TRUE	4	25	0.00484
s113_10398354	TRUE	4	27	0.0104
Mamdc2	FALSE	3	19	0.0208
Zrsr1	TRUE	3	19	0.0288
s116_10564209	TRUE	3	24	0.0192
Mir485,Mirg	TRUE	3	20	0.0164
Ndn	TRUE	2	19	0.00767
Irak1bp1	FALSE	1	18	0.0455
s115_10563915	TRUE	1	17	0.0214
Itga7	FALSE	1	24	0.0425
s115_10563989	TRUE	1	22	0.0324
Eps8l1	FALSE	1	22	0.0408
s115_10563959	TRUE	0	28	0.0259
3300002I08Rik	FALSE	0	24	0.0496

Table A.8: Genes that significantly mediate POE over all 34 behaviors in the aggregate; i.e., genes with a significant Combined Tail Probability$<$.05 (CTP), for POE mediation; the CTP essentially totals a gene's Combined Tail Probabilities over every behavior. The table also includes the number of behaviors for which a given gene is among the 3 most significant mediators of POE (whether or not the POE on each behavior separately is significant), as well as the number of behaviors for which a given gene suppresses rather than contributes to POE. For example, *Carmil1* is one of the 3 most significant mediators of POE on 8 behaviors; for 26 behaviors, its mediation acts to suppress POE on that behavior; its CTP on POE mediation over all 34 behaviors is .00028. *S113/S115/S116* are shorthand for *Snord 113/115/116* respectively; Snord genes may be concatenated with a specific probeset ID when more than 1 probed region within the gene family is a mediator.

Gene	Chr	Probeset Location	Probeset ID	F1 with Higher Expression	Imprinted	-log10	
						p value	q value
Tmem40	6	115729288	10547056	NODxB6	N	5.7	2.7
Eps8l1	7	4460832	10549655	B6xNOD	N	8.2	4.9
s115	7	59364519	10563911	NODxB6	Y	9.3	5.8
s115	7	59368241	10563915	NODxB6	Y	8.9	5.6
s115	7	59414112	10563949	B6xNOD	Y	9.4	5.9
s115	7	59423060	10563959	B6xNOD	Y	10.5	6.8
s115	7	59451136	10563989	B6xNOD	Y	6.4	3.3
s115	7	59470043	10564005	NODxB6	Y	5.2	2.2
s116	7	59904755	10564209	NODxB6	Y	13.2	9.2
Ndn	7	62348358	10553833	B6xNOD	Y	6.1	3.0
Irak1bp1	9	82830047	10587495	B6xNOD	N	5.6	2.6
s113	12	109631074	10398354	B6xNOD	Y	9.1	5.7
s113,Rian	12	109653064	10398360	B6xNOD	Y	7.5	4.3
s113	12	109679951	10398370	NODxB6	Y	9.8	6.2
Mir485,Mirg	12	109734899	10398426	B6xNOD	Y	7.5	4.3
Mir154	12	109738424	10398428	NODxB6	Y	5.1	2.1
Lrrc16a	13	24012417	10408280	NODxB6	N	13.8	9.5
3830406C13Rik	14	12285388	10412701	NODxB6	N	13.0	9.2
Airn	17	12814566	10441787	B6xNOD	Y	7.6	4.4
Pmaip1	18	66458621	10456357	B6xNOD	N	6.1	3.0

Table A.9: Genes whose expression is significantly affected by parent-of-origin, at the FWER threshold level. Chr = chromosome; N = no; Y = Yes; ME = methyl enriched diet; imprinted status determined using Crowley et al. (2015) and mousebook.org (Blake et al., 2010); p-values and q-values (FDR-corrected) are -log10 transformed.

Gene	Chr	Probeset Location	Probeset ID	ME Group Rank	Imprinted	-log10	
						p value	q value
Capn2	1	182467280	10360806	1	N	5.3	2.3
Disp1	1	183135366	10598178	3	N	5.0	2.2
Rif1	2	52073219	10472058	1	N	5.0	2.2
Caprin1	2	103766040	10485514	1	N	5.3	2.3
Usp8	2	126707411	10475665	1	N	6.0	2.4
Snhg11,Gm25187	2	158378222	10478073	4	N	5.0	2.2
Golim4	3	75876220	10498775	1	N	5.5	2.4
B4galt2,Ccdc24	4	117869344	10515500	4	N	5.4	2.3
Gpn2	4	133584404	10508901	4	N	5.2	2.3
Gm26205	5	121205020	10525185	4	N	5.1	2.2
Rny3	6	47781624	10537909	3	N	5.1	2.2
Dact3	7	16875425	10550383	4	N	5.2	2.3
Zfp866,Zfp963	8	69742939	10579089	1	N	5.4	2.3
Dhps	8	85071953	10573566	4	N	5.7	2.4
Trpc6	9	8544203	10583163	4	N	5.1	2.2
Kri1	9	21273705	10591497	1	N	5.5	2.4
Usp28	9	48985437	10585099	1	N	5.3	2.3
Scamp5	9	57441459	10593927	4	N	5.4	2.3
Cnot2	10	116485263	10372534	1	N	7.4	3.1
4632419I22Rik	11	86201483	10379953	1	N	5.4	2.3
Nbr1,NBR2	11	101552329	10381419	1	N	6.1	2.4
Kidins220	12	24986906	10395005	1	N	5.8	2.4
Hist1h1a	13	23763753	10404069	4	N	6.0	2.4
Gm6169	13	97098259	10411359	4	N	5.2	2.3
Ppwd1	13	104205516	10411915	1	N	6.2	2.4
Slc38a1	15	96572325	10431872	1	N	5.8	2.4
Gm6640	16	23717234	10438726	1	N	5.5	2.4
Neu1	17	34931317	10444578	4	N	5.9	2.4
Ik	18	36744733	10454966	1	N	5.5	2.4
Tm7sf2	19	6062921	10465342	4	N	5.5	2.4
Plp2	X	7667967	10603346	4	N	5.7	2.4
Gripap1	X	7790043	10598422	1	N	5.1	2.2
Cul4b	X	38533328	10604199	1	N	5.1	2.2

Table A.10: Genes whose expression is significantly affected by perinatal diet exposure, at the FWER threshold level. Chr = chromosome; N = no; Y = Yes; ME Group rank = This gene's expression rank, in mice exposed to methyl enriched (ME) diet, relative to mice on the other 4 diets— *e.g.*, 1 means ME mice expressed this gene the most, whereas 4 means ME mice expressed this gene the least; imprinted status determined using Crowley et al. (2015) and mousebook.org (Blake et al., 2010); p-values and q-values (FDR-corrected) are -log10 transformed.

Gene	Chr	Probeset Location	Probeset ID	Imprinted	-log10	
					p value	q value
Mir128-1	1	128202355	10349510	N	5.2	1.7
Snhg11	2	158375822	10478066	N	4.7	1.5
Szt2	4	118387131	10515706	N	4.6	1.5
Szt2	4	118388309	10515714	N	5.3	1.8
Zcchc17	4	130316037	10516778	N	6.0	2.1
Cdc42	4	137319843	10517559	N	5.3	1.8
Efcc1	6	87730951	10539810	N	5.0	1.7
Olfml1	7	107567521	10556076	N	4.5	1.5
Plekha1	7	130865914	10558134	N	4.9	1.6
Ginm1	10	7768202	10367717	N	6.0	2.1
Gm10418	11	70540048	10387907	N	4.9	1.6
Mir341	12	109611500	10398350	Y	6.5	2.1
Gdi2	13	3538226	10403258	N	5.1	1.7
Adk	14	21052639	10413086	N	4.5	1.5
Trav13d-1	14	52851305	10414751	N	4.8	1.6
B230359F08Rik	14	53795368	10414973	N	5.5	1.8
Tcp1	17	12916535	10441797	N	4.8	1.6

Table A.11: Genes whose expression is significantly affected by diet-by-POE, at the FWER threshold level. Chr = chromosome; N = no; Y = Yes; imprinted status determined using Crowley et al. (2015) and mousebook.org (Blake et al., 2010); p-values and q-values (FDR-corrected) are -log10 transformed.

BIBLIOGRAPHY

Affymetrix (2017). *Affymetrix Power Tools*. Affymetrix.

Agarwal, V., Bell, G. W., Nam, J.-W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, 4:e05005.

Altobelli, G., Bogdarina, I. G., Stupka, E., Clark, A. J. L., and Langley-Evans, S. (2013). Genome-Wide Methylation and Gene Expression Changes in Newborn Rats following Maternal Protein Restriction and Reversal by Folic Acid. *PLOS ONE*, 8(12):e82989.

Anderson, M. and Braak, C. T. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of statistical computation and simulation*, 73(2):85–113.

Aylor, D. L., Valdar, W., Foulds-Mathes, W., Buus, R. J., Verdugo, R. A., Baric, R. S., Ferris, M. T., Frelinger, J. A., Heise, M., Frieman, M. B., Gralinski, L. E., Bell, T. A., Didion, J. D., Hua, K., Nehrenberg, D. L., Powell, C. L., Steigerwalt, J., Xie, Y., Kelada, S. N. P., Collins, F. S., Yang, I. V., Schwartz, D. A., Branstetter, L. A., Chesler, E. J., Miller, D. R., Spence, J., Liu, E. Y., McMillan, L., Sarkar, A., Wang, J., Wang, W., Zhang, Q., Broman, K. W., Korstanje, R., Durrant, C., Mott, R., Iraqi, F. A., Pomp, D., Threadgill, D., Villena, F. P.-M. d., and Churchill, G. A. (2011). Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Research*, 21(8):1213–1222.

Baran-Gale, J., Fannin, E. E., Kurtz, C. L., and Sethupathy, P. (2013). Beta Cell 5'-Shifted isomiRs Are Candidate Regulatory Hubs in Type 2 Diabetes. *PLoS ONE*, 8(9).

Barlow, D. P. and Bartolomei, M. S. (2014). Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, 6(2):a018382.

Barnett, M. P. G., Bermingham, E. N., Young, W., Bassett, S. A., Hesketh, J. E., Maciel-Dominguez, A., McNabb, W. C., and Roy, N. C. (2015). Low Folate and Selenium in the Mouse Maternal Diet Alters Liver Gene Expression Patterns in the Offspring after Weaning. *Nutrients*, 7(5):3370–3386.

Bartolomei, M. S. and Ferguson-Smith, A. C. (2011). Mammalian Genomic Imprinting. *Cold Spring Harbor Perspectives in Biology*, 3(7):a002592.

Bates, D., Machler, M., Bolker, B., and Walker, S. (2015). Fitting Linear Mixed-Effects Models Using {lme4}. *Journal of Statistical Software*, 67(1):1–48.

Batta, R., Dolan, J. M., and Krishnamurthy, N. N. (1989). The Maximal Expected Covering Location Problem: Revisited. *Transportation Science*, 23(4):277–287.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300.

Berrocal-Zaragoza, M. I., Sequeira, J. M., Murphy, M. M., Fernandez-Ballart, J. D., Abdel Baki, S. G., Bergold, P. J., and Quadros, E. V. (2014). Folate deficiency in rat pups during weaning causes learning and memory deficits. *Br J Nutr*, 112(8):1323–32.

Blake, A., Pickford, K., Greenaway, S., Thomas, S., Pickard, A., Williamson, C. M., Adams, N. C., Walling, A., Beck, T., Fray, M., Peters, J., Weaver, T., Brown, S. D. M., Hancock, J. M., and Mallon, A.-M. (2010). MouseBook: an integrated portal of mouse resources. *Nucleic Acids Research*, 38(suppl 1):D593–D599.

Box, G. E. P. and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2):211–252.

Browning, B. L. and Browning, S. R. (2009). A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, 84(2):210–223.

Burne, T. H., Becker, A., Brown, J., Eyles, D. W., Mackay-Sim, A., and McGrath, J. J. (2004a). Transient prenatal vitamin d deficiency is associated with hyperlocomotion in adult rats. *Behav Brain Res*, 154(2):549–55.

Burne, T. H., Feron, F., Brown, J., Eyles, D. W., McGrath, J. J., and Mackay-Sim, A. (2004b). Combined prenatal and chronic postnatal vitamin d deficiency in rats impairs prepulse inhibition of acoustic startle. *Physiol Behav*, 81(4):651–5.

Burne, T. H., O'Loan, J., McGrath, J. J., and Eyles, D. W. (2006). Hyperlocomotion associated with transient prenatal vitamin d deficiency is ameliorated by acute restraint. *Behav Brain Res*, 174(1):119–24.

Calatayud, F. and Belzung, C. (2001). Emotional reactivity in mice, a case of nongenetic heredity? *Physiology & Behavior*, 74(3):355–362.

Calatayud, F., Coubard, S., and Belzung, C. (2004). Emotional reactivity in mice may not be inherited but influenced by parents. *Physiology & Behavior*, 80(4):465–474.

Chen, M. and Zhou, X. (2017). Controlling for Confounding Effects in Single Cell RNA Sequencing Studies Using both Control and Target Genes. *Scientific Reports*, 7(1):13587.

Cleaton, M. A., Edwards, C. A., and Ferguson-Smith, A. C. (2014). Phenotypic Outcomes of Imprinted Gene Models in Mice: Elucidation of Pre- and Postnatal Functions of Imprinted Genes. *Annual Review of Genomics and Human Genetics*, 15(1):93–126.

Cohen, R. and Katzir, L. (2008). The Generalized Maximum Coverage Problem. *Information Processing Letters*, 108(1):15–22.

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

Cowley, D. E., Pomp, D., Atchley, W. R., Eisen, E. J., and Hawkins-Brown, D. (1989). The Impact of Maternal Uterine Genotype on Postnatal Growth and Adult Body Size in Mice. *Genetics*, 122(1):193–203.

Crabbe, J. C., Wahlsten, D., and Dudek, B. C. (1999). Genetics of mouse behavior: interactions with laboratory environment. *Science (New York, N.Y.)*, 284(5420):1670–1672.

Crider, K. S., Yang, T. P., Berry, R. J., and Bailey, L. B. (2012). Folate and DNA Methylation: A Review of Molecular Mechanisms and the Evidence for Folates Role. *Advances in Nutrition: An International Review Journal*, 3(1):21–38.

Crowley, J. J., Kim, Y., Lenarcic, A. B., Quackenbush, C. R., Barrick, C. J., Adkins, D. E., Shaw, G. S., Miller, D. R., de Villena, F. P.-M., Sullivan, P. F., and Valdar, W. (2014). Genetics of adverse reactions to haloperidol in a mouse diallel: a drug-placebo experiment and Bayesian causal analysis. *Genetics*, 196(1):321–47.

Crowley, J. J., Zhabotynsky, V., Sun, W., Huang, S., Pakatci, I. K., Kim, Y., Wang, J. R., Morgan, A. P., Calaway, J. D., Aylor, D. L., Yun, Z., Bell, T. A., Buus, R. J., Calaway, M. E., Didion, J. P., Gooch, T. J., Hansen, S. D., Robinson, N. N., Shaw, G. D., Spence, J. S., Quackenbush, C. R., Barrick, C. J., Nonneman, R. J., Kim, K., Xenakis, J., Xie, Y., Valdar, W., Lenarcic, A. B., Wang, W., Welsh, C. E., Fu, C.-P., Zhang, Z., Holt, J., Guo, Z., Threadgill, D. W., Tarantino, L. M., Miller, D. R., Zou, F., McMillan, L., Sullivan, P. F., and Pardo-Manuel de Villena, F. (2015). Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. *Nature Genetics*, 47(4):353–360.

Dadd, T., Weale, M. E., and Lewis, C. M. (2009). A critical evaluation of genomic control methods for genetic association studies. *Genetic Epidemiology*, 33(4):290–298.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., and Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15):2156–2158.

Dannemann, M., Lorenc, A., Hellmann, I., Khaitovich, P., and Lachmann, M. (2009). The effects of probe binding affinity differences on gene expression measurements and how to deal with them. *Bioinformatics*, 25(21):2772–2779.

Darvasi, A. (1998). Experimental strategies for the genetic dissection of complex traits in animal models. *Nature Genetics*, 18(1):19–24.

Daskin, M. S. (1983). A Maximum Expected Covering Location Model: Formulation, Properties and Heuristic Solution. *Transportation Science*, 17(1):48–70.

Davies, W., Isles, A. R., and Wilkinson, L. S. (2001). Imprinted genes and mental dysfunction. *Annals of Medicine*, 33(6):428–436.

Davison, J. M., Mellott, T. J., Kovacheva, V. P., and Blusztajn, J. K. (2009). Gestational Choline Supply Regulates Methylation of Histone H3, Expression of Histone Methyltransferases G9a (Kmt1c) and Suv39h1 (Kmt1a), and DNA Methylation of Their Genes in Rat Fetal Liver and Brain. *Journal of Biological Chemistry*, 284(4):1982–1989.

Dent, C. L. and Isles, A. R. (2014). Brain-expressed imprinted genes and adult behaviour: the example of Nesp and Grb10. *Mammalian Genome*, 25(1-2):87–93.

Devlin, B., Roeder, K., and Wasserman, L. (2001). Genomic control, a new approach to genetic-based association studies. *Theoretical Population Biology*, 60(3):155–166.

Didion, J. P., Morgan, A. P., Clayshulte, A. M.-F., Mcmullan, R. C., Yadgary, L., Petkov, P. M., Bell, T. A., Gatti, D. M., Crowley, J. J., Hua, K., Aylor, D. L., Bai, L., Calaway, M., Chesler, E. J., French, J. E., Geiger, T. R., Gooch, T. J., Jr, T. G., Harrill, A. H., Hunter, K., McMillan, L., Holt, M., Miller, D. R., O'Brien, D. A., Paigen, K., Pan, W., Rowe, L. B., Shaw, G. D., Simecek, P., Sullivan, P. F., Svenson, K. L., Weinstock, G. M., Threadgill, D. W., Pomp, D., Churchill, G. A., and Villena, F. P.-M. d. (2015). A Multi-Megabase Copy Number Gain Causes Maternal Transmission Ratio Distortion on Mouse Chromosome 2. *PLOS Genetics*, 11(2):e1004850.

Dobzhansky, T. (1936). Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics*, 21(2):113–135.

Dowle, M. and Srinivasan, A. (2017). *data.table: Extension of 'data.frame'*.

Dudbridge, F. and Koeleman, B. P. C. (2004). Efficient computation of significance levels for multiple associations in large studies of correlated data, including genomewide association studies. *American Journal of Human Genetics*, 75(3):424–35.

Duselis, A. R., Wiley, C. D., O'Neill, M. J., and Vrana, P. B. (2005). Genetic evidence for a maternal effect locus controlling genomic imprinting and growth. *genesis*, 43(4):155–165.

Dykens, E. M., Lee, E., and Roof, E. (2011). Prader-Willi syndrome and autism spectrum disorders: an evolving story. *Journal of Neurodevelopmental Disorders*, 3(3):225–237.

Ferguson, S. A., Berry, K. J., Hansen, D. K., Wall, K. S., White, G., and Antony, A. C. (2005). Behavioral effects of prenatal folate deficiency in mice. *Birth Defects Res A Clin Mol Teratol*, 73(4):249–52.

Ferris, M. T., Aylor, D. L., Bottomly, D., Whitmore, A. C., Aicher, L. D., Bell, T. A., Bradel-Tretheway, B., Bryan, J. T., Buus, R. J., Gralinski, L. E., Haagmans, B. L., McMillan, L., Miller, D. R., Rosenzweig, E., Valdar, W., Wang, J., Churchill, G. A., Threadgill, D. W., McWeeney, S. K., Katze, M. G., Villena, F. P.-M. d., Baric, R. S., and Heise, M. T. (2013). Modeling Host Genetic Regulation of Influenza Pathogenesis in the Collaborative Cross. *PLOS Pathogens*, 9(2):e1003196.

Fisher, R. A. (1925). *Statistical methods for research workers*. Genesis Publishing Pvt Ltd.

Francks, C., Maegawa, S., Lauren, J., Abrahams, B. S., Velayos-Baeza, A., Medland, S. E., Colella, S., Groszer, M., McAuley, E. Z., Caffrey, T. M., Timmusk, T., Pruunsild, P., Koppel, I., Lind, P. A., Matsumoto-Itaba, N., Nicod, J., Xiong, L., Joober, R., Enard, W., Krinsky, B., Nanba, E., Richardson, A. J., Riley, B. P., Martin, N. G., Strittmatter, S. M., Moller, H.-J., Rujescu, D., Clair, D. S., Muglia, P., Roos, J. L., Fisher, S. E., Wade-Martins, R., Rouleau, G. A., Stein, J. F., Karayiorgou, M., Geschwind, D. H., Ragoussis, J., Kendler, K. S., Airaksinen, M. S., Oshimura, M., DeLisi, L. E., and Monaco, A. P. (2007). LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular Psychiatry*, 12(12):1129–1139.

Franzek, E. J., Sprangers, N., Janssens, A. C. J. W., Van Duijn, C. M., and De Wetering, B. J. M. V. (2008). Prenatal exposure to the 1944-45 dutch 'hunger winter' and addiction later in life. *Addiction*, 103(3):433–438.

Fu, C.-P., Welsh, C. E., de Villena, F. P.-M., and McMillan, L. (2012). Inferring ancestry in admixed populations using microarray probe intensities. In *Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine*, pages 105–112. ACM.

Gabory, A., Ripoche, M.-A., Digarcher, A. L., Watrin, F., Ziyyat, A., Forn, T., Jammes, H., Ainscough, J. F. X., Surani, M. A., Journot, L., and Dandolo, L. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development*, 136(20):3413–3421.

Gail, M. H., Tan, W. Y., and Piantadosi, S. (1988). Tests for No Treatment Effect in Randomized Clinical Trials. *Biometrika*, 75(1):57–64.

Gatti, D. M., Svenson, K. L., Shabalin, A., Wu, L.-Y., Valdar, W., Simecek, P., Goodwin, N., Cheng, R., Pomp, D., Palmer, A., Chesler, E. J., Broman, K. W., and Churchill, G. a. (2014). Quantitative Trait Locus Mapping Methods for Diversity Outbred Mice. *G3: Genes, Genomes, Genetics*, 4(9):1623–1633.

Georges, M., Charlier, C., and Cockett, N. (2003). The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends in Genetics*, 19(5):248–252.

Gonzalez-Billault, C., Muoz-Llancao, P., Henriquez, D. R., Wojnacki, J., Conde, C., and Caceres, A. (2012). The role of small GTPases in neuronal morphogenesis and polarity. *Cytoskeleton (Hoboken, N.J.)*, 69(7):464–485.

Gonzalo, M., Vyn, T. J., Holland, J. B., and McIntyre, L. M. (2007). Mapping reciprocal effects and interactions with plant density stress in Zea mays L. *Heredity*, 99(1):14–30.

Good, P. I. (2005). *Permutation, parametric and bootstrap tests of hypotheses*. Springer series in statistics. Springer, New York, 3rd ed edition. OCLC: ocm55473971.

Gordon-Weeks, P. R. and Fournier, A. E. (2014). Neuronal cytoskeleton in synaptic plasticity and regeneration. *Journal of Neurochemistry*, 129(2):206–212.

Gralinski, L. E., Menachery, V. D., Morgan, A. P., Totura, A. L., Beall, A., Kocher, J., Plante, J., Harrison-Shostak, D. C., Schfer, A., Villena, F. P.-M. d., Ferris, M. T., and Baric, R. S. (2017). Allelic Variation in the Toll-Like Receptor Adaptor Protein Ticam2 Contributes to SARS-Coronavirus Pathogenesis in Mice. *G3: Genes, Genomes, Genetics*, 7(6):1653–1663.

Green, R., Wilkins, C., Thomas, S., Sekine, A., Hendrick, D. M., Voss, K., Ireton, R. C., Mooney, M., Go, J. T., Choonoo, G., Jeng, S., de Villena, F. P.-M., Ferris, M. T., McWeeney, S., and Gale, M. (2017). Oas1b-dependent Immune Transcriptional Profiles of West Nile Virus Infection in the Collaborative Cross. *G3 (Bethesda, Md.)*, 7(6):1665–1682.

Guan, Y. and Stephens, M. (2008). Practical Issues in Imputation-Based Association Mapping. *PLoS Genetics*, 4(12).

Gurobi Optimization, Inc. (2015). Gurobi Optimizer Reference Manual.

Hager, R., Cheverud, J. M., and Wolf, J. B. (2008). Maternal Effects as the Cause of Parent-of-Origin Effects That Mimic Genomic Imprinting. *Genetics*, 178(3):1755–1762.

Hall, C. S. (1934). Emotional behavior in the rat. I. Defecation and urination as measures of individual differences in emotionality. *Journal of Comparative psychology*, 18(3):385.

Harms, L. R., Cowin, G., Eyles, D. W., Kurniawan, N. D., McGrath, J. J., and Burne, T. H. (2012). Neuroanatomy and psychomimetic-induced locomotion in c57bl/6j and 129/x1svj mice exposed to developmental vitamin d deficiency. *Behav Brain Res*, 230(1):125–31.

Harms, L. R., Eyles, D. W., McGrath, J. J., Mackay-Sim, A., and Burne, T. H. (2008). Developmental vitamin d deficiency alters adult behaviour in 129/svj and c57bl/6j mice. *Behav Brain Res*, 187(2):343–50.

Hawley, M. E. and Kidd, K. K. K. (1995). HAPLO - A program using the EM algorithm to estimate the frequencies of multisite haplotypes . *Journal Of Heredity*, 86(5):409–411.

Heijmans, B. T., Tobi, E. W., Stein, A. D., Putter, H., Blauw, G. J., Susser, E. S., Slagboom, P. E., and Lumey, L. H. (2008). Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 105(44):17046–17049.

Hickey, J. M., Kinghorn, B. P., Tier, B., van der Werf, J. H. J., and Cleveland, M. A. (2012). A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation. *Genetics, Selection, Evolution*, 44:9.

Howie, B. N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics*, 5(6):e1000529.

Huang, B. E., Clifford, D., and Cavanagh, C. (2013a). Selecting subsets of genotyped experimental populations for phenotyping to maximize genetic diversity. *Theoretical and Applied Genetics*, 126(2):379–388.

Huang, S., Kao, C.-Y., McMillan, L., and Wang, W. (2013b). Transforming genomes using MOD files with applications. In *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics*, page 595. ACM.

Hutter, F., Hoos, H. H., and Leyton-Brown, K. (2010). Automated configuration of mixed integer programming solvers. In *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, pages 186–202. Springer.

Insel, T. (2008). Assessing the economic costs of serious mental illness. *American Journal of Psychiatry*, 165(6):663–665.

Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U., and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 4(2):249–264.

Isles, A. R., Baum, M. J., Ma, D., Keverne, E. B., and Allen, N. D. (2001). Genetic imprinting: Urinary odour preferences in mice. *Nature; London*, 409(6822):783–4.

Isles, A. R. and Wilkinson, L. S. (2000). Imprinted genes, cognition and behaviour. *Trends in cognitive sciences*, 4(8):309–318.

Ivanova, E., Chen, J.-H., Segonds-Pichon, A., Ozanne, S. E., and Kelsey, G. (2012). DNA methylation at differentially methylated regions of imprinted genes is resistant to developmental programming by maternal nutrition. *Epigenetics*, 7(10):1200–1210.

Jannink, J.-L. (2005). Selective Phenotyping to Accurately Map Quantitative Trait Loci. *Crop Science*, 45(3):901.

Jin, C., Lan, H., Attie, A. D., Churchill, G. A., Bulutuglo, D., and Yandell, B. S. (2004). Selective Phenotyping for Increased Efficiency in Genetic Mapping Studies. *Genetics*, 168(4):2285–2293.

Kalish, J. M., Jiang, C., and Bartolomei, M. S. (2014). Epigenetics and imprinting in human disease. *The International Journal of Developmental Biology*, 58(2-3-4):291–298.

Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., Furlotte, N. A., Eskin, E., Nellaker, C., Whitley, H., Cleak, J., Janowitz, D., Hernandez-Pliego, P., Edwards, A., Belgard, T. G., Oliver, P. L., McIntyre, R. E., Bhomra, A., Nicod, J., Gan, X., Yuan, W., van der Weyden, L., Steward, C. A., Bala, S., Stalker, J., Mott, R., Durbin, R., Jackson, I. J., Czechanski, A., Guerra-Assuncao, J. A., Donahue, L. R., Reinholdt, L. G., Payseur, B. A., Ponting, C. P., Birney, E., Flint, J., and Adams, D. J. (2011). Mouse genomic variation and its effect on phenotypes and gene regulation. *Nature*, 477(7364):289–294.

Kemper, K., Bowman, P., Pryce, J., Hayes, B., and Goddard, M. (2012). Long-term selection strategies for complex traits using high-density genetic markers. *Journal of Dairy Science*, 95(8):4646–4656.

Kesby, J. P., Burne, T. H., McGrath, J. J., and Eyles, D. W. (2006). Developmental vitamin d deficiency alters mk 801-induced hyperlocomotion in the adult rat: An animal model of schizophrenia. *Biol Psychiatry*, 60(6):591–6.

Kesby, J. P., Cui, X., O'Loan, J., McGrath, J. J., Burne, T. H., and Eyles, D. W. (2010). Developmental vitamin d deficiency alters dopamine-mediated behaviors and dopamine transporter function in adult female rats. *Psychopharmacology (Berl)*, 208(1):159–68.

Kesby, J. P., O'Loan, J. C., Alexander, S., Deng, C., Huang, X. F., McGrath, J. J., Eyles, D. W., and Burne, T. H. (2012). Developmental vitamin d deficiency alters mk-801-induced behaviours in adult offspring. *Psychopharmacology (Berl)*, 220(3):455–63.

King, E. G., Macdonald, S. J., and Long, A. D. (2012). Properties and power of the Drosophila Synthetic Population Resource for the routine dissection of complex traits. *Genetics*, 191(3):935–49.

Kinghorn, B. P. (2000). The tactical approach to implementing breeding programs. *Animal Breedinguse of new technologies, Post Graduate Foundation in Veterinarian Science of the University of Sydney*, pages 291–308.

Kinghorn, B. P. (2011). An algorithm for efficient constrained mate selection. *Genetics Selection Evolution*, 43(1):4.

Kirkpatrick, M. and Lande, R. (1989). The Evolution of Maternal Characters. *Evolution*, 43(3):485–503.

Koerner, M. V., Pauler, F. M., Huang, R., and Barlow, D. P. (2009). The function of non-coding RNAs in genomic imprinting. *Development*, 136(11):1771–1783.

Konycheva, G., Dziadek, M. A., Ferguson, L. R., Krageloh, C. U., Coolen, M. W., Davison, M., and Breier, B. H. (2011). Dietary methyl donor deficiency during pregnancy in rats shapes learning and anxiety in offspring. *Nutr Res*, 31(10):790–804.

Kutalik, Z., Johnson, T., Bochud, M., Mooser, V., Vollenweider, P., Waeber, G., Waterworth, D., Beckmann, J. S., and Bergmann, S. (2011). Methods for testing association between uncertain genotypes and quantitative traits. *Biostatistics (Oxford, England)*, 12(1):1–17.

Kuznetsova, A., Bruun, B., and Christensen, R. (2015). lmerTest: Tests in Linear Mixed Effects Models. Manual, R. R package version 2.0-29.

Latos, P. A., Pauler, F. M., Koerner, M. V., energin, H. B., Hudson, Q. J., Stocsits, R. R., Allhoff, W., Stricker, S. H., Klement, R. M., Warczok, K. E., Aumayr, K., Pasierbek, P., and Barlow, D. P. (2012). Airn Transcriptional Overlap, But Not Its lncRNA Products, Induces Imprinted Igf2r Silencing. *Science*, 338(6113):1469–1472.

Lawson, H. A., Cheverud, J. M., and Wolf, J. B. (2013). Genomic imprinting and parent-of-origin effects on complex traits. *Nature reviews. Genetics*, 14(9):609–617.

Lee, R. and Avramopoulos, D. (2014). Introduction to Epigenetics in Psychiatry. *Advances in Water Resources*, pages 3–25.

Leek, J. T. (2014). svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Research*, 42(21):00–a–0.

Leek, J. T. and Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics*, 3(9):e161.

Li, Y., Willer, C., Sanna, S., and Abecasis, G. (2009). Genotype Imputation. *Annual review of genomics and human genetics*, 10:387–406.

Li, Y., Willer, C. J., Ding, J., Scheet, P., and Abecasis, G. R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, 34(8):816–34.

Lillycrop, K. A., Slater-Jefferies, J. L., Hanson, M. A., Godfrey, K. M., Jackson, A. A., and Burdge, G. C. (2007). Induction of altered epigenetic regulation of the hepatic glucocorticoid receptor in the offspring of rats fed a protein-restricted diet during pregnancy suggests that reduced dna methyltransferase-1 expression is involved in impaired dna methylation and changes in histone modifications. *Br J Nutr*, 97(6):1064–73.

Linhoff, M. W., Lauren, J., Cassidy, R. M., Dobie, F. A., Takahashi, H., Nygaard, H. B., Airaksinen, M. S., Strittmatter, S. M., and Craig, A. M. (2009). An Unbiased Expression Screen for Synaptogenic Proteins Identifies the LRRTM Protein Family as Synaptic Organizers. *Neuron*, 61(5):734–749.

Liu, E. Y., Zhang, Q., McMillan, L., Pardo-Manuel de Villena, F., and Wang, W. (2010). Efficient genome ancestry inference in complex pedigrees with inbreeding. *Bioinformatics*, 26(12):i199–i207.

MacNeil, L. T. and Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, 21(5):645–657.

Manly, B. F. (2006). *Randomization, bootstrap and Monte Carlo methods in biology*, volume 70. CRC Press.

Marchini, J. and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nature Reviews Genetics*, 11(7):499–511.

Marchini, J., Howie, B. N., Myers, S., McVean, G., and Donnelly, P. (2007). A new multipoint method for genome-wide association studies by imputation of genotypes. *Nature genetics*, 39(7):906–13.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., and Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, 26(16):2069–2070.

McNamara, G. I. and Isles, A. R. (2013). Dosage-sensitivity of imprinted genes expressed in the brain: 15q11q13 and neuropsychiatric illness. *Biochemical Society Transactions*, 41(3):721–726.

Menon, S. and Gupton, S. L. (2016). Building Blocks of Functioning Brain: Cytoskeletal Dynamics in Neuronal Development. *International Review of Cell and Molecular Biology*, 322:183–245.

Meuwissen, T. H., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829.

Miller, B. H., Schultz, L. E., Gulati, A., Su, A. I., and Pletcher, M. T. (2010). Phenotypic Characterization of a Genetically Diverse Panel of Mice for Behavioral Despair and Anxiety. *PLOS ONE*, 5(12):e14458.

Morgan, A. P., Fu, C.-P., Kao, C.-Y., Welsh, C. E., Didion, J. P., Yadgary, L., Hyacinth, L., Ferris, M. T., Bell, T. A., Miller, D. R., Giusti-Rodriguez, P., Nonneman, R. J., Cook, K. D., Whitmire, J. K., Gralinski, L. E., Keller, M., Attie, A. D., Churchill, G. A., Petkov, P., Sullivan, P. F., Brennan, J. R., McMillan, L., and Pardo-Manuel de Villena, F. (2015). The Mouse Universal Genotyping Array: From Substrains to Subspecies. *G3 (Bethesda, Md.)*, 6(2):263–279.

Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., and de Villena, F. P.-M. (2017). Structural variation shapes the landscape of recombination in mouse. *Genetics*, 206(2):603–619.

Morgan, A. P. and Welsh, C. E. (2015). Informatics resources for the Collaborative Cross and related mouse populations. *Mammalian Genome*, 26(9-10):521–539.

Mortensen, O., Lodberg Olsen, H., Frandsen, L., Eigil Nielsen, P., Nielsen, F., Grunnet, N., and Quistorff, B. (2009). *Gestational Protein Restriction in Mice Has Pronounced Effects on Gene Expression in Newborn Offspring's Liver and Skeletal Muscle; Protective Effect of Taurine*, volume 67.

Mosedale, M., Kim, Y., Brock, W. J., Roth, S. E., Wiltshire, T., Scott Eaddy, J., Keele, G. R., Corty, R. W., Xie, Y., Valdar, W., and Watkins, P. B. (2017). Candidate Risk Factors and Mechanisms for Tolvaptan-Induced Liver Injury Are Identified Using a Collaborative Cross Approach. *Toxicological Sciences: An Official Journal of the Society of Toxicology*.

Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., and Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proceedings of the National Academy of Sciences of the United States of America*, 97(23):12649–12654.

Moy, S. S., Nadler, J. J., Young, N. B., Perez, A., Holloway, L. P., Barbaro, R. P., Barbaro, J. R., Wilson, L. M., Threadgill, D. W., Lauder, J. M., Magnuson, T. R., and Crawley, J. N. (2007). Mouse behavioral tasks relevant to autism: phenotypes of 10 inbred strains. *Behavioural Brain Research*, 176(1):4–20.

Moy, S. S., Nikolova, V. D., Riddick, N. V., Baker, L. K., and Koller, B. H. (2012). Preweaning Sensorimotor Deficits and Adolescent Hypersociability in Grin1 Knockdown Mice. *Developmental Neuroscience*, 34(2-3):159–173.

Muller, H. (1942). Isolating mechanisms, evolution, and temperature. In *Biol. Symp.*, volume 6, pages 71–125.

Naber, P. D. D. and Lambert, M. (2009). The CATIE and CUtLASS Studies in Schizophrenia. *CNS Drugs*, 23(8):649–659.

Niculescu, M. D., Craciunescu, C. N., and Zeisel, S. H. (2006). Dietary choline deficiency alters global and gene-specific DNA methylation in the developing hippocampus of mouse fetal brains. *The FASEB Journal*, 20(1):43–49.

Oreper, D., Cai, Y., Tarantino, L. M., de Villena, F. P.-M., and Valdar, W. (2017a). Github repository, https://github.com/danoreper/ISVdb.git.

Oreper, D., Cai, Y., Tarantino, L. M., de Villena, F. P.-M., and Valdar, W. (2017b). Zenodo repository, https://doi.org/10.5281/zenodo.399474.

Oreper, D., Cai, Y., Tarantino, L. M., Villena, F. P.-M. d., and Valdar, W. (2017c). Inbred Strain Variant Database (ISVdb): A Repository for Probabilistically Informed Sequence Differences Among the Collaborative Cross Strains and Their Founders. *G3: Genes, Genomes, Genetics*, 7(6):1623–1630.

Oreper, D., Schoenrock, S., McMullan, R. C., Ervin, R., Farrington, J., Miller, D. R., de Villena, F. P.-M., Valdar, W., and Tarantino, L. M. (2018). *Reciprocal F1 hybrids of two inbred mouse strains reveal parent-of-origin and perinatal diet effects on behavior and expression*.

Palmer, A. A., Brown, A. S., Keegan, D., Siska, L. D., Susser, E., Rotrosen, J., and Butler, P. D. (2008). Prenatal protein deprivation alters dopamine-mediated behaviors and dopaminergic and glutamatergic receptor binding. *Brain Res*, 1237:62–74.

Peripato, A. and Cheverud, J. (2002). Genetic Influences on Maternal Care. *The American Naturalist*, 160(S6):S173–S185.

Pfaff, B. and McNeil, A. (2012). *evir: Extreme Values in R*.

Phillippi, J., Xie, Y., Miller, D. R., Bell, T. A., Zhang, Z., Lenarcic, A. B., Aylor, D. L., Krovi, S. H., Threadgill, D. W., Pardo-Manuel de Villena, F., Wang, W., Valdar, W., and Frelinger, J. A. (2014). Using the emerging Collaborative Cross to probe the immune system. *Genes and Immunity*, 15(1):38–46.

Piazza, P. V., Deminiere, J.-M., Le Moal, M., and Simon, H. (1989). Factors that predict individual vulnerability to amphetamine self-administration. *Science*, 245(4925):1511–1513.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2016). nlme: Linear and Nonlinear Mixed Effects Models. Manual, R. R package version 3.1-124.

Plummer, M. (2003). *JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling*.

Plummer, M. (2016). *rjags: Bayesian Graphical Models using MCMC*.

Prickett, A. R. and Oakey, R. J. (2012). A survey of tissue-specific genomic imprinting in mammals. *Molecular Genetics and Genomics*, 287(8):621–630.

Putterman, S. (1998). Urinary Odour preferences in mice. *Physics world*, 11(5):38.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramasamy, A., Trabzuni, D., Gibbs, J. R., Dillman, A., Hernandez, D. G., Arepalli, S., Walker, R., Smith, C., Ilori, G. P., Shabalin, A. A., Li, Y., Singleton, A. B., Cookson, M. R., for NABEC, Hardy, J., for UKBEC, Ryten, M., and Weale, M. E. (2013). Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies. *Nucleic Acids Research*, 41(7):e88–e88.

Rasmussen, A. L., Okumura, A., Ferris, M. T., Green, R., Feldmann, F., Kelly, S. M., Scott, D. P., Safronetz, D., Haddock, E., LaCasse, R., Thomas, M. J., Sova, P., Carter, V. S., Weiss, J. M., Miller, D. R., Shaw, G. D., Korth, M. J., Heise, M. T., Baric, R. S., de Villena, F. P.-M., Feldmann, H., and Katze, M. G. (2014). Host genetic diversity enables Ebola hemorrhagic fever pathogenesis and resistance. *Science*, 346(6212):987–991.

Risso, D., Ngai, J., Speed, T. P., and Dudoit, S. (2014). Normalization of RNA-seq data using factor analysis of control genes or samples. *Nature Biotechnology*, 32(9):896–902.

Robertson, K. D. (2005). DNA methylation and human disease. *Nature Reviews Genetics*, 6(8):597–610.

Rottner, K., Faix, J., Bogdan, S., Linder, S., and Kerkhoff, E. (2017). Actin assembly mechanisms at a glance. *Journal of Cell Science*, 130(20):3427–3435.

Sachs, G. S., Nierenberg, A. A., Calabrese, J. R., Marangell, L. B., Wisniewski, S. R., Gyulai, L., Friedman, E. S., Bowden, C. L., Fossey, M. D., Ostacher, M. J., and others (2007). Effectiveness of adjunctive antidepressant treatment for bipolar depression. *New England Journal of Medicine*, 356(17):1711–1722.

Sakia, R. M. (1992). The Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 41(2):169–178.

Sargolzaei, M., Chesnais, J. P., and Schenkel, F. S. (2014). A new approach for efficient genotype imputation using information from relatives. *BMC Genomics*, 15(1):478.

Scheet, P. and Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American journal of human genetics*, 78(4):629–44.

Schoenrock, S. A., Oreper, D., Farrington, J., McMullan, R. C., Ervin, R., Miller, D. R., Pardo-Manuel de Villena, F., Valdar, W., and Tarantino, L. M. (2017). Perinatal nutrition interacts with genetic background to alter behavior in a parent-of-origin-dependent manner in adult Collaborative Cross mice. *Genes, Brain, and Behavior*.

Schoenrock, S. A., Oreper, D., Young, N., Ervin, R. B., Bogue, M. A., Valdar, W., and Tarantino, L. M. (2016). Ovariectomy results in inbred strain-specific increases in anxiety-like behavior in mice. *Physiology & Behavior*, 167:404–412.

Schultz, B. M., Gallicio, G. A., Cesaroni, M., Lupey, L. N., and Engel, N. (2015). Enhancers compete with a long non-coding RNA for regulation of the Kcnq1 domain. *Nucleic Acids Research*, 43(2):745–759.

Shapiro, S. S. and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3/4):591–611.

Shorter, J. R., Odet, F., Aylor, D. L., Pan, W., Kao, C.-Y., Fu, C.-P., Morgan, A. P., Greenstein, S., Bell, T. A., Stevans, A. M., and others (2017). Male infertility is responsible for nearly half of the extinction observed in the mouse Collaborative Cross. *Genetics*, 206(2):557–572.

Srivastava, A., Morgan, A. P., Najarian, M. L., Sarsani, V. K., Sigmon, J. S., Shorter, J. R., Kashfeen, A., McMullan, R. C., Williams, L. H., and Giusti-Rodrguez, P. (2017). Genomes of the mouse collaborative cross. *Genetics*, 206(2):537–556.

Sultzer, D. L., Davis, S. M., Tariot, P. N., Dagerman, K. S., Lebowitz, B. D., Lyketsos, C. G., Rosenheck, R. A., Hsiao, J. K., Lieberman, J. A., and Schneider, L. S. (2008). Clinical Symptom Responses to Atypical Antipsychotic Medications in Alzheimers Disease: Phase 1 Outcomes From the CATIE-AD Effectiveness Trial. *American Journal of Psychiatry*, 165(7):844–854.

Sun, W., Lee, S., Zhabotynsky, V., Zou, F., Wright, F. A., Crowley, J. J., Yun, Z., Buus, R. J., Miller, D. R., Wang, J., McMillan, L., Villena, F. P.-M. d., and Sullivan, P. F. (2012). Transcriptome Atlases of Mouse Brain Reveals Differential Expression Across Brain Regions and Genetic Backgrounds. *G3: Genes|Genomes|Genetics*, 2(2):203–211.

Tian, F., Bradbury, P. J., Brown, P. J., Hung, H., Sun, Q., Flint-Garcia, S., Rocheford, T. R., McMullen, M. D., Holland, J. B., and Buckler, E. S. (2011). Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nature Genetics*, 43(2):159–162.

Tobi, E. W., Lumey, L. H., Talens, R. P., Kremer, D., Putter, H., Stein, A. D., Slagboom, P. E., and Heijmans, B. T. (2009). DNA methylation differences after exposure to prenatal famine are common and timing- and sex-specific. *Human Molecular Genetics*, 18(21):4046–4053.

Tong, Z.-B., Gold, L., Pfeifer, K. E., Dorward, H., Lee, E., Bondy, C. A., Dean, J., and Nelson, L. M. (2000). Mater, a maternal effect gene required for early embryonic development in mice. *Nature Genetics; New York*, 26(3):267–8.

Turner, K. M., Young, J. W., McGrath, J. J., Eyles, D. W., and Burne, T. H. (2013). Cognitive performance and response inhibition in developmentally vitamin d (dvd)-deficient rats. *Behav Brain Res*, 242:47–53.

Ubeda, F. and Wilkins, J. F. (2008). Imprinted genes and human disease: an evolutionary perspective. In *Genomic imprinting*, pages 101–115. Springer.

Van der Waerden, B. (1952). Order tests for the two-sample problem and their power. In *Indagationes Mathematicae (Proceedings)*, volume 55, pages 453–458. Elsevier.

Vance, K. W. and Ponting, C. P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. *Trends in Genetics*, 30(8):348–355.

Vanderzalm, P. J., Pandey, A., Hurwitz, M. E., Bloom, L., Horvitz, H. R., and Garriga, G. (2009). C. elegans CARMIL negatively regulates UNC-73/Trio function during neuronal development. *Development (Cambridge, England)*, 136(7):1201–1210.

VanRaden, P. M., O'Connell, J. R., Wiggans, G. R., and Weigel, K. A. (2011). Genomic evaluations with many more genotypes. *Genetics Selection Evolution*, 43(1):10.

Verbyla, A. P., George, A. W., Cavanagh, C. R., and Verbyla, K. L. (2014). Wholegenome QTL analysis for MAGIC. *Theoretical and Applied Genetics*, 127(8):1753–1770.

Vision, T. J., Brown, D. G., Shmoys, D. B., Durrett, R. T., and Tanksley, S. D. (2000). Selective mapping: a strategy for optimizing the construction of high-density linkage maps. *Genetics*, 155(1):407–420.

Vrana, P. B. (2007). Genomic Imprinting as a Mechanism of Reproductive Isolation in Mammals. *Journal of Mammalogy*, 88(1):5–23.

Vrana, P. B., Fossella, J. A., Matteson, P., del Rio, T., O'Neill, M. J., and Tilghman, S. M. (2000). Genetic and epigenetic incompatibilities underlie hybrid dysgenesis in Peromyscus. *Nature genetics*, 25(1):120–124.

Vucetic, Z., Totoki, K., Schoch, H., Whitaker, K. W., Hill-Smith, T., Lucki, I., and Reyes, T. M. (2010). Early life protein restriction alters dopamine circuitry. *Neuroscience*, 168(2):359–70.

Wahlsten, D., Bachmanov, A., Finn, D. A., and Crabbe, J. C. (2006). Stability of inbred mouse strain differences in behavior and brain size between laboratories and across decades. *Proceedings of the national academy of sciences*, 103(44):16364–16369.

Wang, J. R., de Villena, F. P.-M., and McMillan, L. (2012). Comparative analysis and visualization of multiple collinear genomes. In *BMC bioinformatics*, volume 13, page S13. BioMed Central.

Wang, L. P. and Preacher, K. J. (2015). Moderated Mediation Analysis Using Bayesian Methods. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(2):249–263.

Waterland, R. A. and Jirtle, R. L. (2003). Transposable Elements: Targets for Early Nutritional Effects on Epigenetic Gene Regulation. *Molecular and Cellular Biology*, 23(15):5293–5300.

Welsh, C. E., Miller, D. R., Manly, K. F., Wang, J., McMillan, L., Morahan, G., Mott, R., Iraqi, F. A., Threadgill, D. W., and de Villena, F. P.-M. (2012). Status and access to the Collaborative Cross population. *Mammalian Genome: Official Journal of the International Mammalian Genome Society*, 23(9-10):706–712.

Wilkinson, L. S., Davies, W., and Isles, A. R. (2007). Genomic imprinting effects on brain development and function. *Nature Reviews Neuroscience*, 8(11):832–843.

Wolf, J. B., Hager, R., and Cheverud, J. M. (2008). Genomic imprinting effects on complex traits: A phenotype-based perspective. *Epigenetics*, 3(6):295–299.

Wolf, J. B., Oakey, R. J., and Feil, R. (2014). Imprinted gene expression in hybrids: perturbed mechanisms and evolutionary implications. *Heredity*, 113(2):167–175.

Xu, Z., Zou, F., and Vision, T. J. (2005). Improving Quantitative Trait Loci Mapping Resolution in Experimental Crosses by the Use of Genotypically Selected Samples. *Genetics*, 170(1):401–408.

Yalcin, B., Flint, J., and Mott, R. (2005). Using progenitor strain information to identify quantitative trait nucleotides in outbred mice. *Genetics*, 171(2):673–681.

Yang, H., Wang, J. R., Didion, J. P., Buus, R. J., Bell, T. A., Welsh, C. E., Bonhomme, F., Yu, A. H.-T., Nachman, M. W., Pialek, J., Tucker, P., Boursot, P., McMillan, L., Churchill, G. A., and de Villena, F. P.-M. (2011). Subspecific origin and haplotype diversity in the laboratory mouse. *Nature Genetics*, 43(7):648–655.

Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H., and Qu, L.-H. (2013). ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Research*, 41(Database issue):D177–D187.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girn, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R., and Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, 44(D1):D710–D716.

Yuan, Y. and MacKinnon, D. P. (2009). Bayesian Mediation Analysis. *Psychological methods*, 14(4):301–322.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B. D., Shen, Y., Pervouchine, D. D., Djebali, S., Thurman, B., Kaul, R., Rynes, E., Kirilusha, A., Marinov, G. K., Williams, B. A., Trout, D., Amrhein, H., Fisher-Aylor, K., Antoshechkin, I., DeSalvo, G., See, L.-H., Fastuca, M., Drenkow, J., Zaleski, C., Dobin, A., Prieto, P., Lagarde, J., Bussotti, G., Tanzer, A., Denas, O., Li, K., Bender, M. A., Zhang, M., Byron, R., Groudine, M. T., McCleary, D., Pham, L., Ye, Z., Kuan, S., Edsall, L., Wu, Y.-C., Rasmussen, M. D., Bansal, M. S., Keller, C. A., Morrissey, C. S., Mishra, T., Jain, D., Dogan, N., Harris, R. S., Cayting, P., Kawli, T., Boyle, A. P., Euskirchen, G., Kundaje, A., Lin, S., Lin, Y., Jansen, C., Malladi, V. S., Cline, M. S., Erickson, D. T., Kirkup, V. M., Learned, K., Sloan, C. A., Rosenbloom, K. R., de Sousa, B. L., Beal, K., Pignatelli, M., Flicek, P., Lian, J., Kahveci, T., Lee, D., Kent, W. J., Santos, M. R., Herrero, J., Notredame, C., Johnson, A., Vong, S., Lee, K., Bates, D., Neri, F., Diegel, M., Canfield, T., Sabo, P. J., Wilken, M. S., Reh, T. A., Giste, E., Shafer, A., Kutyavin, T., Haugen, E., Dunn, D., Reynolds, A. P., Neph, S., Humbert, R., Hansen, R. S., De Bruijn, M., Selleri, L., Rudensky, A., Josefowicz, S., Samstein, R., Eichler, E. E., Orkin, S. H., Levasseur, D., Papayannopoulou, T., Chang, K.-H., Skoultchi, A., Gosh, S., Disteche, C., Treuting, P., Wang, Y., Weiss, M. J., Blobel, G. A., Good, P. J., Lowdon, R. F., Adams, L. B., Zhou, X.-Q., Pazin, M. J., Feingold, E. A., Wold, B., Taylor, J., Kellis, M., Mortazavi, A., Weissman, S. M., Stamatoyannopoulos, J., Snyder, M. P., Guigo, R., Gingeras, T. R., Gilbert, D. M., Hardison, R. C., Beer, M. A., and Ren, B. (2014). A Comparative Encyclopedia of DNA Elements in the Mouse Genome. *Nature*, 515(7527):355–364.

Zhang, Z., Wang, W., and Valdar, W. (2014). Bayesian Modeling of Haplotype Effects in Multiparent Populations. *Genetics*, 198(1):139–156.

Zheng, C., Boer, M. P., and van Eeuwijk, F. A. (2015). Reconstruction of genome ancestry blocks in multiparental populations. *Genetics*, 200(4):1073–1087.

Zheng, J., Li, Y., Abecasis, G. R., and Scheet, P. (2011). A comparison of approaches to account for uncertainty in analysis of imputed genotypes. *Genetic Epidemiology*, 35(2):102–110.