Winter 12-15-2021

# Genetic Effects Mediated Through Epistatic Networks Onto Metabolic Traits

Juan Francisco Macias-Velasco
*Washington University in St. Louis*

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Computational and Systems Biology

Dissertation Examination Committee:
Heather A. Lawson, Chair
Ting Wang, Co-Chair
Barak Cohen
Brian Finck
Michael Provence

# Genetic Effects Mediated Through Epistatic Networks Onto Metabolic Traits

by

Juan F. Macias-Velasco

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2022
St. Louis, Missouri

# Table of Contents

# List of Figures

# Acknowledgments

I am the fortunate beneficiary of tremendous support. I do not have the space to recognize every person who has contributed to my life. Many doors have been opened for me across the years. I have had so many educators/mentors invest their heart and soul into my future, to them I owe success.

I would like to thank my mentor and friend Dr. Heather Lawson. You have been a constant companion and guide through all the many trials both in and out of lab. You focused and honed my scientific mind. You gave me a space to explore and try hairbrained ideas. I would like to thank the members of the Lawson lab: Alex, Caryn, Jessica, Heather S., and Celine. Without Alex, Caryn, and Brian I could not have made the transition to living away from friends and family without them. To Jessica and Heather S., thank you for supporting me in lab and keeping me sane through experiments. To Celine, thank you for the many conversations and for sharing your feedback on my many hairbrained ideas, but especially I want to thank you for teaching me the power of well thought out Figures and aesthetics.

To my friends in Saint Louis, you added so much value to my life. You gave me a place and a family. I regret the circumstances of my last two years in Saint Louis prevented me from seeing many of you, but we will always have the many parties, cookouts, gaming nights, and nights out dancing. You made my life vibrant and fulfilling.

I would like to thank my family. Your faith and support have lifted me up and made possible much that I could not have done alone. To my mother above all I am grateful. You taught me to always find a way. To see the world with an open and vigilant mind. To be whatever I choose to be.

Lastly, I would like to thank my thesis committee for your guidance, support, and career advice. My work was supported by the Washington University Department of Genetics, the Diabetes Research Center at Washington University, the NIH NIDDK, the NIEHS, and the NIH NHGRI (grant T32-GM007067).

Juan F. Macias-Velasco

*Washington University in St. Louis*

*May 2022*

τ δ ι μ

ABSTRACT OF THE DISSERTATION

Parent-of-Origin Effects Mediated Through Epistatic Networks Onto Metabolic Traits

for Arts & Sciences Graduate Students

by

Juan F. Macias-Velasco

Doctor of Philosophy in Biology and Biomedical Sciences

Computational and Systems Biology

Washington University in St. Louis, 2022

Professor Heather A. Lawson, Chair

Professor Ting Wang, Co-Chair

Predicting variation in complex traits from DNA sequence is a major public health goal, but our understanding of the genotype-to-phenotype relationship is incomplete. It will remain so unless we can adequately integrate genetic, epigenetic, and environmental information into a systems level framework. In a step towards that goal, quantitative trait mapping studies have attempted to account for environmental factors such as sex and diet, and epigenetic factors such as allelic parent-of-origin effects. Several studies used an advanced intercross of the LG/J and SM/J inbred mouse strains to unravel the genetic architecture of multiple metabolic traits. These studies found that parent-of-origin effects are surprisingly common, and that they can be mediated by the environment. This indicates that our ability to predict variation in metabolic phenotypes from genotype alone will be confounded unless environment and allelic parent-of-origin is considered. In this thesis, I explore mechanisms that could explain the prevalence of parent-of-origin effects on metabolic variation and develop tools to predict these effects more generally.

I test the hypothesis that genetic effects can propagate through molecular pathways. I propose that non-imprinted genes can generate complex parent-of-origin effects on metabolic traits through interactions with imprinted genes. I employ data from mouse populations at different levels of intercrossing ($F_0$, $F_1$, $F_2$, $F_{16}$) of the LG/J and SM/J inbred mouse strains to test this hypothesis. Using multiple populations and incorporating genetic, genomic, and physiological data, I leverage orthogonal evidence to identify networks of genes through which parent-of-origin effects propagate. I identify a network comprising 3 imprinted and 6 non-imprinted genes that show parent-of-origin effects. This epistatic network forms a nutritional responsive pathway and the genes comprising it jointly serve cellular functions associated with growth.

While epistatic interactions can explain some of the overabundance of parent-of-origin effects on metabolic traits, other biological explanations for these effects should be explored. Towards this end, I developed a generalized framework for modeling how genetic effects on expression might propagate through a network. I demonstrate that this method detects signatures of pathway structure and produces highly interpretable and actionable candidates for subsequent experimentation. This work lays the foundation for future studies of not only parent-of-origin effects but also other genetic phenomena that can extend to human studies.

# Chapter 1: Background and Introduction

## 1.1 What is a Parent-of-Origin-Effect?

Parent-of-Origin effects (POE) are epigenetic phenomena in which genetically identical individuals differing only in allelic parent-of-origin are phenotypically different[1]. At a given autosomal locus, there are four genotypic classes given two possible alleles, L and S {LL, LS, SL, SS}. By convention, alleles are ordered with the maternal allele followed by the paternal allele. An LS genotype indicates the L allele was inherited maternally and the S allele was inherited paternally. For a given trait associated with some locus when the two heterozygote classes {LS, SL} have unequal genotypic means (mean phenotype for a given genotype), there is a parent-of-origin effect.

As geneticists we want to predict phenotype from genotype, but the genotype-to-phenotype relationship is immensely complex. Parent-of-origin effects, which are not usually considered, add a critical layer to this problem. Parent-of-origin effects can manifest as complex phenotypic patterns. We group these patterns into two broad groups: parental expression patterns (**Figure 1.1 A-B**) and dominance patterns (**Figure 1.1 C-E**). Parental expression occurs when progeny take on the phenotypic value of one the parents {paternal expression and maternal expression}. For example, maternal expression occurs when progeny take on the maternal phenotype regardless of the paternal phenotype (**Figure 1.1 A**). Dominance patterns are more complex and come in three forms: polar over-dominance, polar under-dominance, and bipolar dominance. Polar over-dominance occurs when one heterozygous genotypic class has a larger genotypic mean then all other classes, with the other classes having equal phenotypic values (**Figure 1.1. C**). Similarly,

polar under-dominance occurs when one of the heterozygotes has a lower genotypic mean than all other classes (**Figure 1.1. D**). Lastly, bipolar dominance occurs when one heterozygote has a larger genotypic mean than the mean phenotype across classes and the reciprocal heterozygote has a smaller genotypic mean than the mean phenotype across all classes (**Figure 1.1 E**).

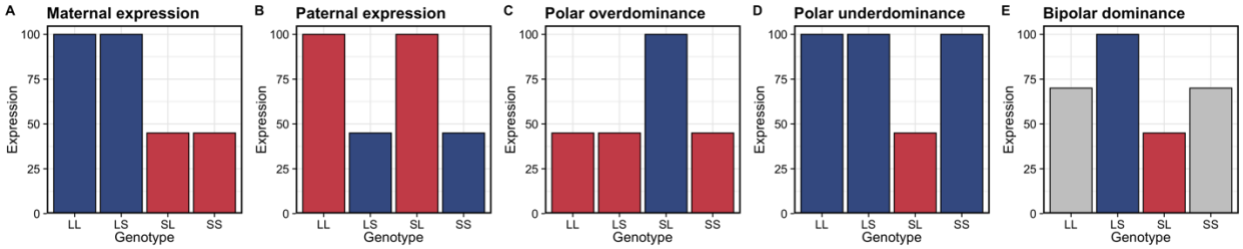


**Figure 1.1. Patterns of parent-of-origin dependent phenotypic (expression) variation.** Four genotypic classes are shown composed of two alleles {L,S}. These phenomena are classified as parental expression (**A & B**) and dominance (**C, D, & E**).

## 1.1.1 Parental Genetic Effects, Random Mono-allelic Expression, and X-inactivation

Other phenomena can produce differences in genotypic means between reciprocal heterozygotes but are not considered true parent-of-origin effects. Parental effects can be defined as the influence of parental phenotype on offspring phenotype [2]. Parental effects can be thought of as a kind of environmental effect where the parental phenotype shapes the individual's 'environment', which influences the offspring's phenotype. For example, a mother's nursing behavior dictates the nutritional environment of the offspring which clearly effects the offspring's phenotype. This is a maternal effect, because the mother shapes the 'environment' of the offspring and alters its phenotype through said 'environment'. Parental effects can be indirect, such as nesting site choice. They can also be direct, such as mRNA pre-loaded into unfertilized eggs [3] or tRNA fragments pre-loaded into sperm [4]. Pre-loaded RNAs shape the early cellular 'environment' which directly alters

an offspring's phenotype. With parental effects, the parent's genotype determines its phenotype which shapes the offspring's environment in turn affecting the offspring's phenotype.

In an individual cell level both alleles are usually expressed, but as transcription is a stochastic process, expression of only one allele occurs at random. This phenomenon is known as random mono-allelic expression (RMAE). If one were to look at a small number of cells, RMAE could lead to detection of a false Parent-of-origin effect. Sufficient tissue sample size is therefore essential to avoiding these artifacts [5].

X-inactivation is a related epigenetic phenomenon in females where one of the X chromosome copies is silenced. This occurs during early development. In placental mammals, which parental copy is silenced is random [6] and is therefore a special case of RMAE and not a Parent-of-origin effect.

### 1.1.2 Known Parent-of-origin effect Machinery

Which parent an allele came from affects the phenotype that allele produces, but what sort of phenotypes are involved and how does that modification occur? The phenotype in question can be at any biological level, including total gene expression, individual allele expression, and gross phenotype. Quantitative genetics tends to deal with gross phenotypes. Molecular genetics tends to deal with the expression of individual alleles, which is the most studied level of Parent-of-origin effect. In an individual cell, both alleles are normally transcribed equally, so there is an approximately 50:50 mixtures of transcripts of the two alleles. This is referred to as biallelic expression. When one allele is more highly expressed than the other you have allele-specific expression (ASE). When an alleles' expression is affected by which parent it came from, leading to unequal expression of alleles within an individual, that is a parent-of-origin effect (for example, 10:90 - ♂:♀). Specifically, this is referred to as parent-of-origin-dependent ASE. This is different

from “ancestry” or sequence-dependent ASE, where alleles are unequally expressed based on the actual identity of the allele (for example, 10:90 - L:S). Genomic imprinting is a special case of ASE where the functional non-equivalence of parental alleles comes from nearly complete silencing of one parent’s allele. Genomic imprinting is by far the best understood mechanism underlying parent-of-origin effects [7]. Genomic imprinting has been observed in therian mammals, plants, insects, and filamentous fungi [8–10], where it is plays an important role in the manifestation of complex traits.

The exact mechanisms of genomic imprinting are complex. Imprinted loci must be individually evaluated, but there are some common trends. Firstly, genomic imprinting tends to occur in clusters. Clusters range in size between 80 kb and 3700 kb in size, containing three to 12 imprinted genes. In mice it is estimated there are between 100 and 600 imprinted genes. To date 219 [11,12] imprinted genes have been identified and mapped to 17 chromosomes, of which 80% are clustered into 16 genomic regions[13,14]. Of those 16 imprinted regions, seven {*Igf2r, KcnqI, Pws, Gnas, Grb10, Igf2, Dlk1*} are well-characterized. The clustering of imprinted genes hints at some shared regulation. Shared genomic features have been found to control genomic imprinting. Such genomic features are called imprinting control elements (ICE). In all the well-characterized cases, ICE are differentially methylated regions (DMR)[15]. So far 26 DMR involved in genomic imprinting have been positively identified. Another commonality of imprinting clusters is that they *always* have one or more long non-coding RNA (lncRNA). Along with other noncoding RNAs, lncRNAs can induce chromatin remodeling[16,17] to repress expression of their “targets”. At such loci, protein-coding genes tend to be expressed from one parental chromosome and the other parental chromosome tends to only the express lncRNAs. Interestingly, when the ICE is deleted from one parental chromosome, imprinting is lost only when the deletion occurs on the parental

chromosome expressing the lncRNA[14,18–21]. This supports the idea that lncRNAs are acting in *cis* to repress expression of protein-coding genes.

Despite how few known imprinted genes and even fewer characterized DMRs there are, a significant excess of parent-of-origin effects have been observed in mapping studies[22–25]. While only about 1% of genes are believed to be imprinted, in these studies 65% of quantitative trait loci (QTL) associated with variation in metabolic traits showed significant parent-of-origin effects in some sex or dietary context. Canonical imprinting mechanisms are not sufficient to explain these phenomena at the phenotypic level.

### 1.1.3 Detecting parent-of-origin effects

There are various ways to detect parent-of-origin effects, depending upon what biological level one is interested in. Methodologies have been developed to evaluate parent-of-origin effects with respect to ASE, total gene expression, gross phenotype, and epigenetic marks/ states. Parent-of-origin effects at the allele specific expression level can be detected by sequencing (next generation, pyrosequencing) [5,11,26–29]. At the gross phenotypic level QTL mapping or some method of variance partitioning is employed[22–25,30–32]. Parent-of-origin effects at the level of epigenetic marks (i.e. DMR) are detected for either methylation patterns or differences in chromatin. Allele-specific methylation is detected using bisulfite sequencing (for example, wgbs, targeted-bs)[33–35].

### 1.1.4 The evolution of parent-of-origin effects

The evolutionary origins for parent-of-origin effects (namely genomic imprinting) are of course unknown, but the best explanation is a combination of the "maternal-offspring co-adaption", "kinship", and "sexual antagonism" theories [11,36–38]. The consensus of these theories is that genomic imprinting evolved to improve specific matri-/patri-lineal fitness. In a system in which resources are limited and multi-paternity litters/communities exist, the management of

nutrient allocation/energy has opposing fitness consequences for parents. A mother's fitness is maximized by having many offspring with a variety of healthy males and for those offspring to share resources amongst themselves. Limited growth, equal distribution of resources, and effective warmth sharing amongst littermates (maternal relatives) improves the mothers fitness. Variants promoting those traits passed along the matriline improve the fitness of maternal relatives. Father's benefit from larger offspring that soak up resources. Since a litter can have a mixture of fathers, a father's fitness is improved if he passes variants that benefit only those in his patriline (including his children). In this way father's benefit from passing on pro-growth variants, if the litter is only made up of his offspring, then no individual soaks up resources more than any other, so nutrients get allocated evenly. If on the other hand, some offspring have a different father with a weaker pro-growth variant, then the offspring with the stronger variant will consume more resources than those with the weaker variant. As a result, any offspring that inherited the stronger variant from their father will have a head start, which improves that father's fitness. In this way there is a conflict between the fathers of a litter. Because the mother's fitness is improved by equal resource sharing (all the offspring are her kin), it creates a conflict of sorts between the mother and the fathers. Mothers must share resources with their offspring (*in utero*) and benefit from sharing amongst the offspring, fathers want their offspring to benefit at the cost of unrelated individuals. In summary: 1) the "maternal-offspring co-adaption" theory asserts that the mother regulates *in utero* growth to balance the nutritional burden she must supply by epigenetically modifying the offspring's genome. 2) The "sexual antagonism" theory asserts that parents have opposed fitness consequences, so what is good for the father is not always good for the mother. 3) The "Kinship" theory somewhat combines these and asserts that the paternal fitness is improved by competition

between the patrilines within a litter and maternal fitness is improved by keeping the benefits of a genetically diverse litter without patriline conflict.

## 1.2 Affected traits

### 1.2.1 Disease

Parent-of-origin effects are associated with a wide range of human diseases generally related to growth/metabolism, neurological function, or some combination of the two.

Growth and metabolism diseases believe to show parent-of-origin effects include: Transient Neonatal Diabetes (ZAC1, HYMAI, ZFP57)[7,39–41], Type-1 Diabetes (DLK1-MEG3 cluster)[42], Type-2 Diabetes (GNAS, KLF14, GRB10)[43,44], various types of Cancer (IGF2/H19, ZAC1)[45,46], Metabolic Syndrome (IGF2, PEG3, DLK1, SLC2A10, KCNK9), Beckwith-Wiedemann Syndrome (IGF2/H19/KCNQ1 cluster), Wilm's tumors (H19/IGF2 cluster)[47], Insulinomas (No known genes)[45], Silver–Russell syndrome (KLF4, GRB10)[48,49], and certain variants of Albright's Hereditary Osteodystrophy (GNAS)[50,51].

Neurological diseases believed to show parent-of-origin effects include: Alzheimers [52–54] (No known genes), Myoclonus-Dystonia Syndrome (SGCE)[55], Jervell and Lange-Nielsen Syndrome (KCNQ1, KCNE1)[56,57].

Diseases which are both metabolic and neurological in nature include: Prader Willi Syndrome (SNRPN, NECDIN, SNORD64, SNORD107, SNORD18, SNORD109, SNORD116, SNORD115) and Angelman Syndrome (UBE3A)[20,58].

### 1.2.2 Metabolic traits

It is well established that parent-of-origin effects play an important role in complex traits. To date 20-60% of phenotypic variance can be explained depending on the trait[22,23,25]. Outside of their role

in fetal development, parent-of-origin effects are also found associated with variation in serum lipid levels, adiposity, and diabetes susceptibility. In a $F_{16}$ advanced intercross of the LG/J and SM/J inbred mice, 52% of metabolic QTL (mQTL) for serum lipids (serum cholesterol, free-fatty acids, and triglycerides) showed parent-of-origin effects in some context (sex, diet, and/ or sex X diet)[23]. This same model showed that 61% of mQTL for obesity traits (reproductive, renal, mesenteric, and inguinal fat, and total depot weights) showed parent-of-origin effects[25]. For diabetes-related traits (glucose tolerance, serum glucose, basal glucose, and serum insulin), 59% of mQTL showed parent-of-origin effects[24].

These traits are involved in metabolic syndrome (MetS), which is the co-occurrence of obesity, dyslipidemia, high blood pressure, and glucose intolerance. MetS is a major risk factor for cardiovascular disease and type-2 diabetes. Significantly, 91% of QTL associated with MetS (pleiotropic QTL for multiple metabolic traits) show parent-of-origin effects[59]. While the exact molecular mechanisms involved are not yet known, some imprinted genes in QTL are strong candidates with functions relevant to the mapped trait.

KLF14 (kruppel-like factor 14) is a maternally expressed transcription factor located in the PEG1/ MEST imprinting cluster (mouse chromosome 6, human chromosome 7)[49]. Large scale GWAS studies have identified SNP variants (rs4731702 and rs972283) upstream of KLF14 as having strong associations with type-2 diabetes and HDL levels[60,61]. These variants have maternally restricted *cis*-regulatory associations with KLF14 expression in adipose tissue[62], implying that they fall in a KLF14 *cis*-regulatory element. Tests for rs4731702 as an expression QTL (eQTL) using human gene expression (Illumina Human HT12 array) have found that rs4731702 *trans*-associations are enriched for low-p-values in subcutaneous white adipose tissue, indicating that KLF14 may be an adipose master transcriptional regulator[43]. Furthermore, the top

46 associated genes for which rs4731702 is an eQTL show a significant enrichment for KLF family transcription factor binding sites. This suggests that KLF14 propagates a parent-of-origin effect in biallelic genes, contributing to variation in expression and subsequently to variation in adipose phenotypes.

Perhaps the best classical example of an imprinting cluster associated with variation in metabolic traits are the IGF2/ H19 gene pair. IGF2 and H19 are reciprocally imprinted, having a functionally antagonistic relationship. The mechanism behind IGF2/ H19 has been identified as a CTCF mediated methylation-sensitive insulator/ enhancer system[63]. IGF2 is upstream of a DMR and H19 is down-stream. Within this system the DMR falls on a CTCF binding site. When un-methylated the DMR can bind CTCF. The bound DMR is then able to interact with downstream enhancers, thereby serving as an insulator and blocking IGF2 transcription, but allowing H19 transcription. Upstream of IGF2 there is another DMR, blocking IGF2 expression. IGF2 is paternally expressed and promotes fetal growth. H19 is maternally expressed and limits fetal growth. This relationship extends to muscle growth and fat deposition[30,64,65]. Surprisingly, IGF2 is negatively correlated with body weight in adult white adipose tissue[66]. The functional switch between time points and tissue remains unknown.

### 1.2.3 Imprinted genes are not enriched in mQTL showing parent-of-origin effects

Metabolic quantitative trait loci (mQTL) showing parent-of-origin effects for multiple metabolic traits were previously identified in an $F_{16}$ advanced intercross between the LG/J and SM/J strains[23–25,59]. Enrichment of known imprinted genes in QTL was measured by permutation tests. Parent-of-origin effect mQTL were randomly shuffled throughout the genome to generate a null distribution of parent-of-origin effect genes in mQTL-sized regions. Diabetes, obesity, and serum

lipids QTL were not enriched for known imprinted genes (p=0.874, 0.143, and 0.404 respectively). This means that the number and location of known imprinted genes alone is not sufficient to explain the number of mQTL showing parent-of-origin effects.

## 1.3 Why are parent-of-origin effects so common?

Quantitative trait mapping has found parent-of-origin effects are observed across a wide range of complex traits. QTL associated with growth, metabolic traits, anxiety, behavior, wound healing, asthma, and immunological function have all shown parent-of-origin effects[22–25,59]. In fact, parent-of-origin effects seem fairly common in complex traits. If genomic imprinting alone does not account for this overabundance, what might? We consider three explanations for this overabundance: 1) Weak and context-specific parent-of-origin dependent ASE is much more common than previously considered; 2) Genomic imprinting evolved at genes along pleiotropic environmental response pathways; and/or 3) Random parental inheritance biases in mapping populations lead to propagated additive genetic effects appearing as parent-of-origin effects.

### 1.3.1 Context-specific and weak ASE

Genomic imprinting produces a strong parent-of-origin dependent ASE (example, a 1:99 allelic bias), but weaker forms of ASE (example, a 30:70 allelic bias) exist. It could be the case that some of these weaker forms contribute to the overall overabundance of observed parent-of-origin effects. Furthermore, sex and environmental contexts are not usually considered when searching for signatures of genomic imprinting. Many QTL showing a significant parent-of-origin effect did so under specific sex and dietary contexts. So perhaps weak context-specific parent-of-origin dependent ASE could explain this overabundance. While weak and context-specific effects are more prevalent than previously believed[67], I do not think this alone could account for the overabundance of parent-of-origin effects. Only 1% of genes show genomic imprinting compared

to 65% of mQTL show significant parent-of-origin effects. There must be some other mechanisms underlying these effects.

### 1.3.2 Imprinting genetic effects propagate through epistasis

Complex traits are multi-locus by nature. The complexity of parent-of-origin effect patterns on complex traits suggests it's a multi-locus mechanism. It is believed that as a class, imprinted genes are highly interactive with each other. Three indirect lines of evidence have been used to argue this point: 1) they cluster together in the genome; 2) they are significantly co-expressed[68] with each other; and 3) the strength of their co-expression has increased over evolutionary time (since the appearance of parent-of-origin effects)[11,36,69,70].

Through interactions, complexity manifests. For example, imprinted gene interactions is the proposed mechanisms for dominance patterns of imprinting. The callipyge trait (buttock hypertrophy first described in sheep) may be caused by two reciprocally imprinted genes interacting, producing the polar over-dominance effect at this locus[71,72] (**Figure 1.2**). However, this working model has not been experimentally validated.

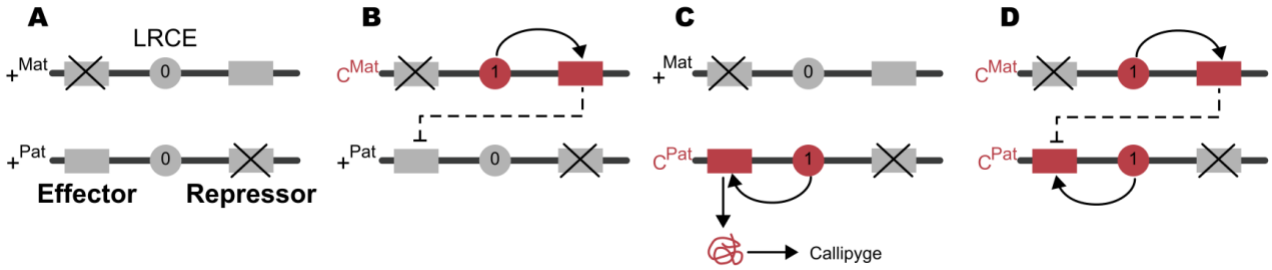


**Figure 1.2. Working model for polar overdominance at callipgye locus.** The locus is composed of two imprinted genes, one of which is maternally expressed, and the other being paternally expressed. Genes repressed by imprinting are denoted by a cross through them. Red denotes activity and grey denotes inactivity. The status of genotypic classes is shown Wt/Wt (**A**), CLPG/Wt (**B**), Wt/CLPG (**C**), CLPG/CLPG (**D**). There are three functional elements: a long-range control element (LRCE), a paternally expressed pro-growth effector, and a maternally expressed *trans*-acting anti-growth repressor. The LCRE is labeled 1 meaning it is active or 0 meaning it is

inactive. In this model the LCRE affects genes in *cis*, but the repressor acts in *trans*. The interaction of imprinting, LCRE *cis* effects, and maternal repressor *trans* effects, means that the paternal pro-growth effector is only active in Wt/CLPG individuals (**C**), this generating a polar overdominance effect.

Imprinted genes and biallelic gene interactions may also underlie some of these patterns. For example, there is evidence that the maternally expressed transcription factor KLF14 regulates the expression of bi-allelic adipose genes[43]. Such interactions might produce complex parent-of-origin effect patterns, but experimental validation has not been performed. Unless imprinted genes are incredibly insular, interactions with biallelic genes are a given, but exactly how many non-imprinted genes interact with imprinted genes? How many are co-expressed with imprinted genes? Under what contexts do interactions occur? Outside of a handful of examples, the question of which non-imprinted genes are interacting with imprinted genes remains largely unanswered.

This explanation has two assertions. The first was discussed above. It says in essence that genetic effects propagate (Imprinted gene → biallelic gene). By propagation, I mean genes interacting in such a way that the effects of epi-/genotype on expression of a "Source" gene affects expression of some "Target" gene.

The second assertion is that pathways in which imprinting has evolved are pleiotropic and affect a wide range of nutrient/energy management growth, metabolism, and thermogenesis are intrinsically tied processes and share pathways. From an evolutionary perspective these processes shape how individuals respond to their nutritional environment; that response in turn affects other mice in the community/litter and can alter the fitness of others. Parent-of-origin effects evolved to maximize the overall fitness of individuals in the same matriline or patriline (section 1.1.3). A thrifty way for evolution to introduce a parental fitness bias through growth would be to introduce genomic imprinting pathways that regulates many aspects of growth in response to nutritional

environment. For example, altering mesenchymal stem cell function in a parent-of-origin dependent way would alter growth, metabolism, and thermogenesis. It could affect many cell types with vastly different functions, but nonetheless contribute to growth. If this explanation is correct, it could explain how pleiotropic mQTL are enriched for parent-of-origin effects.

### 1.3.3 Additive genetic effects propagate and create false positives

Mouse mapping studies are seldom replicated. It would be immensely expensive, and a huge number of animals would need to be generated for the purpose. Justifying replication is further undermined by the difficulty in interpreting what a lack of replication would mean. The complexity of genetic architecture, epistasis, and limited statistical power can mean even a true biological association won't replicate.

Genetic effects propagate in general. My thesis focuses on parent-of-origin effects propagating along epistatic pathways. At a minimum this is a two-locus problem. An imprinted gene manifests a parent-of-origin effect, which alters downstream gene expression in its pathway. When the two loci are sufficiently linked, the parent-of-origin effect on the imprinted locus can be detected at the downstream "Target" locus. How does that detection work? In a mapping study detecting a parent-of-origin effect boils down to comparing the genotypic means of reciprocal heterozygote classes {LS, SL}. A significant difference means there is a parent-of-origin effect[73].

Consider the locus being tested as a "Target" locus downstream (regulated by) of some "Source". In a mapping study we would compare the "Target" genotypic means. But the "Target" genotypic means are a function of the "Source" genotypic means. If the two loci are sufficiently independent, the effect of the "Source" locus should appear as unaccounted for variance when testing the "Target" locus. However, under a two-locus model it becomes quite difficult to ensure groups are sufficiently randomized (**Figure 1.3 A&B**).

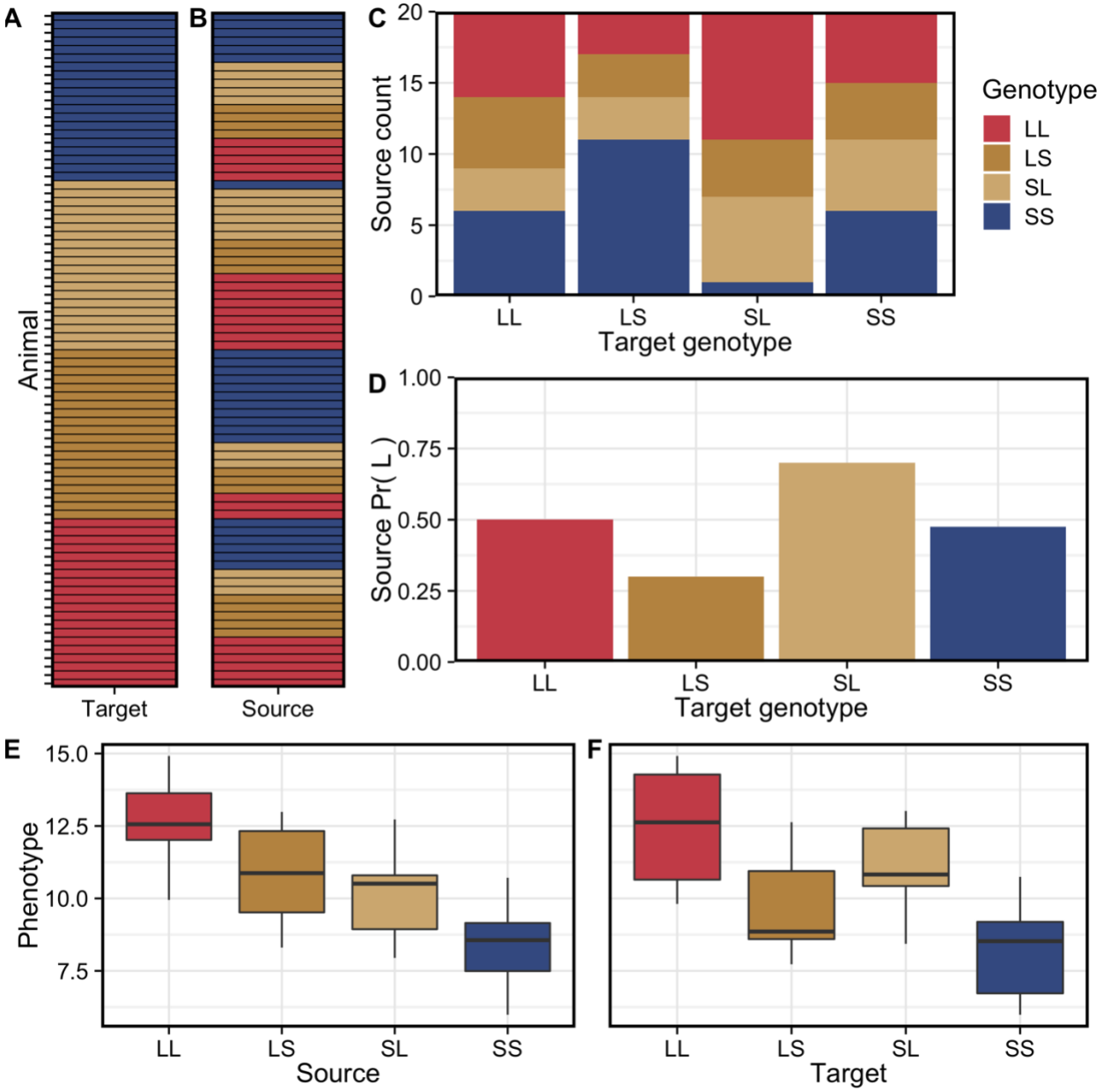


**Figure 1.3. Simulated parental inheritance bias and how they can create false parent-of-origin signals in mapping.** In this example the mapped trait is controlled by two loci in a pathway where the gene at the "Source" locus regulates the gene at the "Target" locus. The two loci are not linked. Individual genotypes for the target locus are displayed as a heatmap in part **A**. Where LL individuals are colored in red, SS individuals are colored in blue, and heterozygotes (LS, SL) are colors in dark and light tan respectively. Corresponding genotypes at the upstream "Source" locus are shown in part **B**. Because these genes function as a pathway, the effect of the Target genotype on the phenotype depends in part on the Source genotype. In part **C** we show the composition between the two loci. The X-axis denotes the genotype at the "Target" locus. The Y-axis shows the composition of genotypes at the "Source" locus for a give "Target" genotype. We calculate the allele frequency at the "Source" locus given a "Target" genotype and display them as part **D**. The

X-axis is the "Target" genotype, and the Y-axis is the corresponding probability of the L allele. In this case individuals with the LS genotype at the mapped "Target" locus are more likely to inherit L alleles at the "Source" locus relative to SL individuals (**D**). In this instance the upstream "Source" locus was encoded to show an additive genetic effect (**E**), where inheriting the L allele at the "Source" locus elicits a larger trait value. The "Target" locus was also encoded to show an additive genetic effect (**F**), where the L allele elicits a larger trait value. It is important to understand that there is no true parent-of-origin mechanism (i.e. genomic imprinting) present at the "Target" locus. However, when mapping at the "Target" locus, a parent-of-origin genetic effect is observed. For this to happen two conditions had to be met: 1) there must be a true additive genetic effect on the upstream "Source" locus which gets propagated to the downstream "Target" locus and 2) there must be a parental inheritance bias on the "Source" locus with respect to the "Target" locus.

The more genotypic classes one deals with, the harder it becomes to ensure sufficient randomization. As a result, mapping parent-of-origin effects can suffer from this technical limitation. When we map parent-of-origin effects, we compare LS and SL genotypic means at the "Target" locus. But what happens if the allele frequencies at the associated "Source" locus is different between the LS and SL populations (**Figure 1.3 C**)? It is possible to have a case where LS ("Target") individuals are more likely to have L alleles at the "Source" locus than SL ("Target") individuals are (**Figure 1.3 D**). We call this a parental inheritance bias. In this case you are more likely to have L alleles at the "Source" locus given you inherited the L allele at the "Target" locus maternally. This can occur randomly or as a result of pedigree substructure rather than by biological mechanism. Consider if the L allele at the "Source" locus gives elevated expression of the "Target" gene relative to the S allele (additive genetic effect) (**Figure 1.3 E**). If one of the "Target" heterozygous genotypic classes is more likely to have an L allele at the upstream "Source" locus, that class would have a higher genotypic mean. The interaction of an additive genetic effect on the "Source" gene with a parental inheritance bias at the "Target" gene would produce what appears as a parent-of-origin effect on the "Target" gene (**Figure 1.3 F**). Unlike the epistatic interaction explanation which is constrained by recombination between the

“Target”/”Source” pair in a mapping study, this additive propagation false positive is not constrained by genomic distance. There is no easy way to control for this possibility but being able to model genetic effect networks in general is an important first step.

## 1.4 Summary of Aims and Motivation

The broad goal of my work was to explore possible mechanisms driving the overabundance of parent-of-origin effects in mouse mapping studies. I focused on imprinted genes in epistatic pathways, but also laid methodological groundwork for future studies. By understanding mechanisms underlying observed complex phenotypic patterns, we advance our ability to predict phenotypic outcomes and ultimately better understand metabolic disease. My aims were to identify and evaluate the role of candidate gene networks underlying the manifestation of parent-of-origin effects at mQTL in mice.

In chapter two, I sought to identify instances in which interactions between imprinted and non-imprinted genes contribute to these complex phenotypic patterns. Identifying instances of these interactions is a proof of principle. Interrogating the associated mechanisms of candidate gene pairs can shed light on the sorts of mechanisms that mediate these phenomena. Stepping back and considering the commonalities of gene pairs, evolutionary theories, and the collective knowledge of the field, we use the results to speak to the general biology of parent-of-origin effects on complex metabolic traits.

This study was a heuristic approach to find candidate cases wherein a specific genetic effect from one gene propagates onto a downstream gene. Chapter three was motivated by the need to consider other types of genetic effects and to make the approach generally accessible. In this study,

I sought to formalize the process into a general framework for pairwise genetic effect propagation at the level of cell types.

My dissertation was motivated by the desire to find potential mechanisms that could shed light on the complex patterns we observe in mapping studies. I was further motivated by potential future exploration of these complex phenomena and a need to integrate molecular biology and quantitative genetics.

# Chapter 2: Parent-of-origin effects propagate through networks to shape metabolic traits

This chapter corresponds to a manuscript, which has been reviewed and revised for publication in the journal eLife.

J**uan F. Macias-Velasco**, Celine L. St Pierre, Jessica P. Wayhart, Li Yin, Larry Spears, Mario A. Miranda, Caryn Carson, Katsuhiko Funai, James M. Cheverud, Clay F. Semenkovich, Heather A. Lawson. Parent-of-origin effects propagate through networks to shape metabolic traits. *eLife*; *under review*

## 2.1 Abstract

Parent-of-origin effects are unexpectedly common in complex traits, including metabolic and neurological diseases. Parent-of-origin effects can be modified by the environment, but the architecture of these gene-by-environmental effects on phenotypes remains to be unraveled. Previously, quantitative trait loci (QTL) showing context-specific parent-of-origin effects on metabolic traits were mapped in the $F_{16}$ generation of an advanced intercross between LG/J and SM/J inbred mice. However, these QTL were not enriched for known imprinted genes, suggesting

another mechanism is needed to explain these parent-of-origin effects phenomena. We propose that non-imprinted genes can generate complex parent-of-origin effects on metabolic traits through interactions with imprinted genes. Here, we employ data from mouse populations at different levels of intercrossing ($F_0$, $F_1$, $F_2$, $F_{16}$) of the LG/J and SM/J inbred mouse lines to test this hypothesis. Using multiple populations and incorporating genetic, genomic, and physiological data, we leverage orthogonal evidence to identify networks of genes through which parent-of-origin effects propagate. We identify a network comprised of 3 imprinted and 6 non-imprinted genes that show parent-of-origin effects. This epistatic network forms a nutritional responsive pathway and the genes comprising it jointly serve cellular functions associated with growth. We focus on 2 genes, *Nnat* and *F2r*, whose interaction associates with serum glucose levels across generations in high fat-fed females. Single-cell RNAseq reveals that *Nnat* expression increases and *F2r* expression decreases in pre-adipocytes along an adipogenic trajectory, a result that is consistent with our observations in bulk white adipose tissue.

## 2.2 Introduction

Parent-of-origin effects, where the phenotypic effect of an allele depends on whether the allele is inherited maternally or paternally, are epigenetic phenomena associated with a wide range of complex traits and diseases [74]. Thus, the functional impact of a specific genetic variant can depend on its parental origin. The best characterized parent-of-origin effect is genomic imprinting, an epigenetic process in which either the maternally or paternally inherited allele is silenced, typically through DNA methylation. In humans there are 107 verified imprinted genes and in mice there are 124, of which ~70% overlap [75]. Despite the rarity of imprinted genes, parent-of-origin effects on complex traits and diseases are relatively common, suggesting that canonical imprinting mechanisms are not sufficient to account for these phenomena [76,77]. With so few imprinted genes,

what mechanisms underlie these parent-of-origin effects? We hypothesize that a small number of imprinted genes can generate a large number of parent-of-origin effects through interactions with non-imprinted genes.

In this study, we use four populations at different levels of intercrossing of the LG/J and SM/J inbred mouse lines to test the hypothesis that non-imprinted genes can contribute to parent-of-origin effects on metabolic phenotypes through epistatic interactions with imprinted genes. Multiple populations ($F_0$, $F_1$, $F_2$, $F_{16}$) allow us to refine our search space and provide orthogonal evidence supporting putative networks of interacting genes. Metabolic traits were previously mapped in a $F_{16}$ generation of an advanced intercross between LG/J and SM/J [23–25,59]. We generated visceral white adipose tissue gene expression profiles from 20 week-old $F_1$ animals in order to match the age of the $F_{16}$ LG/J x SM/J advanced intercross population. $F_1$ reciprocal cross (LxS and SxL) mice were subjected to the same high and low-fat diets and phenotyping protocols as the previously-studied $F_{16}$ mice to keep environmental contexts consistent. We identified genes showing parent-of-origin-dependent allele-specific expression (ASE), characterized interactions among these genes and biallelic genes that are differentially expressed by reciprocal cross (DE), and correlated interacting ASE and DE gene pairs with metabolic phenotypes in the $F_1$ population. Pairs that significantly associated with phenotypic variation were tested for epistasis on correlated traits in the $F_{16}$ population.

We identify an epistatic network that forms a nutritional environment responsive pathway mediated through calcium signaling. This network contributes to metabolic variation by balancing proliferation, differentiation, and apoptosis in adipocytes. The genes comprising this network jointly serve functions associated with growth in multiple tissues, which is consistent with the evolutionary hypothesis that sexual conflict underlies some parent-of-origin effects [69]. We focus

on two key interacting genes: *Nnat* (neuronatin), a canonically imprinted gene, and *F2r* (coagulation factor II receptor), a biallelic gene showing significant DE by cross in $F_1$ high fat-fed female animals. Co-expression of these two genes associates with variation in basal glucose levels, and this association persists across generations. Further, single-cell RNAseq reveals that *Nnat* expression increases and *F2r* expression decreases in pre-adipocytes along an adipogenic trajectory, a pattern consistent both with their expression in bulk white adipose tissue and with their respective roles in adipogenesis. Our results demonstrate that incorporating orthogonal lines of evidence including genotype, allele specific expression, total gene expression, single-cell expression, and phenotype from different populations varying in their degree of intercrossing is a powerful way to identify putative mechanisms and test hypotheses underlying parent-of-origin effects on phenotype.

## 2.3 Results

### 2.3.1 Non-imprinted genes interact with imprinted genes and effect metabolic phenotypes

We test the hypothesis that non-imprinted genes can mediate complex parent-of-origin effects on phenotypes through genetic interactions with imprinted genes using a $F_1$ reciprocal cross model of the LG/J and SM/J inbred mice (LxS and SxL). In this model the effects of parental origin on an allele can be tested directly and isolated from sequence dependent *cis*-regulatory differences. We validated our findings in LG/J and SM/J parentals ($F_0$) as well as in $F_2$ and $F_{16}$ intercrosses of LGxSM (**Figure 2.1**). The parental $F_0$ animals serve to anchor variation in allele-specific expression that is a function of allelic identity (L or S). Incorporating the $F_2$ and $F_{16}$ populations into our validations ensures that the interactions we observe are not solely a function of linkage in the $F_1$ animals. We generated mRNA expression profiles in white adipose tissue from 20-week-

old $F_1$ reciprocal cross animals. These animals were subjected to the same high and low-fat diets and phenotyping protocols as the previously studied $F_{16}$ animals [23–25,59,67,78,79]. We identified two classes of genes: 1) imprinted genes, and 2) non-imprinted genes with parent-of-origin effects on total expression.

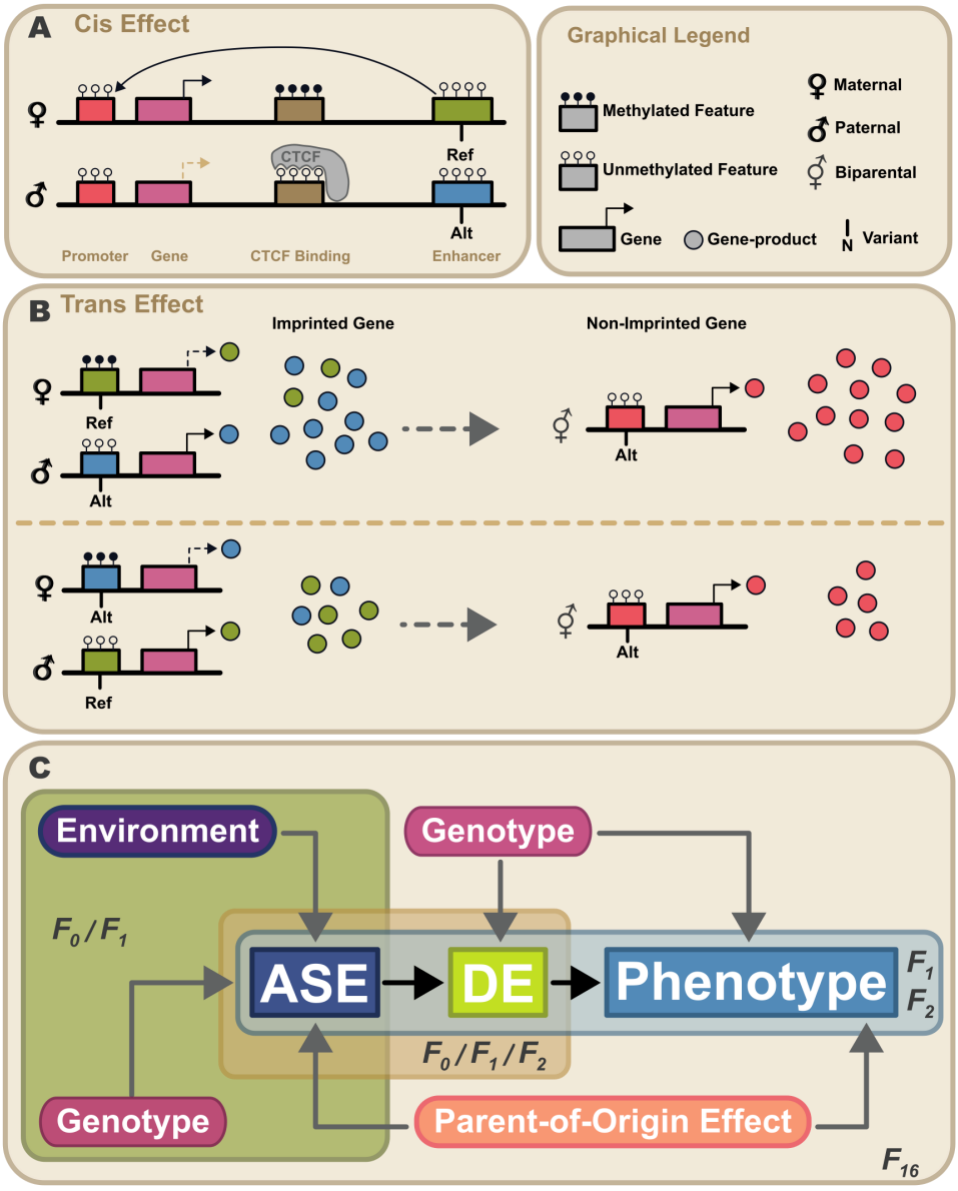


**Figure 2.1: Proposed model for propagation of parent-of-origin effects through gene-gene interactions.** Parent-of-origin effects should be partitioned into *cis* mechanisms and *trans* mechanisms **A**. An example of a *cis* parent-of-origin effect is a system with three regulatory elements: promoter, insulator, and enhancer. Activation of transcription requires the enhancer to

act upon the promoter. Enhancer activity is blocked by the insulator when it has been bound by CTCF. CTCF cannot be bound when methylated. In this system, the insulator is selectively methylated when inherited maternally, so methylation of the maternally inherited insulator blocks CTCF binding, allowing the enhancer to activate transcription. Because the paternally inherited insulator is not methylated, it is bound by CTCF which blocks enhancer activity, silencing transcription. This canonical genomic imprinting mechanism interacts with genetic variation in the three regulatory features. For example, if one allele produces stronger enhancer activity (Alt) than the other, individuals inheriting the Alt allele maternally would have elevated expression compared to those that inherit the same allele paternally. These *cis* genetic effects do not occur in isolation. Due to the highly interconnected nature of biological systems, there are downstream effects. We refer to these as *trans* parent-of-origin effects. **B**. An example of a *trans* parent-of-origin effect is a system with two genes each having its own promoter. The first gene is canonically imprinted, and the activity of the gene promoter is blocked by DNA methylation. The imprinted gene's promoter is methylated when inherited maternally. Consequently, the paternally inherited allele is almost exclusively expressed. As before, when genetic variation in a regulatory feature interacts with these epigenetic mechanisms, we see parent-of-origin effects on expression of the imprinted gene. In this example the imprinted gene regulates expression of a non-imprinted gene. Despite the non-imprinted gene being agnostic to parental origin, its expression nonetheless depends on the parental origin of alleles at the imprinted locus. **C**. Summary of our experimental design. Expression patterns of genes showing allele specific expression (ASE) such as imprinted genes are shaped by parental genotypes and environment (e.g. nutrition). Downstream gene expression is a function of their genotype and the expression of upstream ASE genes. Altered parent-of-origin dependent total gene expression of ASE genes leads to differential expression of downstream genes varying only in allelic parent-of-origin (DE). Phenotype is most directly affected by expression of DE genes. Variation in DE gene expression leads to corresponding variation in phenotype. Mouse populations used to probe parts of this model are labeled $F_0$ (inbred lines), $F_1$ (reciprocal cross of inbred lines), $F_2$ (intercross of $F_1$ mice), and $F_{16}$ (advanced intercross of inbred lines).

To test our model, we identified genes showing parent-of-origin dependent allele specific expression (ASE). We identified 23 genes showing significant ASE (**Figure 2.2 A**; Supplemental Table 2.1). Of these 23 genes, 17 are canonically imprinted genes, two are not reported as imprinted genes but are located in known imprinted domains, and four are novel. Next, we identified genes showing differential total expression between individuals varying only in allelic parent-of-origin (DE between reciprocal crosses, SxL vs LxS). We identified 33 genes that are

significantly DE in at least one sex or dietary context (**Figure 2.2 A**; Supplemental Table 2.2). A larger set of genes show signatures of parent-of-origin effects at the total gene expression level, but do not meet the statistical rigor demanded by the massive multiple tests burden incurred by a genome-wide scan accounting for sex, diet, and parent-of-origin (see Methods).

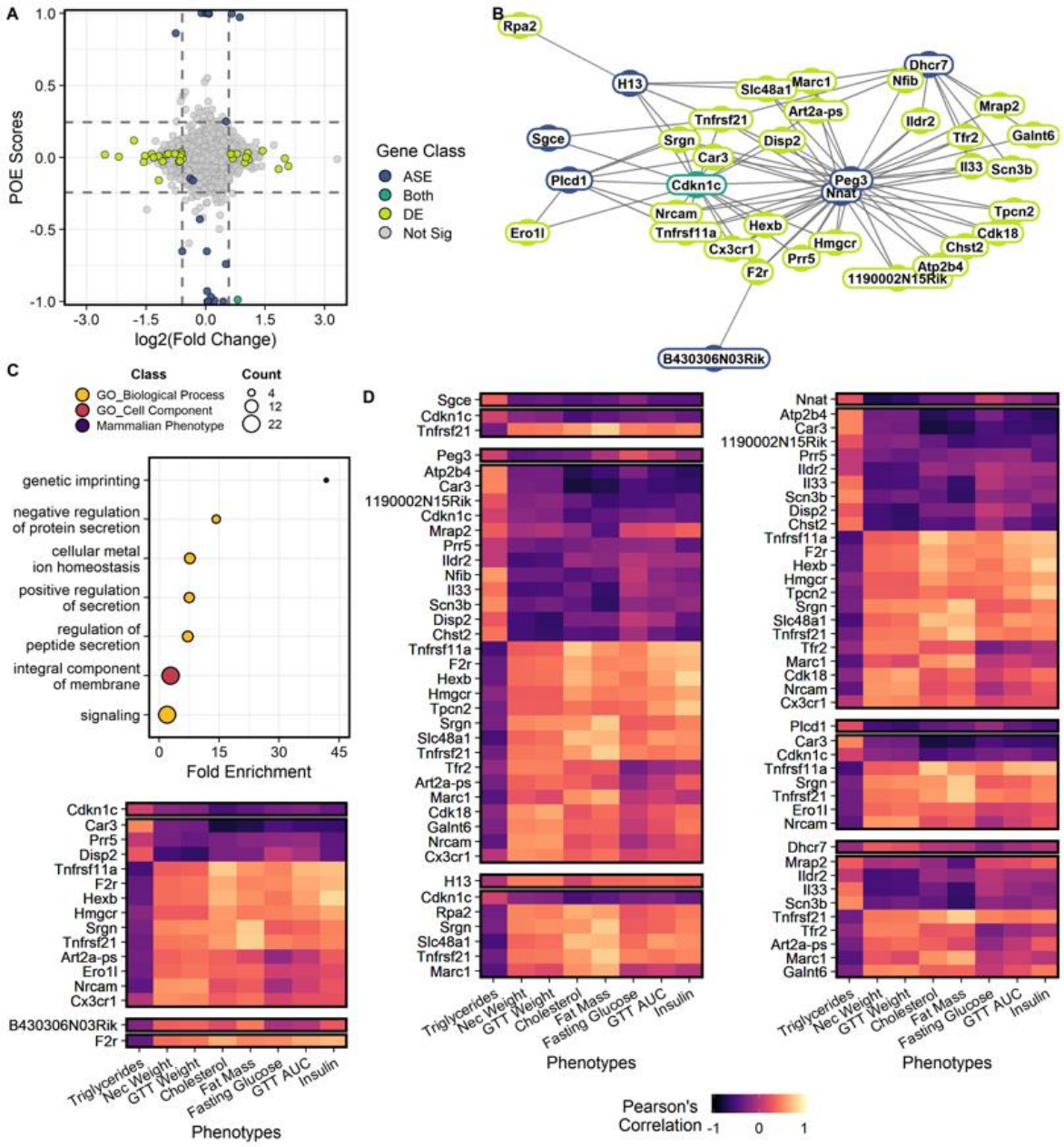


**Figure 2.2: Genes showing parent-of-origin effects at the allele specific and/or total expression levels covary with each other and with metabolic traits. A**. Mean POE score across

contexts. Effect size of ASE is calculated as the mean allelic bias (L / L+S) of SxL animals minus LxS animals. Effect size of DE is measured by $\log_2$(Fold Change) between LxS and SxL crosses. The single context with largest magnitude fold change is plotted for each gene (n=8). Dashed lines represent minimum acceptable effect size cut-offs within a context. Genes showing significant ASE and sufficiently large POE score are shown in blue. Genes showing significant DE and sufficiently large fold change in some sex or dietary context are shown in lime. Genes showing both ASE and DE are shown in teal. Genes not meeting cut-offs are shown in grey. The two genes showing significant ASE but falling short of POE score requirements are a case of context dependent bipolar POE scores (i.e. paternally expressed in one context and maternally expressed in its opposite). **B**. Parent-of-origin effect network constructed from ASE and DE gene pairs (n=32). **C**. Significantly overrepresented ontologies after multiple tests correction in parent-of-origin effect network. Terms are color coded by ontology domain. GO biological process (yellow), GO cellular component (orange), and Mammalian phenotype (purple). Circle size denotes the number of genes with each term. **D**. Correlation of parent-of-origin effect network genes with metabolic traits (n=32). Only genes and phenotypes with at least one significant correlation after multiple test corrections are shown. The heatmap is broken up into subnetworks with the ASE gene as the first separated row followed by associated DE genes in subsequent rows. Columns correspond to metabolic traits. Coloration of each cell denotes the Pearson's correlation coefficient value.

To identify interactions between gene sets, we constructed a network comprised of genes that could initiate a parent-of-origin effect on phenotype (ASE) and genes that may mediate the effect onto phenotype (DE). Interacting gene pairs were predicted by modeling the expression of biallelic genes that are significantly DE by reciprocal cross as a function of the expression of genes showing significant parent-of-origin-dependent ASE, their allelic bias (Lbias), diet, sex, and the diet-by-sex interaction. Genes showing parent-of-origin effects form a highly interconnected network comprised of 52 genes forming 217 gene pairs (**Figure 2.2 B**)(Supplemental Table 2.3). Most of these interactions are *trans*-chromosomal. We identified two genes that could serve as initiation points of propagating parent-of-origin effects through this network. These two genes, *Nnat* (neuronatin) and *Cdkn1c* (cyclin dependent kinase inhibitor 1C), are both canonically imprinted and differentially expressed by cross (Supplemental Table 2.1).

Functional over representation analysis was performed and seven terms were significantly overrepresented at an $FDR \leq 0.05$ (**Figure 2.2 C**) [80]. Enriched terms suggest this network plays a role in signaling and genetic imprinting (Supplemental Table 2.4). In order to identify which phenotypes might be affected by genes in this network, gene expression was correlated with metabolic phenotypes collected for the $F_1$ animals (**Figure 2.2 D**). Seventy-four ASE/DE/phenotype sets were identified as candidates for subsequent testing (Supplemental Table 2.5).

### 2.3.2 Epistasis in an $F_{16}$ advanced intercross identifies a diet-responsive network affecting adipogenesis

To validate the interactions, we identified in $F_1$ animals, we tested for imprinting-by-imprinting epistasis in an $F_{16}$ population. Imprinting-by-imprinting epistasis occurs when the parent-of-origin effect at a locus is dependent on the parent-of-origin of alleles at another locus. This allowed us to determine if the effect of parent-of-origin at DE genotype on phenotype is dependent upon the parent-of-origin at ASE genotype. This orthogonal approach allows us to connect genotype at these loci to phenotype as predicted in the $F_1$ candidates. Nine epistatic interactions replicated in the $F_{16}$ population ($FDR \leq 0.1$; **Figure 2.3 A**; Supplemental Table 2.6). These interactions were comprised of three ASE genes showing parent-of-origin (*Cdknlc*, *Nnat*, *Plcd1*), six genes that are DE by cross (*Car3*, *F2r*, *Hexb*, *Hmger*, *Srgn*, *Tnfrsf11a*) and four phenotypes (basal glucose level, AUC calculated from a glucose tolerance test, serum cholesterol, necropsy weight). Together, these 9 genes form a putative diet-responsive network affecting adipogenesis (**Figure 2.3 B**).

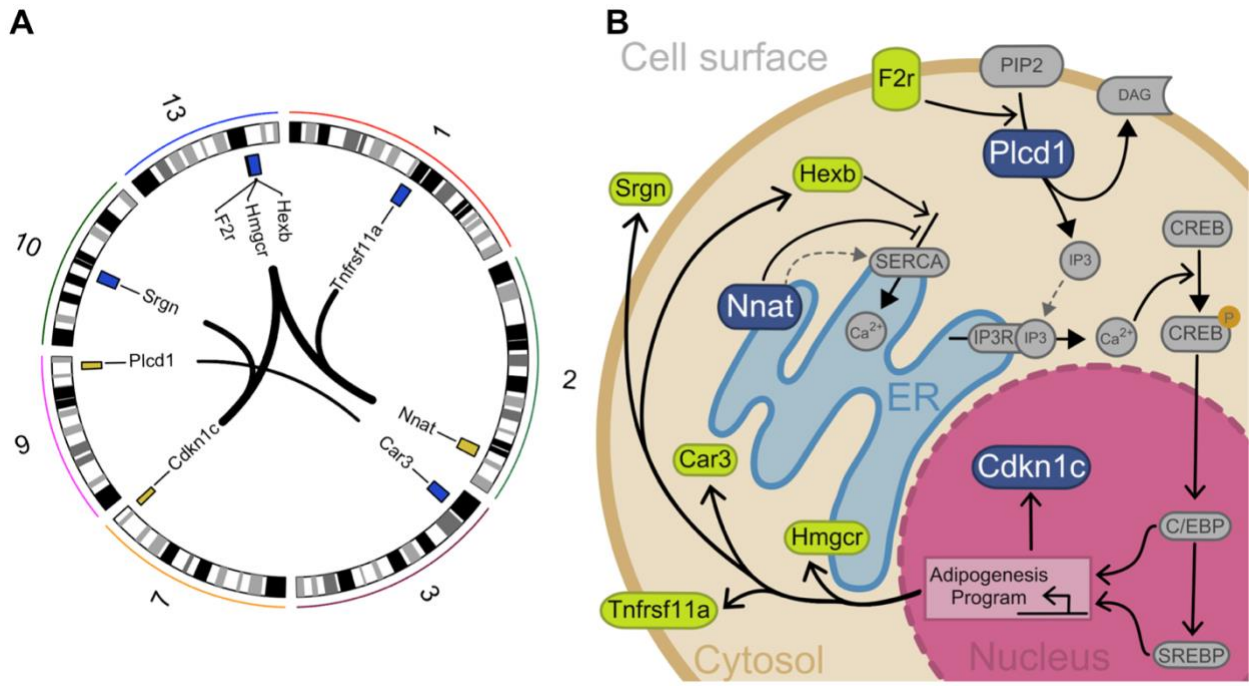


**Figure 2.3: Candidate epistatic network.** There are nine significant imprinting-by-imprinting epistatic ASE/DE/phenotype sets in the $F_{16}$ advanced intercross population (**A**). Interactions are shown as lines connecting ASE (yellow) and DE genes (purple). Chromosome number is shown around the plot. The epistatic parent-of-origin effect network is comprised of key steps in a putative pathway regulating differentiation and survival of adipocytes (**B**). This pathway was constructed by incorporating previously published cellular functions. The pathway members are color coded in blue for ASE genes (*Plcd1*, *Nnat*, and *Cdkn1c*) and green for DE genes (*F2r*, *Hexb*, *Hmgcr*, *Car3*, *Tnfrsf11a*, and *Srgn*). The network breaks down into potentiation, transduction, and response. *Nnat* and *Hexb* potentiate signaling by managing availability and accumulation of calcium necessary for signal transduction. Once a signal is received, *F2r* and *Plcd1* transduce it by activating second messengers to initiate a response. This response initiates an adipogenesis cellular program that affects expression of *Cdkn1c*, *Hmgcr*, *Car3*, *Tnfrsf11a*, and *Srgn*.

The network can be broken down into signal potentiation, transduction, and response. *Nnat* (neuronatin) and *Hexb* (beta-hexosaminidase subunit beta) fall into the potentiation group. These genes play a role in managing the availability and accumulation of calcium necessary for signal transduction. *Nnat* is a paternally expressed canonically imprinted gene which encodes a proteolipid protein that localizes to the ER [81]. *Nnat* is diet-responsive and its overexpression in

3T3L1 pre-adipocytes promotes adipogenesis through increased free cytosolic calcium [82]. In pre-neural stem cells, *Nnat* binds sarco/endoplasmic reticulum $Ca^{2+}$-ATPase (SERCA) to block $Ca^{2+}$ uptake into the ER thereby increasing cytosolic $Ca^{2+}$ levels [83]. In addition to *Nnat*, *Hexb* regulates the uptake and accumulation of $Ca^{2+}$ in the ER via SERCA (Pelled et al. 2003). Upon the arrival of a signal, *F2r* (coagulation factor II receptor) and *Plcd1* (1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-1) in the transduction group initiate the adipogenesis cellular program. *F2r* is a G-protein-bound receptor that promotes phosphoinositide hydrolysis [84]. Variation in the human F2R gene is associated with obesity [85]. G-protein coupled receptors transmit external signals into the cell where they are then propagated by second messenger systems, one of which is mediated by *Plcd1* [86,87]. The downstream effect of PLCD1-mediated signaling is the efflux of calcium into the cytosol from the ER, thereby increasing cytosolic $Ca^{2+}$ levels [88,89]. Increased cytosolic $Ca^{2+}$ in pre-adipocytes promotes phosphorylation of cAMP-response element-binding protein (CREB), which promotes activity of CCAAT/enhancer-binding protein (C/EBP) transcription factors, activating adipogenesis, altering the expression of *Cdkn1c* (cyclin dependent kinase inhibitor 1C), *Hmgcr* (3-hydroxy-3-methylglutaryl-CoA reductase), *Car3* (carbonic anhydrase 3), *Tnfrsf11a* (TNF receptor superfamily member 11a), and *Srgn* (serglycin).

*Cdkn1c* is a canonically imprinted maternally expressed gene that inhibits cell proliferation [90]. Increased expression of *Cdkn1c* is protective against diet-induced obesity in mice [91], and in humans increased caloric intake results in decreased CDKN1C expression [92]. *Hmgcr* is the rate limiting enzyme in cholesterol biosynthesis [93,94] and converts HMG-CoA into mevalonate, which is essential for adipocyte survival [95]. *Srgn* is an adipocytokine thought to be part of a feedback loop with TNFα (tumor necrosis factor alpha), mediating paracrine cross-talk between macrophages and adipocytes [96–99]. *Srgn* is known to play a role in osteoblast-mediated bone

mineralization [100], which along with osteoclast-driven bone deconstruction drives bone remodeling [101]. Osteoblasts share a lineage with adipocytes, and the quantity of osteoblasts is inversely proportional to that of marrow adipose tissue [102–108]. TNFRSF11A is a cell surface protein that regulates differentiation of osteoclasts [109]. Osteoprotegerin (OPG) is a decoy receptor for TNFRSF11A thereby inhibiting osteoclastogenesis and bone resorption [110]. OPG is expressed during differentiation of 3T3L1 adipocytes [111]. Expression of OPG is induced by TNFα in 3T3L1 adipocytes and is associated with obesity in humans [112–114].

The exact function of OPG/TNFRSF11A outside of osteoclastogenesis is unknown, but the function of osteoclasts is to break down bone tissue during bone resorption. Bone resorption regulates the level of blood calcium. The bioavailability of calcium in the blood potentially alters ER calcium stores, creating cross-talk between bone cells and white adipose tissue calcium signaling. Osteoclasts break down bone by acidifying mineralized bone, orchestrated by osteoblasts that have become embedded in the matrix they produce (osteocytes). Oxidative stress on osteocytes from the bone acidification process is prevented by *Car3*. *Car3* is an enzyme that catalyzes the conversion of carbonic acid to $CO_2$ and water. Its expression in white adipose is negatively correlated with, and responsive to, long term obesity in mice and humans [115,116]. *Car3* does not protect against diet induced obesity and is not necessary for fatty acid synthesis [117]. As such its exact function in adipocytes is unknown.

### 2.3.3 *Nnat* and *F2r* covary in white adipose tissue and their interaction associates with variation in basal glucose levels across generations

To better understand how these interactions affect phenotype, we focused on the negative correlation of *Nnat* and *F2r* in the above network. In white adipose tissue, *Nnat* expression significantly covaries with *F2r*, a biallelic gene showing significant DE by cross in $F_1$ high fat-fed

females (FDR=0.05). *Nnat* and *F2r* show significant imprinting-by-imprinting epistasis for basal glucose levels in the $F_{16}$ population (FDR=6.00e-16; **Figure 2.4A and B**). To validate gene expression patterns, we combined $F_1$ biological replicates and $F_0$ high fat-fed female animals ($F_1$ n=13 and $F_0$ n=12) and again observe that *F2r* and *Nnat* are each significantly differentially expressed between reciprocal heterozygotes, i.e. by cross (*F2r* p=0.007 and *Nnat* p=0.026; **Figure 2.4C and D**). Further, the co-expression of *Nnat* and *F2r* also persists in the $F_0/F_1$ population (p=3.00e-4; **Figure 2.4 E**).

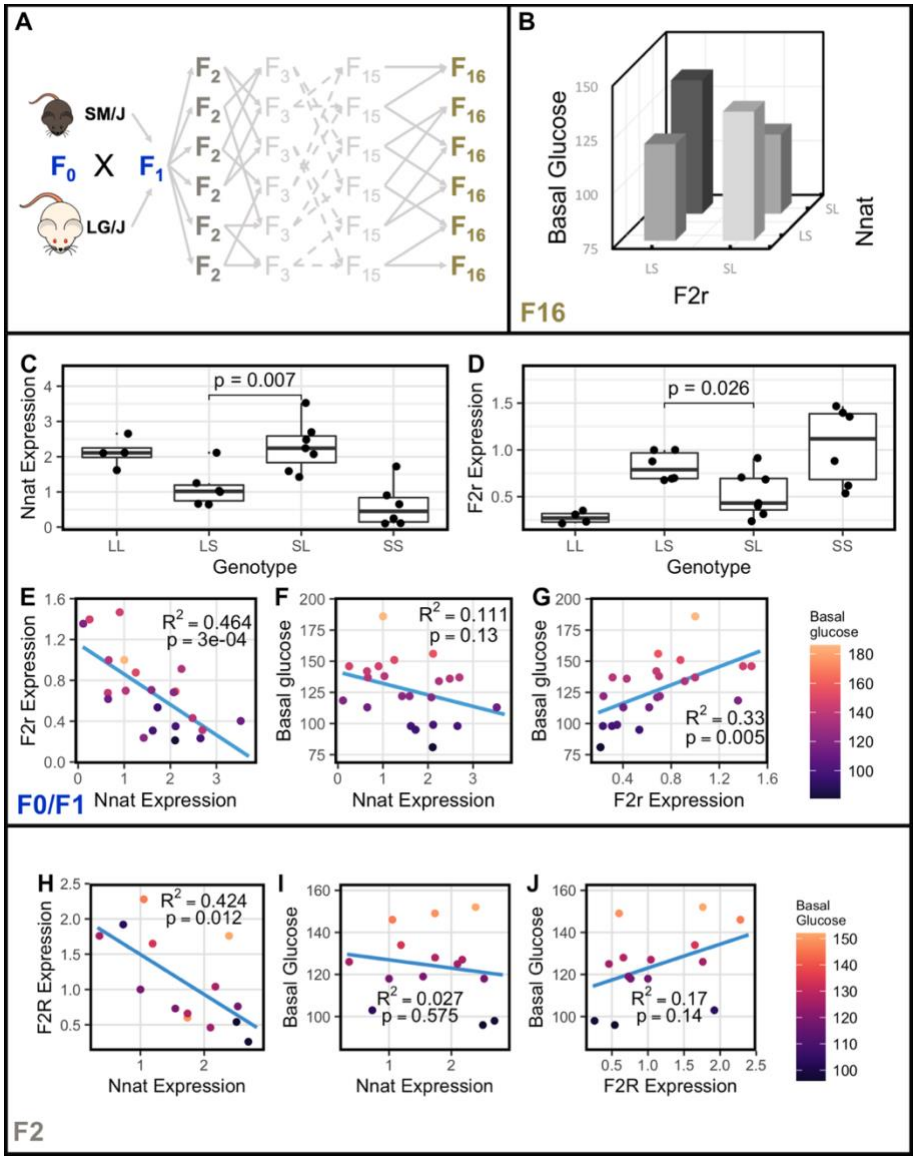

**Figure 2.4. *Nnat* and *F2r* covary across generations. A.** Breeding scheme for the $F_{16}$ Advanced Intercross between the LG/J and SM/J inbred strains. **B.** Significant imprinting-by-imprinting epistasis associated with variation in basal glucose (n=993). The parent-of-origin effects of *F2r* on basal glucose depend on the parent-of-origin effects at *Nnat*. **C.** Expression of *Nnat* across genotypes in a combined $F_0/F_1$ population. **D.** Expression of *F2r* across genotypes in a combined $F_0/F_1$ population. **E.** Significant correlation between *Nnat* and *F2r* expression in the $F_0/F_1$ mice ($F_1$ n=13; $F_0$ n=12). **F and G**. Correlations between basal glucose and *Nnat* and *F2r* in the $F_0/F_1$ mice ($F_1$ n=13; $F_0$ n=12). **H.** Significant correlation between *Nnat* and *F2r* expression in the $F_2$ mice (n=14). **I and J**. Correlations between basal glucose and *Nnat* and *F2r* are not individually significant in the $F_2$ mice. However, the product of *Nnat* and *F2r* expression (*Nnat* x *F2r*) significantly correlates with basal glucose in the $F_2$'s (p=0.045), as predicted by our model of epistasis. Alleles are ordered maternal | paternal within the genotype classes.

A limitation of identifying covariation patterns in $F_1$ and $F_0$ populations is that all loci are linked. This makes it difficult to determine which ASE genes truly co-express with DE genes. While incorporation of orthogonal $F_{16}$ genotypes and phenotypes helps reduce false discoveries, a population with randomized genetic background for which we have expression data is needed to replicate these results. To that end, $F_2$ animals were generated and *Nnat* and *F2r* gene expression levels were measured via qPCR (n=14). We found that *F2r* and *Nnat* are significantly co-expressed in high fat-fed female $F_2$ animals (p= 0.012; **Figure 2.4 H**).

*F2r* expression significantly positively correlates with basal glucose levels in the RNA-sequenced high fat-fed female $F_1$ animals (r=0.514, FDR=0.01; Supplemental Table 2.5). *F2r* expression is also significantly positively correlated with basal glucose in the combined $F_0/F_1$ population (p=0.005; **Figure 2.4 G**). A negative trend between *Nnat* expression and basal glucose level is observed but not statistically significant in the $F_0/F_1$ animals (p=0.130; **Figure 2.4 F**). Correlation of *F2r*'s and *Nnat*'s individual expression with basal glucose in $F_2$ mice follows the same pattern as in the $F_0/F_1$'s. Bootstrapping to calculate confidence intervals shows that the

correlation differences between $F_0/F_1$ and $F_2$ are not significant (**Figure 2.4 I and J**; Supplemental Figure 2.1). However, the product of *Nnat* and *F2r* expression (*Nnat* x *F2r*) is significantly predictive of basal glucose ($p=0.045$, $R^2=0.29$). This indicates that expression of *Nnat* and *F2r*, as a function of their genotypes and allelic parent-of-origin, are not individually sufficient to explain variation in basal glucose levels. But together they are able to explain a significant amount of phenotypic variation. This is precisely what our epistatic model would predict.

Finally, studying the $F_2$ animals allows us to determine if maternal mitochondrial ancestry contributes significantly to *Nnat* or *F2r* expression or to variation in basal glucose. We find mitochondrial genome identity does not significantly covary with *F2r* expression ($p=0.198$), *Nnat* expression ($p=0.365$), or basal glucose ($p= 0.388$).

### 2.3.4 Single-cell RNAseq reveals that *Nnat* expression increases and *F2r* expression decreases in pre-adipocytes along an adipogenic trajectory

To determine what cell types express *Nnat* and *F2r* and whether the directionality of the *Nnat* imprinted → *F2r* target correlation persists along the adipogenic trajectory, we turned to single-cell RNAseq. We used publicly available scRNAseq data collected from stromal vascular cells isolated from C57BL/6J epididymal adipose tissue [118]. Cell type identity was assigned using previously reported markers for this data set (*Adipoq* = differentiating mesenchymal stem cells; *Pdgfra* = mesenchymal stem cells; *Csf1r* = macrophage; *Cdh5* = vascular endothelial cells; *Acta2* = vascular smooth muscle cells; *Cd2* = B cells) (Supplemental Table 2.7; Supplemental Figure 2.2). The adipogenic trajectory refers to cells transitioning from pre-adipocytes (mesenchymal stem cells) to cells differentiating into adipocytes. Clusters along this trajectory were identified by the opposing expression patterns of *Pdgfra* and *Adipoq* (**Figure 2.5 A-D and I**). We found that *Nnat* expression increases along the trajectory while *F2r* expression decreases (**Figure 2.5 E-F**

**and H**). Further there is a negative association between *Nnat* and *F2r* expression within adipocytes along the trajectory (**Figure 2.5 G**). This pattern is consistent with the negative correlation we observe between *Nnat* and *F2r* in the bulk white adipose tissue. Because available scRNAseq data do not match the exact sex/diet/genetic background contexts of the LGxSM mice, there will be unaccounted for differences between the data sets. The observed consistent pattern indicates that the pathway structure persists across sex/diet/genetic backgrounds.

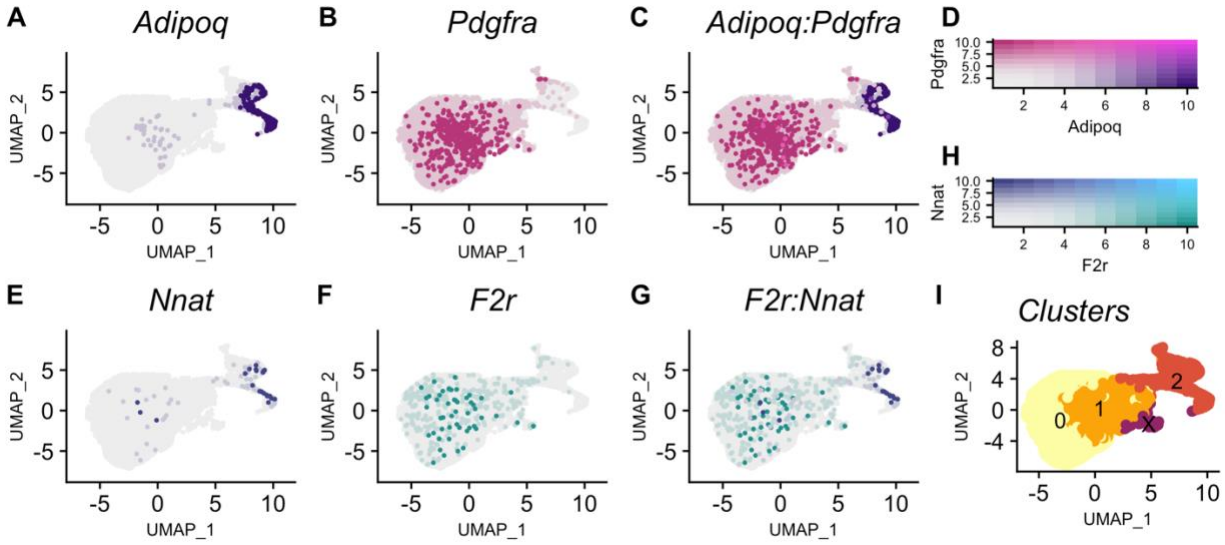


**Figure 2.5: *Nnat* expression increases and *F2r* expression decreases in pre-adipocytes along an adipogenic trajectory. A.** Adipoq is a marker of adipocytes whose expression (purple) increases along the trajectory. **B.** *Pdgfra* is a marker of mesenchymal stem cells whose expression (pink) decreases along the trajectory. **C.** Cells in clusters expressing one or both *Adipoq* and *Pdgfra* fall along an adipogenic trajectory. **D.** Intensity of expression of *Adipoq* and *Pdgfra* indicated by coloration. **E.** *Nnat* expression (blue) increases along the trajectory. **F.** *F2r* expression (teal) decreases along the trajectory. **G.** Negative association between *Nnat* and *F2r* expression within adipocytes along the trajectory. **H.** Intensity of expression of *Nnat* and *F2r* indicated by coloration. **I.** The adipogenic trajectory is broken into subclusters of cells with no *Adipoq* expression (cluster 0) to high *Adipoq* expression (cluster 2).

In addition to interrogating *Nnat* and *F2r* in single cells along an adipogenic trajectory, we found that 8 of the 9 genes comprising the epistatic parent-of-origin effect network described above are differentially expressed along the trajectory, and they associate with cell types that are consistent with their respective roles in adipose tissue (Supplemental Figure 2.3, Supplemental Table 2.8).

## 2.4 Discussion

Epistatic interactions between imprinted and non-imprinted genes can influence complex traits when the genotypic effects of one gene depends on the parent-of-origin of alleles at another [1,119]. Here we examined epistatic interactions associated with parent-of-origin effects on dietary-obesity traits in white adipose using a simple yet powerful $F_1$ reciprocal cross mouse model. Although these parent-of-origin dependent allele-specific expression biases are consistent with imprinting mechanisms, we cannot rule out that maternal and/or paternal effects also contribute to the phenomena we observe [2].

Interactions between imprinted and non-imprinted genes have previously been shown to contribute to variation in metabolic phenotypes. For example, the maternally expressed transcription factor KLF14 (kruppel-like factor 14) regulates biallelic gene expression related to adiposity [43,49]. Mapping studies have identified two SNPs (rs4731702, rs972283) upstream of KLF14 associated with type II diabetes and cholesterol levels [60,61]. Both variants have maternally-restricted *cis*-regulatory associations with KLF14 expression in adipose tissue [62]. eQTL analysis found that rs4731702 is also enriched for *trans*-associations with KLF family transcription factor binding sites in subcutaneous white adipose tissue, suggesting that KLF14 may be a master transcriptional regulator in adipose tissue [43]. Whether additional pairs of imprinted and biallelic genes are similarly co-expressed and associate with phenotypic variation remains an open question

that has not been thoroughly investigated in large landmark functional genomics studies including ENCODE, GTEx, and GWAS, leaving a significant gap in our knowledge. Interactions between imprinted and biallelic genes could explain some of the observed parent-of-origin effect patterns associated with regions lacking obvious candidate genes, as described in a recent survey of 97 complex traits measured in outbred mice [22].

Our model asserts that parent-of-origin effects start at ASE genes and are transduced through DE genes onto phenotype. This is illustrated in the interaction between *Nnat* and *F2r*. If a *cis*-regulatory effect interacts with epigenetic modifications (i.e. imprinting) at *Nnat*, then *Nnat* expression of genotypic classes are affected by paternal allele identity. Between the LG/J and SM/J alleles at *Nnat*, the LG/J allele is more highly expressed. If our model is correct, the downstream DE gene should show a corresponding pattern (**Figure 2.1 B**). In the case of *Nnat* and *F2r*, which have strong negative correlated expression, when the LG/J allele is inherited paternally at *Nnat*, the higher expression of *Nnat* should correspond with lower expression of *F2r*. This is what we observe (**Figure 2.4**). If this relationship is true, we should see persistent co-expression of *Nnat* and *F2r* across genetic backgrounds ($F_0$, $F_1$, $F_2$), which we do (**Figure 2.4**). This supports a biologically meaningful relationship between *Nnat* and *F2r*. Our model further predicts that the DE genes should more closely affect phenotype [120–122]. In the case of *Nnat* and *F2r*, we expect that *F2r* more strongly associates with basal glucose than *Nnat*, which we observe (**Figure 2.4**).

There is a clear relationship between *Nnat* and *F2r* in adipogenesis, but the specifics of how this relationship extends to glucose homeostasis are unclear. One possibility is that by altering SERCA function, *Nnat* affects not only the formation of new adipocytes, but also the beiging of adipocytes. The SERCA channel is uncoupled in beige adipocytes as part of a UCP1-independent

form of non-shivering thermogenesis. Non-shivering thermogenesis consumes a significant amount of energy, thereby altering glucose homeostasis [78]

We identified a putative network coordinated by interactions between ASE and DE genes, and from the literature found that this epistatic network is comprised of key steps in a pathway regulating differentiation and survival of adipocytes in response to nutritional environment (**Figure 2.3 B**). Specifically, there is evidence that it plays a critical role in the induction of adipogenesis. This alone demonstrates how parent-of-origin effects can move through networks along molecular pathways. Beyond proof-of-principle this network provides a clue to the puzzle of the prevalence of parent-of-origin effects.

The constituents of this single network appear to play vastly different physiological roles depending on the tissue. In white adipose the network appears to play some role in balancing proliferation, differentiation, and apoptosis as we describe above. In pancreatic ß-cells, members of this network affect secretion of insulin [123]. In bone, members of this network affect the balance of cartilage/bone growth and reabsorption. These three physiological processes may at first seem unrelated, but they share one key commonality – they are jointly critical to growth. This is consistent with the sexual conflict hypothesis attributed to parent-of-origin effects [11,36]. The of size of progeny in placental mammals can have opposing fitness consequences for mothers/maternal relatives and fathers/paternal relatives. The fitness of fathers and paternal relatives, particularly in the case of multi-paternity litters, is improved with larger progeny [11,36,38,69,124]. This comes at a fitness disadvantage to the mother. The fitness of mothers is improved by progeny of a manageable size, allowing her to produce multiple litters.

According to this model, imprinting evolved in part to allow one parent to hijack parts of a nutritional environment response pathway driving growth in a direction favorable to maximize the fitness of individuals sharing a parental line. Key processes in such a pathway driving growth would include the secretion of growth factors, construction of cartilage and bone, and the accumulation of energy stores. We present a network that appears to play a role in all three processes. If the sexual conflict hypothesis is true, then the most parsimonious place for imprinting to evolve would be in key regulatory points that affect as many aspects of growth as possible. This is consistent with the network we identified, a single pathway affecting many aspects of growth. This hints at the possibility that parent-of-origin effects are common because of the multi-purpose nature of the pathways in which genomic imprinting manifests and parent-of-origin effects propagate.

By leveraging the reciprocal $F_1$ hybrids, we are able to integrate parent-of-origin-dependent allele-specific expression and parent-of-origin-dependent differential expression with $F_{16}$ phenotypes. By doing so, we identify plausible candidates for functional validation and describe discrete molecular networks that may contribute to the observed parent-of-origin effects on metabolic phenotypes. The genes and interactions we present here represent a set of actionable interacting candidates that can be probed to further identify the machinery driving these phenomena and make predictions informed by genomic sequence. The frameworks we have developed account for the genetic, epigenetic, and environmental components underlying these parent-of-origin effects, thereby improving our ability to predict complex phenotypes from genomic sequence. We focused on metabolic phenotypes in this study, but the patterns we identified may translate to other complex traits where parent-of-origin effects have been implicated.

## 2.5 Methods

### Mouse husbandry and phenotyping

LG/J and SM/J founders ($F_0$) were obtained from The Jackson Laboratory (Bar Harbor, ME). $F_1$ reciprocal cross animals were generated by mating LG/J mothers with SM/J fathers (LxS) and the inverse (SxL). $F_2$ reciprocal cross animals were generated by mating LxS mothers with SxL fathers and the inverse. At three weeks of age, animals were both weaned into same-sex cages of 3-5 animals and randomly placed on high-fat (42% kcal from fat; Teklad TD88137) or low-fat (15% kcal from fat; Research Diets D12284) isocaloric diets. Animals were weighed weekly from three weeks until sacrifice. At 19 weeks of age, body composition was determined by MRI and a glucose tolerance test was performed after a 4 hour fast. At 20 weeks of age, animals were given an overdose of sodium pentobarbital after a 4 hour fast and blood was collected via cardiac puncture. Euthanasia was achieved by cardiac perfusion with phosphate-buffered saline. After cardiac perfusion, the reproductive fat pad was harvested, flash frozen in liquid nitrogen, and stored at -80°C.

### Genomes and annotations

LG/J and SM/J indels and SNVs were leveraged to construct strain-specific genomes using the GRC38.72-mm10 reference as a template [125]. This was done by replacing reference bases with alternative (LG/J | SM/J) bases using custom python scripts. Ensembl R72 annotations were adjusted for indel-induced indexing differences for both genomes.

### RNA sequencing

Total RNA was isolated from adipose tissue using the RNeasy Lipid Tissue Kit (QIAgen) (n = 32, 4 animals per sex/diet/cross cohort). RNA concentration was measured via NanoDrop and RNA quality/integrity was assessed with a BioAnalyzer (Agilent). RNA-Seq libraries were constructed

using the RiboZero kit (Illumina) from total RNA samples with RIN scores >8.0. Libraries were checked for quality and concentration using the DNA 1000LabChip assay (Agilent) and quantitative PCR, according to manufacturer's protocol. Libraries were sequenced at 2x100 paired end reads on an Illumina HiSeq 4000. After sequencing, reads were de-multiplexed and assigned to individual samples. RNAseq data are available through the NCBI-SRA, accession: PRJNA753198.

## Library complexity

Complexity was measured by fitting a beta-binomial distribution to the distribution of Lbias values using the VGAM package [126]. The shape parameters (α, β) of beta-binomial distributions were estimated and used to calculate dispersion (ρ). Dispersion values less than 0.05 indicate our libraries are sufficiently complex (Supplemental Figure 2.4). One library was found to have insufficient complexity and was removed from the analyses.

$$\rho_s = \frac{1}{1+\alpha_s+\beta_s} \quad (2.1)$$

## Allele-specific expression

FASTQ files were filtered to remove low quality reads and aligned against both LG/J and SM/J pseudo-genomes simultaneously using STAR with multimapping disallowed [127]. Read counts were normalized via upper quartile normalization and a minimum normalized read depth of 20 was required. Alignment summaries are provided in Supplemental Table 2.9 and Supplemental Figure 2.5.

For each gene in each individual, allelic bias (Lbias) was calculated as the proportion of total reads for a given gene with the LG/J haplotype. Parent-of-origin-dependent allele-specific

expression was detected by ANOVA using one of a number of models in which Lbias is responsive to cross and the interaction of cross with some combination of sex and diet:

$$Model \begin{cases} if\ each\ Cross\ context\ has \geq 2\ samples,\ Lbias \sim Cross \\ if\ each\ Cross{:}Sex\ context\ has \geq 2\ samples,\ Lbias \sim Cross + Cross{:}Sex \\ if\ each\ Cross{:}Diet\ context\ has \geq 2\ samples,\ \ Lbias \sim Cross + Cross{:}Diet \\ if\ each\ context\ has \geq 2\ samples,\ \ Lbias \sim Cross + Cross{:}Sex + Cross{:}Diet + Cross{:}Sex{:}Diet \end{cases} \quad (2.2)$$

Accurately estimating the significance of these effects and correcting for multiple tests is challenging for two reasons: 1) the complexity of the many environmental contexts, and 2) the correlation of allelic bias within and between imprinted domains breaks assumptions of independence. A permutation approach is an effective way to overcome these challenges. The context data was randomly shuffled for each gene independently and analyses were rerun to generate a stable null distribution of p-values (Supplemental Figure 2.6). False discovery rates were estimated for a given significance threshold as the proportion of significant tests under the permutated null model relative to significant tests under the real data model. A value of 1 meaning that 100% of tests at a given significance threshold are likely false positives. An $FDR \leq 0.1$ was considered significant (Supplemental Table 2.1, Supplemental Figure 2.7).

To determine the parental direction and size of expression biases, a POE score was calculated as the difference in mean Lbias between reciprocal crosses (LxS or SxL). POE scores range from completely maternally-expressed (-1), to biallelic (0), to completely paternally-expressed (+1). POE score thresholds were calculated from a critical value of $\alpha = 0.01$, determined from a null distribution created by permutation Genes with significant allele-specific expression and POE scores beyond the critical value were considered to have significant parent-of-origin-dependent allele-specific expression (Supplemental Figure 2.8).

## Differential expression

Differential expression by reciprocal cross was determined by first aligning reads against the LG/J and SM/J genomes simultaneously with multimapping permitted. Reads were normalized by Trimmed mean of M-values (TMM) normalization, which estimates scale factors among samples to allow for differences in RNA composition [128]. A minimum normalized read count of 10 was required. Generalized linear models accounting for diet, sex, and diet-by-sex were fit in EdgeR [129]. Differential expression was detected by a likelihood ratio test. Significance was determined for five models for each gene:

$$1.\ Expression \sim Cross \quad (2.3)$$

$$2.\ \ Expression \sim Cross{:}\,Sex \quad (2.4)$$

$$3.\ Expression \sim Cross{:}\,Diet \quad (2.5)$$

$$4.\ Expression \sim Cross{:}\,Sex{:}\,Diet \quad (2.6)$$

$$5.\ Expression \sim Cross + Cross{:}\,Sex + Cross{:}\,Diet + Cross{:}\,Sex{:}\,Diet \quad (2.7)$$

Multiple test corrections were performed by implementing the "qvalue" R package to estimate false discovery rates (Supplemental Figure 2.9). Genes with a $FDR \leq 0.1$ and a $|fold\ change| \geq 1.5$ were considered significantly differentially expressed by reciprocal cross (Supplemental Figure 2.10 and Supplemental Table 2.2).

## Gene-gene interactions

Networks were constructed in each tissue by pairing genes showing parent-of-origin-dependent allele-specific expression with biallelic genes that are differentially expressed by cross. Pairs were predicted by modeling the expression of biallelic genes as a function of parent-of-origin-dependent

allele-specific expression, Lbias, sex, diet, and sex-by-diet. The strength of a prediction was measured through model fit, which was estimated as a mean test error with 10-fold cross-validation employed to prevent overfitting. Significance was estimated by likelihood ratio test using a chi-square distribution. Given the complexity of contexts, false discovery rates were determined by permuting the context and expression data to generate a stable null-distribution of p-value (Supplemental Figure 2.11). Null distribution stability was evaluated by calculating the critical value for alpha = 0.05 at each genome wide iteration. The standard deviation of critical values was calculated after each iteration for the last 5 iterations. Genome-wide shuffling was done 500 times, with 759 independent randomized tests per iteration, meaning the stable null model is composed of 379,500 randomized observations. Using the null model, the "qvalue" package estimated a $\widehat{\pi_0}$. This estimate was then used to estimate false discovery rates in the real data. MTE score thresholds were calculated from a critical value of $\alpha = 0.01$, determined from a null distribution created by permutation (Supplemental Figure 2.12). Connections with an $FDR \leq 0.1$ (Supplemental Table 2.10) and MTE below the critical value were considered significant (Supplemental Figure 2.13).

## Functional enrichment analysis

Functional enrichment of interacting genes showing parent-of-origin-dependent allele-specific expression with biallelic genes that are differentially expressed by cross was tested by over-representation analysis in the WEB-based Gene Set Analysis Toolkit v2019 [80]. We performed analyses of gene ontologies (biological process, cellular component, and molecular function), pathway (KEGG), and phenotype (Mammalian Phenotype Ontology). The list of all unique interacting genes was analyzed against the background of all unique genes expressed in white adipose. A Benjamini-Hochberg FDR-corrected p-value $\leq 0.01$ was considered significant (Supplemental Table 2.4).

## Phenotype correlation

In order to identify which phenotypes might be affected by genes in the parent-of-origin effect network, gene expression was correlated with metabolic phenotypes collected for $F_1$ animals with the contexts combined. Phenotypes were log transformed when necessary, as determined by Shapiro Wilkes test to approximate normality (Supplemental Figure 2.14). Additionally, the effects of sex and diet were residualized out leaving only the effect of cross. This was done to mirror later residualizing of phenotypes in the $F_{16}$ population when testing for epistasis. Multiple test corrections were performed by implementing the "qvalue" R package to estimate false discovery rates (Supplemental Figure 2.15). The minimum Pearson's correlation coefficient threshold was set to |0.5|. Connections with an $FDR \leq 0.05$ (Supplemental Table 2.5) and MTE below the critical value were considered significant (Supplemental Figure 2.16).

## Epistasis testing

The $F_{16}$ LxS advanced intercross population, phenotypes, genotypes, genotypic scores, and QTL mapping methods are described elsewhere [23–25,59]. We tested for epistasis in interacting pairs between genes showing parent-of-origin-dependent allele-specific expression and biallelic genes that are differentially expressed by cross. We selected $F_{16}$ genotyped markers that fall within 1.5 Mb up- and downstream from the geometric center of each gene, defined as the genomic position halfway between the transcription start and stop position of that gene (Supplemental Table 2.11). For every $F_{16}$ animal, an "imprinting score" was assigned to each marker based on that animal's genotypic values (LL = 0, LS = 1, SL = -1, SS = 0; maternal allele is depicted first). Non-normally distributed phenotypes (as evaluated by a Shapiro-Wilk test) were $\log_{10}$-transformed to approximate normality (Supplemental Figure 2.17). Because of the number of epistasis tests performed and the number of contexts represented in the data, we removed the effects of sex, diet

and their interaction from each $F_{16}$ phenotype with a covariate screen. We tested for epistasis on the residualized data using the following generalized linear model:

$$R_{pheno} \sim BDE_{IMP} + ASE_{IMP} + BDE_{IMP}: ASE_{IMP} \qquad (2.8)$$

Where $R_{pheno}$ is the residual phenotype, $BDE_{IMP}$ is the imprinted genotypic score for the biallelic gene that is differentially expressed by cross, $ASE_{IMP}$ is the imprinted genotypic score for the gene showing parent-of-origin-dependent allele-specific expression bias, and $BDE_{IMP}$:$ASE_{IMP}$ is the interaction between the two genes' imprinted genotypic score. We employed a permutation approach to accurately estimate significance given the linkage of proximal markers. Imprinted genotypic values were randomly shuffled to generate a stable null model of p-values (Supplemental Figure 2.18). False discovery rates were estimated for a given significance threshold as the proportion of significant tests under the permutated null model relative to significant tests under the real data model (Supplemental Figure 2.19). An $FDR \leq 0.1$ was considered significant. Epistasis was considered significant if the $BDE_{IMP}$:$ASE_{IMP}$ interaction term met the significance threshold (Supplemental Table 2.6).

## Validation of *Nnat* and *F2r* expression patterns

Expression patterns of *Nnat* and *F2r* in white adipose were validated by qRT-PCR in high fat-fed female LG/J and SM/J mice and in biological replicates of high fat-fed female $F_1$ reciprocal cross animals (n = 6 LG/J homozygotes, n = 10 LxS and 10 SxL reciprocal heterozygotes, n = 6 SM/J homozygotes). Total RNA was extracted from adipose samples using the Qiagen Rneasy Lipid Kit. High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher) was used for reverse transcription. Quantitative RT-PCR was performed with an Applied Biosystems (USA) QuantStudio 6 Flex instrument using SYBR Green reagent. Results were normalized to *L32*

expression using the ΔΔCt method. *Nnat* forward primer – CTACCCCAAGAGCTCCCTTT and reverse primer – CAGCTTCTGCAGGGAGTACC. *F2r* forward primer – TGAACCCCCGCTCATTCTTTC and reverse primer – CCAGCAGGACGCTTTCATTTTT. *L32* forward primer – TCCACAATGTCAAGGAGCTG and reverse primer – GGGATTGGTGACTCTGATGG. Data points were considered outliers if they led to violation of normality assumptions or were considered outliers by box and whisker plots. ANOVA was used to estimate significance of differential expression by cross (1), paternal allele identity (2), mitochondrial ancestry (3).

$$1.\, Expression \sim Cross \;\in \begin{cases} LL, 0 \\ LS, -1 \\ SL, 1 \\ SS, 0 \end{cases} \qquad (2.9)$$

$$2.\, Expression \sim Paternal\ Allele \;\in \begin{cases} LL, 0 \\ LS, 1 \\ SL, 0 \\ SS, 1 \end{cases} \qquad (2.10)$$

$$3.\, Expression \sim Mitochondrial\ ancestry \;\in \begin{cases} LxS\ x\ SxL, 0 \\ SxL\ x\ LxS, 1 \end{cases} \qquad (2.11)$$

Expression patterns were also validated by qRT-PCR in high fat-fed female $F_2$ animals (n = 14). Co-expression was determined by fitting a general linear model and estimating significance using the Wald test approximation of the LR test. Correlation with basal glucose was determined by fitting a general linear model and estimating significance using the Wald test approximation of the LR test. Pearson's correlation coefficients were calculated for each gene with basal glucose. To test whether patterns in these correlations was significantly different between $F_0/F_1$ and $F_2$ populations, bootstrapping was used to calculate 90% confidence intervals for the Pearson's

correlation coefficients. 5,000 iterations were run with 10 individuals randomly selected with replacement.

## scRNA analysis of *Nnat* and *F2r*

scRNAseq data was downloaded from SRA: SRP145475 [118]. Data were processed and aligned to the C57BL/6J reference (mm10) using Cell Ranger [130]. Analysis and cell quality control was performed using the Seurat (3.2.2)[131]package in R (3.6.1)[132]. Cell quality was controlled using three metrics [133]1) number of features, 2) number of counts, 3) covariation of features and counts. High quality cells were required to have between 500 and 3,000 features and read counts between 1,000 and 30,000. As sequencing is a process of random sampling, the number of features and the number of counts should covary. This relationship was fit to a generalized additive model. Deviation from this relationship (residuals) were computed for each cell. High quality cells were required to have a residual within 3 standard deviations of the mean residual of all cells (Supplemental Figure 2.20).

Seurat normalization with a scale factor of 10,000 was performed. Dimensionality reduction (UMAP) was performed (dims = 1:10, resolution = 0.15). Resolution was chosen using the clustree (0.4.3) package [134]A range of resolutions from 0.06 to 0.18 were tested, and the highest resolution with stable clustering was chosen (Supplemental Figure 2.21). Cell type markers were identified by differential expression analysis using the "MAST" hurdle-model test [135]. Genes overexpressed in a given cell type relative to all other cell types were considered cell type "markers". Cell type identity was assigned using previously reported markers for this data set (Supplemental Figure 2.2).

Cells along the adipogenic trajectory were subset and subjected to dimensionality reduction (UMAP, dims=1:10, resolution=0.17). A range of resolutions from 0.01 to 0.25 were tested. Using

*Adipoq* as a marker of differentiation, we sought to identify the set of clusters that would best encapsulate the stages of differentiation. To this end for every level of resolution we calculated the mean count variance ($\overline{C_\sigma}$). This is done by calculating the standard deviation ($\sigma$) of *Adipoq* expression I within each cluster ($G$), referred to as the count variance ($C_\sigma$). Cells with no expression of *Adipoq* were excluded. The mean of count variances for all clusters is calculated. This process is similar to k-means clustering, where the goal is to find that parameters which minimize the within group variation.

$$Count\ Variance = \frac{\sum_{G=1}^{n} \sigma(E_G)}{n} \qquad (2.12)$$

We also calculated the percent expressing variance ($\overline{P_\sigma}$). This was taken as the mean of the standard deviation in the percent of cells expressing *Adipoq*.

$$Percent\ Expressing\ Variance = \frac{\sum_{G=1}^{n} \sigma(\%E{>}0_G)}{n} \quad (2.13)$$

The resolution 0.17 was chosen as the lowest resolution where variation is minimized and no longer significantly changes (Supplemental Figure 2.22). Using *Adipoq* as a marker of adipogenesis, clusters 1 and 2 were identified as pre- and post-differentiated cells, respectively. Differential expression was analyzed using the "MAST" test. Expression was compared between clusters 1 and 2 only. Multiple tests correction was performed using the Bonferroni method. We required changes in expression to show either a sufficiently large fold change ( $|\log_2 FoldChange| \geq 0.3$ ) OR a sufficiently large change in the percent of cells expressing the gene in question ( $pct.\Delta \geq 0.4$ ). The change in percent of cells expressing a gene was calculated as the difference in percent of cells expressing the gene between the clusters and scaled by dividing by the larger percentage.

$$pct.\Delta = \frac{pct.2 - pct.1}{\max(pct.1, pct.2)} \qquad (2.14)$$

# Chapter 3: A general framework to model of genetic effects within networks

In this chapter, I use ASE to find candidate "Source" genes and total expression to find "Target" genes, then incorporate additional data (multiple generations, mapping data, single-cell RNAseq) to refine a pairwise network.The results are included in a manuscript to be submitted to the journal *Molecular Biology and Evolution*.

**Juan F Macias-Velasco** and Heather A Lawson. Cell Type Specific Genetic Effect Networks Reveal Mechanisms Underlying Complex Murine Traits. *in preparation*

## 3.1 Abstract

Metabolic dysregulation poses serious risks to public health. Understanding the genetic underpinning of metabolic traits may aid in the management of metabolic dysregulation. The genotype-to-phenotype problem is confounded by our inability to dissect and contextualize genetic and epigenetic information. Addressing the genotype-to-phenotype problem requires a paradigm shift from a single locus naïve model to a biologically interpretable multi-locus model accounting for genetic, parent-of-origin, and environmental effects. Here we present a statistical framework and general approach to construct and dissect expression level genetic effects as cell type-specific pairwise genetic effects networks. We call the model at the heart of this approach the bridge model. With the bridge model, we use total expression to measure total genetic effects on a target gene.

By using ASE data, we can isolate the *cis* effects on a target gene. Any observed genetic effects which cannot be accounted for by *cis*-effects, must be the result of a *trans* effect from some unobserved source gene(s). We construct a pairwise network where the *cis*-effects of a source gene matches the total genetic effects of a target gene. We use the LG/ J and SM/J inbred mouse strains to model gene expression in white adipose tissue and found wide-spread evidence of genetic effect propagation with 38% of genes showing signatures of additive *trans* effects. We demonstrated that the bridge model captures genetic effects just as effectively as a GEM and effectively uses ASE data. We put forth Bridge as a viable model to related total expression, ASE, and genetic effects. We identified a genetic effects network with 8,996 gene pairs representing 325 "Source" genes and 299 "Target" genes. Integrating single-cell RNAseq we split this network into cell type-specific networks, representing 199 unique genes. Within this we identified an additive genetic effect propagating from *Msr1* to *Gpnmb* in macrophages. *Msr1* and *Gpnmb* are overwhelmingly expressed by macrophages. The estimated *cis* genetic effect on *Gpnmb* can only account for 13% of it's heritability. Whereas, the *cis* effect from *Msr1* can account for 62% of the heritability of *Gpnmb*. *Msr1* is expressed in macrophages and is required for macrophages to take on lipid-laden phenotype. Accumulation of lipids in macrophages induces expression *Gpnmb*. By promoting the lipid laden phenotype elevated *Msr1* expression promotes expression of *Gpnmb*, an established biomarker of lipid laden macrophages. The exact cascade of events between *Msr1* activation and *Gpnmb* expression are unknown, but a there is a clear connection between them. Our model showed not only that they are related, but also the causality (*Msr1* → *Gpnmb*) and the directionality (positive correlation). All of which is supported by previous studies.

## 3.2 Introduction

Metabolic dysregulation poses serious risks to public health. Not only in terms of direct health consequences, but also as a risk factor for other diseases. Understanding the genetic underpinning of metabolic traits may aid in the management of metabolic dysregulation. Improvements in sequencing technology have made large amounts of data available. However, utilizing these data will be limited until major advances can be made in the prediction/modeling of complex traits. The genotype-to-phenotype problem is confounded by our inability to dissect and contextualize genetic and epigenetic information.

The full significance of genetic phenomena can only be tabulated at the level of networks and pathways. Addressing the genotype-to-phenotype problem requires a paradigm shift from a single locus naïve model to a biologically interpretable multi-locus model accounting for genetic, parent-of-origin, and environmental effects. The need to shift to a systems genetics paradigm is well accepted, after all biological systems are in fact *systems* and should be studied as such.

Single locus frameworks are designed to find reproducible associations between phenotype ($P$) and genotype ($G$) (Eq. 3.1). However, they have limited interpretability and can only explain a small proportion of phenotypic variation. This is in part because biological systems are inherently complex, stochastic, and multi-locus in nature.

$$P = \beta_1 G_1 + \beta_0 \qquad (3.1)$$

With sufficient sample size we could simply expand the single locus paradigm to include many or all loci (Eq. 3.2).

$$P = \beta_1 G_1 + \boldsymbol{\beta_2 G_2} + \boldsymbol{\beta_3 G_3} + \cdots + \boldsymbol{\beta_i G_i} + \beta_0 \qquad (3.2)$$

This approach would increase the amount of phenotypic variation explained. But it would require unrealistic sample sizes and overfitting would seriously undermine the generalizability of findings. It also wouldn't be any more interpretable. Further, such an expansive additive model still has no real way to account for genetic interactions. It is possible to include interaction terms (Eq. 3.3).

$$Eq.3. \quad P = \beta_1 G_1 + \beta_2 G_2 + \beta_3 G_3 + \cdots + \beta_i G_i + \boldsymbol{\beta_{1,2} G_{1,2}} + \boldsymbol{\beta_{1,3} G_{1,3}} + \cdots + \boldsymbol{\beta_{i,j} G_{i,j}} + \beta_0 \qquad (3.3)$$

This would capture genetic interactions, but would also dramatically increase the number of terms leading to overfitting. We need a framework that is resistant to overfitting, able to explain a greater degree of phenotypic variation, and lends itself to interpretation.

Our goal is to take an incremental step towards addressing the genotype-to-phenotype problem by modeling the trait values of a population with a shared genetic background as a distribution $\mathcal{N}(P)$. So, while we can't accurately predict the trait value of an individual ($\hat{P}$), we may be able to calculate the mean trait value for a population of similar individuals ($\bar{P}$).

## Genetic effects

A classic way to model population traits as a function of genotype is by genetic effects models[32,73,136] (**Figure 3.1 A**). Genetic effects models are additive linear models which reliably describe traits as a function of genotypic classes (LL, LS, SL, SS) when there are two possible alleles (here L and S). By convention, the maternal allele precedes the paternal allele. Heterozygotes are subdivided by allelic parent-of-origin, where individuals who inherit the L allele maternally and S allele paternally (LS) are treated as a separate class from those with reversed allelic origins (SL). This model partitions the effect of genotype on phenotype into genetic effects coefficients (**Figure 3.1 B**): reference (*R*), Additive (*A*), Dominance (*D*), and imprinting (*I*). These four coefficients capture the variance *between* genotypic classes. These coefficients can be directly

calculated from genotypic means and vice versa. For example, additivity is the difference between homozygote means divided by two ($\frac{\overline{LL}-\overline{SS}}{2}$). This framework is a robust and established approach to measuring genetic effects. It lends itself quite well to mapping studies, which have employed it to map quantitative trait loci (QTL) for a variety of complex metabolic traits [23–25,59].

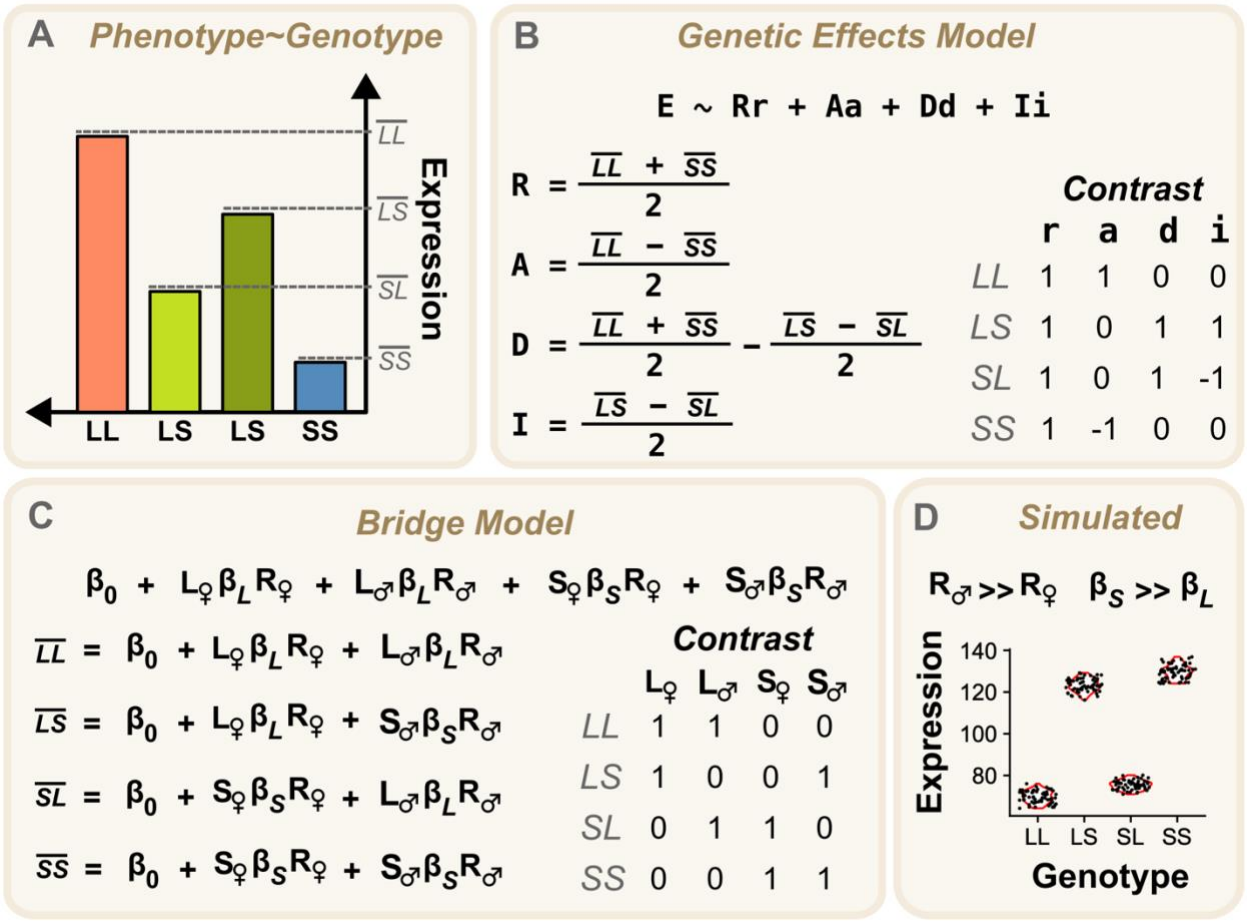


**Figure 3.1. Population level phenotypes can be predicted using genetic effects models.** A classical way to model population traits as a function of genotype is the genetic effects model. Genetic effects models describe traits as a function of genotype split into four genotypic class populations {LL, LS, SL, SS} when there are two possible alleles {L & S}(**A**). By convention for the nomenclature of genotypes, the maternal allele precedes the paternal allele. Genetic effect models partition the effect of genotype on phenotype into genetic effects (**B**); Reference (*R*), Additivity (*A*), Dominance (*D*), and Imprinting (*I*). These four coefficients accurately capture the relationship between genotypic means. Coefficients can be directly calculated from genotypic

means $\{\overline{LL}, \overline{LS}, \overline{SL}, \overline{SS}\}$. Genotype is encoded using four genetic effects scores: reference (*r*), additive (*a*), dominance (*d*), and imprinting (*i*).

We present a statistical framework to construct and dissect expression level genetic effects as pairwise genetic effects networks. Our approach relies on using the bridge model which re-frames genetic effects as pseudo-molecular parameters (**C**). Bridge models expression as a function of genotype and five "bridge parameters" $\{\beta_L, \beta_S, R_{♀}, R_{♂}$, and $\beta_0\}$. Genotype is encoded using allelic scores $\{L_{♀}, L_{♂}, S_{♀}, S_{♂}\}$ which take on the value 1 when the corresponding allele is inherited from the indicated parent and takes on the value 0 otherwise. An LS individual would have the following allelic scores: $\{L_{♀} = 1, L_{♂} = 0, S_{♀} = 0, S_{♂} = 1\}$. Bridge parameters can be used to estimate population genotypic means. Genotypic mean estimates can then be used to calculate classic genetic effects coefficients (*R, A, D, I*). In this way bridge parameters can be used to calculate genetic effects.

We showcase an example of how these parameters relate to expression using simulated data in which a gene shows strong preferential expression of the paternally inherited allele and much stronger expression of the S allele (**D**).

The LG/ J and SM/J inbred mouse strains are established models in the study of metabolic traits. Intercrossed lines of these strains were developed to uncover the genetic architecture of metabolic dysregulation. Ultimately, the goal of these studies is to find the molecular circuitry/pathways through which genotype influences phenotype. Unfortunately, mapping like this has several major limitations: 1) Identified QTL are comprised of vast genomic spaces (~3-6 Mb) often containing many genes; 2) It reveals little or no mechanistic leads and doesn't lend itself well to mechanistic interpretation by experimentalist; and 3) It provides no information as to what tissues or cell types the observed genetic effects are occurring in.

Our aim is to introduce a complimentary system to better address these specific limitations. Here we present a statistical framework and general approach to construct and dissect expression

level genetic effects as cell type-specific pairwise genetic effects networks. We call the model at the heart of this approach the bridge model (**Figure 3.1 C**).

The bridge model attempts to re-frame genetic effects as pseudo-molecular parameters. It models expression as a function of genotype and five "bridge parameters" ($\beta_L$, $\beta_S$, $R_{♀}$, $R_{♂}$, and $\beta_0$). Genotype is encoded as allelic scores ($L_{♀}$, $L_{♂}$, $S_{♀}$, $S_{♂}$) which take on the value 1 when the corresponding allele is inherited from the indicated parent and takes on the value 0 otherwise. For example, an LS individual would have allelic scores as follows: $\{L_{♀} = 1, L_{♂} = 0, S_{♀} = 0, S_{♂} = 1\}$. Bridge parameters can be used to calculate population genotypic means ($\overline{LL}$, $\overline{LS}$, $\overline{SL}$, $\overline{SS}$). Bridge captures complex patterns and describes them using bridge parameters. We showcase an example of this using simulated data where a gene shows strong preferential expression of the paternally inherited allele and strong expression of the S allele (**Figure 3.1 D**). Like the genetic effects model, genotypic mean estimates can be calculated from bridge parameters, which can in turn be used to calculate the coefficients of a classic genetic effects model ($R, A, D, I$). Thus, bridge parameters can be used to calculate genetic effects. Bridge parameters are more intuitive than classic genetic effects coefficients.

Most studies employing RNAseq measure the total expression of a gene agnostic to its allelic identity or it's parental origin. By measuring the expression of each allele individually, we learn how the identity and parental origin of an allele affects expression (ASE). ASE occurs in two forms, sequence dependent and parent-of-origin dependent[5,11,67]. Sequence dependent ASE happens when alleles within an individual are unequally expressed based on the identity of the alleles (L vs S). Parent-of-origin dependent ASE happens when alleles are unequally expressed based on which parent they were derived from (maternal vs paternal). This form of ASE is an

epigenetic phenomenon primarily driven by DNA methylation that marks which parent a chromosome came from [11,34].

Unlike the genetic effects model, the bridge model was designed with ASE in mind. Its parameters are meant to be intuitively reflective of ASE machinery. The parental repression parameters ($R_{♀}$, $R_{♂}$) correspond to parental epigenetic mechanisms. The allelic expression parameters ($\beta_L$, $\beta_S$) correspond to the amount of expression attributable to each allele based on its identity. Since bridge parameters can be used to calculate genetic effects, it allows us to "bridge" ASE phenomena to genetic effects.

## Genetic effects propagate through systems

Genetic effects are the effects of genotype on traits. The total genetic effect on a trait at a given locus is the sum of all *cis* and *trans* effects. The genotype of an autosomal locus is composed of two alleles inherited either maternally or paternally. In our $F_0/F_1$ model we have four possible allele by parental combinations $\{L_{♀}, L_{♂}, S_{♀}, S_{♂}\}$. The total expression of a gene is the combined expression of these two alleles. We define a *cis* genetic effect as the effect of an allele's identity and parental origin on its own expression. A *cis* effect is specific, it acts only on the expression of the allele on the same chromosomal copy. This is different from a *trans* effect which is non-specific, acting across chromosomal copies and between genes. For example, the effect of allele identity on the maternal copy of a promoter that controls expression of its gene on the maternal copy would be a *cis* effect. The effect the same allele has on expression of other genes outside of its genomic domain would be a *trans* effect. In this way genetic effects propagate through regulatory pathways in complex ways.

We can use quantitative genetics to begin to detect signatures of pathways. In a genetic effects model, additivity measures the difference between homozygote genotypic means and can be thought of as the difference in expression between alleles (L vs S) and between genotypes (LL vs SS)[74,137]. Additivity is calculated from total gene expression (L+S), which measures the total additive genetic effect. ASE measures the difference in expression between alleles within heterozygous individuals (L vs S). Crucially, measuring ASE within an individual means that both alleles are exposed to the same cellular environment, and therefore both alleles are exposed to the same *trans* genetic effects. This means that differences observed as ASE can only result from *cis*-regulatory differences. Thus, sequence dependent ASE is a measure of the *cis* additive genetic effect.

For example, if there is sequence dependent ASE at a locus within heterozygotes (L > S), LL homozygotes should have higher overall expression relative to SS homozygotes (L+L > S+S). However, this assumes that genes function in isolation. The network structure of biological systems means that genetic effects on one gene affect other genes within its network[138]. We call this genetic effect propagation. Accounting for genetic effect propagation allows us to better understand the structure and nature of genetic networks. Resolving genetic network structure is the first step to contextualizing genetic information across networks and thereby advancing our understanding of the genotype-to-phenotype problem.

## Genetic effect partitioning reveals widespread genetic effect propagation

How important might genetic effect propagation be? The more interconnected biological systems are, the more widespread and important propagation becomes. Total genetic effects are the sum of *cis* genetic effects and *trans* genetic effects. Genetic effects on total expression measures total

genetic effects, whereas ASE measures only the *cis* genetic effect. Differences between observed *cis* versus total effects indicates the presence of some *trans* effect(s). By comparing additivity on total expression to sequence dependent ASE, we can infer the prevalence of *trans* effects.

We measured genetic effects on total gene expression in high fat fed female mice (n=16) using a genetic effects model. We considered an additive genetic effect with an $FDR \leq 0.1$ to be significant (**Figure 3.2A**). We used an $F_1$ hybrid with matching context (n=8) to measure sequence dependent ASE. We considered a $|\text{Sequence Dependent Allelic Bias}| \geq 0.25$ to be significant, equating to >75% of expression coming from one allele.

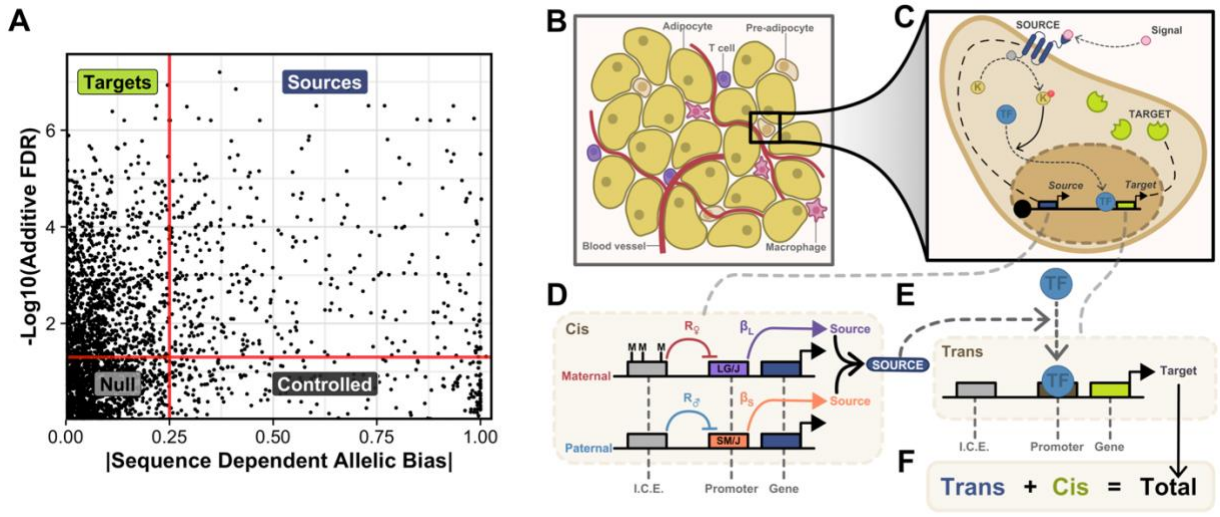


**Figure 3.2. Genetic effects propagate within biological networks along molecular pathways.** We wanted to see how much of the observable genetic effects might arise from network structure (*trans* effects) as opposed to widespread *cis*-regulatory difference between strains. Differences between observed *cis* and Total genetic effects indicate the presence of *trans* effects. Additivity in a genetic effects model measures the difference between homozygote genotypic means and can be taken as the difference in expression between alleles (L vs S) between populations (LL vs SS). ASE measures the difference in expression between alleles within heterozygous individuals (L vs S). Sequence dependent ASE is a measure of the *cis* additive genetic effect. Additivity calculated from total gene expression (LL vs SS) is a measure of total genetic effects (*cis* + *trans*).

We measured genetic effects on total gene expression in high fat fed females using a genetic effects model. We considered an additive genetic effect with an $FDR \leq 0.1$ to be significant (**A**). Using $F_1$ hybrids with matching context we measured sequence dependent ASE. We considered $|\text{Sequence Dependent Allelic Bias}| \geq 0.25$ to be significant. Comparing Additivity and sequence dependent ASE, we saw four groups. 1] The "Null" group (41.7%) genes lack both additivity and sequence dependent ASE [$Total = 0, Cis = 0$]. 2) The "Target" group (38.8%) have additivity but lack sequence dependent ASE [$Total \neq 0, Cis = 0$]. 3) The "Source" group (12.5%) have both additivity and sequence dependent ASE [$Total \neq 0, Cis \neq 0$]. Lastly 4), the "Controlled" group (7.0%) lack additivity but have sequence dependent ASE [$Total = 0, Cis \neq 0$].

We present a representative molecular mechanism to aid in understanding how these concepts and parameters interrelate. White adipose is composed of a variety of cell types including, Adipocytes, pre-adipocytes/mesenchymal stem cells (MSC), T cells, Macrophages, and cell types found in blood vessels (**B**). For our example there is a hypothetical pathway we will focus on in pre-adipocytes (**C**). It is composed of two genes (*Source* and *Target*). The protein product SOURCE (dark blue) is a cell surface receptor activated upon binding of an extracellular signal (pink). Activated SOURCE promotes activation of kinase K (yellow) via phosphorylation. Activated K promotes activation and translocation of the transcription factor TF (light blue) to the nucleus. TF binds the *Target* promoter, initiating transcription and ultimately production of TARGET protein (lime green). Under this this mechanism, genetic effects which increase or decreasing the amount of available SOURCE affect cellular response to signal in turn increasing or decreasing *Target* expression.

We focus in further to aid in understanding how bridge parameters relate to cellular machinery. We show a model for how genomic machinery can produce *cis* genetic effects/ASE (**D**). For this mechanism there are three genomic features. An imprinting control element (ICE - grey) which is selectively marked via DNA methylation based on which parent that chromosome was inherited from. In this example, the maternally inherited ICE is selectively methylated. ICE is bound by a co-factor which when bound blocks enhancer function, effectively repressing transcription of *Source*. Methylation prevents co-factor binding, and preserves enhancer activity, permitting full expression of *Source*. Bridge encodes selective repression of transcription based on parental origin (genomic imprinting) as $R_{♀}$and $R_{♂}$. The gene *Source* and its promoter make up the remaining features. In this example, a variant (LG/J - Purple, SM/J - Orange) alters promoter activity. How much expression is explained by promoter allele identity is encoded by bridge as $\beta_L$ for the LG/J allele and $\beta_S$ for the SM/J allele. SOURCE levels regulate TF activation of the *Target* gene (**E**). Target has no meaningful epigenetic ($R_{♀} = R_{♂}$) or *cis*-regulatory differences ($\beta_L = \beta_S$), and therefore no cis genetic effect from the *Target* locus.

The total genetic effect acting on *Target* is the sum of its *cis* effects (lime green), in this case zero, and *trans* effects from upstream genes (**F**). In this example, the *Target cis* effect is zero,

which means that the observed (total) genetic effects are the result of trans effects which propagate from *Source*. Through this mechanism, *Source cis* effects would be predictive of the genetic effects on *Target*.

Comparing these two metrics gives are four possible scenarios: 1) The "Null" group (41.7%) genes lack both additivity and sequence dependent ASE [$Total = 0, Cis = 0$]; 2) The "Target" group (38.8%s additivity but lacks sequence dependent ASE [$Total \neq 0, Cis = 0$]; 3) The "Source" group (12.5%) has both additivity and sequence dependent ASE [$Total \neq 0, Cis \neq 0$]; or 4), the "Controlled" group (7.0%) lacks additivity but has sequence dependent ASE [$Total = 0, Cis \neq 0$].

The "Target" group represents a case where *cis* genetic effects do not account for the total observed genetic effects. Total genetic effects are the sum of *cis* and *trans* effects, if *cis* effects cannot explain the total, there must be some *trans* effects at play. We call the process of comparing *cis* and total effects to infer the presence of *trans* effects "genetic effect partitioning". These "missing" *trans* effects must come from somewhere. The simplest explanation is that they are derived from interactions with genes in the "Source" group. The process of genetic effects manifesting at a "Source" gene (*cis* effect) and in turn affecting the expression of a "Target" gene is what we call genetic effect propagation.

## Mechanistically, how might genetic effects propagate?

Genetic effects and their propagation are abstract and difficult to intuit by nature. It is helpful to view these phenomena with a simplistic, but realistic mechanism in mind. We present a mechanism using expression data is derived from murine $F_0$ and $F_1$ reproductive white adipose (**Figure 3.2 B**). Adipose is composed of a variety of cell types including adipocytes, mesenchymal stem cells (MSC), T cells, macrophages, and various cell types found in blood vessels. Let us say the example

cellular mechanisms occurs within a specific cell type, focusing on a hypothetical pre-adipocyte (**Figure 3.2 C**).

This mechanism is composed of two genes (*Source* and *Target*). The protein product SOURCE is a cell surface receptor activated when bound by an extracellular signal. Activated SOURCE promotes the activation of kinase K via phosphorylation. Activated K promotes activation and translocation of the transcription factor TF to the nucleus. TF binds the *Target* promoter, initiating transcription and ultimately production of the TARGET protein.

We can demonstrate genetic effect propagation using this mechanism. If a genetic effect at *Source* increases or decreases available SOURCE levels, it would enhance or dampen cellular response to signal, in turn increasing or decreasing expression of *Target*. Which means a genetic effect on *Source* expression can produce a genetic effect on *Target* expression.

## Mechanisms of allele specific expression

We present an example model for how genomic machinery could produce *cis* genetic effects/ASE and how it relates to the bridge model (**Figure 3.2 D**) [14]. This example has three genomic features. First, an imprinting control element (ICE - grey) which is a genomic feature selectively marked via DNA methylation based on which parent it was inherited from. For this example, the maternally inherited ICE is selectively methylated. ICE function in several different ways [13,139,140], but for this example the ICE functions as an insulator such as a CTCF binding site. The ICE is bound by a co-factor, which blocks enhancer function, effectively repressing transcription of *Source*. DNA methylation of ICE blocks co-factor binding, which preserves enhancer activity, allowing full expression of *Source*. In the bridge model, we encode the selective repression of transcription based on parental origin (genomic imprinting) as $R_{♀}$and $R_{♂}$. The gene *Source* and its

promoter make up the remaining features. For this example, a variant exists (L - Purple, S - Orange) in the promoter and alters promoter activity. In the bridge model, the amount of expression explained by (promoter) allele identity is encoded as $\beta_L$ for the LG/J -derived allele and $\beta_S$ for the SM/J – derived allele. In this example, expression of L and S *Source* alleles combine to produce the SOURCE protein. SOURCE levels regulate TF activation of the *Target* gene (**Figure 3.2 E**). Target has the same type of features as *Source*. In this example, there is no meaningful epigenetic ($R_{♀} = R_{♂}$) or *cis* regulatory differences ($\beta_L = \beta_S$), and therefore no *cis*-genetic effect from the *Target* locus. The total genetic effect acting on the *Target* locus is the sum of its *cis* effects, in this case zero, and *trans* effects from upstream genes (**Figure 3.2 F**). In this example, the *Target cis*-effect is zero, which means that the observed (total) genetic effects are *trans* effects coming from *Source*. Therefore, the *Source cis*-effect produces the *Target* total effect.

## General approach

We use total expression to measure total genetic effects on a target gene (**Figure 3.3 A**). By using ASE data, we can isolate the *cis* effects on a target gene (**Figure 3.3 B**). Any observed genetic effects which cannot be accounted for by *cis*-effects, must be the result of a *trans* effect from some unobserved source gene(s) (**Figure 3.3 C**). We construct a pairwise network where the *cis*-effects of a source gene matches the total genetic effects of a target gene (**Figure 3. 3D and E**).

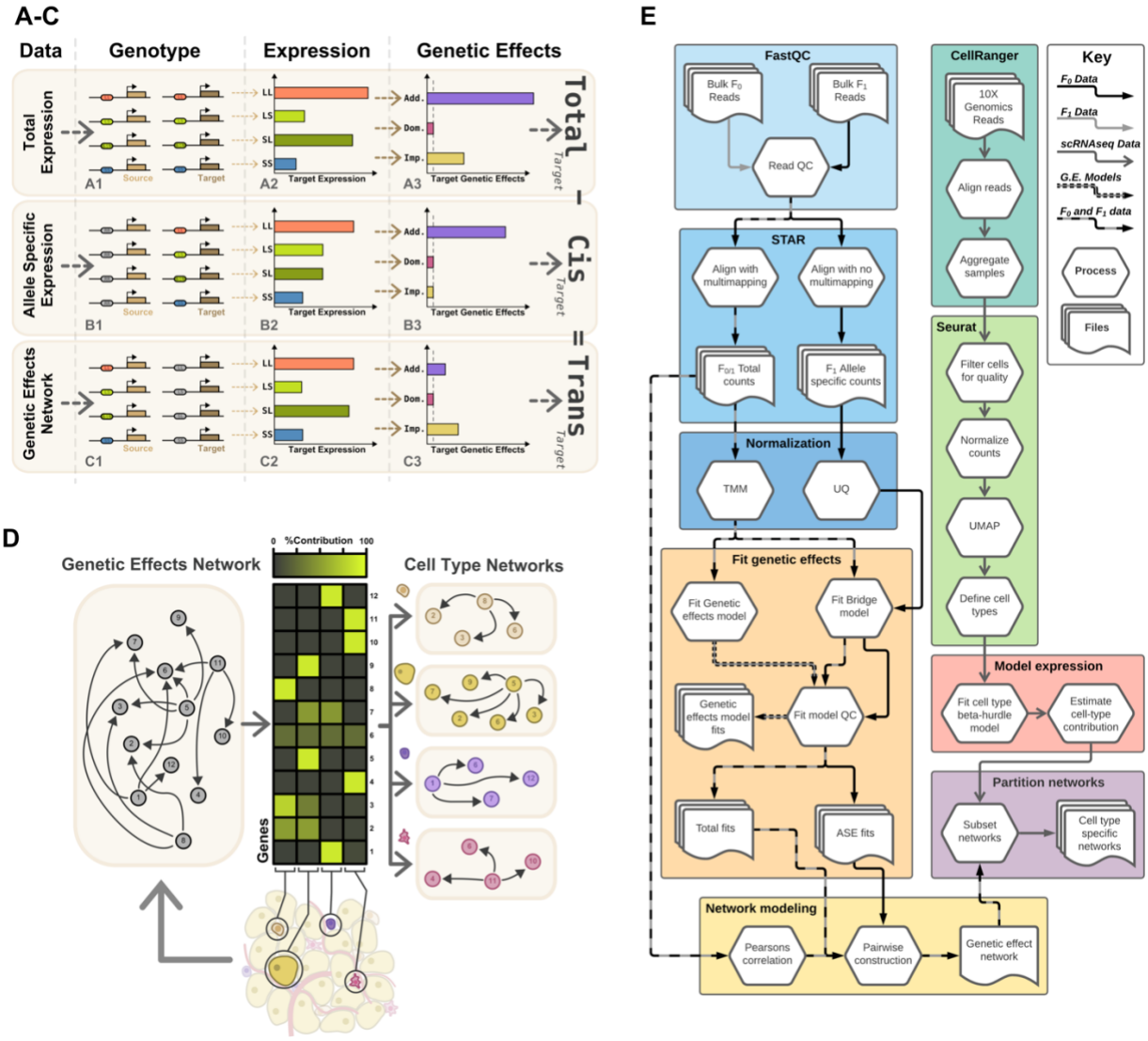


**Figure 3.3. General approach for genetic effect partitioning, network construction, and detangling cell type specific networks.** For some potential target gene, we use total expression to measure total genetic effects on the "Target" gene (**A1-3**). Using ASE data we isolate the *cis* effects on the "Target" gene (**B1-3**). Observed genetic effects that cannot be accounted for by *cis* effects, must be the result of a *trans* effect from some unobserved source gene(s) (**C1-3**). We construct a pairwise network where the *cis* effects of "Source" genes match the total genetic effect of "Target" genes. The resulting genetic effects network is a highly tangled composite of the networks functioning withing each cell type present in that tissue. By using single cell RNAseq we can calculate the cell type specificity of expression for each gene in a network (**D**). This allows us to split the genetic effects network into subnetworks including only the genes expressed within a given cell type. In this way we can estimate the genetic effects networks within a cell. This methodology integrates bulk total expression, bulk allele specific expression, and single cell expression data to identify genetic effects networks within cell types (**E**).

## 3.3 Results

### Total expression alignment

We tested three strategies for total expression alignments. The first was to align against just the SM/J genome. The second was to align against LG/J and SM/J simultaneously allowing for multi-mapping and taking the mean expression. The third strategy was to align against SM/J and LG/J separately, then align against LG/J and SM/J simultaneously but disallowing multi-mapping. We ran a linear model comparing total expression counts to ASE counts for each gene and calculated an $R^2$, which quantifies how well Total/ASE counts match. To compare alignment strategies, we compare $R^2$ distributions. The most appropriate strategy has the strongest agreement between ASE and total counts (Supplemental Figure 3.1 D-F), which translates to the highest proportion of $R^2$ values close to one. We chose the second alignment strategy, because it had the strongest agreement between count types. We saw no relationship between expression magnitude and agreement strength (Supplemental Figure 3.1 G) and we see strong agreement between ASE and total expression genotypic means (Supplemental Figure 3.1 H).

### Modeling total genetic effects

We fit the bridge model using total gene expression data from all four genotypic classes (LL = 4, LS = 4, SL = 4, SS = 4). There were 4,995 genes that met our QC requirements to perform fitting. 3,780 genes were considered significant ($FDR \leq 0.1$). We evaluated how well the bridge model fits using total expression to predict genotypic means by fitting a linear model relating estimated to actual genotypic means. The model reliably captures genotypic means ($R^2$=0.77, Supplemental Figure 3.3).

We tested if the bridge model can capture genetic effects as well as an established model[73]. We implemented a classical genetic effects model (GEM) to the same total gene expression data

(Supplemental Figure 3.4). There were 3,778 genes fit of which 1,942 showed at least one significant effect (imprinting 3.85%, dominance 56.5%, additive 61.5%). To quantify how well bridge can recapitulate genetic effects captured by the GEM, we compare GEM coefficients to bridge estimates (Supplemental Figure 3.5 A-C). The GEM reference coefficient (*R*) is reliably estimated by the bridge model ($R^2$=0.71) as well as imprinting ($R^2$=0.96) and additive ($R^2$=0.99) coefficients. To handle dominance effects the bridge model sums reference and dominance coefficients. We determined this by taking the residuals for reference, imprinting, and additive coefficients and then compared them to the dominance coefficients. Neither imprinting nor additive residuals can account for dominance ($R^2$=0 and 0 respectively, Supplemental Figure 3.5 E & F). Thus the reference coefficient residuals account for the dominance coefficients ($R^2$=0.93, Supplemental Figure 3.5 D).

## Modeling *cis* genetic effects

We fit the bridge model to allele specific expression data from the heterozygote genotypic classes (LS = 4, SL = 4). There were 4,856 genes which met our QC requirements to perform fitting, of which 4,846 had a significant overall fit ($FDR \leq 0.1$). We are interested in genes with significant fits for both total and ASE data. We were able to fit 4,855 genes using both total and allele specific expression data, of which 3,778 were significant for both total and ASE.

Using the bridge model with ASE data requires that we can be sure the bridge model is able to reliably capture ASE metrics. We tested this requirement using two established metrics for ASE [5,67]. We calculated ASE metrics for all genes fit and compared them to the bridge estimates (Supplemental Figure 3.7). The bridge model estimates reliably capture POE scores ($R^2$=0.98) and ancestry scores ($R^2$=0.95).

## Construct Genetic Effects Networks

We identified genes showing a total genetic effect on expression, some of which also show a *cis* genetic effect. Using these genes, we use Pearson's correlation tests to construct a first pass network (Supplemental Figure 3.8). There were 34,758 pairs which met our threshold with $\rho \geq 0.7$, representing 1,317 unique genes. Combining first pass network structure, total, and *cis* genetic effects we can refine and improve on the first pass network. By partitioning genetic effects into *cis* and *trans* we can identify "Source" genes which have a *cis* effect sufficient to explain the observed total effect. In a pathway we would expect that the genetic effects on a "Source" gene should be in some way reflected in downstream genes. We might expect that if a gene is truly a downstream "Target" gene, it will show a total genetic effect but lack a *cis* effect. This discrepancy between *cis* and total effects must be the result of some unobserved or "missing" *trans* effect. Such a *trans* effect would be the result of pathway structure and the presence of a *cis* effect on an upstream regulator ("Source" gene). For a given gene pair, we can test if the total genetic effects on a "Target" gene are better explained by its *cis* effects or if they are better explained by the *cis* effects present in the upstream "Source" gene. Running all pairwise tests from the first pass network, we can construct a genetic effects network in which edges represent the flow of genetic effects along unobserved pathway structure. We perform pairwise genetic effect partitioning using five models and analysis of variance (supplemental Figure 3.9). We tested 34,758 gene pairs representing 1,317 unique genes. After thresholding we were left with 8,996 gene pairs representing 325 source genes and 299 target genes, of which 8,561 genes pairs met our significance thresholds ($FDR \leq 0.05$).

## Identifying cell type specific genetic effects in networks

Pathways occur within cells, but our genetic effects network represents data from whole tissue, which is comprised of a variety of different cell types. We integrate publicly available epididymal adipose tissue single cell RNA-seq data to further refine our network. Cell types were defined

using marker genes reported previously[118] (**Figure 3.4 B**). The cell single data set is composed of 23.5% macrophages, 35% ASC, 6.3% Differentiating ASC, 11% VEC, 8.9% VSMC, 8.1% T cell, and 7.3% B cell (**Figure 3.4 D**). We use single cell data to estimate the fraction of a gene's expression that is derived from different cell types.

To this end we needed to model expression of each gene within each cell type as a beta-hurdle model. Modeling expression in this way lets us estimate the expected amount of expression from a given cell type if we know cell composition (Islet paper). There were 11,837 genes fit across 7 cell types. As a quality control, we plot expected values of each gene in each cell type against goodness of fit (ks statistic) and colored by dropout rates (Supplemental Figure 3.11). The beta models were good fits, given the small ks values. Goodness of fit appears largely to be affected by the probability of dropout. The lower the dropout rate, the better the model can be fit. Beta hurdle model statistic distributions are shown in supplemental Figure 3.12. Examples of genes this approach reports are overwhelmingly expressed by a particular cell type and are listed in in supplemental Figure 3.13.

We took the white adipose tissue genetic effects network we identified in the bulk RNAseq data and subset out all genes for which we had single cell data (**Figure 3.4 A**) to partition the network into subnetworks corresponding to the 7 cell types in the single cell data (**Figure 3.4 B**).

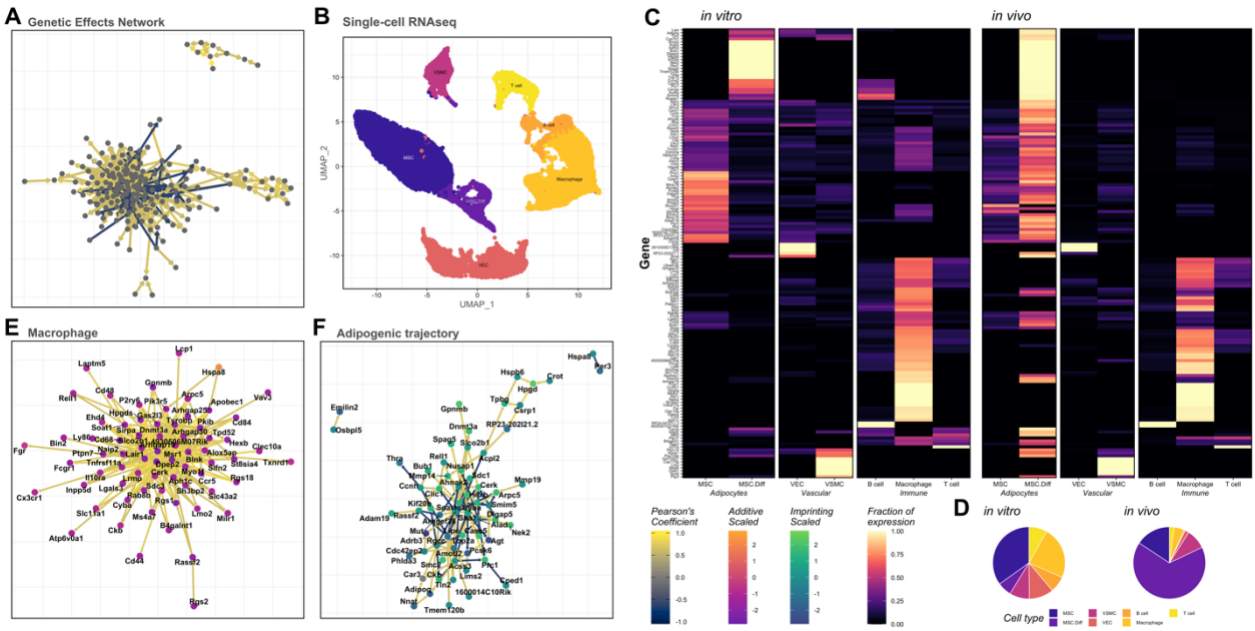


**Figure 3.4. Genetic effect partitioning, *trans*-network construction, and integration of cell type contribution data reveal cell type specific genetic effects networks.** Genetic effects network across to cell types (**A**). Each node represents a gene and each represents a potential *trans* effect between a "Source" and a "Target". Line color indicates correlation magnitude and direction. UMAP representation of single cell RNAseq data showing relevant cell types (**B**). Cell type specific contributions to expression of network genes measured *in vitro* as well as estimated contributions using *in vivo* cell type composition (**C**). Cell color denotes the fraction of total expression is expected to come from that cell type. Comparison of *in vitro* versus *in vivo* cell type composition (**D**). Macrophage additive genetic effects network (**E**). Node color indicates scaled total additive genetic effect magnitude and direction. Adipogenic imprinting genetic effects network (**F**). Node color indicates scaled total Imprinting genetic effect magnitude and direction.

For each gene in each cell type, we estimate the fraction of bulk expression we would expect to come from that cell type (**Figure 3.4 C**). This was done for all genes in the final white adipose tissue genetic effects network. The fraction of expression is dependent on cell composition expected in the bulk. For this reason, we used two sets of cell compositions; *in vitro* observed composition and estimated *in vivo* composition (**Figure 3.4 D**). We do this to get a sense for how bulk *in vivo* data might differ from the *in vitro* single cell data. Using the *in vitro* composition, we subset the networks into cell type networks. To include a given gene in a cell type-specific network

we required that cell type contribute at least 25% of that gene's *in vitro* expression. We selected 0.25 as our cutoff because at that threshold most included genes were expressed in 1-2 cell types (supplemental Figure 3.12 E). We found 1,104 gene pairs had sufficient single cell data and could be split into cell type-specific networks, representing 199 unique genes. Split networks are as follows: Macrophage network = 314 pairs and 73 genes (**Figure 3.4 E**); Adipogenic network = 147 pairs and 74 genes (**Figure 3.4 F**); B cell network = 9 pairs and 10 genes; T cell network = 5 pairs and 6 genes; VEC network = 6 pairs and7 genes; VSMC network = 6 pairs and 8 genes.

## *Cis* genetic effects on *Msr1* influence *Gpnmb* expression

Our approach identifies gene pairs in which a genetic effect on a "Source" gene drives the expression pattern we observe on a "Target" gene. For example, *Msr1* is a candidate "Source" gene. It shows a total (*cis* + *trans*) genetic effect (**Figure 3.5 A**) where the S allele is more highly expressed than the L allele. By isolating the *cis* effect on expression, we see that the *cis* expression pattern (**Figure 3.5 B**) matches the total expression pattern (**Figure 3.5 A**). Random sampling from fit distributions allows us to compare additive and imprinting genetic effect coefficients from either total or *cis* expression (**Figure 3.5 G**). The observed *Msr1* genetic effects coefficients (dashed lines) match the estimated coefficient distributions (boxplot) for both total and *cis* effects. Together this makes *Msr1* a candidate "Source" gene because it has a *cis* effect that matches its total genetic effect.

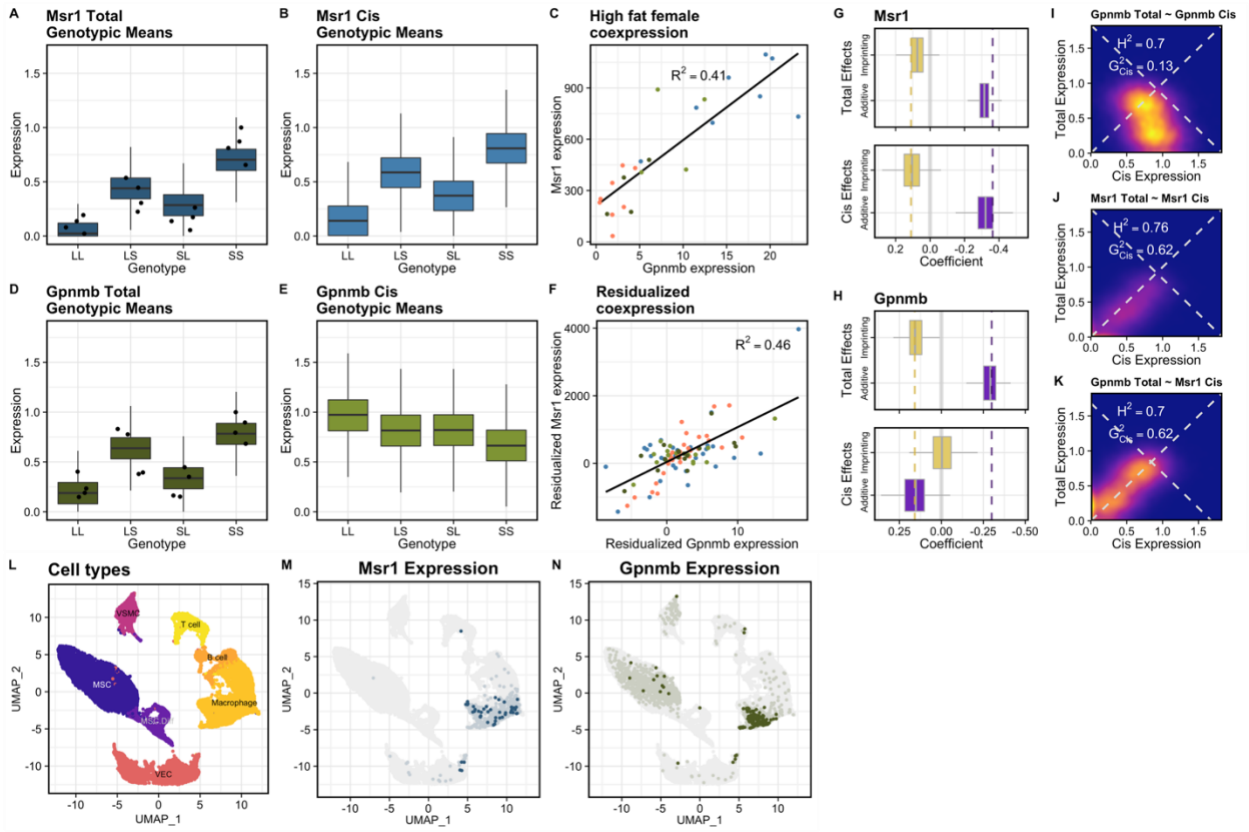


**Figure 3.5. Additive genetic effect propagates from *Msr1* to *Gpnmb* through macrophage lipid accumulation.** Total expression of "Source" gene *Msr1* (**A**) and "Target" gene *Gpnmb* (**D**). Bridge model estimation of total expression distribution for each genotypic class visualized as boxplots. Bridge model estimation of *cis* expression distribution for each genotypic class visualized as boxplots (**B & E**). *Msr1* expression is strongly predictive of *Gpnmb* expression in both high fat fed females (**C**) and across contexts when effects of sex, diet, and genotype are removed (**F**). *Msr1* total genetic effect coefficients estimated by Bridge (boxplot) match genetic effects model fit coefficients (dashed line) for both additive (purple) and imprinting (yellow) effects (**G**). *Cis* genetic effect coefficients for *Msr1* match total genetic effect estimates, suggesting the absence of *trans* effects. *Gpnmb* total genetic effect coefficients estimated by Bridge (boxplot) match genetic effects model fit coefficients (dashed line) for both additive (purple) and imprinting (yellow) effects (**H**). *Cis* genetic effect coefficients for *Msr1 do not* match total genetic effect estimates, suggesting the presence of *trans* effects. Genetic effects partitioning shows *Gpnmb* has a high heritability, but only 13% of heritability can be accounted for by bridge estimated *cis* effects suggesting the presence of some "missing" *trans* effect(s) (**I**). *Msr1* on the other hand has both a high heritability, 62% of which can be accounted for by bridge estimated *cis* effects (**J**). Given the strong genotype independent co-expression of *Msr1* and Gpnmb (**F**), *Msr1* is a candidate that may account for missing *Gpnmb trans* effects. Genetic effects partitioning shows that 62% of *Gpnmb* heritability can be accounted for by bridge estimated *Msr1 cis* effects (**I**). Suggesting *Msr1* regulates *Gpnmb* expression and genetic effects originating from *Msr1* propogate onto *Gpnmb*

expression. To further strengthen this candidate pair, co-expression is examined in white adipose single-cell RNAseq (**L**). which shows that *Msr1* and *Gpnmb* are co-expressed in macrophages (**M & N**).

We identified initial "Target" candidates by Pearson's correlation testing on residualized expression of all gene pairs showing total genetic effects. Residualized *Gpnmb* expression is significantly correlated with *Msr1* ($R^2$=0.48, **Figure 3.5 F**). Co-expression of *Msr1* and *Gpnmb* persists in unresidualized expression within high fat-fed females ($R^2$=0.41, **Figure 3.5 C**). *Gpnmb* shows a total genetic effect on its expression (**Figure 3.5 D**) and its expression pattern matches that of *Msr1*. Unlike *Msr1*, when we isolate the *cis* genetic effects on *Gpnmb*, we see its expression pattern does not match its total expression pattern (**Figure 3.5 E**). While the predicted additive and imprinting coefficients match the observed true coefficients (**Figure 3.5 H**), the *cis* predicted coefficients do not match. In other words, there is a significant additive effect in *Gpnmb*'s total genetic effects, but this cannot be accounted for by a *cis* effect. This indicates a *trans* effect on *Gpnmb* expression. We quantify this comparison (**Figure 3.5 I**) and observe that *Gpnmb* has a sufficiently high $H^2$. The estimated *cis* genetic effect on *Gpnmb* can only account for 13% of heritability. For *Msr1* (**Figure 3.5 J**) we find the *cis* genetic effect can account for 61% of its heritability. Can the *cis* genetic effect on *Msr1* be serving as a *trans* effect on *Gpnmb*? We test this (**Figure 3.5 K**) and see that the *cis* effect from *Msr1* can account for 62% of the heritability of *Gpnmb*. This makes *Gpnmb* a candidate "Target" gene of *Msr1*. Single cell data indicates *Msr1* and *Gpnmb* are overwhelmingly expressed by macrophages (**Figure 3.5 L-N**).

## Macrophage network enriched for immune activation

We wanted to know what this network might reveal about SM/J and LG/J physiology. In the macrophage network we see that except for *Hspa8*, all other genes show an additive effect where

the SM/J allele is more highly expressed. We performed over representation analysis on these genes and found that leukocyte activation was significantly enriched (FDR=0.003)[80].

## 3.4 Discussion

In this study we sought to model genetic effects on expression in the context of networks. We demonstrated that genetic effects propagate through networks shaped by molecular pathways. The more interconnected biological systems are, the more widespread and important propagation becomes. By comparing additivity on total expression to sequence dependent ASE, inferred the prevalence of trans effects on white adipose tissue gene expression (**Figure 3.2 A**). A large quantity of genes (~38.8%) showed clear evidence of possible trans genetic effects. The implication is that expression regulation in white adipose tissue is highly interconnected and complex.

We demonstrated that the bridge model captures genetic effects just as effectively as a GEM and effectively uses ASE data. We put forth Bridge as a viable model to related total expression, ASE, and genetic effects. The standard bridge model is very much focused on additive and imprinting genetic effects, but we demonstrate how dominance can be captured as well (Supplemental 5D-F Figure).

Interestingly, the adipogenic cell and macrophage networks were much larger than those for other cell types. This could mean that genetic difference between SM/J and LG/J affect most directly affect these cell types. But macrophages and adipogenic cells also have a higher proportion of cell type specific genes in this data set overall (Supplemental figure 3.12), which would increase the likelihood of capturing sufficient expression of genes specific to those cell types. This highlights a limitation to the viability of the cell type genetic effects network approach in other

settings. The methodology is likely to be more informative in samples with moderate cell type complexity unless a particularly large number of cells can be sequenced.

Where candidate gene pairs supported by the literature? We highlighted *Msr1* and *Gpnmb* as an example of this method recapitulating previous studies. Our study predicted that *Msr1* regulates *Gpnmb* expression in macrophages. *Msr1* is expressed in macrophages and is required for macrophages to take on lipid-laden phenotype [141–144]. Accumulation of lipids in macrophages induces expression *Gpnmb* [143]. By promoting the lipid laden phenotype elevated *Msr1* expression promotes expression of *Gpnmb*, an established biomarker of lipid laden macrophages. The exact cascade of events between *Msr1* activation and *Gpnmb* expression are unknown, but a there is a clear connection between them. Our model showed not only that they are related, but also the causality (*Msr1* → *Gpnmb*) and the directionality (positive correlation). All of which is supported by previous studies.

The enrichment of the macrophage network for leukocyte activation suggests that compared to the LG/J mouse line, SM/J animals have elevated expression of genes involved in macrophage activation which should translate to stronger activation of macrophages. We turn to a study which evaluated macrophage activation across a variety of mouse lines including both LG/J and SM/J[145]. It revealed that SM/J mice have a very strong activation response while LG/J mice have a very low response, which is consistent with our observations of the genetic effects in this pathway.

Candidate gene pairs/networks identified in this way are primed for experimental validation. Not only does this approach suggest causality/directionality in candidate networks but is able to identify the putative cell type(s) of interest. Using these predictions, Trained

experimentalists can then isolate the specific cell type(s) and perform over/under expression assays for candidate genes to begin to dissect the validity and specific mechanisms underlying predictions.

## 3.5 Methods

### Mouse husbandry

LG/J and SM/J founders ($F_0$) were obtained from The Jackson Laboratory (Bar Harbor, ME). $F_1$ reciprocal cross animals were generated by mating LG/J mothers with SM/J fathers (LxS) and the inverse (SxL). At three weeks of age, animals were weaned into same-sex cages and randomly placed on high-fat (42% kcal from fat; Teklad TD88137) or low-fat (15% kcal from fat; Research Diets D12284) isocaloric diets. At 20 weeks of age, animals were given an overdose of sodium pentobarbital after a 4 hour fast and blood was collected via cardiac puncture. Euthanasia was achieved by cardiac perfusion with phosphate-buffered saline. After cardiac perfusion, the reproductive fat pad was harvested, flash frozen in liquid nitrogen, and stored at -80°C.

### RNA sequencing

Total RNA was isolated from adipose tissue using the RNeasy Lipid Tissue Kit (QIAgen) (n = 32, 4 animals per sex/diet/cross cohort). RNA concentration was measured via NanoDrop and RNA quality/integrity was assessed with a BioAnalyzer (Agilent). RNA-Seq libraries were constructed using the RiboZero kit (Illumina) from total RNA samples with RIN scores >8.0. Libraries were checked for quality and concentration using the DNA 1000LabChip assay (Agilent) and quantitative PCR, according to manufacturer's protocol. Libraries were sequenced at 2x100 paired end reads on an Illumina HiSeq 4000. After sequencing, reads were de-multiplexed and assigned to individual samples. RNAseq data are available through the NCBI-SRA, accession: PRJNA753198 & PRJNA661764

.

## Genomes and annotations

LG/J and SM/J indels and SNVs were leveraged to construct strain-specific genomes using the GRC38.72-mm10 reference as a template [125]. This was done by replacing reference bases with alternative (LG/J | SM/J) bases using custom python scripts. Ensembl R72 annotations were adjusted for indel-induced indexing differences for both genomes.

## Alignment

We aligned the RNAseq reads against LG/J and SM/J genomes simultaneously allowing for multi-mapping and taking the mean expression. The ASE alignment was done as reported in [67]The total gene count (L+S) from the ASE data was compared to count from the three total expression alignment strategies (Supplemental Figure 3.1 A-C).

## Modeling expression of genotypic classes

The bridge model is intended to model genetic effects on expression. We can use maximum likelihood estimation to fit bridge model parameters ($\beta_0, \beta_L, \beta_S, R_♀, R_♂$) and within genotypic class variation ($\sigma$). We describe the expression distribution for each genotypic class with mean ($\mu_{GG}$) and standard deviation ($\sigma$).

$$Expression = Bridge(\boldsymbol{\beta_0}, \boldsymbol{\beta_L}, \boldsymbol{\beta_S}, \boldsymbol{R_♀}, \boldsymbol{R_♂}, L_♀, S_♀, L_♂, S_♂) \quad (3.4)$$

$$Expression_{GG} = \mathcal{N}(\mu_{GG}, \sigma) \quad (3.5)$$

The MLE procedure directly fits $\sigma$. We assume roughly equal variances amongst genotypic classes. The mean for each class is calculated using bridge parameters (**Figure 3.1 C**).

$$\sigma_{LL} \approx \sigma_{LS} \approx \sigma_{SL} \approx \sigma_{SS} \quad (3.6)$$

$$\mu_{LL} = \beta_0 + \beta_L R_{♀} + \beta_L R_{♂} \qquad (3.7)$$

## Capturing total genetic effects with bridge

We implemented the bridge model to capture total genetic effects. We fit the bridge model using total gene expression data from all four genotypic classes (LL = 4, LS = 4, SL = 4, SS = 4). For each gene, expression is scaled to fall in the range 0 to 1. There are 6 parameters in the bridge model. We use maximum likelihood estimation with a grid search to fit parameters. Overall fit significance was estimated by likelihood ratio tests assuming wilks theorem.

$$H_0: \ y \sim \beta_0 \qquad (3.7)$$

$$H_1: \ y \sim \beta_0 + \beta_L L_m R_m + \beta_L L_p R_p + \beta_S S_m R_m + \beta_S S_p R_p \qquad (3.8)$$

$$LR = \frac{\mathcal{L}(H_0|X)}{\mathcal{L}(H_1|X)} \quad (3.9)$$

Multiple tests correction was performed by estimating false discovery rate (FDR). We required a model to have an $FDR \leq 0.1$ to consider it significant. We evaluated how well bridge model fits using total expression predicts genotypic means, by fitting a linear model relating estimated to actual genotypic means (Supplemental Figure 3.3).

## Capturing total genetic effects with classic genetic effects model

We implemented GEM as an established way to capture genetic effects. We do this because we want to be sure that the bridge model can capture genetic effects as well as the more traditional approach. We fit the GEM using total gene expression data. Genetic effect coefficients are fit as this general linear model. Overall model significance was estimated by LRT. Multiple tests corrections were performed using the FDR method. Genetic effects were considered significant if their $FDR \leq 0.1$ (Supplemental Figure 3.4).

## Compare Bridge model to classic genetic effects model

To evaluate how well bridge can recapitulate genetic effects captured by the GEM, we compare GEM coefficients to bridge estimates (Supplemental Figure 3.5). The GEM reference coefficient (*R*) is estimated by bridge parameters using the following equation.

$$\hat{R} = \frac{2\beta_0 + (R_p + R_m)(\beta_L + \beta_S)}{2} \quad (3.10)$$

The GEM imprinting coefficient (*I*) is estimated by bridge parameters using the following equation.

$$\hat{I} = \frac{(R_p - R_m)(\beta_L - \beta_S)}{2} \quad (3.11)$$

The GEM additive coefficient (*A*) is estimated by bridge parameters using the following equation.

$$\hat{A} = \frac{(R_p + R_m)(\beta_L - \beta_S)}{2} \quad (3.12)$$

## Capturing *cis* genetic effects with bridge model

We implement the bridge model to capture *cis* genetic effects. We fit the bridge model to allele specific expression data from the heterozygote genotypic classes (LS = 4, SL = 4). For each gene, expression is scaled to the range from 0 to 1. Alleles within individuals are treated as separate observations, meaning that with 8 individuals we fit the model using 16 observations. We use maximum likelihood estimation with a grid search to fit parameters.

Overall fit significance was estimated by LRT. Multiple tests correction was performed by the FDR method. Models were considered significant with an FDR<=0.1.

Ancestry and POE scores are useful and established metrics for ASE, but we need to capture ASE using metrics that can be directly compared to total expression. The bridge model was built for integrating allele specific and total gene expression data for the purpose of partitioning genetic effects. We estimate ASE scores from bridge parameters like so:

$$POE\ Score \approx R_{♀} - R_{♂} \qquad (3.13)$$

$$Ancestry\ Score \approx \beta_S - \beta_L \quad (3.14)$$

We compare ASE metrics for to bridge estimates for ASE metrics (Supplemental Figure 3.7). We fit a linear model comparing the estimated to the actual metrics [67].

$$POE\ Score = \left(\frac{L_{LS}}{L_{LS}+S_{LS}}\right) - \left(\frac{L_{SL}}{L_{SL}+S_{SL}}\right) \in [-1,1] \qquad (3.15)$$

$$Ancestry\ Score = \left(\frac{L_{LS}}{L_{LS}+S_{LS}}\right) - \left(\frac{S_{SL}}{L_{SL}+S_{SL}}\right) \in [-1,1] \quad (3.16)$$

## Initial Network Construction

We identify genes showing a total genetic effect on expression, some of which also show a *cis* genetic effect. Using these genes, we use Pearson's correlation tests to construct a first pass network. We wanted this initial network to be as robust as possible, so we used RNAseq data from all combinations of Sex, Diet, and Genotype (n=64) to construct it. However, Pearson's correlation requires data not have the sort of substructure that Sex, Diet, and Genotype contexts produce. To meet this requirement and to yield the strongest gene pairs, we residualized out the effects of Sex, Diet, and Genotype. Pearson's correlation tests were then performed using data with context effects removed for all gene pairs. We required a pair to have a correlation value |p|>=0.7 to be considered.

## Modeling expression of genotypic classes

The bridge model is intended to model genetic effects on expression. We can use maximum likelihood estimation to fit bridge model parameters ($\beta_0, \beta_L, \beta_S, R_{♀}, R_{♂}$) and within genotypic class variation ($\sigma$) to describe the expression distribution for each genotypic class with mean ($\mu_{GG}$) and standard deviation ($\sigma$). The MLE procedure directly fits $\sigma$. We assume roughly equal variances amongst genotypic classes. The mean for each class is calculated using bridge parameters.

## Estimating bridge derived genetic effect coefficient distributions

By fitting bridge using both ASE and total expression data, we get two theoretical distributions for each genotypic class, one shaped by *cis* + *trans* effects and a second shaped by *cis* effects alone: with a total of 8 distributions per gene.

$$Expression\ distributions \begin{cases} Cis + Trans, \begin{cases} LL,\ \mathcal{N}(\mu_{LL,Total}, \sigma_{Total}) \\ LS,\ \mathcal{N}(\mu_{LL,Total}, \sigma_{Total}) \\ SL,\ \mathcal{N}(\mu_{LL,Total}, \sigma_{Total}) \\ SS,\ \mathcal{N}(\mu_{LL,Total}, \sigma_{Total}) \end{cases} \\ Cis\ only, \begin{cases} LL,\ \mathcal{N}(\mu_{LL,Cis}, \sigma_{Cis}) \\ LS,\ \mathcal{N}(\mu_{LL,Cis}, \sigma_{Cis}) \\ SL,\ \mathcal{N}(\mu_{LL,Cis}, \sigma_{Cis}) \\ SS,\ \mathcal{N}(\mu_{LL,Cis}, \sigma_{Cis}) \end{cases} \end{cases} \quad (3.17)$$

We randomly sample from these theoretical distributions 1,000 times for each genotype (4,000 total). This allows us to plot the theoretical distributions for each genotypic class. We then bootstrap taking subsamples (N=16, fully balanced classes) from these data. At each iteration we calculate additivity and imprinting genetic effects (500 iterations).

## Quantifying genetic effect partitioning

Total expression parameters from bridge are used to calculate the shape parameters of expression distributions as described above. Theoretical distributions for each genotypic class can be

compared to the observed total expression of individuals. The same procedure is conducted using allele specific parameters. This gives us shape parameters for the expression distributions we would observe if we could control for *trans* effects. In other words, we can describe expression patterns between genotypic classes as they would be if there were no *trans* effects at play.

We can compare genetic effects measured using total expression (*cis* + *trans*) to those measured using allele specific expression (*cis* only). If there is a difference, we conclude there must be some "missing" *trans* effect which should account for the difference. But how do we quantify comparisons?

Using MLE, we can estimate the distributions of expression or each genotypic class at both the level of *cis* expression and total expression. For a two-locus system we want to identify cases of genetic effects propagating as *trans* effects from a "Source" gene to a "Target" gene. We use analysis of variance to identify and classify pair-wise relationships. Data are generated by randomly sampling from the 16 *cis* and total distributions describing the two genes.

To describe a pair-wise relationship ("Source" x "Target"), we fit 5 models on either *cis* or total expression. Model 1 and Model 2 are used to calculate the fraction of total expression variance that genotype explains ($R^2$) for "Source" and "Target" genes, respectively. This approximates heritability ($H^2$). These two measures ($H^2_{source}$, $H^2_{target}$) are the maximum proportions of explainable variance. These serves as references when making classifications.

Model 3 ("Source") and Model 4 ("Target") are used to determine directionality and to partition *cis*/*trans* effects. Each model is used to calculate the fraction of variation in total expression for the response gene that can be accounted for by its own *cis* expression plus the *cis* expression of the other gene. Model 4 measures what proportion of variation in target gene total

expression is explained by its own *cis* expression (I) versus what proportion is explained by source gene *cis* expression (K). We expect *cis* expression of the source gene will explain a large proportion of its total expression (J), while *cis* expression of the target gene will explain very little of source gene total expression. *Cis* expression of the source gene should also explain a large proportion of target gene total expression (K), while *cis* expression of the target gene should explain less of its own total expression (I). In model 5, we calculate how much "Target" gene Total expression variation can be explained by "Source" gene total expression. For co-expressed genes this should be quite high. Each of these metrics is represented as $R^2$, which we scale by the $H^2$ of the response gene. We call this scaled percent of variance explained $G^2$. For example, a $G^2_{Target,Source}$ of 1 means that 100% of target gene heritability can be accounted for by Source gene *cis* expression.

## Refine network by genetic effect partitioning

Combining first pass network structure, total, and *cis* genetic effects we can refine and improve on the first pass network. We are interested in modeling and detecting genetic effects within networks. By partitioning genetic effects into *cis* and *trans* we can identify "Source" genes which have a *cis* effect sufficient to explain the observed total effect. In a pathway we would expect that the genetic effects on a "Source" gene should be in some way reflected in downstream genes. We might expect if a gene is truly a downstream "Target" gene, that it will show a total genetic effect but lack a *cis* effect. This discrepancy between *cis* and total effects must be the result of some unobserved or "missing" *trans* effect. Such a *trans* effect would be the result of pathway structure and the presence of a *cis* effect on an upstream regulator ("Source" gene).

For a given gene pair, we can test if the total genetic effects on a "Target" gene are better explained by its *cis* effects or if they are better explained by the *cis* effects present in the upstream

"Source" gene. Running all pairwise tests from the first pass network, we can construct a genetic effects network in which edges represent the flow of genetic effects along unobserved pathway structure. We perform pairwise genetic effect partitioning using five models and analysis of variance (supplemental Figure 3.9).

The first two models serve to determine how much variation in total expression can be accounted for by genotype. This represents a rough estimate of heritability ($H^2$) and serves as the upper limit of accountable variance. Because we are interested in variation between genotypic classes, rather than within them, we scaled the fractions of accountable variance ($R^2$) for each term by heritability. As such, terms are evaluated by the fraction of accountable heritability ($G^2$).

Model 5 serves to see if the gene pair is significantly co-expressed when genotype is not residualized out. Models 3 and 4 are where the pairwise genetic effect partitioning occurs. They differ in that the "Source" gene is the response variable for model 3, whereas the "Target" gene is the response variable in model 4. The terms of both models are the same. The terms are *cis* expression of the "Source" gene and *cis* expression of the "Target" gene. Analysis of variance is performed for each model and an $R^2$ is calculated for each term.

The $R^2$ from models 3, 4, and 5 are scaled by $H^2$ for the response variable from models 1 and 2 to yield $G^2$ for each term in models 3, 4, and 5. Significance of these $G^2$ was estimated by randomization tests in which expression values are randomly sampled from fit expression distributions of the 4 genotypic classes and then genotype is shuffled. The $G^2$ are calculated for this randomized data. The procedure is then repeated 100 times, which produces a null distribution of $G^2$ values generated for each gene pair.

Multiple tests corrections were then performed by estimating a false discovery rate for each term. An $FDR \leq 0.05$ was considered significant (supplemental Figure 3.10). Pairs were thresholded by requiring the following. Model 3 source *cis* $G^2$ ($G^2_{3,source}$)≥0.15, $G^2_{4,target}$ <0.15, $G^2_{5,source}$ ≥0.4, $H^2_{source}$ ≥0.4, and $H^2_{target}$ ≥0.4 (Supplemental Figure 3.9). We also calculate the relative *cis* contributions for models 3 and 4. We call this ratio the genetic partition coefficient ($g$).

$$g_3 = \frac{G^2_{3,source}}{G^2_{3,target}} \qquad (3.18)$$

$$g_4 = \frac{G^2_{4,target}}{G^2_{4,source}} \qquad (3.19)$$

This quantifies the relative contribution of *cis* versus *trans* effects for each response gene. A value greater than 0 means the *cis* effect is larger than the candidate *trans* effect. We require $Log_2(\ g_3\ ) \geq 1$ and $Log_2(\ g_4\ ) \leq 0$. Which means we require a target gene have greater than half of its accountable heritability accounted for by the *trans* effect. We also require a source gene have less than 1/3 of it's accountable heritability accounted for by the *trans* effect (supplemental Figure 3.10). By thresholding network edges by significance and partition coefficients we are left with a directional genetic effects network for white adipose tissue.

## Integrating single cell RNAseq

Our goal is to model how genetic effects flow through pathways. Pathways occur within cells, but our genetic effects network represents the whole tissue, which is composed of a variety of different cell types. We integrate publicly available epididymal adipose tissue single cell RNA-seq data to further refine our network. Single cell RNAseq data was processed as reported in. Cell types were defined using marker genes reported previously [118].

Our goal was to use single cell data to estimate what fraction of a gene's expression is derived from different cell types. To this end we needed to model expression of each gene within each cell type. Expression was scaled to the range of 0 to 1 across all cell types. Scaled expression within a subtype is approximately Beta distributed amongst cells expressing the gene. Due to gene dropout in the sequencing process and cell to cell variability, some cells have zero expression a given gene that is expressed by other cells of the same type. To best deal with this data, we implement a beta-hurdle model as reported in [146]. Briefly, the model has two parts, it models the probability of a cell being a dropout and if it is not a dropout, it models expression as a beta distribution.

$$Expression \sim f(\ \Pr(Dropout|0)\ ,\ Beta(\alpha,\beta)\ ) \quad (3.20)$$

For each gene we calculated the $\Pr(0)$, then we iterate through all possible $\Pr(Dropout|0)$. In each iteration we use $\Pr(Dropout|0)$ and the ks package to fit $\alpha$ and $\beta$ parameters. This package performs a maximum goodness of fit procedure by minimizing the deviation between real and fit cumulative distributions (ks statistic).

The lower the ks statistic, the better the fit. The set of $\Pr(Dropout|0)$, $\alpha$, and $\beta$ parameters that produces the lowest ks statistic was selected as the optimal fit. $\alpha$ and $\beta$ parameters were used to calculate the expected expression value of a cell given it is not a dropout.

$$Expected\ macrophage\ expression: [\#Macrophages] * \Pr(Dropout|0) * \mathrm{E}(\alpha_{Macrophage}, \beta_{Macrophage}) \quad (3.21)$$

## Splitting genetic effects networks

We take the white adipose tissue genetic effects network and subset out all genes for which we had single cell data. We want to subset this network into subnetworks corresponding to the 6 cell

types in the single cell data. For each gene in each cell type, we estimate the fraction of bulk expression we would expect to come from that cell type. This was done for all genes in the final white adipose tissue genetic effects network. The fraction of expression is dependent on cell composition expected in the bulk. For this reason, we used two sets of cell compositions; *in vitro* observed composition and estimated *in vivo* composition. We do this to get a sense for how bulk *in vivo* data might differ from the *in vitro* single cell data. Using the *in vitro* composition, we subset the networks into cell type networks. To include a given gene in a cell type of network we simply require that cell type contribute at least 25% of that gene's *in vitro* expression. We selected 0.25 as our cutoff because at that threshold most included genes were expressed in 1-2 cell types (supplemental Figure 3.12 E).

## Functional enrichment analysis

Functional enrichment of interacting genes showing parent-of-origin-dependent allele-specific expression with biallelic genes that are differentially expressed by cross was tested by over-representation analysis in the WEB-based Gene Set Analysis Toolkit v2019 [80]. We performed analyses of gene ontologies (biological process, cellular component, and molecular function), pathway (KEGG), and phenotype (Mammalian Phenotype Ontology). The list of all unique interacting genes was analyzed against the background of all unique genes expressed in white adipose. A Benjamini-Hochberg FDR-corrected p-value $\leq 0.01$ was considered significant.

## Capturing dominance genetic effects

In the context of $F_0/F_1$ genetics, dominance is a special case trans effect that can occur in two ways. At a given locus there are two alleles, one from each parent (on different chromosomes). A trans effect between such corresponding alleles can produce a dominance genetic effect. It is a trans affect between one copy of a gene and it's mate. Alternatively, the regulatory feedback of a gene's expression could behave like a sigmoidal dose response curve. If either of the alleles is strong enough to illicit a plateaued response, that would produce a dominance genetic effect. While interesting and biologically relevant, special care needs to be taken when interpreting dominance.

$$y \sim \beta_0 + \beta_L L_m R_m + \beta_L L_p R_p + \beta_S S_m R_m + \beta_S S_p R_p + \delta\left(1 - \left[\left(L_m - S_p\right)\left(S_m - L_p\right)\right]^2\right) \quad (3.22)$$

$$\widehat{D} = \delta$$

# Chapter 4: Conclusions and future directions

One day, we would like to build a predictive model of metabolic phenotypes that accounts for parent-of-origin effects and is generalizable across tissues. This is a major challenge because parent-of-origin effects can depend on genetic, epigenetic, and environmental factors. To unravel these intertwined effects, we developed a robust system for quantifying and modeling parent-of-origin effects on metabolic traits in mice. In line with the pleiotropic pathway explanation, was the extent to which non-imprinted genes contribute to parent-of-origin effects through interactions with imprinted genes.

Untangling the genetic, epigenetic, sex, and environmental components of these interactions will improve our ability to predict complex phenotypes from genomic sequence and will nominate variants for functional testing. The patterns of correlations we observe, in conjunction with the supporting mouse QTL, are strong evidence that sequence variants in *Nnat* and/or its non-imprinted interacting genes can result in parent-of-origin expression biases that may impact adipose function. But how exactly do these genes interact? how exactly do their interactions affect phenotypic variation? If we had a molecular interpretation of these phenomena, we could better understand the phenotypic consequences of these interactions and build better predictive models. For example, *Nnat* function enhances or attenuates cellular response to external signals by altering $Ca^{2+}$ kinetics. Could this suggest something about the role imprinted genes fill at the molecular level? Perhaps the search for parent-of-origin effects should focus on receptors and ligands?

We identified robust candidates, but this work would benefit greatly from molecular validation. Both to validate our predictions and to learn more about the molecular function of

candidate genes. An efficient platform should be developed to test candidate pairs and mechanisms. Implementing CRISPR technology, gene expression can be altered, promoters can be swapped, genetic effects can be isolated.

Future work will doubtless integrate whole-genome sequence with methylation, transcriptomic and phenotypic data. This will hopefully shed light on regulatory features underlying parent-of-origin dependent and context specific ASE. In cases where regulatory features are known, targeted DNA methylation could be used to introduce or remove parent-of-origin dependent ASE[147]. We believe that by understanding *how* parent-of-origin effects manifest at imprinted loci, how they are transduced through interactions with non-imprinted genes and propagated to phenotype we will improve our ability to predict phenotype from genotype across a wide-range of tissues and complex traits.

In an age of rampant obesity and diabetes, evolution may have given us a boon in the study of these traits. If evolution truly acted through parent-of-origin effects to alter a specific class of traits and we are motivated to understand the genetic basis of those same traits, then we can be reasonably certain that any cellular machinery showing parent-of-origin effect must be important to those traits.

We would predict the function of mesenchymal stem cells and their lineage show a parent-of-origin effect. If this is true in humans, it would have serious implications on human health. I would be particularly interested in whether wounding healing shows a parent-of-origin effect. That knowledge could give clinicians a stronger estimate of risk when considering intensive surgery.

Our catalog of genes in networks where genetic effects propagate could provide a focused molecular interpretation of genetic interactions that are conserved between mouse and human, as

well as a broad systems perspective that can be leveraged to understand parent-of-origin effects in general. We believe a fully comprehensive catalogue of such genes and networks in multiple mouse metabolic tissues should be generated. We could then correlate their effects with variation in metabolic phenotypes.

Systematically characterizing these patterns in mice is the first step towards developing a framework for detecting and characterizing parent-of-origin effects in human studies, particularly for complex traits that are deeply homologous between species [148,149]. Humans are notably less cooperative than mice. Parental genotypes and consequently parental origin of alleles is generally unknown. This makes detecting parent-of-origin effects extremely difficult. Expanding on the work of others [11,150], I was able to develop a framework to detect parent-of-origin dependent ASE when parental origins are unknown. Unlike previous approaches it was designed to jointly estimate parent-of-origin and sequence dependent ASE. However, like with previous approaches it is unknown how environment might confound the prediction of these effects. Future work should use the sex and dietary contexts of our data to evaluate how sensitive this approach might be to environmental factors. If sequence and parent-of-origin dependent ASE can be jointly estimated in humans without parental information, that would significantly improve the study of these phenomena in humans.

Modeling pairwise genetic effect networks is the first step in evaluating alternative explanations for the overabundance of parent-of-origin effects. Re-analyzing $F_{16}$ mapping data with genetic effects networks in mind may have shed light on how parental inheritance biases and genetic effect propagation could relate to QTL. For example, how often do we see parental inheritance biases? Of the QTL that show a significant parent-of-origin effect how many have a parental inheritance bias relationship with other loci? Are any of those loci also QTL and if so do

they show a significant additive effect? A complete answer to the overabundance of parent-of-origin effects will remain unobtainable until a comprehensive map of genetic effect networks has been generated and vetted. Until then, we are limited to finding proof-of-principle examples, but cannot say whether they explain the overall overabundance.

Perhaps most importantly, this works takes an incremental step towards bridging the worlds of quantitative genetics and molecular biology. While I crafted the bridge model to suit my needs and to maximize the chances I would find strong candidates, I also put an emphasis on designing it to be useful. Doubtless more advanced machine learning approaches could be implemented that would work more efficiently and more effectively. But I wanted the outputs of this work to be as interpretable and ready for experimental validation as possible. My approach should identify which gene is the source of a genetic effect, which genes are downstream, and what cell type that effect is happening in.

## 4.4 Parting thoughts

Parent-of-origin effects are essentially a special case of a more general phenomena. Genetic effects flow and blend and negate one another. Parent-of-origin effects can only manifest in the presence of some other form of genetic effect (i.e. additivity). Understanding these phenomena and the genotype-to-phenotype problem more generally requires that we understand multiple types of genetic effects.

Parent-of-origin effects have not been thoroughly investigated in large landmark functional genomics studies including ENCODE, GTEx, and GWAS, leaving a significant gap in our knowledge. Future work should leverage the patterns identified in mice to identify similar patterns in human data. If classical computational models of gene regulatory networks can be used to

capture parent-of-origin effects, an extended model can be trained using comprehensive data generated using the mouse and applied to model human data. Such an endeavor would enhance the accuracy of phenotype prediction and improve the functional annotation of genetic variation.

# References


1 Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. Nat. Rev. Genet. 2013. doi:10.1038/nrg3543.

2 Hager R, Cheverud JM, Wolf JB. Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting. *Genetics* 2008; **178**: 1755–1762.

3 Schier AF. The Maternal-Zygotic Transition: Death and Birth of RNAs. *Science (80- )* 2007; **316**: 406–407.

4 Sharma U, Conine CC, Shea JM, Boskovic A, Derr a. G, Bing XY *et al.* Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. *Science (80- )* 2015; **6780**: science.aad6780-.

5 Wang X, Clark a G. Using next-generation RNA sequencing to identify imprinted genes. *Heredity (Edinb)* 2014; **113**: 1–11.

6 Clerc P, Avner P. Random X-chromosome inactivation: skewing lessons for mice and men. *Curr Opin Genet Dev* 2006; **16**: 246–253.

7 Strogantsev R, Krueger F, Yamazawa K, Shi H, Gould P, Goldman-Roberts M *et al.* Allele-specific binding of ZFP57 in the epigenetic regulation of imprinted and non-imprinted monoallelic expression. *Genome Biol* 2015; **16**: 112.

8 Martienssen RA, Colot V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* 2001; **293**: 1070–1074.

9 Feil R, Berger F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet* 2007; **23**: 192–199.

10 Ferguson-Smith AC. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* 2011; **12**: 565–575.

11 Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS *et al.* Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nat Publ Gr* 2015; **47**: 544–549.

12 Schulz R, Woodfine K, Menheniott TR, Bourc'his D, Bestor T, Oakey RJ. Wamidex: A web atlas of murine genomic imprinting and differential expression. *Epigenetics* 2008; **3**: 89–96.

13 Wan L Ben, Bartolomei MS. Chapter 7 Regulation of Imprinting in Clusters: Noncoding RNAs Versus Insulators. *Adv Genet* 2008; **61**: 207–223.

14 Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. *Cold Spring Harb Perspect Biol* 2014; **6**. doi:10.1101/CSHPERSPECT.A018382.

15 Barlow DP. Genomic imprinting: a mammalian epigenetic discovery model. *Annu Rev Genet* 2011; **45**: 379–403.

16 Sanli I, Feil R. Chromatin mechanisms in the developmental control of imprinted gene expression. *Int J Biochem Cell Biol* 2015; **67**: 139–147.

17 Ishihara K, Oshimura M, Nakao M. CTCF-Dependent Chromatin Insulator Is Linked to Epigenetic Remodeling. *Mol Cell* 2006; **23**: 733–742.

18 Edwards CA, Ferguson-Smith AC. Mechanisms regulating imprinted genes in clusters. *Curr Opin Cell Biol* 2007; **19**: 281–289.

19 Fitzpatrick G V, Soloway PD, Higgins MJ. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat Genet* 2002; **32**: 426–31.

20 Lin MS, Zhang A, Fujimoto A. Asynchronous DNA-replication between 15q11.2q12 homologs - cytogenetic evidence for maternal imprinting and delayed replication. *Hum Genet* 1995; **96**: 572–576.

21 Zwart R. Bidirectional action of the Igf2r imprint control element on upstream and downstream imprinted genes. *Genes Dev* 2001; **15**: 2361–2366.

22 Mott R, Yuan W, Kaisaki P, Gan X, Cleak J, Edwards A *et al.* The architecture of parent-of-origin effects in mice. *Cell* 2014; **156**: 332–342.

23 Lawson H a, Zelle KM, Fawcett GL, Wang B, Pletscher LS, Maxwell TJ *et al.* Genetic, epigenetic, and gene-by-diet interaction effects underlie variation in serum lipids in a LG/JxSM/J murine model. *J Lipid Res* 2010; **51**: 2976–84.

24 Lawson HA, Lee A, Fawcett GL, Wang B, Pletscher LS, Maxwell TJ *et al.* The importance of context to the genetic architecture of diabetes-related traits is revealed in a genome-wide scan of a LG/J × SM/J murine model. *Mamm Genome* 2011; **22**: 197–208.

25 Cheverud JM, Lawson HA, Fawcett GL, Wang B, Pletscher LS, R Fox A *et al.* Diet-dependent genetic and genomic imprinting effects on obesity in mice. *Obesity (Silver Spring)* 2011; **19**: 160–70.

26 Lagarrigue S, Martin LJ, Hormozdiari F, Roux P-F, Pan C, van Nas A *et al.* Analysis of Allele Specific Expression in Mouse Liver by RNA-Seq: A Comparison with 'cis'-eQTL Identified Using Genetic Linkage. *Genetics* 2013. doi:10.1534/genetics.113.153882.

27 Zou F, Sun W, Crowley JJ, Zhabotynsky V, Sullivan PF, de Villena FPM. A novel statistical approach for jointly analyzing RNA-Seq data from F1 reciprocal crosses and inbred lines. *Genetics* 2014; **197**: 389–399.

28 Wang H, Elbein SC. Detection of allelic imbalance in gene expression using pyrosequencing. *Methods Mol Biol* 2007; **373**: 157–76.

29 DeVeale B, van der Kooy D, Babak T. Critical evaluation of imprinted gene expression by RNA-seq: A new perspective. *PLoS Genet* 2012; **8**. doi:10.1371/journal.pgen.1002600.

30 Nezer C, Moreau L, Brouwers B, Coppieters W, Detilleux J, Hanset R *et al.* An imprinted QTL with major effect on muscle mass and fat deposition maps to the IGF2 locus in pigs. *Nat Genet* 1999; **21**: 155–156.

31 Jeon JT, Carlborg O, Törnsten a, Giuffra E, Amarger V, Chardon P *et al.* A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the IGF2 locus. *Nat Genet* 1999; **21**: 157–158.

32 Mantey C, Brockmann GA, Kalm E, Reinsch N. Mapping and exclusion mapping of genomic imprinting effects in mouse F 2 families. *J Hered* 2005; **96**: 329–338.

33 Jeong S, Hahn Y, Rong Q, Pfeifer K. Accurate quantitation of allele-specific expression patterns by analysis of DNA melting. *Genome Res* 2007; **17**: 1093–1100.

34 Xie W, Barr CL, Kim A, Yue F, Lee AY, Eubanks J *et al.* Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 2012; **148**: 816–831.

35 Diep D, Plongthongkum N, Gore A, Fung H-L, Shoemaker R, Zhang K. Library-free methylation sequencing with bisulfite padlock probes. *Nat Methods* 2012; **9**: 270–272.

36 Patten MM, Ross L, Curley JP, Queller DC, Bonduriansky R, Wolf JB. The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity (Edinb)* 2014; **113**: 119–28.

37 Haig D. Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proc R Soc London Ser B Biol Sci* 1997; **264**: 1657–1662.

38 Wilkins JF, Haig D. What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* 2003; **4**: 359–368.

39 Arima T, Yamasaki K, John RM, Kato K, Sakumi K, Nakabeppu Y *et al.* The human HYMAI/PLAGL1 differentially methylated region acts as an imprint control region in mice. *Genomics* 2006; **88**: 650–658.

40 Smith RJ, Arnaud P, Konfortova G, Dean WL, Beechey C V., Kelsey G. The mouse Zac1 locus: Basis for imprinting and comparison with human ZAC. *Gene* 2002; **292**: 101–112.

41 Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res* 2001; **29**: 275–276.

42 Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat Genet* 2010; **42**: 68–71.

43 Small KS, Hedman ÅK, Grundberg E, Nica AC, Thorleifsson G, Kong A *et al.* Identification of an imprinted master trans regulator at the KLF14 locus related to multiple metabolic phenotypes. *Nat Genet* 2011; **43**: 561–564.

44 Morita S, Horii T, Kimura M, Arai Y, Kamei Y, Ogawa Y *et al.* Paternal allele influences

high fat diet-induced obesity. *PLoS One* 2014; **9**: 1–9.

45 Bhatti TR, Ganapathy K, Huppmann AR, Conlin L, Boodhansingh KE, MacMullen C *et al.* Histologic and molecular profile of pediatric insulinomas: Evidence of a paternal parent-of-origin effect. *J Clin Endocrinol Metab* 2016; **101**: 914–922.

46 Theodoropoulou M, Stalla GK, Spengler D. ZAC1 target genes and pituitary tumorigenesis. *Mol Cell Endocrinol* 2010; **326**: 60–65.

47 Moulton T, Chung WY, Yuan L, Hensle T, Waber P, Nisen P *et al.* Genomic imprinting and Wilms' tumor. *Med Pediatr Oncol* 1996; **27**: 476–83.

48 Bartholdi D, Krajewska-Walasek M, Õunap K, Gaspar H, Chrzanowska KH, Ilyana H *et al.* Epigenetic mutations of the imprinted IGF2-H19 domain in Silver-Russell syndrome (SRS): Results from a large cohort of patients with SRS and SRS-like phenotypes. *J Med Genet* 2009; **46**: 192–197.

49 Parker-Katiraee L, Carson AR, Yamada T, Arnaud P, Feil R, Abu-Amero SN *et al.* Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet* 2007; **3**: 665–678.

50 Davies SJ, Hughes HE. Imprinting in Albright's hereditary osteodystrophy. *J Med Genet* 1993; **30**: 101–3.

51 Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. *Trends Genet* 2005; **21**: 457–465.

52 Bassett SS, Avramopoulos D, Fallin D. Evidence for parent of origin effect in late-onset Alzheimer disease. *Am J Med Genet - Neuropsychiatr Genet* 2002; **114**: 679–686.

53 Bassett SS, Avramopoulos D, Perry RT, Wiener H, Watson B, Go RCP *et al.* Further evidence of a maternal parent-of-origin effect on chromosome 10 in late-onset Alzheimer's disease. *Am J Med Genet Part B Neuropsychiatr Genet* 2006; **141**: 537–540.

54 Luedi PP, Hartemink AJ, Jirtle RL. Genome-wide prediction of imprinted murine genes. *Genome Res* 2005; **15**: 875–884.

55 Grabowski M, Zimprich A, Lorenz-Depiereux B, Kalscheuer V, Asmus F, Gasser T *et al.* The epsilon-sarcoglycan gene (SGCE), mutated in myoclonus-dystonia syndrome, is maternally imprinted. *Eur J Hum Genet* 2003; **11**: 138–144.

56 Casimiro MC, Knollmann BC, Ebert SN, Vary JC, Greene AE, Franz MR *et al.* Targeted disruption of the Kcnq1 gene produces a mouse model of Jervell and Lange- Nielsen Syndrome. *Proc Natl Acad Sci* 2001; **98**: 2526–2531.

57 Tycko B, Morison I. Physiological Functions of Imprinted Genes. *J Cell Physiol* 2002; **192**: 245–258.

58 Cassidy SB, Schwartz S. Prader-Willi and Angelman syndromes. Disorders of genomic imprinting. *Medicine (Baltimore)* 1998; **77**: 140–51.

59 Lawson HA, Cady JE, Partridge C, Wolf JB, Semenkovich CF, Cheverud JM. Genetic effects at pleiotropic loci are context-dependent with consequences for the maintenance of genetic variation in populations. *PLoS Genet* 2011; **7**. doi:10.1371/journal.pgen.1002256.

60 Voight BF, Scott LJ, Steinthorsdottir V, Morris ADP, Dina C, Welch RP *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010; **42**: 579–89.

61 Teslovich TM, Musunuru K, Smith A V, Edmondson AC, Stylianou IM, Koseki M *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–13.

62 Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* 2009; **462**: 868–874.

63 Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 2000; **405**: 486–489.

64 Van Laere A-S, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L *et al.* A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* 2003; **425**: 832–836.

65 Soubry A, Schildkraut JM, Murtha A, Wang F, Huang Z, Bernal A *et al.* Paternal obesity is associated with IGF2 hypomethylation in newborns: results from a Newborn Epigenetics Study (NEST) cohort. *BMC Med* 2013; **11**: 29.

66 Morita S, Nakabayashi K, Kawai T, Hayashi K, Horii T, Kimura M *et al.* Gene expression profiling of white adipose tissue reveals paternal transmission of proneness to obesity. *Sci Rep* 2016; **6**: 21693.

67 Pierre CL St., Macias-Velasco JF, Wayhart JP, Yin L, Semenkovich CF, Lawson HA. Genetic, epigenetic, and environmental mechanisms govern allele-specific gene expression. *bioRxiv* 2021; : 2021.09.09.459642.

68 Al Adhami H, Evano B, Le Digarcher A, Gueydan C, Dubois E, Parrinello H *et al.* A systems-level approach to parental genomic imprinting: The imprinted gene network includes extracellular matrix genes and regulates cell cycle exit and differentiation. *Genome Res* 2015; **25**: 353–367.

69 Mochizuki A, Takeda Y, Iwasa Y. The evolution of genomic imprinting. *Genetics* 1996; **144**: 1283–1295.

70 Sleutels F, Barlow DP. *The origins of genomic imprinting in mammals*. Elsevier Inc., 2002 doi:10.1016/S0065-2660(02)46006-3.

71 Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowder GD *et al.* Polar overdominance at the ovine callipyge locus. *Science* 1996; **273**: 236–8.

72 Georges M, Charlier C, Cockett N. The callipyge locus: Evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet* 2003; **19**: 248–252.

73 Cui Y, Lu Q, Cheverud JM, Littell RC, Wu R. Model for mapping imprinted quantitative trait loci in an inbred F 2 design. *Genomics* 2006; **87**: 543–551.

74 Lawson HA, Cheverud JM, Wolf JB. Genomic imprinting and parent-of-origin effects on complex traits. *Nat Rev Genet* 2013; **14**: 609–617.

75 Jirtle RL. No Title. http://www.geneimprint.com/site/genes-by-species. 2012.

76 Mozaffari S V., DeCara JM, Shah SJ, Sidore C, Fiorillo E, Cucca F *et al.* Parent-of-origin effects on quantitative phenotypes in a large Hutterite pedigree. *Commun Biol* 2019. doi:10.1038/s42003-018-0267-4.

77 Zeng Y, Amador C, Xia C, Marioni R, Sproul D, Walker RM *et al.* Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nat Commun* 2019; **10**: 1383.

78 Carson C, Macias-Velasco JF, Gunawardana S, Miranda MA, Oyama S, St. Pierre CL *et al.* Brown Adipose Expansion and Remission of Glycemic Dysfunction in Obese SM/J Mice. *Cell Rep* 2020; **33**: 108237.

79 Miranda MA, Carson C, St. Pierre CL, Macias-Velasco JF, Hughes JW, Kunzmann M *et al.* Spontaneous restoration of functional β-cell mass in obese SM/J mice. *Physiol Rep* 2020; **8**: 1–12.

80 Zhang B, Kirov S, Snoddy J. WebGestalt: An integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 2005. doi:10.1093/nar/gki475.

81 Li X, Thomason PA, Withers DJ, Scott J. Bio-informatics analysis of a gene co-expression module in adipose tissue containing the diet-responsive gene Nnat. *BMC Syst Biol* 2010; **4**: 1–11.

82 Young HS, Won HK, Moon C, Yun HH, Eun SY, Joo HL *et al.* Ectopic expression of Neuronatin potentiates adipogenesis through enhanced phosphorylation of cAMP-response element-binding protein in 3T3-L1 cells. *Biochem Biophys Res Commun* 2005. doi:10.1016/j.bbrc.2005.09.078.

83 Lin H-H, Bell E, Uwanogho D, Perfect LW, Noristani H, Bates TJD *et al.* Neuronatin Promotes Neural Lineage in ESCs via Ca2+ Signaling. *Stem Cells* 2010; **28**: 1950–1960.

84 Soh UJ, Dores MR, Chen B, Trejo J. Signal transduction by protease-activated receptors. *Br J Pharmacol* 2010; **160**: 191–203.

85 Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK *et al.* Leveraging Polygenic Functional Enrichment to Improve GWAS Power. *Am J Hum Genet* 2019; **104**: 65–75.

86 Nakamura Y, Hamada Y, Fujiwara T, Enomoto H, Hiroe T, Tanaka S *et al.* Phospholipase

C-1 and-3 Are Essential in the Trophoblast for Placental Development. *Mol Cell Biol* 2005; **25**: 10979–10988.

87 McDonald LJ, Mararack MD. Phosphoinositide hydrolysis by phospholipase C modulated by multivalent cations La3, Al3+, neomycin, polyamines, and melittin. *J Lipid Mediat Cell Signal* 1995; **11**: 81–91.

88 Thatcher JD. The Inositol Trisphosphate (IP3) Signal Transduction Pathway. *Sci Signal* 2010; **3**: tr3–tr3.

89 Berridge MJ. The Inositol Trisphosphate/Calcium Signaling Pathway in Health and Disease. *https://doi.org/101152/physrev000062016* 2016; **96**: 1261–1296.

90 Kang JW, Choi Y, Park JH, Kim JS, Park KD, Baek DH *et al.* The effects of cyclin-dependent kinase inhibitors on adipogenic differentiation of human mesenchymal stem cells. *Biochem Biophys Res Commun* 2008; **366**: 624–630.

91 Pette M Van de, Tunster SJ, John RM. Loss of Imprinting of Cdkn1c Protects against Age and Diet-Induced Obesity. *Int J Mol Sci* 2018; **19**. doi:10.3390/IJMS19092734.

92 Franck N, Gummesson A, Jernås M, Glad C, Svensson P-A, Guillot G *et al.* Identification of Adipocyte Genes Regulated by Caloric Intake. *J Clin Endocrinol Metab* 2011; **96**: E413–E418.

93 Burg JS, Espenshade PJ. Regulation of HMG-CoA reductase in mammals and yeast. *Prog Lipid Res* 2011; **50**: 403–410.

94 Jo Y, DeBose-Boyd RA. Control of cholesterol synthesis through regulated ER-associated degradation of HMG CoA reductase. *Crit Rev Biochem Mol Biol* 2010; **45**: 185–198.

95 Yeh Y-S, Jheng H-F, Iwase M, Kim M, Mohri S, Kwon J *et al.* The Mevalonate Pathway Is Indispensable for Adipocyte Survival. *iScience* 2018; **9**: 175.

96 Lemire JM, Chan CK, Bressler S, Miller J, LeBaron RG, Wight TN. Interleukin-1β selectively decreases the synthesis of versican by arterial smooth muscle cells. *J Cell Biochem* 2007; **101**: 753–766.

97 Imoto-Tsubakimoto H, Takahashi T, Ueyama T, Ogata T, Adachi A, Nakanishi N *et al.* Serglycin is a novel adipocytokine highly expressed in epicardial adipose tissue. *Biochem Biophys Res Commun* 2013; **432**: 105–110.

98 Schick BP, Gradowski JF, Antonio JDS. Synthesis, secretion, and subcellular localization of serglycin proteoglycan in human endothelial cells. *Blood* 2001; **97**: 449–458.

99 Zernichow L, Åbrink M, Hallgren J, Grujic M, Pejler G, Kolset SO. Serglycin Is the Major Secreted Proteoglycan in Macrophages and Has a Role in the Regulation of Macrophage Tumor Necrosis Factor-α Secretion in Response to Lipopolysaccharide. *J Biol Chem* 2006; **281**: 26792–26801.

100 Bigdeli N, Peppo GM de, Lennerås M, Sjövall P, Lindahl A, Hyllner J *et al.* Superior

Osteogenic Capacity of Human Embryonic Stem Cells Adapted to Matrix-Free Growth Compared to Human Mesenchymal Stem Cells. *https://home.liebertpub.com/tea* 2010; **16**: 3427–3440.

101 Aubin JE. Perspectives: Osteoclast adhesion and resorption: The role of podosomes. *J Bone Miner Res* 1992; **7**: 365–368.

102 Rodriguez J, Astudillo P, Rios S, Pino A. Involvement of Adipogenic Potential of Human Bone Marrow Mesenchymal Stem Cells (MSCs) in Osteoporosis. *Curr Stem Cell Res Ther* 2008; **3**: 208–218.

103 Prockop DJ. Marrow Stromal Cells as Stem Cells for Nonhematopoietic Tissues. *Science (80- )* 1997; **276**: 71–74.

104 Ali AA, Weinstein RS, Stewart SA, Parfitt AM, Manolagas SC, Jilka RL. Rosiglitazone Causes Bone Loss in Mice by Suppressing Osteoblast Differentiation and Bone Formation. *Endocrinology* 2005; **146**: 1226–1235.

105 Akune T, Ohba S, Kamekura S, Yamaguchi M, Chung U Il, Kubota N *et al.* PPARγ insufficiency enhances osteogenesis through osteoblast formation from bone marrow progenitors. *J Clin Invest* 2004; **113**: 846–855.

106 Cho SW, Yang J-Y, Her SJ, Choi HJ, Jung JY, Sun HJ *et al.* Osteoblast-targeted overexpression of PPARγ inhibited bone mass gain in male mice and accelerated ovariectomy-induced bone loss in female mice. *J Bone Miner Res* 2011; **26**: 1939–1952.

107 Rosen CJ, Bouxsein ML. Mechanisms of Disease: is osteoporosis the obesity of bone? *Nat Clin Pract Rheumatol 2006 21* 2006; **2**: 35–43.

108 RT T, SA M, UT I. Metabolic Coupling Between Bone Marrow Adipose Tissue and Hematopoiesis. *Curr Osteoporos Rep* 2018; **16**: 95–104.

109 Nakagawa N, Kinosaki M, Yamaguchi K, Shima N, Yasuda H, Yano K *et al.* RANK Is the Essential Signaling Receptor for Osteoclast Differentiation Factor in Osteoclastogenesis. *Biochem Biophys Res Commun* 1998; **253**: 395–400.

110 Matsuo FS, Cavalcanti de Araújo PH, Mota RF, Carvalho AJR, Santos de Queiroz M, Baldo de Almeida B *et al.* RANKL induces beige adipocyte differentiation in preadipocytes. *Am J Physiol Metab* 2020; **318**: E866–E877.

111 An J-J, Han D-H, Kim D-M, Kim S-H, Rhee Y, Lee E-J *et al.* Expression and Regulation of Osteoprotegerin in Adipose Tissue. *Yonsei Med J* 2007; **48**: 765.

112 Holecki M, Zahorska-Markiewicz B, Janowska J, Nieszporek T, Wojaczyńska-Stanek K, Żak-Gołąb A *et al.* The Influence of Weight Loss on Serum Osteoprotegerin Concentration in Obese Perimenopausal Women*. *Obesity* 2007; **15**: 1925–1929.

113 Erol M, Gayret OB, Nacaroglu HT, Yigit O, Zengi O, Akkurt MS *et al.* Association of Osteoprotegerin with Obesity, Insulin Resistance and Non-Alcoholic Fatty Liver Disease in Children. *Iran Red Crescent Med J* 2016; **18**: 41873.

114 Zaky DS, Ali AA, Abd-Elraheem SE, Abdel-Moniem SH. Circulating Osteoprotegerin Level in Relation to Obesity in Middle Aged Females. doi:10.5923/j.ijpt.20190802.02.

115 Stanton LW, Ponte PA, Coleman RT, Snyder MA. Expression of CA III in Rodent Models of Obesity. *Mol Endocrinol* 1991; **5**: 860–866.

116 Font-Clos F, Zapperi S, Porta CAM La. Integrative analysis of pathway deregulation in obesity. *NPJ Syst Biol Appl* 2017; **3**. doi:10.1038/S41540-017-0018-Z.

117 Renner SW, Walker LM, Forsberg LJ, Sexton JZ, Brenman JE. Carbonic anhydrase III (Car3) is not required for fatty acid synthesis and does not protect against high-fat diet induced obesity in mice. *PLoS One* 2017; **12**: e0176502.

118 Burl RB, Ramseyer VD, Rondini EA, Pique-Regi R, Lee Y-H, Granneman JG. Deconstructing Adipogenesis Induced by β3-Adrenergic Receptor Activation with Single-Cell Expression Profiling. *Cell Metab* 2018; **28**: 300-309.e4.

119 Wolf JB, Cheverud JM. A framework for detecting and characterizing genetic background-dependent imprinting effects. *Mamm Genome* 2009; **20**: 681–698.

120 Pierce BL, Tong L, Chen LS, Rahaman R, Argos M, Jasmine F *et al.* Mediation Analysis Demonstrates That Trans-eQTLs Are Often Explained by Cis-Mediation: A Genome-Wide Analysis among 1,800 South Asians. *PLoS Genet* 2014; **10**. doi:10.1371/JOURNAL.PGEN.1004818.

121 Shan N, Wang Z, Hou L. Identification of trans-eQTLs using mediation analysis with multiple mediators. *BMC Bioinformatics* 2019; **20**: 14–16.

122 Lutz SM, Hokanson JE. Mediation analysis in genome-wide association studies: current perspectives. *Open Access Bioinformatics* 2015; **7**: 1–5.

123 Millership SJ, Da Silva Xavier G, Choudhury AI, Bertazzo S, Chabosseau P, Pedroni SMA *et al.* Neuronatin regulates pancreatic β cell insulin content and secretion. *J Clin Invest* 2018; **128**: 3369–3381.

124 Fowden AL, Moore T. Maternal-fetal resource allocation: Co-operation and conflict. *Placenta* 2012; **33**: e11–e15.

125 Nikolskiy I, Conrad DF, Chun S, Fay JC, Cheverud JM, Lawson HA. Using whole-genome sequences of the LG/J and SM/J inbred mouse strains to prioritize quantitative trait genes and nucleotides. *BMC Genomics* 2015; **16**: 415.

126 Yee TW. The **VGAM** Package for Categorical Data Analysis. *J Stat Softw* 2010. doi:10.1007/BF01016332.

127 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S *et al.* STAR: ultrafast universal RNAseq aligner. *Bioinformatics* 2013; : 1–7.

128 Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010; **11**: 1–9.

129 Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010; **26**: 139–40.

130 Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun 2017 81* 2017; **8**: 1–12.

131 Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM *et al.* Comprehensive Integration of Single-Cell Data. *Cell* 2019; **177**: 1888-1902.e21.

132 R Core Team. R: A Language and Environment for Statistical Computing. 2013.http://www.r-project.org/.

133 Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol Syst Biol* 2019; **15**: e8746.

134 Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience* 2018; **7**: 1–9.

135 Finak G, McDavid A, Yajima M, Deng J, Gersuk V, Shalek AK *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol 2015 161* 2015; **16**: 1–13.

136 Haley CS, Knott SA, Elsen JM. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 1994; **136**: 1195–1207.

137 Haley CS, Knott SA. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Hered 1992 694* 1992; **69**: 315–324.

138 Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. doi:10.1016/j.cell.2017.05.038.

139 Lee JT, Bartolomei MS. X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 2013; **152**: 1308–1323.

140 Barlow DP, Bartolomei MS. Genomic Imprinting in Mammals. *Cold Spring Harb Perspect Biol* 2014; **6**: a018382–a018382.

141 Moore KJ, Kunjathoor V V., Koehn SL, Manning JJ, Tseng AA, Silver JM *et al.* Loss of receptor-mediated lipid uptake via scavenger receptor A or CD36 pathways does not ameliorate atherosclerosis in hyperlipidemic mice. *J Clin Invest* 2005; **115**: 2192.

142 Kong FQ, Zhao SJ, Sun P, Liu H, Jie J, Xu T *et al.* Macrophage MSR1 promotes the formation of foamy macrophage and neuronal apoptosis after spinal cord injury. *J Neuroinflammation* 2020; **17**: 1–15.

143 van Eijk M, Aerts JMFG. The Unique Phenotype of Lipid-Laden Macrophages. *Int J Mol Sci 2021, Vol 22, Page 4039* 2021; **22**: 4039.

144 Govaere O, Petersen SK, Martinez-Lopez N, Wouters J, Van Haele M, Mancina RM *et al.*

Macrophage Scavenger Receptor 1 mediates lipid-induced inflammation in non-alcoholic fatty liver disease. *J Hepatol* 2021. doi:10.1016/J.JHEP.2021.12.012.

145 Buscher K, Ehinger E, Gupta P, Pramod AB, Wolf D, Tweet G *et al.* Natural variation of macrophage activation as disease-relevant phenotype predictive of inflammation and cancer survival. *Nat Commun 2017 81* 2017; **8**: 1–10.

146 Miranda MA, Macias-Velasco JF, Schmidt H, Lawson HA. Integrated transcriptomics identifies β-cell subpopulations and genetic networks associated with obesity and glycemic control in SM/J mice. *bioRxiv* 2021; : 2021.07.15.452524.

147 Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM *et al.* Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet* 2019; **51**: 611–617.

148 Attie AD, Churchill GA, Nadeau JH. How mice are indispensable for understanding obesity and diabetes genetics. Curr. Opin. Endocrinol. Diabetes Obes. 2017. doi:10.1097/MED.0000000000000321.

149 Wang X, Pandey AK, Mulligan MK, Williams EG, Mozhui K, Li Z *et al.* Joint mouse-human phenome-wide association to test gene function and disease risk. *Nat Commun* 2016. doi:10.1038/ncomms10464.

150 Pirinen M, Lappalainen T, Zaitlen NA, Dermitzakis ET, Donnelly P, McCarthy MI *et al.* Assessing allele-specific expression across multiple tissues from RNA-seq read data. *Bioinformatics* 2015; **31**: 2497–2504.