

THE UNIVERSITY OF CHICAGO

PARENT OF ORIGIN EFFECTS ON GENE EXPRESSION AND QUANTITATIVE
TRAITS IN A FOUNDER POPULATION

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

COMMITTEE ON GENETICS, GENOMICS & SYSTEMS BIOLOGY

BY
SAHAR VICTORIA MOZAFFARI

CHICAGO, ILLINOIS
AUGUST 2018

Copyright © 2018 by Sahar Victoria Mozaffari

All Rights Reserved

Freely available under a CC-BY 4.0 International license

For my family.

Especially my maternal grandmother and paternal grandfather,

Shamsi Joon (Shamsi Karimkhanzand) and Bababozorg (Hossein Mozaffari).

For encouraging us to pursue education and knowledge, and moving to a new country and

enduring loss and hardship to make it possible for us to do so.

Table of Contents

| | |
|---|-----|
| LIST OF FIGURES | vii |
| LIST OF TABLES | ix |
| ACKNOWLEDGMENTS | x |
| ABSTRACT | xiv |
| 1 INTRODUCTION | 1 |
| 1.1 Human Genetics and the Genetics of Complex Traits | 1 |
| 1.2 On The Origin of Genomic Imprinting | 3 |
| 1.3 The Search for Parent of Origin Effects | 6 |
| 1.3.1 Dissertation Overview | 7 |
| 2 PARENT OF ORIGIN EFFECTS ON QUANTITATIVE PHENOTYPES IN A FOUNDER POPULATION | 9 |
| 2.1 Abstract | 9 |
| 2.2 Introduction | 9 |
| 2.3 Results | 11 |
| 2.3.1 GWAS | 11 |
| 2.3.2 Parent of Origin GWAS | 12 |
| 2.3.3 Maternal and Paternal GWAS | 12 |
| 2.3.4 GWAS for Differential Parent of Origin Effects | 15 |
| 2.3.5 Parent of Origin Effects on Gene Expression | 19 |
| 2.4 Discussion | 21 |
| 2.5 Methods | 24 |
| 2.5.1 Sample Composition | 24 |
| 2.5.2 Genotype Data | 24 |
| 2.5.3 Phenotype Data | 25 |
| 2.5.4 GWAS | 25 |
| 2.5.5 Maternal and Paternal GWAS | 25 |
| 2.5.6 Differential Effect GWAS (PO-GWAS) | 26 |
| 2.5.7 Parent of Origin eQTL studies | 27 |
| 2.5.8 Maternal and Paternal Parent of Origin eQTL | 28 |
| 2.5.9 Differential Parent of Origin eQTL | 28 |
| 2.6 Supplementary Information | 29 |
| 2.6.1 Supplementary Figures | 29 |
| 3 PARENT OF ORIGIN GENE EXPRESSION IN A FOUNDER POPULATION IDENTIFIES TWO NOVEL IMPRINTED GENES AT KNOWN IMPRINTED REGIONS. | 47 |
| 3.1 Abstract | 47 |

| | | |
|-------|---|----|
| 3.2 | Introduction | 47 |
| 3.3 | Results | 49 |
| 3.3.1 | Mapping Transcripts to Parental Haplotypes | 49 |
| 3.3.2 | Imprinted Genes in Lymphoblastoid Cell Lines (LCLs) | 50 |
| 3.3.3 | Validation of Imprinted Genes in PBLs | 53 |
| 3.3.4 | Methylation at Imprinting Control Regions | 54 |
| 3.4 | Discussion | 55 |
| 3.5 | Methods | 58 |
| 3.5.1 | Genotypes | 58 |
| 3.5.2 | RNA-seq in Lymphoblastoid Cell Lines (LCLs). | 58 |
| 3.5.3 | RNA-seq in Peripheral Blood Leukocytes (PBLs) | 59 |
| 3.5.4 | Identifying Imprinted Genes | 59 |
| 3.5.5 | DNA methylation profiling and processing in PBLs | 59 |
| 4 | PARENT OF ORIGIN EFFECTS ON GENE EXPRESSION | 67 |
| 4.1 | Abstract | 67 |
| 4.2 | Introduction | 67 |
| 4.3 | Results | 68 |
| 4.3.1 | Opposite Parent of Origin eQTL (oeQTL) | 69 |
| 4.3.2 | Single Parent eQTL (mat-eQTL, pat-eQTL) | 71 |
| 4.3.3 | Parent of Origin (PO) - ASE Test | 73 |
| 4.3.4 | Modified ASE Test on Symmetrically Expressed Genes | 77 |
| 4.4 | Discussion | 81 |
| 4.5 | Methods | 83 |
| 4.5.1 | Genotypes and Sample Information | 83 |
| 4.5.2 | RNA-seq QC | 83 |
| 4.5.3 | Parent of Origin Expression QC | 83 |
| 4.5.4 | Informative Genes | 84 |
| 4.5.5 | Opposite Parent of Origin eQTL | 86 |
| 4.5.6 | Single Parent eQTL | 86 |
| 4.5.7 | PO ASE Test | 86 |
| 4.5.8 | Not Asymmetrically Expressed Genes | 87 |
| 5 | CONCLUSION | 88 |
| 5.1 | A novel method to detect opposite effects of parentally inherited variants on cardiovascular disease and asthma associated traits | 88 |
| 5.2 | Identifying two novel imprinted genes in known imprinted regions using parent of origin gene expression | 89 |
| 5.3 | Can genetic variation by parent of origin influence gene expression on the same haplotype? | 90 |
| 5.4 | Future Directions | 92 |
| 5.5 | Concluding remarks | 93 |

| | |
|----------------------|----|
| REFERENCES | 95 |
|----------------------|----|

List of Figures

| | | |
|-------|---|----|
| 1.1 | Asthma GWAS Manhattan Plot. | 2 |
| 2.1 | Maternal and Paternal GWAS results for Age of Menarche. | 14 |
| 2.2 | Maternal and Paternal GWAS results for LDL Cholesterol. | 15 |
| 2.3 | Opposite Effect Parent of Origin GWAS Result for Age of Menarche. | 17 |
| 2.4 | Opposite Effect Parent of Origin GWAS Result for BMI. | 18 |
| 2.5a | Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes. | 29 |
| 2.5b | Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes (Continued). | 30 |
| 2.5c | Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes (Continued). | 31 |
| 2.6 | Maternal and Paternal GWAS results for CIMT. | 32 |
| 2.7 | Maternal and Paternal GWAS results for LVMI. | 32 |
| 2.8 | Maternal and Paternal GWAS results for FEV ₁ | 33 |
| 2.9 | Maternal and Paternal GWAS results for Systolic Blood Pressure. | 33 |
| 2.10 | Maternal and Paternal GWAS results for Total Cholesterol. | 34 |
| 2.11 | Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes. | 35 |
| 2.11a | Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes (Continued). | 36 |
| 2.11b | Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes (Continued). | 37 |
| 2.12 | Opposite Effect Parent of Origin GWAS Result for LDL. | 38 |
| 2.13 | Opposite Effect Parent of Origin GWAS Result for LVMI. | 38 |
| 2.14 | Opposite Effect Parent of Origin GWAS Result for Triglycerides. | 39 |
| 2.15 | Opposite Effect Parent of Origin GWAS Result for Total Cholesterol. | 39 |
| 2.16 | Opposite Effect Parent of Origin GWAS Result for Blood Eosinophil Count. | 40 |
| 2.17 | Opposite Effect Parent of Origin GWAS Result for FEV ₁ | 40 |
| 2.18 | Opposite Effect Parent of Origin GWAS Result for Neutrophil Count. | 41 |
| 2.19 | Opposite Effect eQTL for rs7033776. | 41 |
| 3.1 | Plot of maternal (x-axis) and paternal (y-axis) gene expression for four genes. | 51 |
| 3.2 | Validation in PBLs. | 54 |
| 3.3 | Methylation at ICRs. | 56 |
| 3.4 | Plots of maternal and paternal expression for remaining genes with parent of origin asymmetry. | 61 |
| 3.5 | Density plot for DMRs for all imprinted genes. | 62 |
| 4.1 | Opposite effect eQTLs driven by one individual's genotype. | 70 |
| 4.2 | Maternal eQTL Associations Driven by Null Values. | 71 |
| 4.3 | Paternal eQTL Associations Driven by Null Values. | 72 |
| 4.4 | Similar effect sizes across mat-eQTL and pat-eQTL. | 73 |

| | | |
|------|--|----|
| 4.5 | Parent specific eQTLs significant in one parent and not the other. | 74 |
| 4.6 | Significant PO-ASE association with gene <i>SNHG14</i> | 75 |
| 4.7 | Significant PO-ASE association with gene <i>ERAP2</i> | 76 |
| 4.8 | Significant PO-ASE association with gene <i>ZDBF2</i> | 76 |
| 4.9 | Significant PO-ASE association with gene <i>PEG10</i> | 77 |
| 4.10 | Significant Association from PO-ASE test with <i>SEC22B</i> | 79 |
| 4.11 | Significant Association from PO-ASE test with <i>ZMAT3</i> | 80 |
| 4.12 | Number of Individuals with Gene Expression. | 85 |

List of Tables ¹

| | | |
|------|---|----|
| 1.1 | Imprinted Gene Disorders. | 5 |
| 2.1 | Phenotypes with significant single parent of origin associations. | 13 |
| 2.2 | Significant Opposite Parent of Origin Effect GWAS Associations. | 16 |
| 2.3 | Parent of Origin eQTLs in LCLs. | 20 |
| 2.4 | Opposite Parent of Origin eQTLs in LCLs. | 20 |
| 2.5 | Summary of the Hutterite Phenotypes and Sample Composition. | 42 |
| 2.6 | Summary of the Hutterite Phenotypes and Sample Composition (Continued). . | 43 |
| 2.7 | GWAS Results. | 44 |
| 2.8 | Candidate Genes for Parent of Origin eQTL. | 45 |
| 2.9 | Candidate Genes for Parent of Origin Differential eQTL. | 45 |
| 2.10 | Maternal GWAS results with p-value <5x10-08. | 46 |
| 2.11 | Paternal GWAS results with p-value <5x10-08. | 46 |
| 2.12 | Differential Effect GWAS results with p-value <5x10-08. | 46 |
| 3.1 | Summary Statistics for Parental Origin of Transcripts. | 50 |
| 3.2 | Results for Genes with Parent of Origin Expression Asymmetry. | 52 |
| 3.3 | Genes with only Maternal/ only Paternal gene expression. | 63 |
| 3.4 | Genes with only Maternal/ only Paternal gene expression (Continued - 1). . . | 64 |
| 3.5 | Genes with only Maternal/ only Paternal gene expression (Continued - 2). . . | 65 |
| 3.6 | Genes with only Maternal/ only Paternal gene expression (Continued - 3). . . | 66 |
| 4.1 | Top Ten Significant Genes from PO-ASE Test. | 75 |
| 4.2 | Top Ten Significant Genes from PO-ASE Test after Filtering Asymmetrically Expressed Genes. | 78 |
| 4.3 | Setup for PO-ASE test. | 87 |

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

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Carole Ober. Thanks for providing a rigorous but nurturing environment to grow into a scientist. Throughout my PhD, I have grown to respect and look up to you even more which I didn't think was possible. Thanks for providing the perfect amount of mentorship for us to succeed. Thanks to Dan Nicolae for teaching me so much and encouraging me to think outside of the box and meet with him to discuss all sorts of progress and ideas. Thanks to my committee for taking the time out of their busy schedules to provide me with lots of direction and support, both in one on one and in committee meetings: Yoav Gilad, John Novembre, and Vincent Lynch.

My decision to pursue Genetics in college came about after my program advisor, Sarah Dempsey, rightly so, refused to let me take both genetics and cell biology lab courses in the same semester. After crying in her office, I was forced to make a decision and chose the one I enjoyed learning about more and told myself I'd worry about the pre-optometry requirements later. It was through these genetics courses that I fell in love with the field and met Daniel Rokhsar. Dan took a chance on me and let me work in his lab for a senior honors thesis during my last year at UC Berkeley. He has since encouraged me and been a supporter of me and my science. Many thanks to Eric Edsinger and Jess Lyons for teaching me how to science, and Adam Session, Therese Mitros, Simon Prochnik, and Jessen Bredeson for teaching me how to code and work with genomic data. Special thanks to my undergraduate friends in the lab who are now pursuing their own graduate degrees: Cindy Ha and Konnor La.

With only one year of science research under my belt, I was admitted into one graduate university - University of Chicago. Thank you to the GGSB faculty who fought for me to come here. I've had the great opportunity to work with some strong amazing women in science, Nancy Cox, Elizabeth McNally, and Hae Kyung Im, among many others, who have helped me along my graduate career from the beginning, from writing me strong letters of

recommendation, supporting conference attendance, enrolling me in R Master classes with Hadley Wickham, and convincing me to become active in the science community on Twitter - all of which have contributed to make me the scientist I am today.

I would like to thank the opportunities and outside activities that have helped keep me sane and motivated. Thanks to my fellow volunteers at the Museum of Science and Industry and Matt, Rene, and Rudy for giving me the opportunity to teach adults and children about DNA for the past three years. I had the privilege of answering random questions, having engaging discussions, and correcting misconceptions which I find important as a scientist and more crucial in our current political climate.

It should be acknowledged that none of this happened without any struggles. It's hard to forget the teaching assistant during my first year of graduate school, who instead of helping me, told me I didn't know anything and said the faculty questioned my presence in graduate school; or the peers who forgot about the female computational biologists - when we can all learn from each other; and other glitches in the system, both financial and academic, that have caused unnecessary stresses on many a graduate student.

I would not have made it through graduate school without the support of my friends and colleagues. Katie Igartua and Michelle Stein, thanks for welcoming me to the lab, teaching me a lot of what I know today and leaving the bar quite high. Thanks to the past, current, and future members of the lab who have helped make it the great environment that it is and provided help with everything all these years. Thanks to my cohort of Molecular Bioscience students who have gone through the same things with me since Day 1- especially Lindsey Montefiori, Ana Beringer, Lindsey Mao, Erin Boyle Anderson, Chris Katanski, Phil McGilvray, Chris Craddock, and Beth Pollard. Thanks to the older GGSB students who welcomed me in when I was the only GGSB student in my cohort : Kenneth Barr, Nick VanKuren, John Blischak, Bryce van de Geijn, Aashish Jha, Ziyue Gao, Talia Karasov, Patrick Landback, and Jason Pitt. Thanks for all the advice through the years, all

sorts of computational help, and most importantly the invitation to play soccer with The Repressors. The GGSB students who have made our program exceptional and especially those I am lucky to call my friends: UnJin Lee, Charlie Lang, Leila Reyes Ruiz, Briana Mittleman, and Linsin Smith. Thanks to the Human Genetics students who included me in their department, especially Erin Fry, Ittai Eres, Katie Mika, Lauren Blake, Alex Gileta, and Natalia Gonzales. Thanks to my many friends from other programs and departments who I have come to be really close with: Dana Gilmore & Brendan Horton, Amelia Joslin & Andrew Tremain, Keston & Ayelet Aquino Michaels, Larischa de Wet. Also Kyle Cron, Sangman Kim, Toufic Mayassi, Chris Stamper, Jaime Chao, Chanie Howard, Ryan Duncombe, and Steven Erickson, Colles Price, Jason Torres, Kevin Lei, Joel Smith, and Ali Ekrem Yesilkanal.

I have to thank all the administrators and advisors that made my graduate career path go smoothly as well as help me achieve my dream job, especially Sue Levison, Briana Konnick, and Abby Stayart.

Thanks to the fuzzy butts that have provided emotional support and furry comfort along the way: Efe, Audrey, Penny (dog), Meatball, Darth Vader, Penny (cat), and Panda.

Thanks to Alan Chang for being my partner these past four years and making every day brighter. Thanks for exploring Chicago in food and drink with me and being my favorite adventure buddy. I can't imagine having gone through any of this without your encouragement, warmth, and love by my side.

Thanks to my entire family for their loving support. Thanks to my Aunt Peymaneh and Uncle Mahmood for sending me letters full of love, Halloween packages full of my favorite candy, and a whole slough of Christmas gifts for when I wasn't home to celebrate. Thanks to my grandma for always asking to FaceTime with me - but only for a short time to let me get back to work- and for sending me too many raw pistachios, pomegranates from her backyard, and homemade Persian sweets I miss during Persian New Year every year for the past five

years. Thanks to my sisters, my best friends, who will talk to me early in the morning (California is two hours behind Chicago) and let me rant about any and everything. None of this would have been possible without the support of my parents, who didn't question my decision to move halfway across the country for the past five years but continued to visit (albeit only in the summer) and tell me how proud they are of me. Thanks for pushing me as hard as you did and taking care of me all of these years.

ABSTRACT

Variants can affect traits differently depending on whether they are inherited from the mother or the father, but genome wide association studies (GWAS) treat maternal and paternal alleles as equivalent. In addition, the variants identified by GWAS do not account for a significant portion of the heritability for the corresponding trait and the “missing heritability” could be due to underlying biological mechanisms that are not yet well understood. My thesis addresses these limitations by disentangling the effects of maternal and paternal alleles on gene expression as well as on disease-associated phenotypes in the Hutterites, a founder population of European descent. With phased genotype data we can ask questions about parent of origin effects in this population. First, we tested for maternal and paternal genetic associations on cardiovascular disease and asthma associated traits and developed a novel method to detect variants that have opposite effects on the trait of interest depending on the parent of origin of the variant. We identified variants that have maternal-only or paternal-only effects, as well as variants that have opposite effects on traits, which would not be detected in a standard GWAS. This is the largest family based study of parent of origin effects on quantitative traits and the first to look for opposite parental effects. In the second chapter, we map RNA-seq reads from lymphoblastoid cell lines (LCLs) to parental haplotypes in 306 Hutterites and detect known imprinted genes and two novel imprinted genes (*PXDC1* and *PWAR6*). These imprinted gene patterns are validated using parent of origin expression from peripheral blood leukocytes (PBL) from 99 different Hutterites; imprinting control regions near the novel genes were validated using PBL methylation in the same 99 Hutterites. Finally, we explore searching for parent of origin effects on gene expression or parent of origin eQTLs, first for opposite effects and then for maternal and paternal specific effects.

CHAPTER 1

INTRODUCTION

1.1 Human Genetics and the Genetics of Complex Traits

A central goal of genetics is to understand the contribution of genetic variation to phenotypic variation. The mechanism by which genetic variants contribute to a phenotype is determined by the genetic architecture of the phenotype, however, the underlying rules that determine how genetic variants contribute to phenotype diversity are still not fully known.

Monogenic traits are determined by genetic variation in one gene whereas complex traits do not have predictable patterns of inheritance and have heterogenous phenotypes. Genetic variation in many genes, as well as interaction of genes with environmental factors typically contribute to complex trait phenotypes.

Genome wide association studies (GWAS) have been effective in detecting associations between common variants and common diseases since 2005, with the publication of the first large GWAS with good coverage of the genome in 2007 from the Wellcome Trust Case Control Consortium [112]. Although GWAS have resulted in the discovery of thousands of novel associations to hundreds of phenotypes, the loci identified by GWAS explain a small proportion of the estimated heritability of the trait, or the fraction of phenotypic variation in a population that is due to genetic variation. There are many explanations that could account for this “missing heritability,” or the proportion of heritability not accounted for by these significant GWAS loci, including gene-environment interactions, epistatic interactions, inflated heritability estimates, rare variants or structural variants not tagged by GWAS SNPs, common variants with small effect sizes, and parent of origin effects [113, 33, 39, 116]. Heritability of parent of origin effects on traits have been extensively studied in mice [9, 73, 10, 8] but is only beginning to be studied in humans [56]. The significant GWAS associations, or “low-hanging fruit” have been further investigated to determine causality

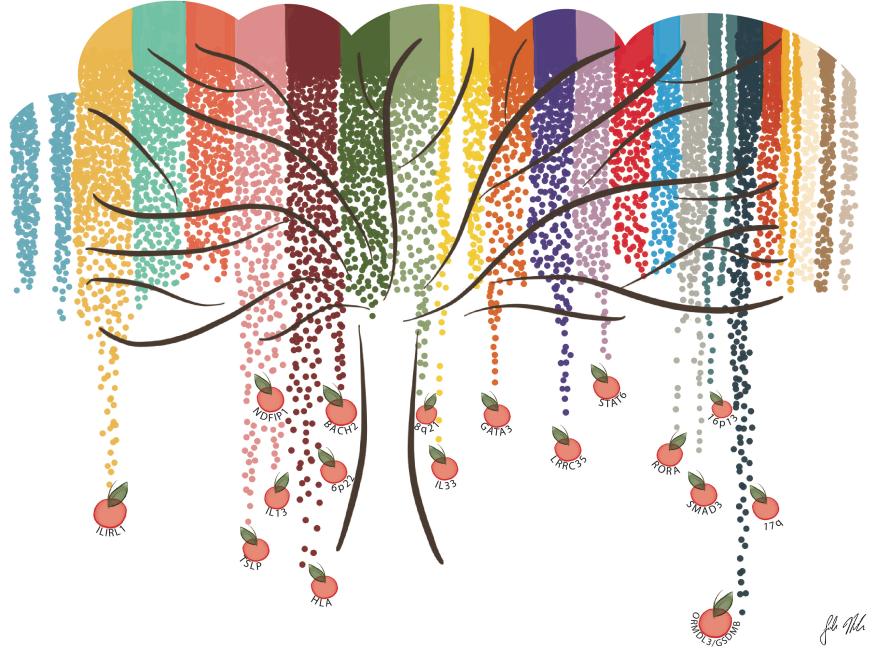


Figure 1.1: Asthma GWAS Manhattan Plot. Inverted manhattan plot for asthma GWAS highlighting “low hanging fruit” as apples hanging from the manhattan plot “tree.” Figure from Ober, 2016 [79] updated to reflect results from Demenais, 2018 [28].

but there still remains a lot of the genome, or “mid-hanging fruit,” that could contribute to a trait. Identifying which variants are causal among those with small p-values that are not genome wide-significant is challenging (Figure 1.1).

Additionally, in GWAS, the impact of parental origin of associated alleles has been largely ignored, and maternal and paternal alleles are treated as equivalent. Sequence variants could affect disease susceptibility or a quantitative trait differently depending on whether the variant was inherited from the father or the mother. Parent of origin effects include phenomena such as imprinting where epigenetic modifications determined by parental origin allows for differential gene expression of genes on homologous chromosomes [62, 57].

The classic examples of parent of origin effects are imprinted genes. More than 80% of imprinted genes in humans are found in genomic clusters, and at least thirteen clusters have been identified on eight chromosomes [57, 87, 88, 2]. These clusters contain both maternally

and paternally expressed genes as well as non-coding RNA genes [87, 2]. Parent-specific expression of the genes within a cluster is determined by cis-acting imprinting control regions (ICRs). ICRs show parental allele-specific DNA methylation and chromatin modifications. ICRs methylated in females during oogenesis typically contain the promoters of long non-coding RNA that are antisense to a protein- coding gene in the cluster and silence it. In contrast, ICRs that acquire methylation in the male germ line are located in intergenic regions.

The testing of maternal and paternal alleles separately can disentangle parent of origin effects. Parent of origin effects can alter gene expression levels that can ultimately affect other phenotypic traits including disease [57, 87] . Moreover, parent of origin eQTLs can provide insight into the molecular mechanisms that may underlie genetic associations with both rare and common diseases [57, 87, 55, 99, 35].

1.2 On The Origin of Genomic Imprinting

Genomic imprinting in its broadest sense suggests that a phenotype observed for a particular gene or genes depends on the sex of the parent from which the gamete containing that gene or genes originated [95]. A particular gene is imprinted if it results in a different phenotype when it is maternally inherited versus paternally inherited. The phenotype can be cytological, morphological, behavioral or biochemical [95].

The first use of the term “imprinting” was used in reference to the recognition and selective elimination of the paternal chromosomes in *Sciara* [25, 95]. “The ‘imprint’ a chromosome bears is unrelated to the genic constitution of the chromosome and is determined only by the sex of the germ line through with the chromosome has been inherited.” [25]

The preferential inactivation of the paternally-derived X chromosomes in mouse was the first demonstration of a functional imprint in mammalian genomes [103, 65, 20]. Imprinting on autosomes was first suggested by a deletion on mouse chromosome 17 that showed a

different phenotype based on which parent the deletion was inherited from [47, 91]. It was not until the development of the pronuclear transplantation technique that allowed for the creation of mice zygotes which contained only maternal or only paternal genetic contributions that there was any evidence that the maternal and paternal genomes are not equal. The differential imprinting on the parental chromosomes prevented complete embryonic development in these mice with complete uniparental disomy [95, 72]. Parental chromosomes have different regions silenced, or imprinted, such that one parental copy is expressed and having either both parental copies or neither expressed results in genetic and developmental abnormalities.

Further experiments suggested that imprinting occurs during gametogenesis and is necessary for full term development. An egg with a male pronucleus developed to term, but, an egg with two female pronuclei (gynogenetic embryos) or two male pronuclei (androgenetic) developed poorly[101, 72]. This provides evidence that input from both parents are required for normal development and the genome of the egg and sperm nuclei are not equal. Non-complementation in genetic crosses of translocated chromosomes provided a way to refine the imprinted regions of the genome[19].

Genetic characterization of Prader-Willi syndrome (PWS) was the first human genetic disease to be associated with maternal heterodisomy of chromosome 15q11-13[75]. It suggested that clinical phenotype of PWS arises from the absence of paternal contribution of 15q11-13 as opposed to a specific genetic mutation. Conversely the absence of maternal contribution to the same region should result in Angelman syndrome (AS) [75, 90]. This provided more evidence that, at “imprinted” regions, the functional differences depend on the sex of the transmitting parent and genetic input from both parents are required for normal human development [75]. Various other human imprinted syndromes due to loss or gain of expression of imprinted genes have been summarized in Table 1.1.

| Human syndrome | Syn- | Location | Major features | Causes |
|---|------|------------------------------|--|--|
| Transient neonatal diabetes mellitus type 1[66] | | 6q24 | Neonatal hyperglycaemia and intrauterine growth restriction | Overexpression of <i>PLAG1</i> and <i>HYMAI</i> |
| Silver-Russell syndrome[32, 106] | | 11p15.5 (65%), MatUPD7 (10%) | Dysmorphism, intrauterine growth restriction and postnatal growth retardation | Complex: 11p15.5: hypomethylation of <i>H19</i> DMR, silencing of <i>IGF2</i> and biallelic expression of <i>H19</i> ; <i>MEST</i> and <i>GRB10</i> are candidates for MatUPD7 cases |
| Beckwith-Wiedemann syndrome[22] | | 11p15.5 | Prenatal and or postnatal overgrowth, enlarged tongue, abdominal wall defects, placental overgrowth and predisposition to embryonal tumours. | Complex: mostly epigenetic errors- silencing of <i>CDKN1C</i> or biallelic expression <i>IGF2</i> and silencing of <i>H19</i> ; inactivating mutations in <i>CDKN1C</i> ; PatUPD11 |
| Temple Syndrome[46, 52] / MatUPD14 syndrome | | 14q32 | Prenatal and postnatal growth retardation, premature puberty and obesity | Loss of paternal expression of <i>DLK1</i> and <i>RTL1</i> |
| Kagami-Ogata syndrome[51, 52, 81] / PatUPD14 syndrome | | 14q32 | Dysmorphism, placentomegaly and excessive amniotic fluid | Increased expression of <i>RTL1</i> |
| Prader-Willi syndrome[17] | | 15q11-13 | Developmental delay, obesity, hypogonadism, cognitive impairment | Loss of paternal expression up to 11 genes in 15q11-13 : paternal deletion of MatUPD15 |
| Angelman syndrome[17] | | 15q11-13 | Developmental delay, microcephaly, absent or limited speech, gait ataxia, characteristic EEG and behavioral profile with happy demeanour | loss of maternal expression of <i>UBE3A</i> , <i>UBE3A</i> mutation or patUPD15 |
| Mulchandani-Bhoj-Conlin syndrome[74] | | chr15 | Prenatal growth restriction, severe short stature with proportional head circumference, and profound feeding difficulty | MatUPD20 |
| Schaaf-Yang syndrome[36] | | chr15 | Delayed psychomotor development, intellectual disability, hypotonia, and behavioral abnormalities | inactivation of <i>MAGEL2</i> on paternal allele |
| Central precocious puberty 2[3] | | chr15 | Development of secondary sexual characteristics before age 8 in girls and age 9 in boys. | inactivation of <i>MKRN3</i> on paternal allele |
| Pseudo-hypoparathyroidism type 1a and type 1b[69, 34] | | 20q13.3 | Dysmorphism, obesity, cognitive impairment, end-organ resistance to parathyroid hormone (which results in hypocalcemia and hyperphosphatemia) and resistance to other hormones | Inactivation/lack of maternal <i>GNAS</i> |

Table 1.1: **Imprinted Gene Disorders.** Adapted from Peters (2014) and Mackay and Temple (2017) [87, 67], ordered by chromosome location of causal mutation. UPD = Uniparental Disomy

1.3 The Search for Parent of Origin Effects

Parent of origin effects and imprinted genes have been most elegantly studied in mice, where two inbred strains are bred reciprocally to identify parent of origin effects on gene expression in progeny that have the same genotypes but different patterns of inheritance [8]. Such studies are obviously more challenging in humans. Previous studies have attempted to identify parent of origin alleles using different approaches, addressing parent of origin effects on gene expression and phenotypic traits.

In one study of gene expression, investigators examined Hardy Weinberg Equilibrium estimates using genotypes derived from RNA-seq data and considered imprinted loci to be those with no or fewer than expected heterozygotes (C.T. Watson, ASHG 2014). Garg et al. used gene expression in lymphoblastoid cell lines (LCLs) from 29 CEU and 30 YRI HapMap trios to identify 30 imprinting eQTLs with parent of origin specific effects on expression by first comparing maternal alleles and paternal alleles associated with gene expression, and then comparing reciprocal heterozygotes [38]. A study from the GTEx Consortium used RNA-seq data to determine allele specific expression (ASE) in 45 different tissues from various numbers of individuals to identify new imprinted genes [12]. By considering genes with monoallelic expression that were evenly distributed to both the reference and alternate alleles across individuals as evidence for imprinting, they identified 42 imprinted genes, both known and novel, and used family studies to confirm imprinting of 5 novel genes. Most recently, Santoni et al. identified nine novel imprinted genes using single-cell allele-specific gene expression and identified genes with mono-allelic expression in fibroblasts from 3 unrelated individuals and probands of 2 family trios, and then used the trios to confirm parent of origin of the alleles [94].

Not many studies have searched for parent of origin effects on binary and quantitative traits. In a study on 38,167 Icelanders with known status for 7 diseases, investigators identified variants that were associated with breast cancer only when paternally inherited and

variants associated with type 2 diabetes only when maternally inherited [55]. Parent of origin associations with height in the same Icelandic population ($n=88,835$) identified four associations of which three were in known imprinted regions, one of which was also replicated in the Sardinia population [115].

1.3.1 Dissertation Overview

Large pedigrees are ideal for identifying parent of origin effects [12]. The advantages that large family studies have for these studies include: 1) formally proving parent of origin effects detected from ASE, and 2) detecting subtle imprinting that does not lead to strictly monoallelic expression [12]. The Hutterite population is ideally suited for these studies. The $>1,400$ Hutterite individuals studied by our group are related to each other in a 13-generation pedigree that includes 3,671 individuals, all of whom are descendants of only 64 founders. Ninety-eight Hutterites were initially selected for whole genome sequencing; alleles were phased using Affymetrix framework markers in the 98 individuals and then imputed to the remaining 1,532 Hutterites who had been previously genotyped with the Affymetrix framework markers [61]. After quality control, parent of origin was assigned to more than 10 million variants. Of the 1,532 with genotype data, 431 also had RNA-seq expression data from LCLs, and between 600-1300 individuals have been phenotyped for cardiovascular disease (CVD) associated and asthma associated quantitative traits.

In this dissertation, I use a novel variation on GWAS to detect parent of origin effects on quantitative disease-related traits in the Hutterites that would be normally missed in standard GWAS. I am able to find maternal- and paternal-only effects, as well as opposite parent of origin effects. This method can be applied to any quantitative trait for which we know parent of origin of alleles, including gene expression. The method and results of testing this method with quantitative disease related traits is in Chapter 2. Using LCL gene expression and parent of origin allele information in the Hutterites, I develop a new method

of mapping RNA-seq reads to parental haplotypes and detect known and novel imprinted genes in Chapter 3. The patterns of imprinted genes are validated in a different sample of Hutterite individuals for which we have peripheral blood leukocyte (PBL) RNA-seq and DNA methylation. In Chapter 4, I use methods, including one from Chapter 2, to try and find parent of origin variants that have opposite effects on gene expression. Additionally, we tested for maternal and paternal effects on maternal and paternal gene expression, respectively, and used a parent of origin ASE (PO-ASE) test to identify differences in maternal and paternal gene expression among reciprocal heterozygotes.

CHAPTER 2

PARENT OF ORIGIN EFFECTS ON QUANTITATIVE PHENOTYPES IN A FOUNDER POPULATION

2.1 Abstract¹

The impact of the parental origin of associated alleles in GWAS has been largely ignored. Yet sequence variants could affect traits differently depending on whether they are inherited from the mother or the father. To explore this possibility, we studied 21 quantitative phenotypes in a large Hutterite pedigree. We first identified variants with significant single parent (maternal-only or paternal-only) effects, and then used a novel statistical model to identify variants with opposite parental effects. Overall, we identified parent of origin effects (POEs) on 11 phenotypes, most of which are risk factors for cardiovascular disease. Many of the loci with POEs have features of imprinted regions and many of the variants with POE are associated with the expression of nearby genes. Overall, our results indicate that POEs, which can be opposite in direction, are relatively common in humans, have potentially important clinical effects, and will be missed in traditional GWAS.

2.2 Introduction

Genome-wide association studies (GWAS) typically treat alleles inherited from the mother and the father as equivalent, although variants can affect traits differently depending on whether they are maternal or paternal in origin. In particular, parent of origin effects (POEs) can result from imprinting, where epigenetic modifications allows for differential gene expression on homologous chromosomes that is determined by the parental origin of the chromosome. Mutations in imprinted genes or regions can result in diseases. For example,

1. Citation for chapter: Mozaffari SV, DeCara JM, Shah SJ, Lang RM, Nicolae DL, Ober C. Parent of Origin Effects on Quantitative Phenotypes in a Founder Population bioRxiv (2017).

two very different diseases, Prader-Willi Syndrome (PWS) and Angelman Syndrome (AS), are due to loss of function alleles in genes within an imprinted region on chromosome 15q11-13. Inheriting a loss of function mutation for the *SNRPN* gene from the father results in PWS but inheriting a loss of function mutation for the *UBE3A* gene from the mother results in AS [87, 35]. Long noncoding RNA genes at this and other imprinted regions act to silence (i.e. imprint) genes in cis. Imprinted genes are often part of imprinted gene networks, suggesting regulatory links between these genes [85, 37, 105]. More than 150 imprinted loci have been described in humans [14] but there are likely many other, as yet undiscovered, imprinted loci.

Previous studies have utilized pedigrees to test maternal and paternal alleles separately for association with phenotypes or with gene expression to uncover new imprinted loci [55, 12, 38, 86, 14]. Kong *et al* [55] discovered one locus associated with breast cancer risk only when the allele is inherited from the father and another locus associated with type 2 diabetes risk only when the allele is inherited from the mother. Garg et al. reported parent-of-origin cis-eQTLs with known or putative novel imprinted genes affecting gene expression [38]. Two additional studies by Zoledziewsk et al. and Benonisdottir et al. identified opposite POEs on adult height at known imprinted loci [115, 14]. Both studies reported associations with variants at the *KCNQ1* gene, and one showed additional opposite POEs with height at two known imprinted loci (*IGF2-H19* and *DLK1-MEG3*) [14]. These studies provide proof-of-principle that alleles at imprinted loci can show POEs, some with opposite effects, with common phenotypes.

Many existing studies and methods identify parent of origin effects use case/parent trios or case/mother duos [23, 44, 4, 110, 109]. Similar to Kong *et al.* [55], our method does not require data on the parent and only uses the parent of origin informative alleles which were assigned and phased using PRIMAL [61]. In contrast to Kong *et al.* [55] which used binary traits, our method tests for parent of origin effects on quantitative traits, similar to

Benonisdottir *et al.* [14] which tested for parent of origin effects on height.

No previous study has included a broad range of human quantitative phenotypes or has studied genome-wide variants with effects in different directions depending on the parent of origin. To address this possibility, we developed a statistical model that directly compares the effects of the maternal and paternal alleles to identify effects that are different, including those that are opposite. We applied this model in a study of 21 common quantitative traits that were measured in the Hutterites, a founder population of European descent for which we have phased genotype data [61]. We identified variants with maternally inherited or paternally inherited effects only and variants with opposite POEs. Some of the identified regions have characteristics similar to known imprinted genes. Overall, we show that this model can identify putative novel imprinted regions with POEs for a broad range of clinically relevant quantitative phenotypes.

2.3 Results

2.3.1 GWAS

We first performed standard genome-wide association studies (GWAS) of 21 traits in the Hutterites (Table 2.5). These studies identified one genome wide significant association ($p < 5 \times 10^{-8}$) with each of five of the 21 traits: low density lipoprotein level (LDL)-cholesterol, triglycerides, carotid artery intima media thickness (CIMT), left ventricular mass index (LVMI), and monocyte count. The results of all 21 GWAS are summarized in Table S2 and Supplementary Figure 2.5a. Results for all variants for all GWAS are deposited in dbGaP (phs000185 - submission in progress).

2.3.2 Parent of Origin GWAS

We considered two possible mechanisms of POEs. In the first, the effect size of one parent's allele is close to zero and the effect size of the other parent's allele is significantly different from zero. For these cases, we performed a paternal only or maternal only GWAS. In other cases, the maternal and paternal alleles may both have effect sizes different from zero, but the effects are significantly different from each other or opposite in direction. To detect these types of POEs, we developed a model that tests for differences between parental effects (see Methods). This model is especially powerful to identify variants with parental effects in opposite directions.

2.3.3 Maternal and Paternal GWAS

Using the same phenotypes, genotypes, pedigree, and criteria for significance as in the standard GWAS, we tested for maternal and paternal effects on each trait by testing each parentally inherited allele with the trait of interest, similar to previous studies [55, 115, 38]. Variants were considered to have POEs if they had a p-value less than 5×10^{-8} in only one parent and were not significant in the standard GWAS (i.e., the LDL association on chromosome 19 and the triglycerides association chromosome 11 were not considered to have POEs; see Table 2.7). The most significant parent of origin associations are summarized in Table 2.1. All significant results of the parent of origin GWAS for all 21 phenotypes are included in Table 2.10.

Overall, seven phenotypes had genome-wide significant parent of origin associations: four in the maternal only GWAS and three in the paternal only GWAS. Three cardiovascular disease (CVD)-associated phenotypes (age at menarche, CIMT, LVMI) and one lung function phenotype (forced expiratory volume in one second [FEV₁]) were associated with maternally-inherited alleles only.

A maternally inherited allele at rs7184983 (G) on chromosome 16 was associated with

| Phenotype | rsid | chr:loc | Variant Location | Nearest Gene | MAF | N | Beta (SE) | Paternal GWAS p-value | Maternal GWAS p-value | Standard GWAS p-value |
|---------------------------------|-------------|-------------|------------------|-------------------|-------|------|----------------|-----------------------|-----------------------|-----------------------|
| A. Maternal Associations | | | | | | | | | | |
| Age at Menarche | rs7184983 | 16:56554709 | Upstream (A/G) | <i>BBS2</i> | 0.059 | 336 | 0.862 (0.154) | 5.01e-01 | 3.11e-08 | 6.75e-03 |
| CIMT | rs4077567 | 2:216703202 | Intronic (G/A) | <i>LINC00607*</i> | 0.30 | 429 | 0.047 (0.008) | 5.72E-01 | 3.02E-08 | 4.21E-06 |
| FEV ₁ | rs9849387 | 3:77764243 | Intergenic (A/G) | <i>ROBO2</i> | 0.39 | 1029 | -0.089 (0.015) | 3.87E-01 | 4.10E-09 | 4.38E-04 |
| | rs6791779 | 3:74996505 | Intergenic (C/G) | <i>MIR4444-1*</i> | 0.24 | 879 | -0.102 (0.021) | 6.88E-02 | 1.48E-08 | 4.52E-02 |
| LVMI | rs574232282 | 1:41662388 | Intronic (G/A) | <i>SCMH1</i> | 0.018 | 537 | 0.239 (0.042) | 5.52E-01 | 1.39E-08 | 1.05E-03 |
| B. Paternal Associations | | | | | | | | | | |
| LDL | rs12024326 | 1:227146433 | Intronic (A/G) | <i>ADCK3</i> | 0.175 | 686 | -0.295 (0.048) | 8.06E-10 | 4.21E-01 | 4.24E-05 |
| | rs4843650 | 16:87683486 | Intronic (A/G) | <i>JPH3</i> | 0.448 | 621 | 0.211 (0.036) | 6.57E-09 | 2.21E-01 | 1.50E-04 |
| SBP | rs1536182 | 13:46275415 | Upstream (A/G) | <i>LINC01055*</i> | 0.2 | 684 | -0.028 (0.005) | 1.53E-08 | 1.78E-01 | 6.93E-04 |
| Total cholesterol | rs113588203 | 1:228979156 | Intergenic (G/T) | <i>RHOU</i> | 0.099 | 703 | -0.341 (0.060) | 1.76E-08 | 7.43E-02 | 8.08E-03 |

Table 2.1: **Phenotypes with significant single parent of origin associations.** *The most significant variant ($P < 5 \times 10^{-8}$) at each locus for the (A) maternal and (B) paternal associations associated with each phenotype is shown. *non-coding RNA genes

younger age of menarche ($P = 3.11 \times 10^{-8}$) (Figure 2.1). This SNP, rs7184983, is located upstream of the *BBS2* gene and is associated with increased expression of *OGFOD1* in transformed fibroblast cells and tibial nerve [40]. The maternally inherited allele at rs4077567 (G) on chromosome 2 was associated with decreased CIMT ($P = 3.02 \times 10^{-8}$) (Figure 2.6). This SNP is in the intron of a long intergenic noncoding gene, *LINC00607*, that is expressed in aorta, coronary, and tibial artery, all tissues potentially relevant to CIMT and atherosclerosis [40]. A maternally inherited allele at rs574232282 (G) in the intron of *SCMH1* on chromosome 1 was associated with increased LVMI ($P = 1.39 \times 10^{-8}$) (Supplementary Figure 2.7). *SCMH1* is expressed in aorta, coronary, and tibial artery [40]. *SCMH1* protein associates with the polycomb group multiprotein complexes required to maintain the transcriptionally repressive state of certain genes [40]. Lastly, maternally inherited alleles at rs9849387 (A) and rs6791779 (C) on chromosome 3 were both associated with reduced FEV₁ ($P = 4.10 \times 10^{-9}$ and 1.48×10^{-8} , respectively) (Supplementary Figure 2.8). The nearest gene to rs9849387 is *ROBO2* (65kb, downstream), which is expressed in the lung as well as in brain, and ovary

[40]. The nearest gene to rs6791779 is *MIR4444-1*(267kb) whose expression has not been characterized.

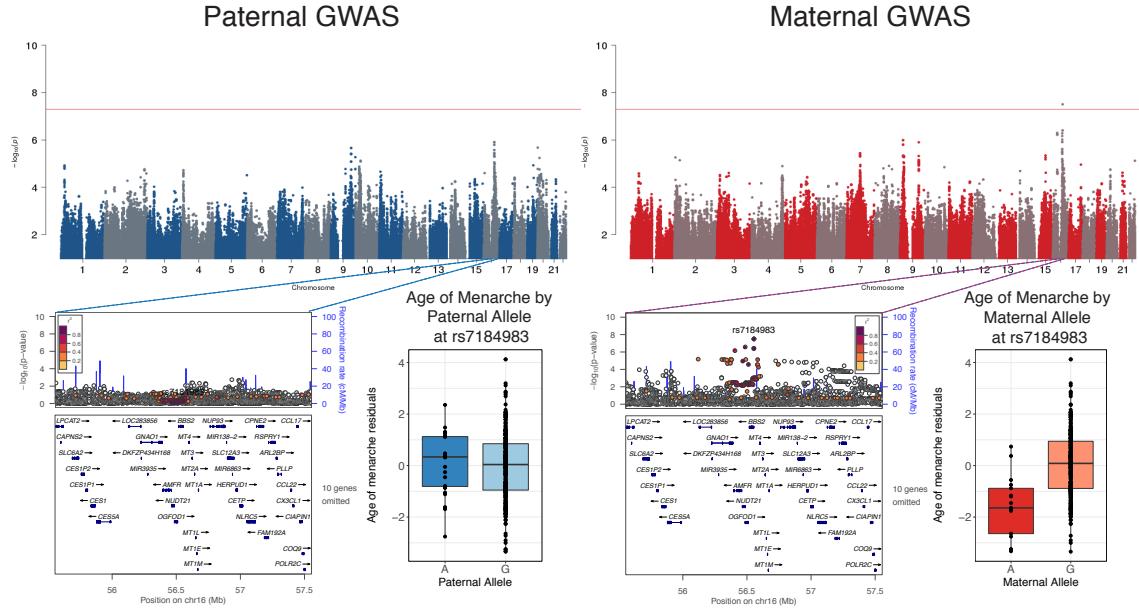


Figure 2.1: Maternal and Paternal GWAS results for Age of Menarche. The top panel shows the Manhattan plots from the maternal (left) and maternal (right) GWAS. LocusZoom plots for both GWAS are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of age of menarche residuals (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $+/-1.5 \times \text{IQR}$.

Three other CVD-related phenotypes (systolic blood pressure, LDL-C, and total cholesterol) had associations with paternally inherited alleles only. The paternally inherited allele at rs12024326 (A) on chromosome 1 was associated with lower LDL-cholesterol levels ($P = 8.06 \times 10^{-10}$) (Figure 2.2). rs12024326 is in the intron of gene *ADCK3*, and the same allele was associated with increased expression of *ADCK3* in whole blood, as well as decreased expression of a neighboring gene, *CDC42BPA* in brain (cerebellum), heart (left ventricle), esophagus, and tibial artery [40]. The paternal G allele at rs4843650 on chromosome 16 was associated with increased LDL-C and is located in the intron of *JPH3*, which is expressed predominantly in the brain [40]. A SNP on chromosome 13 (rs1536182) was associated with

systolic blood pressure levels when it was inherited from the father (Figure 2.9). The paternally inherited A allele at this SNP was associated with decreased systolic blood pressure, as well as decreased expression of its closest gene, *LINC01055*, a long intergenic noncoding gene, in testis [40]. A paternally inherited allele at rs113588203 (G) on chromosome 1 was associated with lower total cholesterol ($P = 1.76 \times 10^{-8}$) (Figure 2.10). This SNP is intergenic between *RHOU* (96kb, downstream), which is expressed across multiple tissues, and *MIRR4454* (331kb), which is expressed in adipose, kidney and heart tissues [40].

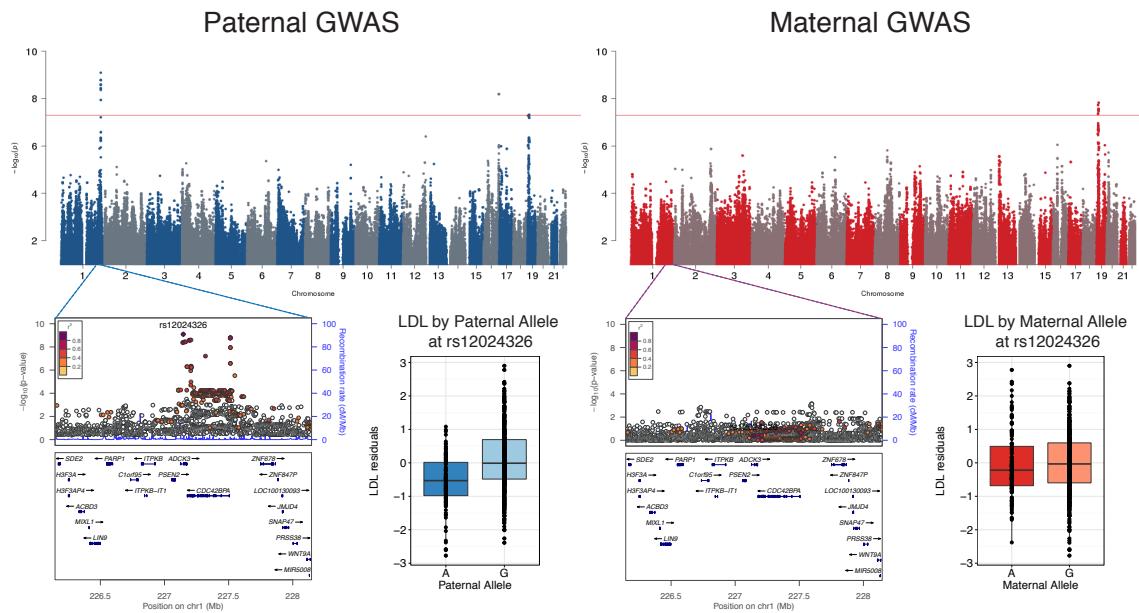


Figure 2.2: Maternal and Paternal GWAS results for LDL Cholesterol. The top panel shows the Manhattan plots from the maternal (left) and maternal (right) GWAS. LocusZoom plots for both GWAS are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of LDL residuals (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $\pm 1.5 \times$ IQR.

2.3.4 GWAS for Differential Parent of Origin Effects

Because some imprinted regions include genes that have both maternal or paternal specific tissue expression, we next tested for such differential effects with these 21 phenotypes. In

| Phenotype | rsid | chr:loc | Variant Location | Nearest Gene | MAF | $\beta_M - \beta_P$ (SE) | Opposite Effect GWAS | Paternal GWAS | | Maternal GWAS | | GWAS p-value |
|-------------------|------------------------|-------------|------------------|----------------------|-------------------|-----------------------------|----------------------|------------------|---------------------|---------------------|----------|--------------|
| | | | | | | | | P-value | Beta(SE) | P-value | Beta(SE) | |
| Age of menarche | rs12447191 | 16:62199299 | Intergenic | <i>CDH8</i> | 0.17 | -0.654 (0.109) | 5.27E-09 | 5.20E-06 | 0.391 (0.085) | 1.85E-05 (0.085) | -0.368 | 8.68E-01 |
| BMI | rs77785972 | 5:97415767 | Intergenic | <i>LINC01340</i> * | 0.025 | 0.154 (0.025) | 5.12E-10 | 5.84E-07 | -0.094 (0.019) | 1.58E-05 (0.019) | 0.081 | 5.39E-01 |
| | rs17605739 | 6:22962798 | Intronic | <i>RP1-209A6.1</i> * | 0.17 | 0.053 (0.010) | 3.01E-08 | 6.99E-05 | -0.032 (0.008) | 1.42E-06 (0.007) | 0.034 | 1.56E-01 |
| Eosinophil count | rs2355879 | 1:18732860 | Intergenic | <i>IGSF21</i> | 0.14 | 0.091 (0.016) | 1.69E-08 | 5.83E-08 | -0.065 (0.012) | 5.59E-04 (0.012) | 0.043 | 2.53E-01 |
| FEV1 | rs12714812 | 3:74813002 | Intergenic | <i>CNTN3</i> | 0.45 | -0.119 (0.021) | 4.52E-08 | 1.78E-03 | 0.052 (0.017) | 6.35E-06 (0.016) | -0.073 | 9.58E-01 |
| LDL | rs1032596 | 16:86281537 | Intronic | <i>LINC01081</i> * | 0.30 | -0.310 (0.056) | 3.69E-08 | 1.05E-06 | 0.201 (0.041) | 4.56E-04 (0.042) | -0.148 | 2.71E-01 |
| LVMI | rs16853098 | 2:168013281 | Intronic | <i>XIRP2</i> | 0.12 | -0.091 (0.053) | 4.18E-08 | 5.29E-06 | 0.064 (0.014) | 2.04E-04 (0.013) | -0.048 | 9.26E-01 |
| Neutrophil count | rs14203084118:34371947 | Intonic | <i>TPGS2</i> | 0.042 | -0.224 (0.041) | 4.40E-08 | 2.25E-03 | 0.078 (0.025) | 1.30E-07 (0.035) | -0.188 | 5.77E-01 | |
| Triglycerides | rs7525463 | 1:218860879 | Intronic | <i>MIR548F3</i> * | 0.16 | -0.401 (0.071) | 2.51E-08 | 1.14E-03 | 0.195 (0.060) | 5.52E-08 (0.049) | -0.267 | 2.84E-02 |
| Total cholesterol | rs7033776 | 9:36704465 | Intergenic | <i>MELK</i> | 0.41 | 0.230 (0.041) | 4.12E-08 | 5.60E-08 | -0.183 (0.034) | 2.28E-03 (0.032) | 0.099 | 6.70E-02 |

Table 2.2: **Significant Opposite Parent of Origin Effect GWAS Associations.** The most significant variant at each locus for each phenotype is shown. $\beta_M - \beta_P$ represents difference in parental effect size. *non-coding RNA genes

these analyses, we compared the effect and direction of the association between maternal and paternal alleles to identify variants that have different effects, including opposite effects, on the phenotype. Such loci would be completely hidden in standard GWAS in which paternally and maternally inherited alleles are combined. These opposite effect GWAS revealed 11 independent loci with opposite POEs for nine different traits, at least six of which are associated with CVD risk (Table 2.2, Figure 2.4).

A locus on chromosome 16, near the *CDH8* gene (128kb, upstream), was associated with opposite POEs with age of menarche (Figure 2.3). *CDH8* is highly expressed in the brain, as well as in the aorta artery and pituitary gland. Two loci on chromosomes 5 and 6 were associated with opposite POEs on body mass index (BMI) (Figure 2.4). The most significant variant on chromosome 5 (rs77785972) is near a long intergenic noncoding gene, *LINC01340* (409kb, downstream), whose expression has not been well characterized. The SNP on chromosome 6 (rs17605739) is also in a long intergenic noncoding gene, *RP1-209A6.1*, which is expressed in low levels in the tibial artery, bladder, spleen, lung, pituitary gland, as well as testis.

A SNP on chromosome 16 (rs1032596) was associated with opposite POEs on LDL-

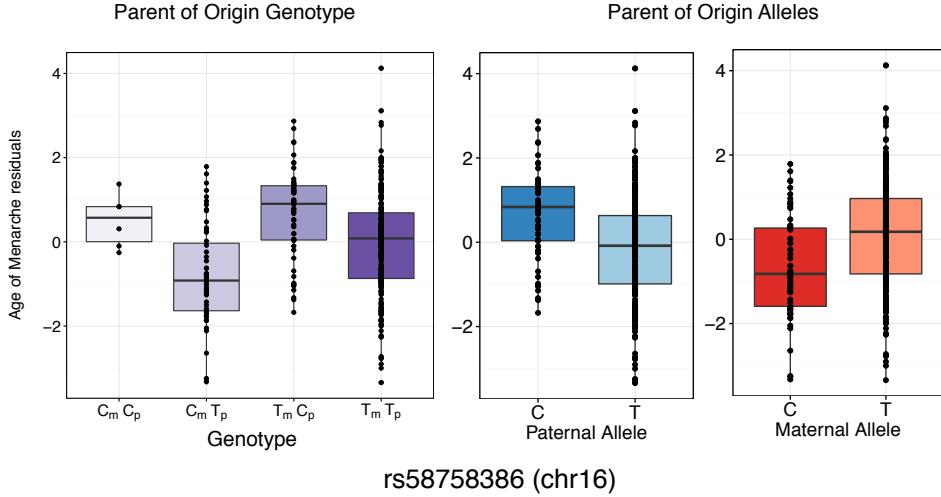


Figure 2.3: Opposite Effect Parent of Origin GWAS Result for Age of Menarche. Box plots of age of menarche residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased age of menarche. The paternal C allele is associated with increased and the paternal T allele with decreased age of menarche. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

cholesterol (Figure 2.12). This SNP lies in the intron of another long noncoding RNA gene, *LINC01081*, which has been suggested to be imprinted because its downstream genes have also been shown to have parent- and tissue-specific activity [102]. A region on chromosome 2 has opposite effects associated with LVMI (Figure 2.13). The associated SNPs are in the intron of *XIRP2*, a cardiomyopathy associated protein that is expressed in skeletal muscle and heart left ventricle, suggesting that this gene could play a role in determining left ventricular mass [107, 78, 40]. In addition, the most significant SNP at this region, rs17616252 (and multiple SNPs in LD) is a strong eQTL ($P = 1.8 \times 10^{-13}$) for the gene *XIRP2* in skeletal muscle, *XIRP2-AS1* in testis, and *B3GALT1* in transformed fibroblast cells [40]. Four variants in a region on chromosome 1 in a microRNA gene, *MIR548F3*, were associated with opposite POEs on triglyceride levels (Figure 2.14). The expression of *MIR548F3* has not been characterized. SNP rs7033776 near *MELK* (27kb, downstream) on chromosome 9 was

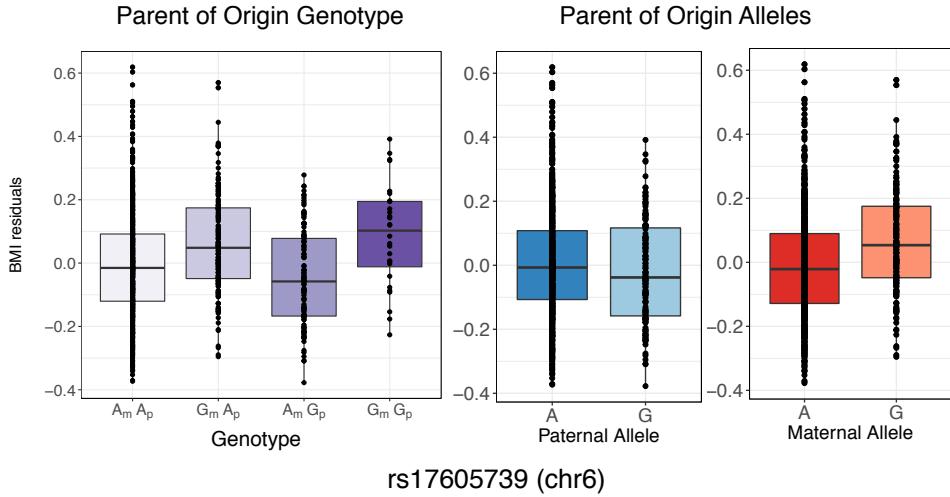


Figure 2.4: Opposite Effect Parent of Origin GWAS Result for BMI. Box plots of two significant loci plot BMI residuals (y-axes) for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. For the (A) SNP on chromosome 5 the maternal A allele is associated with decreased and maternal G allele with increased BMI. The paternal A allele is associated with increased and the paternal G allele with decreased BMI. For the (B) SNP on chromosome 6 the maternal A allele is associated with decreased and maternal G allele with increased BMI. The paternal A allele is associated with increased and the paternal G allele with decreased BMI. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

associated with opposite effects on total cholesterol (Figure 2.15). *MELK* is expressed in the colon and esophagus in addition to transformed lymphocytes and fibroblasts [40].

Nine linked variants on chromosome 1 were associated with opposite POEs of blood eosinophil count (Figure 2.16). These variants are near the gene *IGSF21* (27kb, downstream) which is a member of the immunoglobulin superfamily and likely acts as a receptor in immune response pathways [83]. A variant on chromosome 3, rs12714812, was associated with opposite POEs for FEV₁ (Figure 2.17). This variant has been shown to regulate the expression of a gene *CNTN3* (45kb, upstream) in heart and brain [40]. Studies in mice have suggested that this gene is imprinted and maternally expressed in the murine placenta [16]. Variant rs142030841 in the intron of the gene *TPGS2* on chromosome 18 has opposite POEs with neutrophil levels (Figure 2.18). This SNP is an expression quantitative trait

locus (eQTL) for the noncoding RNA gene *RP11-95O2.5* in skin, testis, breast, thyroid and adipose tissue, for *CELF4* in tibial nerve and lung, and for *TPGS2* in tibial artery and transformed fibroblast cells [40].

2.3.5 Parent of Origin Effects on Gene Expression

To determine if any of the associated variants also showed POEs on gene expression in the Hutterites, we used RNA-seq gene expression data from lymphoblastoid cell lines (LCLs) collected from 430 of the individuals in the GWAS sample. We first tested for association of maternal and paternal variants with genes detected as expressed in the LCLs and whose transcript start site was within 1Mb of each associated SNP. There were no significant associations after multiple testing correction, similar to a previous study [14]. However, because we considered this to be exploratory analyses, we show results for the five most significant parent of origin eQTLs (Table 2.3). We next used the opposite effect model for each SNP in Table 2.2 and expression of all genes that were detected as expressed in LCLs and whose transcript start site was within 1Mb of the associated SNP. This resulted in 57 tests (1 SNP for each of 8 phenotypes, and 57 genes). The five most significant opposite effect eQTLs, none of which passed the Bonferroni threshold of 8.77×10^{-4} , are shown in Table 2.4. The most significant opposite effect eQTL was for *POLR1E* expression with the SNP on chromosome 9 (rs7033776) that was associated with total cholesterol (opposite effect eQTL $P = 9.86 \times 10^{-4}$) (Figure 2.19). *POLR1E* is involved in the purine metabolism pathway as well as DNA-directed polymerase activity. The same SNP, rs7033776, had modest opposite effects with the expression of three other genes in the region (*PAX5*, *FBXO10*, and *FRMPD1*), a signature consistent with an imprinted region. Another SNP with opposite POEs on LVMI, rs16853098, was an opposite effect eQTL for *STK39*, a gene that has been previously associated with hypertension [108].

| Phenotype | Sample Size | rsid | chr:loc | Gene | Beta (SE) | Maternal eQTL p-value | Paternal eQTL p-value |
|---------------------------------|-------------|-------------|-------------|-----------------|-------------------|-----------------------|-----------------------|
| A. Maternal Associations | | | | | | | |
| CIMT | 334 | rs4077567 | 1:216703202 | <i>ABCA12</i> | 0.039 (0.017) | 2.14E-02 | 1.53E-02 |
| Age at menarche | 336 | rs7184983 | 16:56554709 | <i>POLR2C</i> | -0.085 (0.039) | 2.91E-02 | 7.93E-01 |
| Age at menarche | 336 | rs7184983 | 16:56554709 | <i>SLC12A3</i> | -0.064 (0.031) | 3.77E-02 | 2.28E-01 |
| CIMT | 334 | rs4077567 | 1:216703202 | <i>RPL37A</i> | 0.030 (0.016) | 5.72E-02 | 5.90E-01 |
| LVM | 457 | rs74232282 | 1:41662388 | <i>SMAP2</i> | 1.40 (0.159) | 5.82E-02 | 1.12E-01 |
| B. Paternal Associations | | | | | | | |
| Total cholesterol | 352 | rs113588203 | 1:228979165 | <i>HIST3H2A</i> | 0.560 (0.308) | 8.81E-01 | 6.85E-02 |
| Total cholesterol | 352 | rs113588203 | 1:228979165 | <i>SPHAR</i> | 0.073 (0.047) | 6.01E-01 | 1.20E-01 |
| LDL | 352 | rs1110603 | 16:87687317 | <i>MAP1LC3B</i> | -0.024 (0.015) | 4.35E-01 | 1.25E-01 |
| LDL | 352 | rs1110603 | 16:87687317 | <i>FBXO31</i> | -0.027 (0.018) | 1.56E-01 | 1.36E-01 |
| Total cholesterol | 357 | rs113588203 | 1:228979165 | <i>RAB4A</i> | -0.039 (0.028) | 6.16E-01 | 1.59E-01 |

Table 2.3: **Parent of Origin eQTLs in LCLs.** The most significant SNP for each phenotype (Table 2.1) was tested for association with gene expression for genes with TSS within 1Mb of the SNP. The effect sizes correspond to the maternal (A) or paternal (B) effect sizes.

| Phenotype | Sample Size | rsid | chr:loc | Gene | $\beta_M - \beta_P$ (SE) | Opposite Effect p-value |
|-------------------|-------------|------------|-------------|---------------|--------------------------|-------------------------|
| Total cholesterol | 381 | rs7033776 | 9:36704465 | <i>POLR1E</i> | 0.0603 (0.399) | 9.86E-04 |
| Total cholesterol | 381 | rs7033776 | 9:36704465 | <i>PAX5</i> | 0.0608 (0.0253) | 0.0162 |
| Total cholesterol | 381 | rs7033776 | 9:36704465 | <i>FBXO10</i> | 0.0789 (0.0337) | 0.019 |
| LVMI | 355 | rs16853098 | 2:168013281 | <i>STK39</i> | -0.238 (0.124) | 0.055 |
| Total cholesterol | 381 | rs7033776 | 9:36704465 | <i>FRMPD1</i> | 0.185 (0.0988) | 0.060 |

Table 2.4: **Opposite Parent of Origin eQTLs in LCLs.** The most significant SNP for each phenotype (Table 2.2) was tested for opposite effect eQTLs with genes with TSS within 1Mb of the SNP. The effect size corresponds to the difference in maternal and paternal effect sizes.

2.4 Discussion

In this study, we introduced a novel statistical method that allows assessment of standard GWAS signals along with measures of differential POEs on common quantitative phenotypes. Similar to previous parent of origin studies of fewer phenotypes [55, 14, 38], we tested for associations of maternally- or paternally-derived alleles with each phenotype. We then extended this method to identify variants for which maternally- and paternally-derived alleles have different, including opposite, effects on phenotypic values. The focus on 21 common disease-associated phenotypes in a single large pedigree allowed us to broadly survey physiological effects of putative imprinted regions and the candidate genes at each associated locus. In contrast to previous studies, our new model can identify variants with opposite POEs that would be missed in traditional GWAS (Table 2.2).

Our studies of >1,000 Hutterites who are related to each other in a single pedigree allowed us to detect POEs, even when few genome-wide significant associations were detected in standard GWAS of the same phenotypes. Our method revealed parent of origin specific genome-wide significant associations for seven of the 21 phenotypes examined, with maternally-inherited alleles associated with four phenotypes, paternally-inherited alleles with three phenotypes (Table 2.1), and opposite parent of origin alleles with nine phenotypes, of which five also showed single POEs at different loci (Table 2.2). Overall, 11 of the 21 phenotypes examined showed genome-wide significant evidence of POEs with alleles at one or more loci. In contrast, standard GWAS of these same phenotypes and using the same markers in these same individuals revealed genome-wide significant association for only five traits.

It is notable that four of the nine significant opposite parent of origin effects (one each with LDL-C and triglycerides, and two with BMI) lie in or near long intergenic non-coding RNA genes (lincRNAs). LincRNAs are a feature of imprinted regions [87], where they can silence the expression of genes on the opposite chromosome [13, 85]. One of the variants, rs1032596, with an opposite parent of origin effect on LDL-C is located in the *LINC01081*

gene. This noncoding RNA, along with *LINC01082*, regulates the *FOXF1* enhancer resulting in *FOXF1* parent- and tissue-specific activity[102] providing experimental support for tissue specific expression, a feature of imprinted regions.

Another variant with POEs in our study has been suggested to be imprinted in previously published work. The variant associated with opposite POEs for FEV₁ is an eQTL for the gene *CNTN3*. *CNTN3* was shown to have exclusive maternal allele-specific expression in murine placentas[16], although this finding may have been due to contaminating maternal cells [82, 89].

Other regions associated with POEs harbor genes involved in transcriptional repression (e.g., *SCMH1* with LVMI on chromosome 1) or the associated SNPs are reported as eQTLs in GTEx with expression in tissues relevant to the phenotype under investigation (e.g., the LVMI-associated SNPs are eQTLs for *XIRP2*, which is expressed in skeletal muscle and heart left ventricle) [40]. Overall, these patterns of expression provide additional support that the parent of origin associations in our study are flagging imprinted regions or regions involved in the regulation of gene expression. Finally, we used gene expression in LCLs from the Hutterites to directly test for parent of origin eQTLs among SNPs associated with phenotypes in the parent of origin GWAS. Although none of the parent of origin eQTLs met criteria for significance after correcting for multiple testing, the SNP on chromosome 9 with opposite POEs on total cholesterol levels was borderline significant as an opposite parent of origin eQTL for *POLR1E*, and possible for three other genes at the same locus (*PAX5*, *FBXO10*, and *FRMPD1*). The presence of multiple genes with potential parent of origin expression patterns is further supportive of an imprinted locus. The availability of gene expression only in LCLs from the Hutterites limits the inferences we can draw about effects on expression because imprinted regions are often tissue-specific and sometimes developmentally regulated [87, 35]. Despite this limitation, the fact that many of the SNPs associated with POEs on phenotypes are themselves eQTLs in relevant tissues (GTEx) and some are suggestive

of having POEs on expression in LCLs from the Hutterites is generally supportive of the suggestion that some of the regions identified in this study are imprinted or have network interactions with imprinted genes[21] in humans. Additionally, our data suggest that loci with POEs influence a broad spectrum of quantitative phenotypes that are themselves risk factors for common diseases.

In particular, the discovery of POEs for eight traits that are associated cardiovascular disease risk is intriguing. These include metabolic phenotypes, such as BMI, total cholesterol, triglycerides, LDL, and age of menarche, that have indirect effects on cardiac health, as well as LVMI and CIMT, which more directly reflect cardiac health. Some of these phenotypes showed associations with paternally inherited alleles only (systolic blood pressure, LDL-C, total cholesterol), maternally inherited alleles only (LVMI, CIMT, and age at menarche), and/or with opposite effect variants (BMI, LDL-C, triglycerides, total cholesterol, LVMI, age at menarche). It has been suggested that genomic imprinting evolved in the mammalian lineage as a way to regulate maternally and paternally expressed genes in the placenta during pregnancy and modulate metabolic functions related to growth, where the parental interests may be in conflict — paternal alleles favoring growth of the fetus at the expense of the mother while maternal alleles favor restricting resources to the fetus to ensure preservation of her nutritional needs [41, 13, 85]. Our data show some effects that are consistent with this theory. For example, three independent paternally inherited alleles on chromosome 1 are associated with increased LDL-C (Fig. 2) and total cholesterol (Figure); a paternal allele on chromosome 13 is also associated with increased systolic blood pressure (Figure). However, it is not always possible to interpret our results in light of this model, such as the association of maternal allele on chromosome 2 with decreased CIMT (Figure 2.7), or the maternal allele on chromosome 16 associated with decreased age of menarche (Figure 2.1), which confers increased cardiovascular risk [18]. However, because many of the traits associated with POEs in this study were measured in adults, and none were measured in neonates, we

are likely observing the downstream effects of processes that occurred in utero. Nonetheless, this kinship theory, or parent-conflict hypothesis, could account for the enrichment of parent of origin associations, particularly those with opposite effects, among metabolic and CVD-associated traits [87].

Finally, we note that the parent of origin GWAS for 21 phenotypes in the Hutterites revealed overall twice as many genome-wide significant loci compared to standard GWAS of the same phenotypes in the same individuals, suggesting that variation at imprinted loci may represent some of the "missing heritability" of these phenotypes and potentially for the disease for which they confer risk. This idea is consistent with observations in both mice and humans [56]. POEs in mice contribute disproportionately to the heritability of 97 traits, including those related to total cholesterol, weight, HDL, and triglycerides [73]. Exactly how much loci with POEs in humans contribute to phenotypic variation and disease risk overall remains to be determined, but our study provides compelling evidence that it is likely to be significant for many important traits.

2.5 Methods

2.5.1 Sample Composition

The individuals in this study have participated in one or more of our studies on the genetics of complex traits in the Hutterites [26, 111, 1]. The more than 1,500 Hutterites in our study are related to each other in a 13-generation pedigree including 3,671 individuals.

2.5.2 Genotype Data

Variants detected in the whole genome sequences of 98 Hutterites were previously imputed to an additional 1,317 individuals using PRIMAL, a high-accuracy pedigree based imputation method [61]. PRIMAL was used to phase alleles and assign parent of origin for 83% of about

12 million autosomal SNPs. For these studies, we selected SNPs that had a MAF 1% and genotype call rate 85%. This yielded 5,891,982 autosomal SNPs. Parent of origin allele call rates differed among individuals and between phenotypes (Table 2.5).

2.5.3 *Phenotype Data*

We included 21 quantitative phenotypes that were previously measured in the Hutterites. Descriptions for each phenotype, as well as exclusion criteria, transformations, and covariates used with each phenotype in the GWAS, are available in the Supplementary Methods (Table 2.5).

Descriptions for 18 of the 21 phenotypes can be found in Cusanovich et al [26]. The remaining three are described here. Height was measured in cm on a stadiometer with shoes removed. BMI was calculated using weight (kg, measured on scale) divided by height (m) squared. Age at menarche was collected retrospectively by interview.

2.5.4 *GWAS*

We used a linear mixed model as implemented in GEMMA to test for genome wide association with 21 phenotypes using an additive model. We corrected for relatedness, as well as relevant covariates (Table 2.5).

2.5.5 *Maternal and Paternal GWAS*

To evaluated the evidence for POEs, we tested maternal and paternal alleles separately with each phenotype, comparing phenotypic differences between the maternally inherited alleles and between the paternally inherited alleles. We used a linear mixed model as implemented in GEMMA, which allows us to correct for relatedness as a random effect, as well as sex, age, and other covariates as fixed effects [114].The linear mixed model for the parent of

origin GWAS for testing maternal alleles and paternal alleles is shown in Equation 2.1 and Equation 2.2, respectively.

$$Y = W\alpha + X_M\beta_M + g + \epsilon \quad (2.1)$$

$$Y = W\alpha + X_P\beta_P + g + \epsilon \quad (2.2)$$

n is the number of individuals, Y is an $n \times 1$ vector of quantitative traits, W is an $n \times c$ matrix of covariates (fixed effects) including intercept 1. α is a $c \times 1$ vector of covariate coefficients. X_M is an $n \times 1$ vector of maternal alleles, and X_P an $n \times 1$ vector of paternal alleles. β_M and β_P are the effect sizes of maternal and paternal alleles, respectively. g is a vector of genetic effects with $g \sim N(0, A(\sigma_g)^2)$ where A is the genetic relatedness matrix; ϵ is a vector of non-genetic effects with $\epsilon \sim N(0, I(\sigma_e)^2)$.

2.5.6 Differential Effect GWAS (PO-GWAS)

To test for a difference in the same allele inherited from each parent, including opposite effects, we re-parameterized the test model (Equation 2.3) from Garg et al [38]. The null model (Equation 2.4) is a standard GWAS model, ignoring parent of origin of alleles. The test model (Equation 2.3) is more significant when maternal and paternal alleles have differential effects on gene expression.

$$Y = W\alpha + X_M\beta_M + X_P\beta_P + g + \epsilon \quad (2.3)$$

$$Y = W\alpha + X_{MP}\beta_{MP} + g + \epsilon \quad (2.4)$$

This new model allows us to measure the difference in parental effect of the same allele

when the genotype is a covariate in Equation 2.5.

$$Y = W\alpha + \frac{X_M - X_P}{2}(\beta_M - \beta_P) + X_{MP}\frac{(\beta_P + \beta_M)}{2} + g + \epsilon \quad (2.5)$$

X_{MP} is a $n \times 1$ vector of genotypes with possible values [0,1,2], equivalent to $X_P + X_M \cdot (\beta_M - \beta_P)$ is the difference in parental effect size. If the difference in parental effect size is large and significantly different from 0 it suggests a parent of origin effect exists at this variant. $((X_M - X_P))/2$ is a $n \times 1$ vector of genotypes with possible values [-1,0,1]. $((\beta_P + \beta_M))/2$ is the average parental effect size that is captured in normal GWAS using genotypes. The average genotypes are added in as a covariate, with the average parental effect size the corresponding covariate coefficient. This differential effect GWAS was tested in GEMMA using BIMBAM format to use average genotype values [96].

2.5.7 Parent of Origin eQTL studies

RNA-seq data from LCLs were available from a previous study in the Hutterites [26]. For this study, sequencing reads were reprocessed as follows. Reads were trimmed for adaptors using Cutadapt (with reads <5 bp discarded) then remapped to hg19 using STAR indexed with gencode version 19 gene annotations [29, 70]. To remove mapping bias, reads were processed using WASP mapping pipeline [104]. Gene counts were collected using HTSeq-count [6]. VerifyBamID was used to identify sample swaps to include individuals that were previously excluded [50]. Genes mapping to the X and Y chromosome were removed; genes with a Counts Per Million (CPM) value of 1 (expressed with less than 20 counts in the sample with lowest sequencing depth) were also removed. The R/Bioconductor package edgeR was used to convert the RNA-seq counts to log2 TMM-normalized CPM values [93, 92]. Technical covariates that showed a significant association with any of the top principal components were regressed out (RNA Integrity Number and RNA concentration).

2.5.8 Maternal and Paternal Parent of Origin eQTL

LCL RNA-seq data was used to test the single parent model for the most significant SNP from the maternal or paternal only GWAS for each phenotype. We selected all genes detected as expressed in the LCLs and residing within 1Mb of each most significant associated SNP. Summary of the SNPs and genes tested are in Table 2.8.

2.5.9 Differential Parent of Origin eQTL

LCL RNA-seq data was used to test the opposite effect model for the most significant SNP in each region that was associated with a phenotype in the parent of origin opposite effects GWAS. We selected all genes detected as expressed in the LCLs and residing within 1Mb of each associated SNP. Summary of the SNPs and genes tested are in Table 2.9.

2.6 Supplementary Information

2.6.1 Supplementary Figures

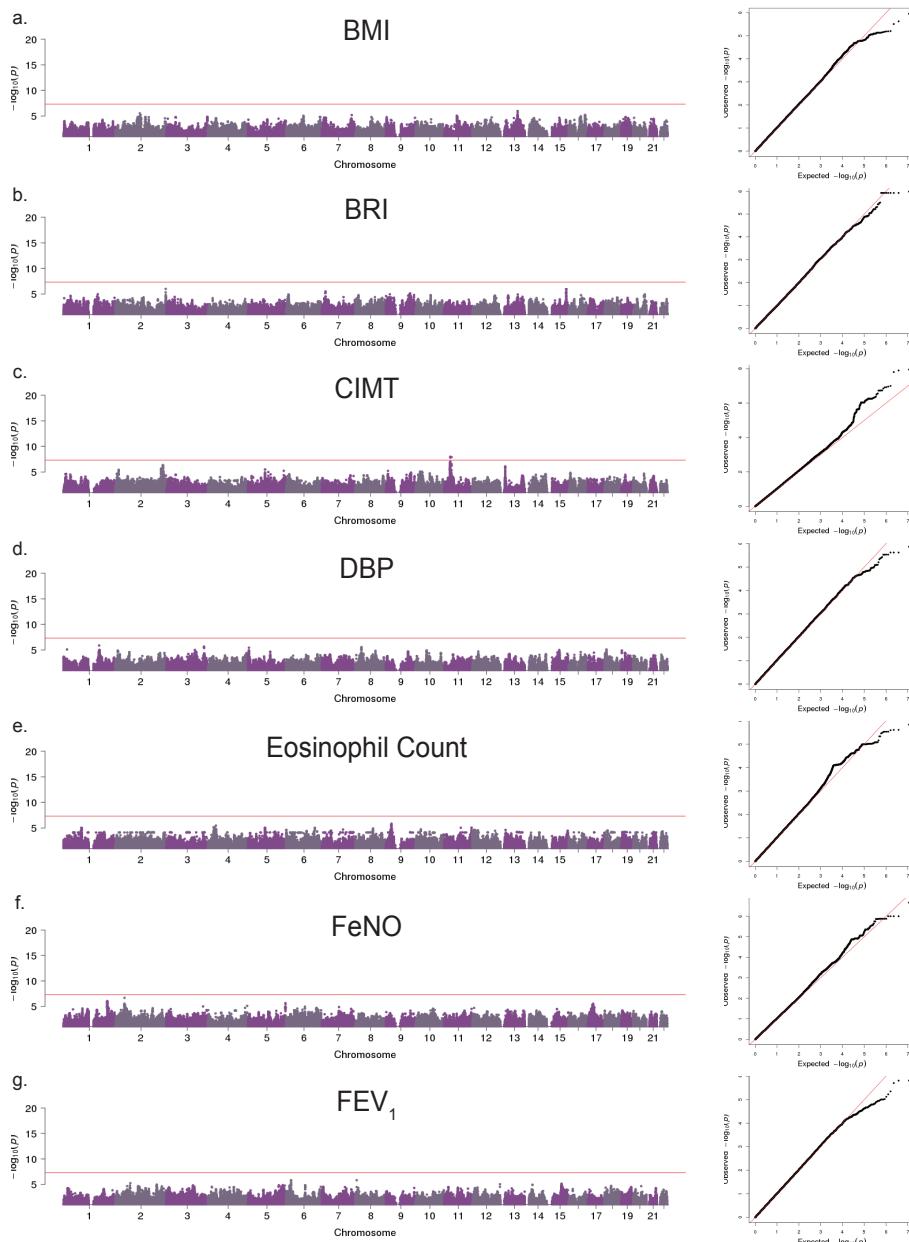


Figure 2.5a: Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes.

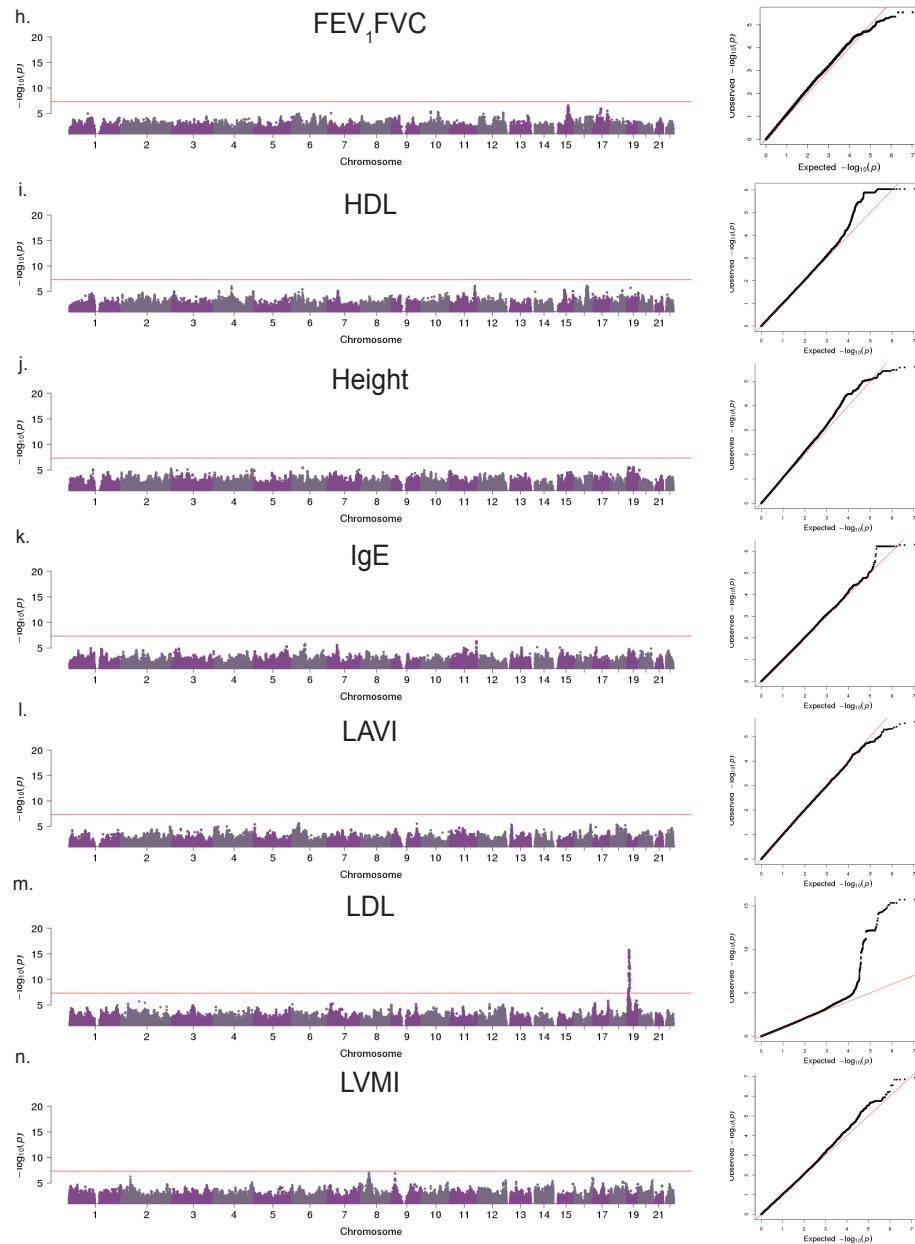


Figure 2.5b: Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes (Continued).

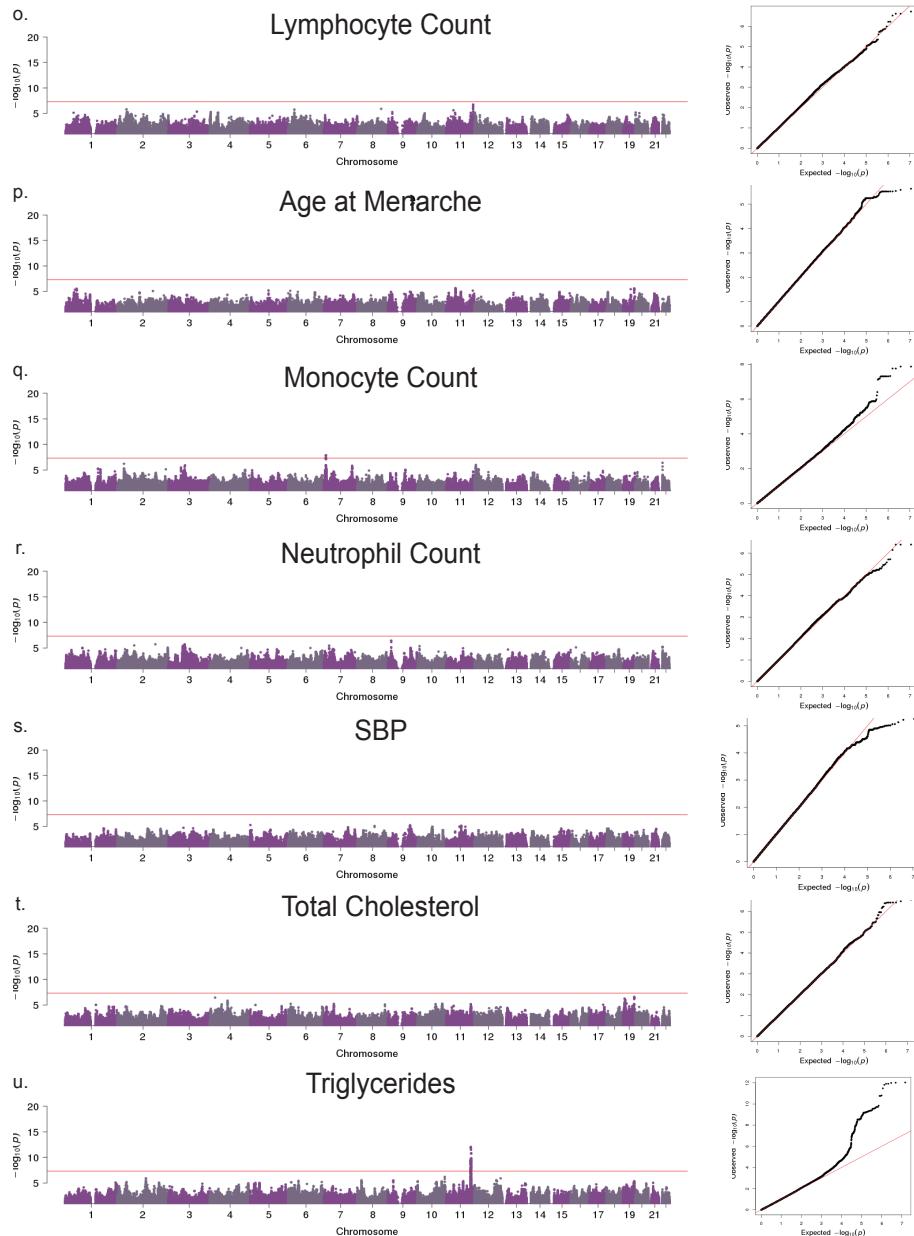


Figure 2.5c: Manhattan and QQ Plots from Standard GWAS of 21 Quantitative Phenotypes(Continued).

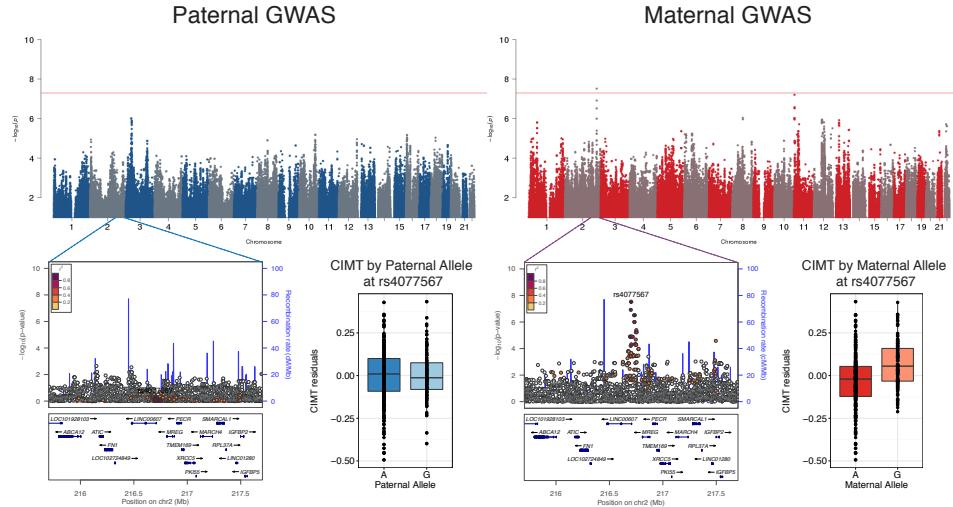


Figure 2.6: Maternal and Paternal GWAS results for CIMT. The top panel shows the Manhattan plots from the paternal (left) and maternal (right) GWAS. LocusZoom plots are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of CIMT residuals (the residuals correspond to the inverse of raw CIMT values) (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $\pm 1.5 \times$ IQR.

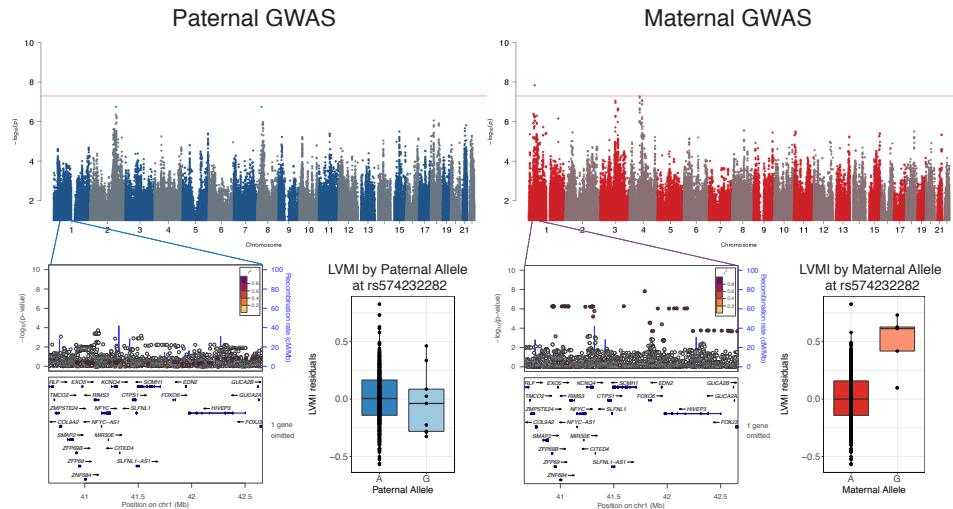


Figure 2.7: Maternal and Paternal GWAS results for LVMI. The top panel shows the Manhattan plots from the paternal (left) and maternal (right) GWAS. LocusZoom plots are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of LVMI residuals (the residuals correspond to the inverse of raw CIMT values) (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $\pm 1.5 \times$ IQR.

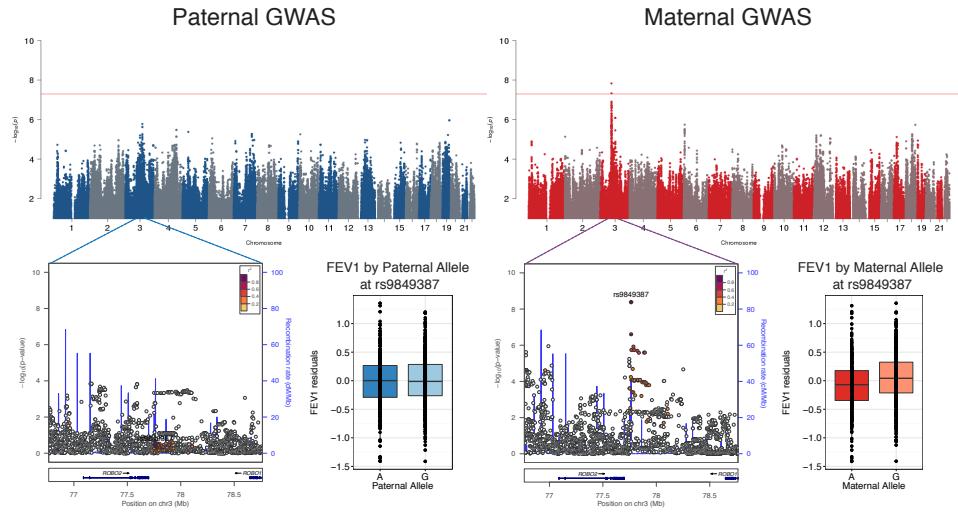


Figure 2.8: Maternal and Paternal GWAS results for FEV_1 . The top panel shows the Manhattan plots from the paternal (left) and maternal (right) GWAS. LocusZoom plots for both GWAS are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of FEV_1 residuals (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $1.5 \times \text{IQR}$.

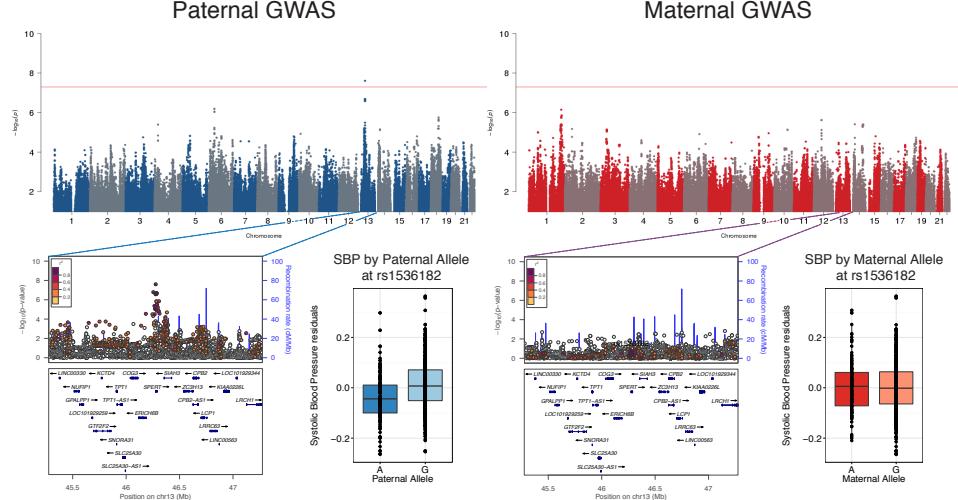


Figure 2.9: Maternal and Paternal GWAS results for Systolic Blood Pressure. The top panel shows the Manhattan plots from the paternal (left) and maternal (right) GWAS. LocusZoom plots for both GWAS are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of systolic blood pressure residuals (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $1.5 \times \text{IQR}$.

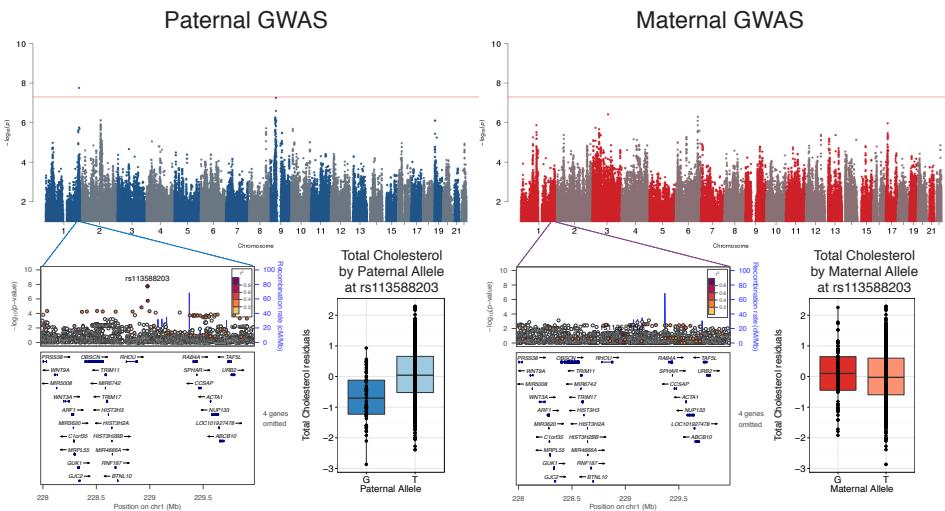


Figure 2.10: Maternal and Paternal GWAS results for Total Cholesterol. The top panel shows the Manhattan plots from the paternal (left) and maternal (right) GWAS. LocusZoom plots for both GWAS are shown in the lower panel for the associated region in the GWAS. Boxplots show the distribution of total cholesterol residuals (y-axes) by the corresponding maternal and paternal alleles at this SNP (x-axes). The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show 1.5 x IQR.

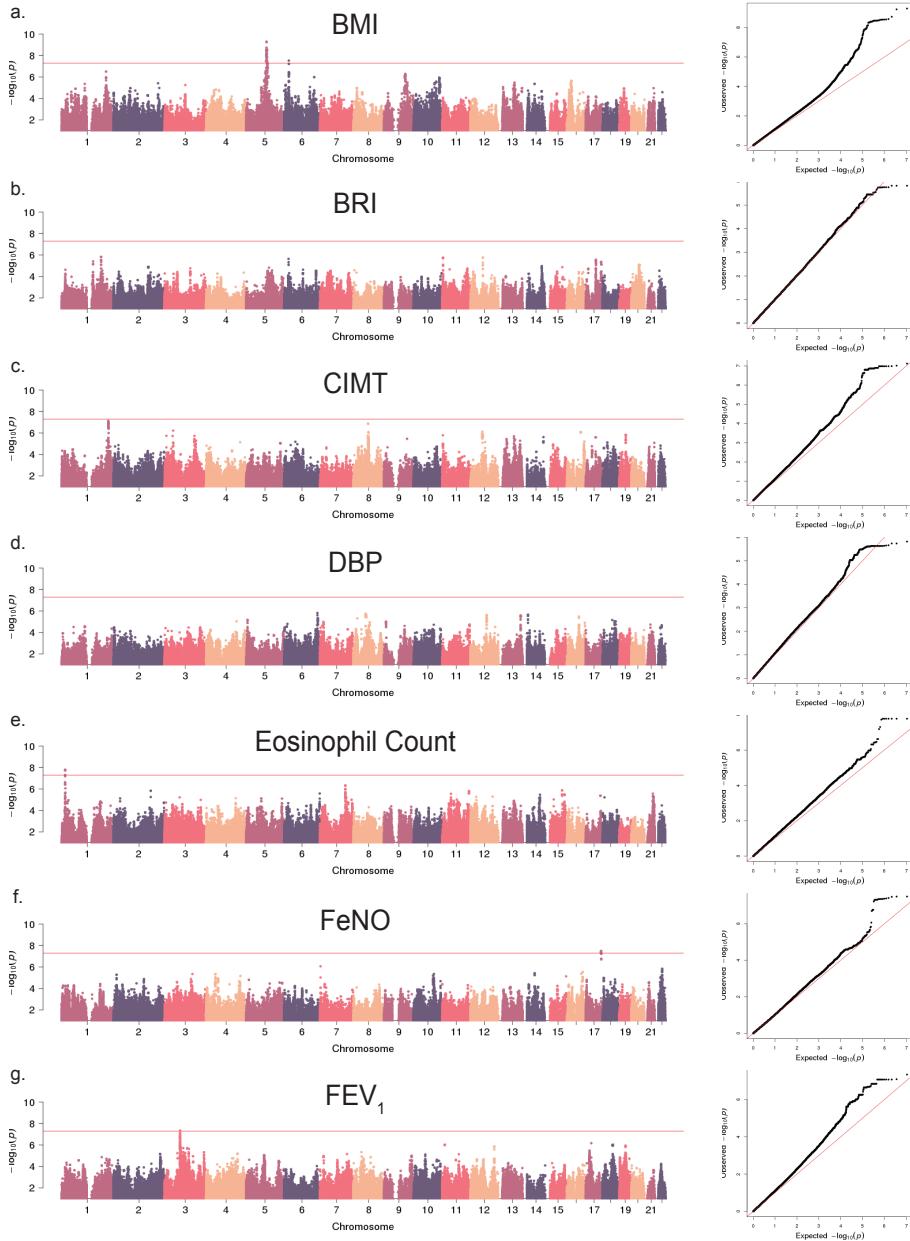


Figure 2.11: Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes .

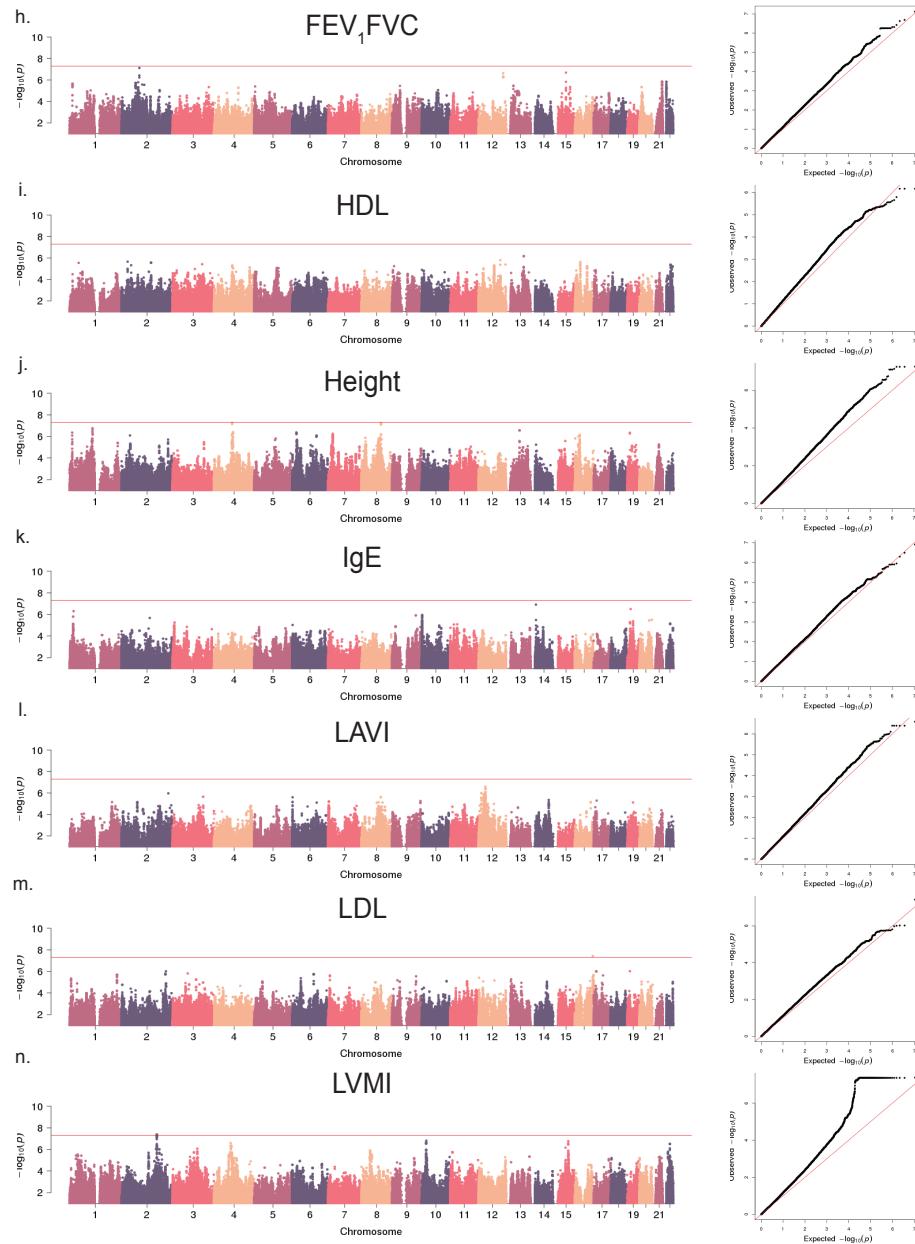


Figure 2.11a: Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes (Continued).

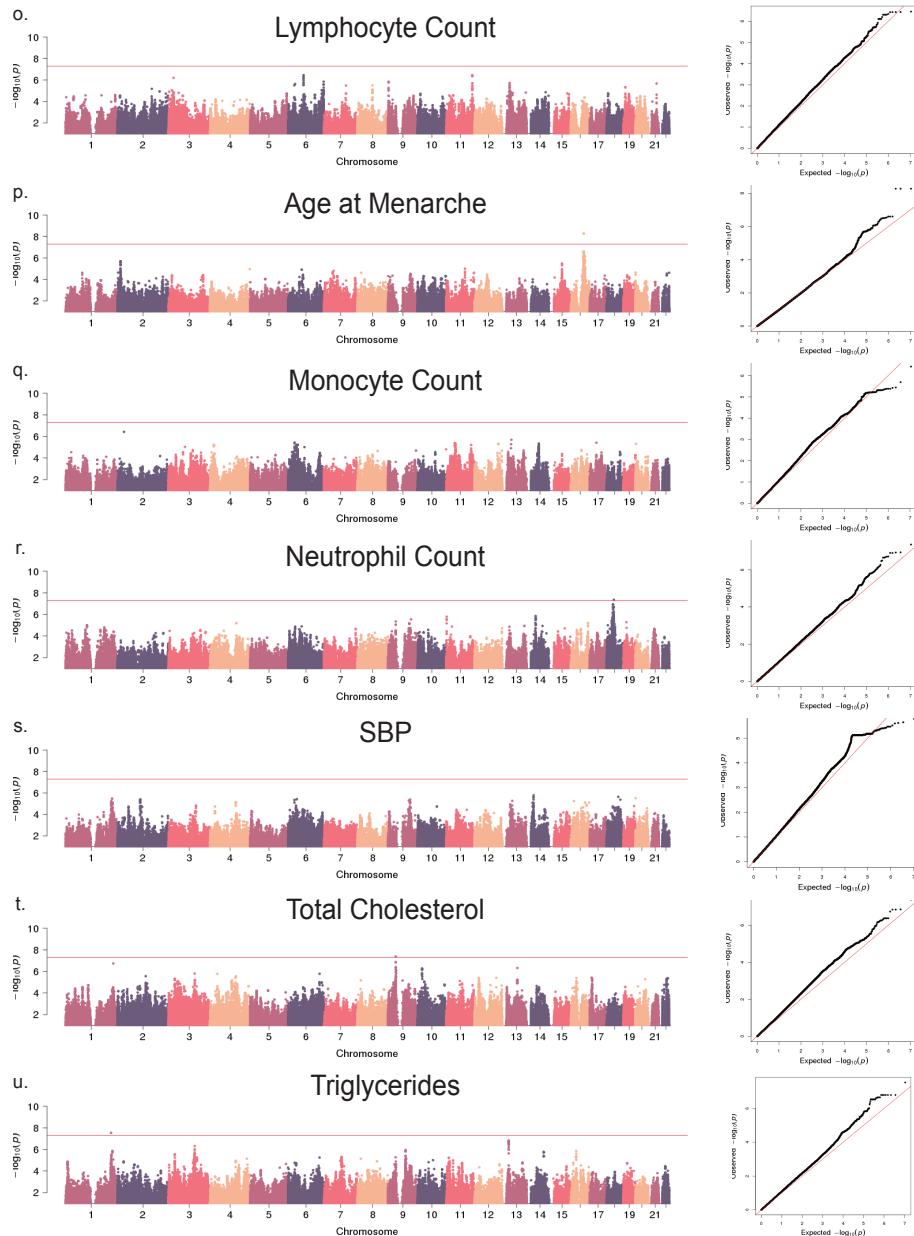


Figure 2.11b: Manhattan and QQ Plots from Differential Effect GWAS of 21 Quantitative Phenotypes (Continued).

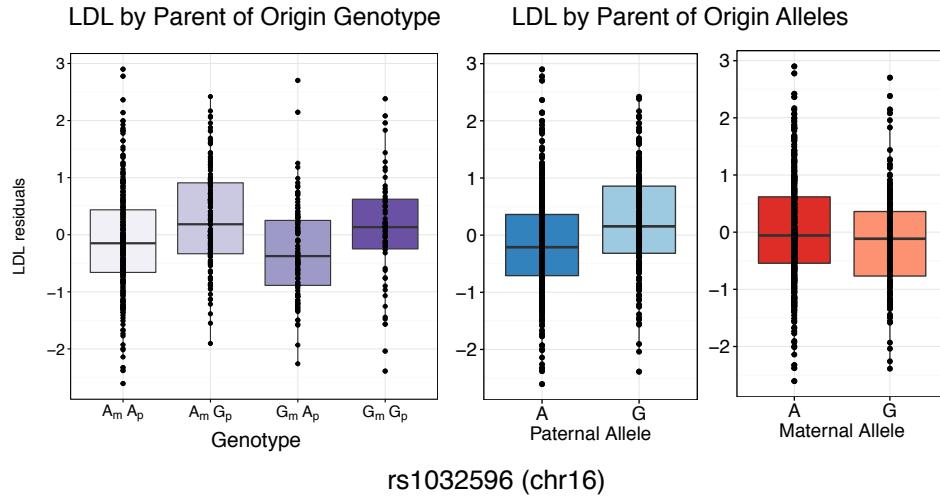


Figure 2.12: Opposite Effect Parent of Origin GWAS Result for LDL. Box plots of LDL residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased LDL. The paternal C allele is associated with increased and the paternal T allele with decreased LDL. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

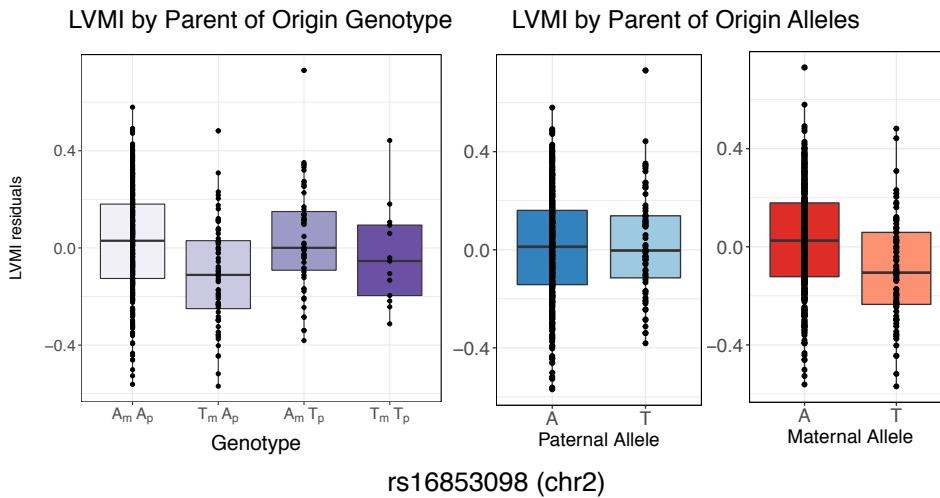


Figure 2.13: Opposite Effect Parent of Origin GWAS Result for LVMI. Box plots of LVMI residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased LVMI. The paternal C allele is associated with increased and the paternal T allele with decreased LVMI. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

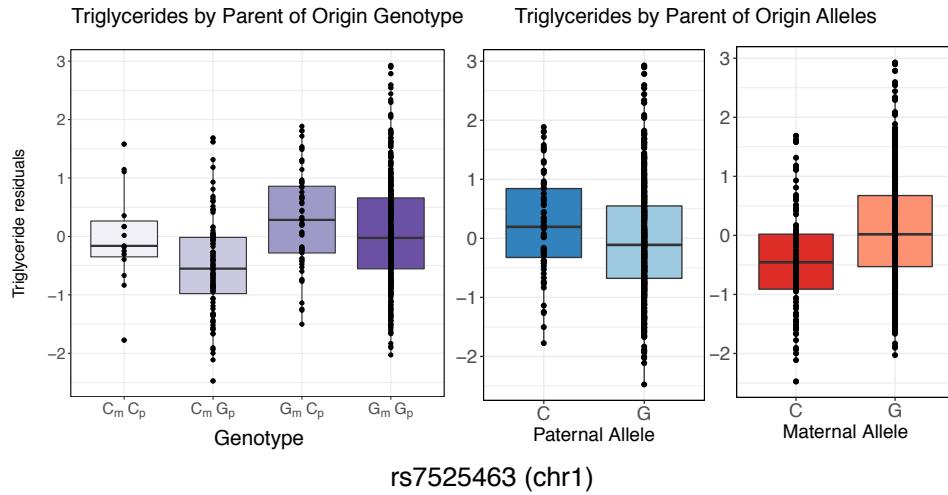


Figure 2.14: Opposite Effect Parent of Origin GWAS Result for Triglycerides. Box plots of triglyceride residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased triglycerides. The paternal C allele is associated with increased and the paternal T allele with decreased triglycerides. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

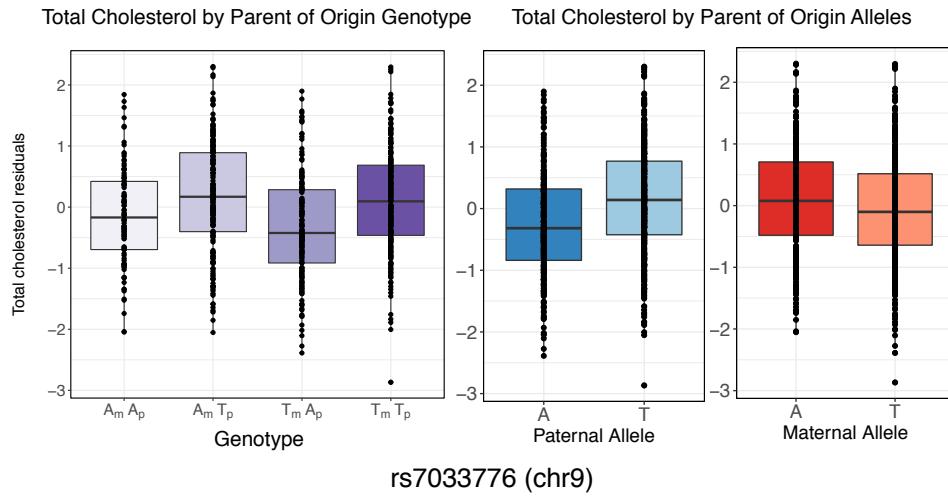


Figure 2.15: Opposite Effect Parent of Origin GWAS Result for Total Cholesterol. Box plots of total cholesterol residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal A allele is associated with decreased and maternal T allele with increased total cholesterol. The paternal A allele is associated with increased and the paternal T allele with decreased total cholesterol. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

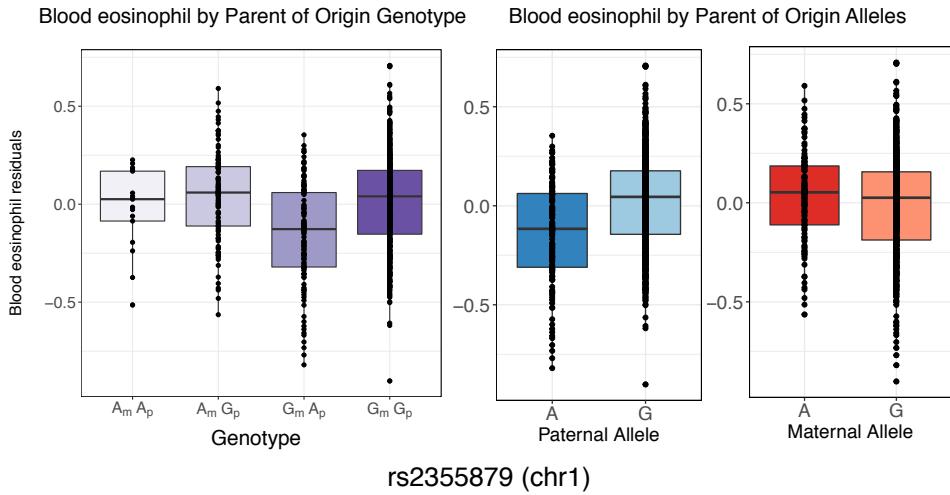


Figure 2.16: Opposite Effect Parent of Origin GWAS Result for Blood Eosinophil Count. Box plots of eosinophil residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased eosinophil count. The paternal C allele is associated with increased and the paternal T allele with decreased eosinophil count. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

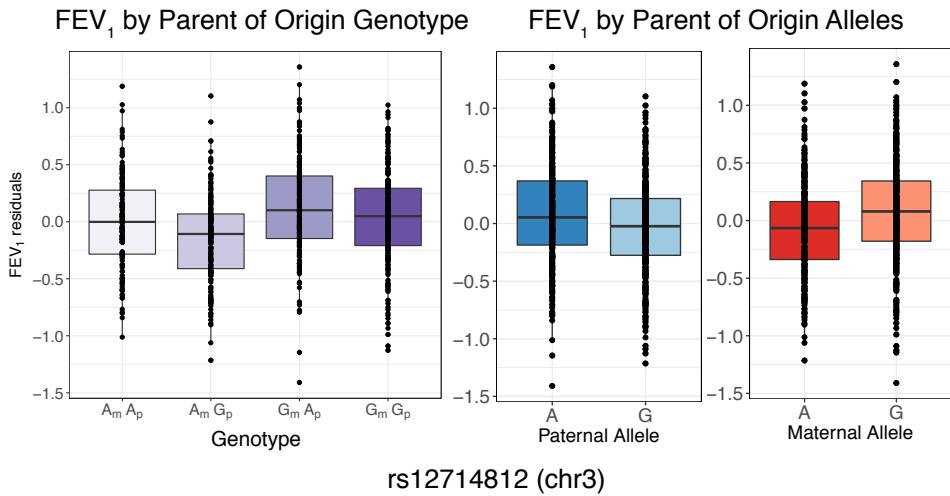


Figure 2.17: Opposite Effect Parent of Origin GWAS Result for FEV₁. Box plots of FEV₁ residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal C allele is associated with decreased and maternal T allele with increased FEV₁. The paternal C allele is associated with increased and the paternal T allele with decreased FEV₁. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show +/-1.5 x IQR.

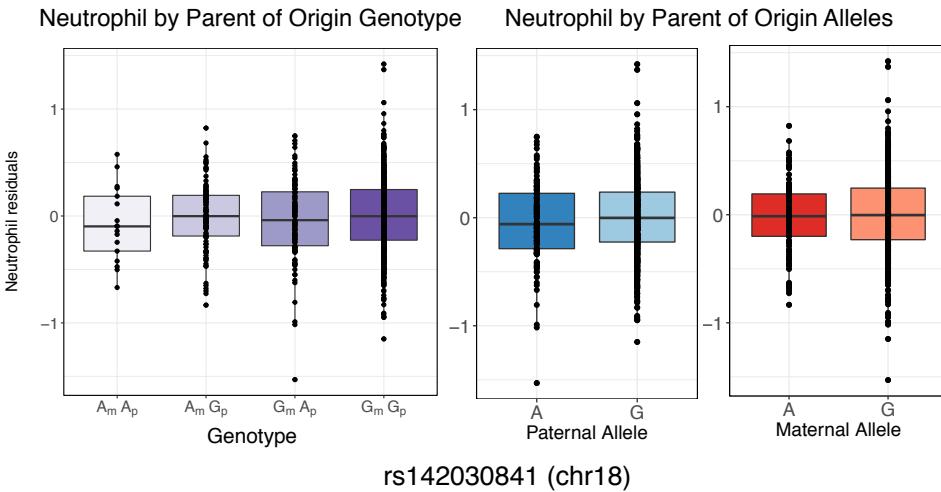


Figure 2.18: Opposite Effect Parent of Origin GWAS Result for Neutrophil Count. Box plots of neutrophil residuals (y-axes) are shown for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal G allele is associated with decreased and maternal A allele with increased neutrophil count. The paternal G allele is associated with increased and the paternal A allele with decreased neutrophil count. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $\pm 1.5 \times \text{IQR}$.

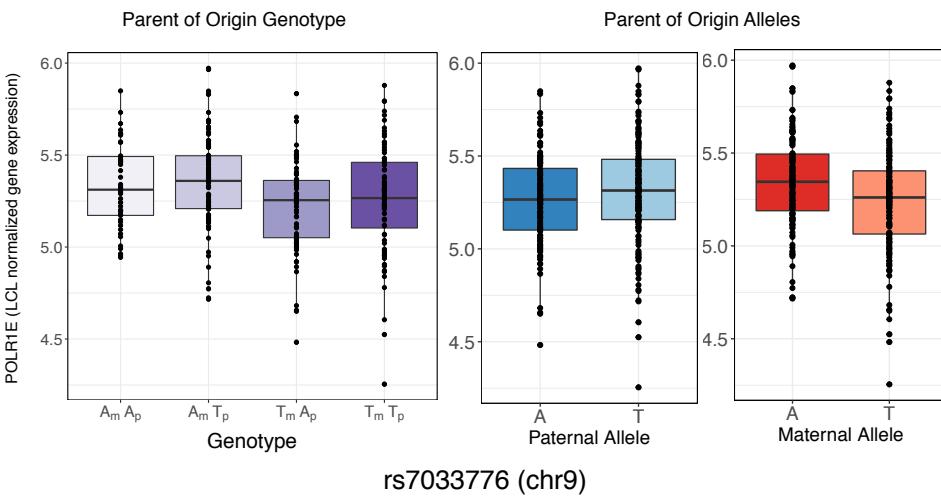


Figure 2.19: Opposite Effect eQTL for rs7033776. Box plots of two significant loci plot *POLR1E* gene expression residuals (y-axes) for each of the four genotypes (left panel; x-axis), and for paternal (center panel; x-axis) and maternal (right panel; x-axis) alleles. The maternal T allele is associated with decreased and maternal A allele with increased *POLR1E* expression. The paternal T allele is associated with increased and the paternal A allele with decreased *POLR1E* expression. The horizontal bar of the boxplot shows the median, the box delineates the first and third quartile, and the whiskers show $\pm 1.5 \times \text{IQR}$.

| Quantitative Trait | Sample Size | M/F Ratio | Age range (years) | GWAS call rate | PO call rate | Transformation | Covariates | Exclusions |
|-----------------------|-------------|-----------|-------------------|----------------|--------------|--------------------|--------------------------------|--|
| SBP | 807 | 371/436 | 6-85 | 0.94 | 0.82 | log | age, sex, age*sex, technician | Anti-hypertensive medication |
| DBP | 807 | 371/436 | 6-85 | 0.94 | 0.82 | log | age, sex, age*sex, technician | Anti-hypertensive medication |
| HDL | 828 | 381/447 | 14-85 | 0.94 | 0.80 | cube root | age, sex | Anti-hypercholesterolemia medication, hormone replacement therapy, birth control, diagnosis of sitosterolemia |
| LDL | 807 | 367/440 | 14-85 | - | 0.80 | cube root | age, sex | Anti-hypercholesterolemia medication, hormone replacement therapy, birth control, diagnosis of sitosterolemia |
| Triglycerides | 828 | 382/446 | 0.94 | 0.80 | log | age, sex | age, sex, breeding, technician | Anti-hypercholesterolemia medication, hormone replacement therapy, birth control, diagnosis of sitosterolemia |
| Total cholesterol/HDL | 828 | 381/447 | 14-85 | 0.94 | 0.80 | log | age, sex | Anti-hypercholesterolemia medication, hormone replacement therapy, birth control, diagnosis of sitosterolemia |
| Monocyte count | 1069 | 503/566 | 5.47-85.10 | 0.94 | 0.81 | log | age, sex, age*sex | Antibiotics, immunosuppressants, and/or steroids was an exclusion for all WBC count phenotypes. Antifungal medication was an exclusion for the Eosinophil count phenotype. |
| Lymphocyte count | 1079 | 507/572 | 6-85 | 0.94 | - | log | age | Pregnant, history of heart valve replacement, or poor quality echocardiography images |
| Eosinophil count | 1068 | 502/566 | 5.47-85.10 | 0.94 | 0.81 | square root(log()) | sex | Aortic stenosis by history or echocardiogram Cholesterol and/or thyroid medication, or poor quality echocardiography images |
| Neutrophil count | 1070 | 503/567 | 5.47-85.10 | 0.94 | 0.81 | log | age, sex, age*sex, phase | Cholesterol and/or thyroid medication, history of liver disease, or poor quality imaging |
| LAVI | 637 | 296/341 | 14-88 | 0.94 | 0.80 | log | age, sex | Poor quality measurements |
| LVMI | 621 | 286/335 | 14-88 | 0.94 | 0.79 | log | age, sex | |
| CIMT | 547 | 248/299 | 14-86 | 0.94 | 0.80 | inverse | age, sex | |
| FeNO | 825 | 381/444 | 6-85 | 0.95 | 0.82 | log | age, sex, technician | |

Table 2.5: **Summary of the Hutterite Phenotypes and Sample Composition.** *SBP: Systolic Blood Pressure, DBP: Diastolic Blood Pressure, HDL: High Density Lipoprotein, LDL: Low Density Lipoprotein, LAVI: Left Atrial Volume Index, LVMI: Left Ventricular Mass Index, CIMT: Carotid Intima Media Thickness, FeNO: Fraction of Exhaled Nitric Oxide, FEV₁: Forced Expiratory Volume at 1 s, FVC: Forced Vital Capacity, BRI: Bronchial Responsiveness Index

| Quantitative Trait | Sample M/F Ratio Size | Age range (years) | GWAS PO call rate | Transformation | Covariates | Exclusions |
|-----------------------|--------------------------|----------------------|-------------------------|----------------|------------|------------|
| BRI | 950 | 445/505 | 6.11-78.22 | 0.94 | 0.82 | none |
| FEV ₁ | 1102 | 509/593 | 5.47-8.47 | 0.94 | 0.82 | none |
| FEV ₁ /FVC | 1106 | 512/594 | 5.47-8.47 | 0.94 | 0.82 | none |
| Total serum IgE | 1219 | 562/657 | 6-91 | 0.94 | 0.81 | log |
| BMI | 1188 | 577/652 | 5.5-89.2 | 0.94 | 0.82 | log |
| Height | 1199 | 576/669 | 5-89 | 0.94 | 0.82 | none |
| Age at menarche | 477 | 0/719 | 9-17 | 0.92 | 0.71 | none |

Table 2.6: **Summary of the Hutterite Phenotypes and Sample Composition (Continued).** *SBP: Systolic Blood Pressure, DBP: Diastolic Blood Pressure, HDL: High Density Lipoprotein, LDL: Low Density Lipoprotein, LAVI: Left Atrial Volume Index, LVMI: Left Ventricular Mass Index, CIMT: Carotid Intima Media Thickness, FeNO: Fraction of Exhaled Nitric Oxide, FEV₁: Forced Expiratory Volume at 1 s, FVC: Forced Vital Capacity, BRI: Bronchial Responsiveness Index

| Phenotype | Sample size | Most significant GWAS p-value | SNP Size | Sample rsid | (Ef-chr:loc) | Beta (SE) | Variant | (Nearest) Gene | MAF |
|-----------------------|-------------|-------------------------------|----------|-------------------|--------------------------|----------------------|--|----------------|-----|
| BMI | 1188 | 8.04E-07 | 1016 | rs139659764 (A/G) | 13:81:274241 (1.64E-02) | -8.16E-02 (6.47E-03) | intergenic <i>LINC00377</i> | 0.07555 | |
| BRI | 950 | 3.05E-07 | 935 | rs7498042 (T/G) | 15:91:119014 (6.47E-03) | 3.34E-02 (6.47E-03) | intron <i>CRTC3</i> | 0.2921 | |
| CIMT | 547 | 1.51E-09 | 470 | rs116908536 (A/G) | 11:33:69981 (2.34E-02) | -0.144 (2.34E-02) | intron <i>LOC105376617</i> | 0.05183 | |
| DBP | 807 | 9.66E-07 | 999 | rs11619649 (T/C) | 18:90:14768 (9.82E-03) | -4.84E-02 (9.82E-03) | intergenic <i>NDUFV2</i> | 0.1312 | |
| Eosinophil count | 1068 | 1.44E-07 | 939 | rs2544465 (A/G) | 9:260:45558 (1.68E-02) | 8.17E-02 (1.68E-02) | intergenic <i>LOC100506422</i> | 0.147 | |
| FeNO | 825 | 2.15E-07 | 753 | rs1121780 (T/C) | 2:4296:1321 (2.28E-02) | 0.157 (2.28E-02) | intron <i>MTA3</i> | 0.4824 | |
| FEV ₁ | 1102 | 1.22E-06 | 1102 | rs13281444 (C/T) | 8:652:20628 (2.28E-02) | 1.11E-01 (2.28E-02) | intron <i>LOC100507530</i> | 0.358 | |
| FEV ₁ /FVC | 1106 | 2.49E-07 | 1099 | rs8028898 (A/G) | 15:71:679543 (5.31E-02) | -2.75E-01 (5.31E-02) | intron <i>THSD4</i> | 0.2049 | |
| HDL | 828 | 8.91E-07 | 828 | rs18613312 (A/T) | 16:58:850424 (9.73E-02) | -0.483 (9.73E-02) | intergenic <i>GOT2</i> | 0.0687 | |
| Height | 1199 | 7.62E-07 | 1009 | rs1318252 (G/A) | 11:1653200 (0.278) | 1.38 (0.278) | intergenic <i>KRTAP5-5</i> | 0.4239 | |
| IgE | 1219 | 3.99E-07 | 1187 | rs66498879 (C/T) | 11:125495211 (5.04E-02) | -2.25E-01 (5.04E-02) | 5'UTR <i>CHEK1</i> | 0.3243 | |
| LAVI | 637 | 3.07E-06 | 573 | rs17205373 (G/C) | 6:32626210 (2.81E-02) | -0.1144 (2.81E-02) | intergenic <i>HLA-DQB1</i> | 0.09724 | |
| LDL | 807 | 2.83E-17 | 714 | rs507778817 (T/C) | 19:11:30534 (1.3 (0.15)) | intron <i>KANK2</i> | 0.03382 | | |
| LVMII | 621 | 1.23E-09 | 517 | rs113389683 (A/G) | 8:360267788 (2.47E-02) | 0.153 (2.47E-02) | intergenic <i>UNC5D</i> , <i>KCNU1</i> | 0.08624 | |
| Lymphocyte count | 1079 | 5.67E-08 | 894 | rs3238874 (G/A) | 11:129110347 (1.49E-02) | 8.19E-02 (1.49E-02) | intergenic <i>ARHGAP32</i> | 0.4395 | |
| Age at Menarche | 477 | 3.17E-07 | 463 | rs785474 (G/C) | 1:46609106 (8.72E-02) | 4.35E-01 (8.72E-02) | intron <i>PIK3R3</i> | 0.4268 | |
| Monocyte count | 1069 | 1.36E-08 | 902 | rs31317434 (G/T) | 7:9466593 (4.73E-02) | -0.27 (4.73E-02) | intergenic <i>NXPH1</i> , <i>PERR1</i> | 0.03361 | |
| Neutrophil count | 1070 | 2.1E-07 | 1070 | rs1215134 (C/T) | 9:15347518 (1.71E-02) | -8.97E-02 (1.71E-02) | intergenic <i>TTC39B</i> | 0.4307 | |
| SBP | 807 | 5.23E-06 | 774 | rs28742608 (G/T) | 5:1440934 (1.05E-02) | -4.8E-02 (1.05E-02) | intron <i>LC6A3</i> | 0.07549 | |
| Total Cholesterol | 828 | 1.20E-07 | 828 | rs11084211 (A/G) | 19:53460861 (5.39E-02) | 0.287 (5.39E-02) | intron <i>ZNF816</i> , <i>ZNF816-2</i> | 0.3071 | |
| Triglycerides | 828 | 7.12E-13 | 707 | rs184333869 (T/C) | 11:117947268 (0.175) | -1.28 (0.175) | intron <i>TMPRSS4</i> | 0.02399 | |

Table 2.7: **GWAS Results.** Phenotype definitions, exclusions, transformations and covariates are summarized in Table 2.5

| Phenotype | SNP | genes +/- 1Mb expressed in LCLs |
|--------------------------|-------------------------|---|
| A. Maternal Associations | | |
| Age at Menarche | rs7184983 | <i>AMFR, ARL2BP, BBS2, CCDC102A, CCL17, CCL22, CES5A, CETP, CIAPIN1, COQ9, CPNE2, CX2CL1, DOK4, FAM192A, GNAO1, HERPUD1, MR1E, MT1L, MT2A, NLRC5, NUDT21, NUP93, OGFOD1, PLLP, POLR2C, RSPRY1, SLC12A3</i> |
| CIMT | rs4077567 | <i>ABCA12, ATIC, FN1, IGFBP2, MREG, PECR, RPL37A, SMARCAL1, TMEM169, XRCC5</i> |
| FEV1 | rs9849387 rs6791779 | <i>ROBO1 ZNF717</i> |
| LVMI | rs574232282 | <i>CITED4, COL9A2, FOXJ3, HIVEP3, KCNQ4, NFYC, RIMS3, RLF, SCMH1, SMAP2, ZMPSTE24, SNF684</i> |
| B. Paternal Associations | | |
| LDL | rs12024326 rs4843650 | <i>ADAMTS10, ANGPTL4, AHGEF18, CAMSAP3, CCL25, CD209, CD320, CERS4, CLEC4G, CTXN1, ELAVL1, EMR4P, EVI5L, FCER2, HNRNPM, INSR, KANK3, MAP2K7, MARCH2, MCOLN1, MYO1F, NDUFA, PCP2, PEX11G, PNPLA6, RAB11B, RAB11B-AS1, RPS28, SNAPC2, STXBP2, TIMM44, TRAPPC5, XAB2, ZNF358, ZNF414, ZNF557, ZNF558</i> <i>BANP, C16orf95, FBXO31, KLHDC4, MAP1LC3B, SLC7A5, ZC3H18, ZCCHC14, ZFPM1, ZNF469 SBP rs1536182 COG3, CTF2F2, KIAA0226L, LCP1, LRCH1, NUFIP1, SLC25A30, SLC25A30-AS1, TPT1, TPT1-AS1, ZC3H13</i> |
| Total Cholesterol | rs113588203 | <i>ABCB10, ARF1, C1orf35, GUK1, HIST3H2A, IBA57, IBA57-AS1, MRPL55, NUP133, OBSCN, RAB4A, RHOU, RNF187, SNORA51, SPHAR, TAF5L, TRIM11, TRIM17, URB2, WNT3A</i> |

Table 2.8: Candidate Genes for Parent of Origin eQTL.

| Phenotype | SNP | Genes +/- 1Mb expressed in LCLs |
|-------------------|--------------------------|--|
| Total Cholesterol | rs7033776 | <i>CLTA, CREB3, FBXO10, FRMPD1, GBA2, GLIPE2, GNE, GRHPR, HINT2, MELK, MSMP, NRP2, PAX5, POLR1E, RECK, RGP1, SPAG8, TLN1, TMEM8B, TOMM5, ZBTB5, ZCCHC7</i> |
| BMI | rs77785972 rs17605739 | <i>CHD1, RGMB, RGMB-AS1, RIOK2</i> - |
| LDL | rs1032596 | <i>C16orf74, C16orf95, COX4I1, GINS2, IRF8, MTHFSD</i> |
| Triglycerides | rs7527236 | <i>LYPLA1, RRP15, TGFB2</i> |
| LVMI | rs16853098 | <i>STK39</i> |
| Age at menarche | rs58758366 | - |
| Eosinophil count | rs2355879 | <i>AKR7A2, ALDH4A1, ARHGEF10L, CAPZB, IFFO2, MRTO4, PQLC2, RCC2, UBR4</i> |
| Neutrophil count | rs142030841 | <i>C18orf21, CELF4, ELP2, FHOD3, KIAA1328, MCOS, RPRD1A, SLC39A6, TPGS2</i> |
| FEV ₁ | rs12714812 | <i>FAM86DP</i> |

Table 2.9: Candidate Genes for Parent of Origin Differential eQTL.

Table 2.10: **Maternal GWAS results with p-value <5x10-08.** Significant results from the Maternal GWAS, not pruned for LD.

Table 2.11: **Paternal GWAS results with p-value <5x10-08.** Significant results from the Paternal GWAS, not pruned for LD.

Table 2.12: **Differential Effect GWAS results with p-value <5x10-08.** Significant results from the Differential Effect GWAS, not pruned for LD.

CHAPTER 3

PARENT OF ORIGIN GENE EXPRESSION IN A FOUNDER POPULATION IDENTIFIES TWO NOVEL IMPRINTED GENES AT KNOWN IMPRINTED REGIONS.

3.1 Abstract¹

Genomic imprinting is the phenomena that leads to silencing of one copy of a gene inherited from a specific parent. Mutations in imprinted regions have been involved in diseases showing parent of origin effects, such as Prader-Willi and Angelman syndrome, among others. Identifying genes with evidence of parent of origin expression patterns in family studies allows the detection of more subtle imprinting. Here we use allele specific expression in lymphoblastoid cell lines from 306 Hutterites related in a single pedigree to provide formal evidence for parent of origin effects. Our approach identified known imprinted genes, two putative novel imprinted genes, and 14 genes with asymmetrical parent of origin gene expression. We used gene expression in peripheral blood leukocytes (PBL) to validate our findings, and then confirmed imprinting control regions (ICRs) using DNA methylation levels in the PBLs.

3.2 Introduction

Imprinted genes have one allele silenced in a parent of origin specific manner. In humans, approximately 105 imprinted loci have been identified, many of which play important roles in development and growth [35, 87]. Dysregulation of imprinted genes or regions can cause diseases that show parent of origin effects, such as Prader-Willi or Angelman syndrome, among others [87]. Imprinted regions have also been associated with complex traits, such as

1. Citation for chapter: Mozaffari SV, Stein MM, Magnaye KM, Nicolae DL, Ober C. Gene Expression and Methylation of Imprinted Genes in the Hutterites bioRxiv (2018).

height and age of menarche [14, 115], as well as common diseases such as obesity and some cancers [87]. More than 80% of imprinted genes in humans are clustered in genomic regions that contain both maternally and paternally expressed genes, as well as genes that encode non-coding RNAs. Parent-specific expression of the genes within a cluster are maintained by complex epigenetic mechanisms at cis-acting imprinting control regions (ICRs) [53], which show parent of origin specific DNA methylation patterns and chromatin modifications.

Using RNA-seq and allele specific expression (ASE) we can map genes to parental haplotypes and identify those that are expressed when inherited from only the father or only from the mother, a hallmark feature of imprinted loci. Parent of origin effects and imprinted genes have been most elegantly studied in mice, where two inbred strains are bred reciprocally to identify parent of origin effects on gene expression in progeny that have the same genotypes but different patterns of inheritance [10]. Additionally, uniparental inheritance of imprinted regions in mice were associated with abnormal developmental phenotypes[19] before it was shown that imprinting defects are associated with human disease [75]. One approach to identifying imprinted loci in humans has been to test for parent of origin effects on gene expression and phenotypes in pedigrees [55, 14]. For example, Garg et al. used gene expression in LCLs from HapMap trios to identify 30 imprinting eQTLs with parent of origin specific effects on expression including two imprinted genes [38]. A study from the GTEx Consortium used RNA-seq data and allele specific expression to identify allelic imbalance in 45 different tissues. By considering genes with monoallelic expression that was evenly distributed to both the reference and alternate alleles across individuals as evidence for imprinting, they identified 42 imprinted genes, both known and novel, and used family studies to confirm imprinting of 5 novel imprinted genes [12]. Santoni et al. identified nine novel imprinted genes using single-cell allele-specific gene expression and identifying genes with mono-allelic expression in fibroblasts from 3 unrelated individuals and probands of 2 family trios, and then using the trios to confirm parent of origin of the alleles [94].

Here, we perform a parent of origin ASE study in a large pedigree to characterize parent of origin specific gene expression in the Hutterites, a founder population of European descent, for which we have phased genotype data [61]. We use RNA-seq from lymphoblastoid cell lines (LCLs) to map transcripts to parental haplotypes and identify known and two not previously reported imprinted genes. We validated the two putative imprinted genes by showing the same patterns of parent of origin expression PBLs from different Hutterite individuals, and show DNA methylation signatures of imprinting in the PBLs at these regions.

3.3 Results

3.3.1 *Mapping Transcripts to Parental Haplotypes*

For each of 306 individuals, the total number of transcripts at each gene was assigned as maternally inherited, paternally inherited, or unknown parent of origin. The last group included transcripts without heterozygote SNPs or SNPs without parent of origin information. Transcripts were assigned to the parentally inherited categories using SNPs in the reads and matching alleles to either the known maternally or paternally inherited alleles. All the genes analyzed had some transcripts of unknown origin (average 97.8%, range 8.3-100%). For each gene we assigned parental origin to an average of 1.8% of transcripts (range: 0-34.7%), and for each individual we assigned parental origin to an average of 1.4% of transcripts (range: 0-1.7%). On average, about 40 SNPs per gene were used to assign the transcripts of a gene to parent (range 1-1839 SNPs).

After quality control (see Methods), transcripts in 15,889 genes were detected as expressed in 306 individuals. Some transcripts for 14,791 of those genes could be assigned to a parent. Of these, 75 genes were only expressed on the paternally-inherited allele in at least one individual and not on the maternally inherited allele in any individuals. Similarly, 64 genes were only expressed on the maternally-inherited allele in at least one individual and not on

| | Mean | Standard Deviation | Range |
|--|-------|--------------------|------------|
| Proportion of transcripts from each gene assigned to transcripts of unknown origin | 0.978 | 0.031 | (0.083, 1) |
| Proportion of transcripts from each gene assigned to parental origin | 0.018 | 0.019 | (0, 0.347) |
| Proportion of transcripts for each individual assigned to parental origin | 0.014 | 0.0015 | (0, 0.017) |

Table 3.1: **Summary Statistics for Parental Origin of Transcripts.**

the paternally inherited allele in any individuals (S1 Table).

3.3.2 *Imprinted Genes in Lymphoblastoid Cell Lines (LCLs)*

Among the 139 genes with only paternally inherited expression or only maternally inherited expression, there are three known imprinted genes (*CDKN1C*, *NDN*, *SNRPN*) and one predicted to be imprinted (*IFITM1*) [63]. *CDKN1C* showed patterns opposite of what has been reported[42, 71], which could be due to the small sample (only three individuals showed expression from one parent) or to the different cell types used here (LCLs) and in previous studies (developing brain and embryonal tumors for *CDKN1C*).

We expect some imprinted genes to have “leaky” expression, such that there is some expression from the parental chromosome that is mostly silenced. To detect these genes, we used a binomial test to find patterns of gene expression asymmetry by parental transcript levels. This analysis identified 28 genes with an FDR <5% (Table 3.2). The 11 genes that showed the most asymmetry are known imprinted genes: *ZDBF2*, *PEG10*, *SNHG14*, *NHP2L1*, *L3MBTL1*, *ZNF331*, *LPAR6*, *FAM50B*, *KCNQ1*, *NAP1L5*, and *IGF1R*. Parent of origin expression for *ZDBF2* and *KCNQ1* are shown in Fig 3.1A and 3.1B, respectively. We identified two additional genes that showed asymmetry in parental expression from mostly one parent (*PXDC1*, *PWAR6*), which we consider potentially new imprinted genes. The remaining fourteen genes showed significant patterns of asymmetry but had expression from both maternal and paternal transcript levels. These genes are likely not imprinted but could

have asymmetry in expression due to an expression quantitative trait loci (eQTL).

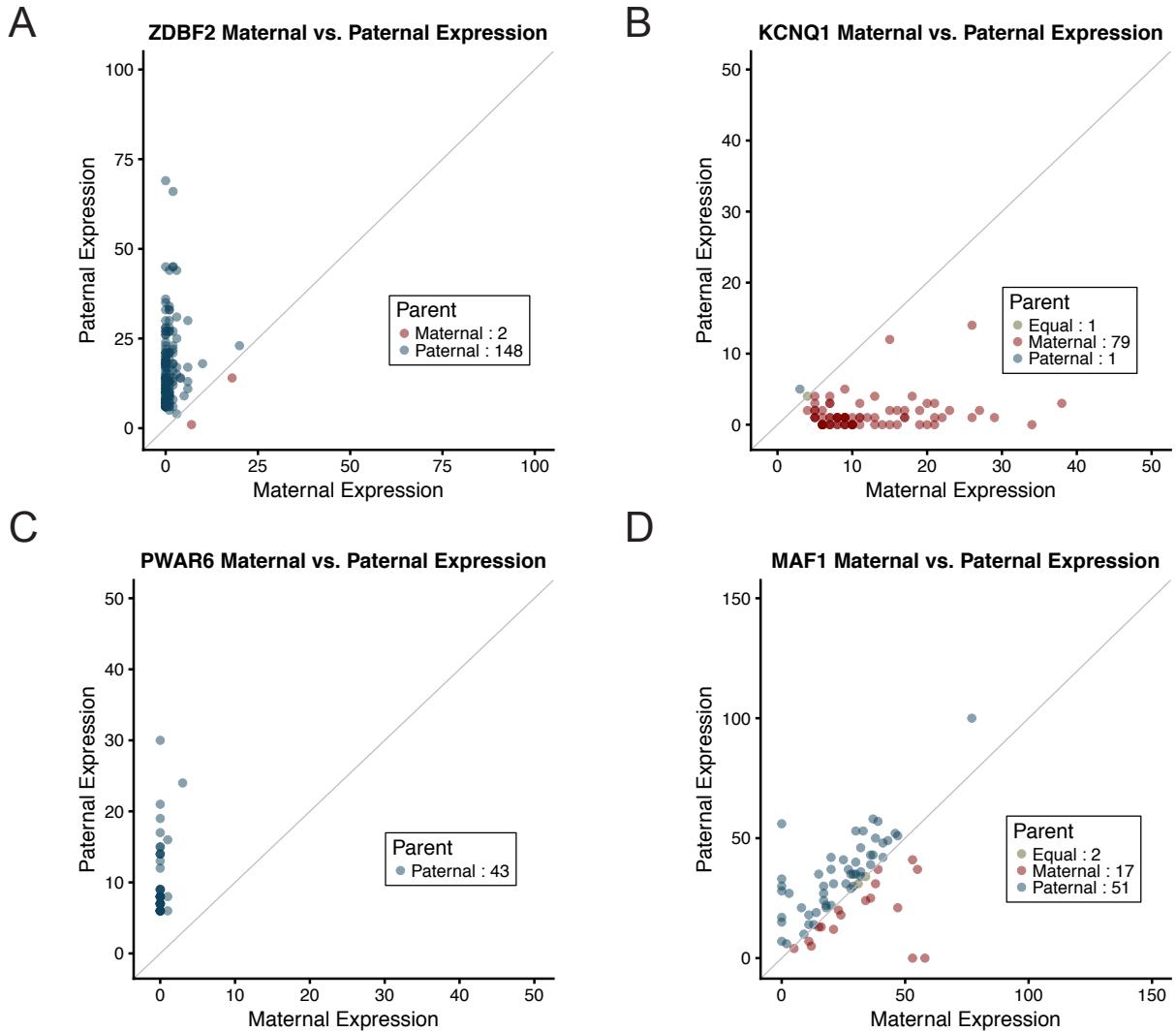


Figure 3.1: Plot of maternal (x-axis) and paternal (y-axis) gene expression for four genes. Plot of maternal (x-axis) and paternal (y-axis) gene expression for four genes. (A) maternally imprinted gene *ZDBF2* (paternally expressed), (B) paternally imprinted gene *KCNQ1* (maternally expressed), (C) novel maternally imprinted gene *PWAR6* (paternally expressed), (D) gene with asymmetry in parental expression *MAF1*. Each point represents one individual. Numbers in the legend represent the number of individuals with equal maternal and paternal expression, more maternal expression, or more paternal expression.

Two genes showed gene expression signatures consistent with imprinting but have not previously been recognized as imprinted genes. The first potentially new imprinted gene is *PXDC1*, which is in the same region and next to (<100kb) a known imprinted gene, *FAM50B*.

| Gene | p-value | Number of individuals with more maternal expression than paternal expression | Number of individuals with more paternal expression than maternal expression | References |
|---|----------|--|--|---|
| A. Known Imprinted | | | | |
| <i>ZDBF2</i> | 1.59e-41 | 2 | 148 | geneimprint.com, Baran et al[12], and Babak et al.[10] |
| <i>PEG10</i> | 5.51e-38 | 2 | 136 | geneimprint.com, Baran et al[12], and Babak et al.[10] |
| <i>SNHG14</i> | 1.64e-36 | 2 | 131 | Baran et al[12] |
| <i>NHP2L1</i> | 1.24e-33 | 23 | 189 | Babak et al.[10] and Docherty et al. [30] |
| <i>L3MBTL1</i> | 6.72e-31 | 2 | 107 | geneimprint.com, and Li et al. [59] |
| <i>ZNF331</i> | 4.05e-25 | 36 | 184 | Daelemans et al.[27] and Baran et al[12] |
| <i>LPAR6</i> | 2.65e-23 | 9 | 76 | Baran et al[12] |
| <i>FAM50B</i> | 5.29e-23 | 0 | 75 | geneimprint.com, Baran et al[12] |
| <i>KCNQ1</i> | 1.34e-22 | 79 | 1 | geneimprint.com, Baran et al[12] |
| <i>NAP1L5</i> | 3.76e-09 | 0 | 29 | geneimprint.com |
| <i>IGF1R</i> | 1.11e-05 | 14 | 49 | geneimprint.com, Sun et al. [100], Kang et al. [54], Boucher et al. [15], and Al Adhami et al.[5] |
| B. Conflicting Evidence for Imprinting Status in the Literature | | | | |
| <i>PRIM2</i> | 5.53e-05 | 30 | 71 | geneimprint.com, Santoni et al. [94] |
| C. New Imprinted Genes | | | | |
| <i>PXDC1</i> | 9.83e-14 | 12 | 81 | - |
| <i>PWAR6</i> | 2.27e-13 | 0 | 43 | - |
| D. Genes with Asymmetrical Parent of Origin Expression | | | | |
| <i>SNHG17</i> | 6.2e-08 | 113 | 45 | - |
| <i>ZNF813</i> | 8.7e07 | 63 | 132 | - |
| <i>DAAM1</i> | 1.78e-05 | 66 | 126 | - |
| <i>RP11-379H18.1</i> | 2.09e-05 | 52 | 106 | - |
| <i>HMGN1P38</i> | 2.09e-05 | 52 | 106 | - |
| <i>MTX2</i> | 3.05e-05 | 0 | 16 | - |
| <i>ZNF714</i> | 4.61e-05 | 35 | 79 | - |
| <i>MAF1</i> | 4.45e-05 | 17 | 51 | - |
| <i>IL16</i> | 5.71e-05 | 61 | 115 | - |
| <i>CPNE1</i> | 5.56e-05 | 111 | 58 | - |
| <i>ATP6V0D1</i> | 7.03e-05 | 32 | 7 | - |
| <i>FAHD1</i> | 9.34e-05 | 68 | 29 | - |
| <i>CNN2</i> | 1.18e-04 | 127 | 72 | - |
| <i>HSP90AB3P</i> | 1.16e-04 | 7 | 31 | - |

Table 3.2: Results for Genes with Parent of Origin Expression Asymmetry. Genes listed by category of imprinting status: (A) Known Imprinted, (B) Conflicting Evidence for Imprinted Status, (C) New Imprinted Genes, (D) Genes with Asymmetrical Parent of Origin Expression. Genes are ordered by significance within each category.

The second potentially novel imprinted gene is *PWAR6*, or Prader Willi Angelman Region RNA6, a gene encoding regulatory class of RNA. Although this gene is located within the intron of a known imprinted region, *SNHG14*, this noncoding RNA has not previously been recognized as having parent of origin specific expression (Fig 1C).

The remaining fourteen genes show significant asymmetry using the binomial test but do not have expression from mostly one parental chromosome. One of these genes, *SNHG17*, is a noncoding RNA. Another gene with parent of origin asymmetry, *ZNF813*, is next to a known imprinted gene, *ZNF331*. The remaining genes with asymmetrical parent origin expression have expression from both parental chromosomes, unlike imprinted genes. These genes include *DAAM1*, which is involved in cytoskeleton, specifically filopodia formation [43, 64], and has a suggested role for cytoskeleton organization during Mammalian testis morphogenesis and gamete progression [84]; *RP11-379H18.1*, a noncoding RNA gene; *HMGN1P38* [98]; *MTX2*, a nuclear gene that interacts with mitochondrial membrane protein metaxin 1 and is involved in mitochondrial protein import and metabolism of proteins in mice; *MAF1* a negative regulator of RNA polymerase 2; *ZNF714*, *CPNE1*, *IL16*, *ATP6V0D1*, *FAHD1*, *HSP90AB3P*, and *CNN2* are the remaining genes that show parent of origin asymmetry but not with a pattern consistent with imprinting (S1 Figure).

3.3.3 Validation of Imprinted Genes in PBLs

Using the same methods described above, we assigned parent of origin to transcripts in PBLs from 99 Hutterite individuals not included in the LCL studies. Maternal and paternal expression in PBLs for all 28 genes identified in LCLs showed similar trends of asymmetry as in LCLs (Figure 3.2).

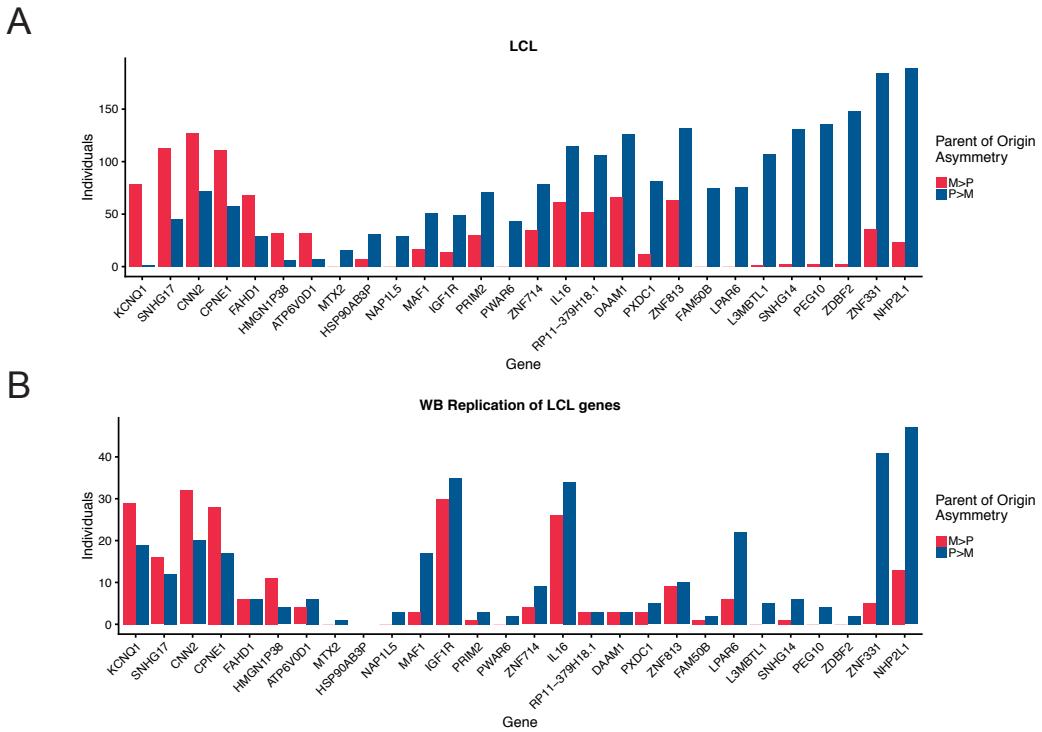


Figure 3.2: Validation in PBLs. Histogram showing the number of individuals with more maternal expression (M > P) or more paternal expression (P > M) for the 28 genes showing parent of origin asymmetry in LCLs (A) and PBLs (B). Genes are ordered by the magnitude of the difference in the number of individuals with more maternal expression than paternal expression in LCLs.

3.3.4 Methylation at Imprinting Control Regions

One of the mechanisms underlying parent of origin effects on expression at imprinted loci is differential methylation at cis-acting imprinting control regions (ICRs). DNA methylation from the Illumina HumanMethylation 450K array was available in PBLs from the same individuals included in the validation studies described above. To determine the expected patterns of methylation at known imprinted loci, we first looked at previously characterized methylated regions at known imprinted regions from Court et al. and Joshi et al. [24, 49].

The methylation patterns at the two potentially novel imprinted genes identified in this study, *PXDC1* and *PWAR6*, lie in or near known imprinted regions that contain previously

characterized ICRs. These previously characterized ICRs show about 50% methylation (beta value of between 0.25 and 0.75) in our DNA methylation data, which likely reflect methylation at only one parental chromosome in all the cells in the sample. Methylation patterns in PBLs at these two ICRs fall within this hemi-methylation range, further suggesting that these two genes are indeed imprinted (Fig 3.3).

3.4 Discussion

Dysregulation of imprinted genes can have a large impact on mammalian development and has been associated with significant diseases in humans. Studies aimed at identifying imprinted genes at genome-wide levels have used allele specific expression and imbalance to infer parent of origin. Here we used a large pedigree with assigned parent of origin alleles to map transcripts to chromosomes with known parent of origin and identify imprinted genes.

Using this approach, we found genes with expression primarily from either the maternal or paternal haplotype. Because gene silencing at imprinted loci may be incomplete, we used a binomial test on parent of origin gene expression and identified 11 known imprinted genes and two potentially novel imprinted genes. Both of these novel genes, *PWAR6* and *PXDC1*, lie in known imprinted regions but have not themselves been characterized as imprinted. The remaining genes that have significant parent of origin asymmetry in gene expression do not show clear imprinting expression patterns. To validate these findings, we mapped gene expression in PBLs from Hutterite individuals not included in the LCL study. The same genes showed similar patterns of asymmetry in these different cell sources (transformed B cells and peripheral blood leukocytes) from different individuals.

In addition to validating gene expression, we characterized methylation patterns near genes showing asymmetry. Using results from studies that had previously characterized ICRs in patients with uniparental disomy at many imprinted regions [49, 24], we estimated regions for defining hemi-methylation near the genes identified in our study. Using this

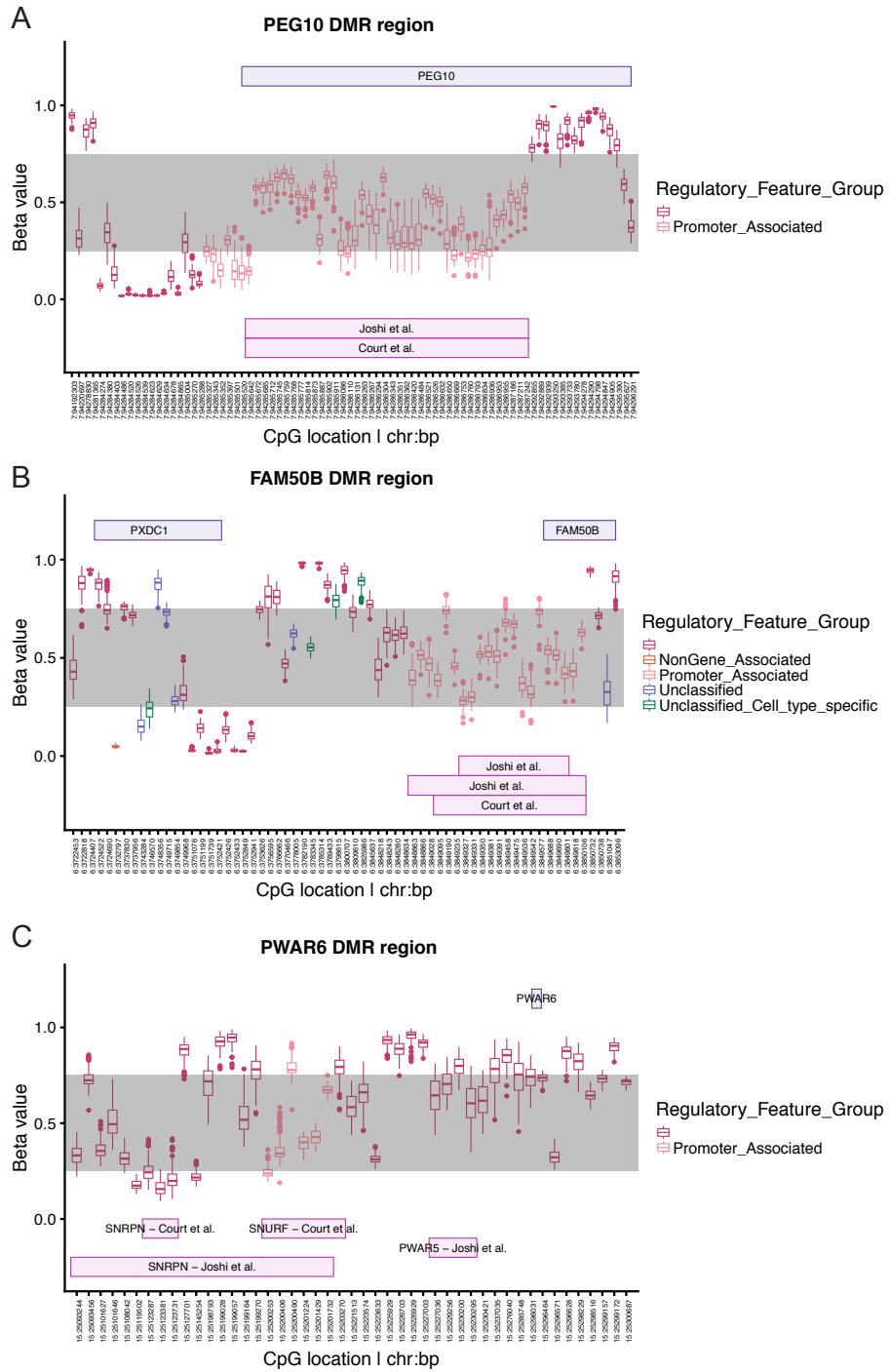


Figure 3.3: **Methylation at ICRs.** DNA methylation levels near known and novel imprinted genes previously defined by Joshi et al. and Court et al. (A) *PEG10*, (B) *PXDC1* and *FAM50B*, (C) *PWAR6*.

approach, we were able to provide additional supportive data for the two potentially novel imprinted genes to be true imprinted genes regulated by previously characterized ICRs.

Although our study is the largest pedigree-based study to date to search genome-wide for imprinted genes, it has limitations. First, we are able to determine the parent of origin for many transcripts in the Hutterites but we could not assign every RNA sequencing read to a parent due to lack of heterozygous sites or missing parent of origin information for alleles. Second, we conducted these studies in lymphoblastoid cell lines, and therefore could only study genes imprinted in this cell type and would miss the many imprinted genes that are tissue-specific and/or developmentally regulated. Third, while we can verify previously characterized ICRs, our study is not designed to identify novel ICRs because DNA methylation values from an array cannot be assigned to parental haplotype. Lastly, although we characterized the gene expression and methylation patterns for two potentially novel imprinted genes, replication of these genes in a different population and in different tissues, and functional characterization of these genes are required to confirm their status as imprinted genes. Similarly, some of the other genes with parent of origin asymmetry in the blood cells examined in this study may show more clear-cut evidence for imprinting in other tissues or at specific periods of development.

In summary, we have identified two new imprinted genes using gene expression from a founder population. The genes with asymmetrical parental expression had similar patterns of asymmetry in a different source of blood cells and in different individuals, and we were able to replicate the methylation patterns in known ICRs near the known and novel imprinted genes in this study. Our method and study population allowed us to map reads to parental haplotypes and uncovered *PWAR6* and *PXDC1* as new imprinted genes that could potentially impact disease risk and development.

3.5 Methods

3.5.1 Genotypes

Hutterite individuals ($n=1,653$) were genotyped using one of three Affymetrix genotype arrays, as previously described[61], of which 121 underwent whole genome sequencing by Complete Genomics, Inc (CGI) ($n=98$) or Illumina whole genome sequencing ($n=27$). A total of 10,235,233 variants present in the sequenced individuals were imputed and phased to the remaining 1532 genotyped individuals using PRIMAL [61]. Parent of origin was assigned to 89.85% of the alleles with call rate 81.6842% after QC. For this study, we included individuals with genotyped parents in the primary analyses in LCLs. Written consents for these studies were obtained from the adult participants and parents of children under 18; written assents were obtained from all children. This study was approved by the University of Chicago Institutional Review Board.

3.5.2 RNA-seq in Lymphoblastoid Cell Lines (LCLs).

RNA-seq was performed in LCLs as previously described [26]. For this study, sequencing reads were reprocessed as follows. Reads were trimmed for adaptors using Cutadapt (reads less than 5 bp discarded) then remapped to hg19 using STAR indexed with gencode version 19 gene annotations [29, 70]. To remove mapping bias, reads were processed and duplicate reads removed using WASP [104]. We used a custom script modified from WASP to separate reads that overlap maternal alleles or paternal alleles. Reads without informative SNPs (homozygous, or no parent of origin information) were categorized as unknown where the unknown, maternal, and paternal make up the total gene expression. Gene counts were quantified using STAR for each category. VerifyBamID was used to identify sample swaps [50]. Genes mapping to the X and Y chromosome were removed; genes with a CPM log transformed value less than 1 in less than 20 individuals were also removed.

3.5.3 RNA-seq in Peripheral Blood Leukocytes (PBLs)

RNA-seq was performed in whole blood as previously described [97]. For this study, sequencing reads were reprocessed as described above for the studies in LCLs. For all analyses, we excluded 32 individuals who were also in the LCL study.

3.5.4 Identifying Imprinted Genes

We used a binomial test to detect asymmetry in parent of origin gene expression. We generated a binomial Z-score for each individual for each gene (Z_i) and excluded those where $Z_i = 0$. For each gene, the number of subjects with $Z_i > 0$ can be modeled by a Binomial distribution with probability 1/2. For imprinted genes that show patterns of asymmetry, we expect a distribution of Z-scores that are skewed to one direction: right-skewed for genes asymmetrically maternally expressed and left-skewed for genes asymmetrically paternally expressed. Because we are only asking whether there are more individuals with more maternal expression or more paternal expression and not gene expression measures there is no need to model over-dispersion.

3.5.5 DNA methylation profiling and processing in PBLs

One milliliter of whole blood from 145 Hutterites was drawn into TruCulture (Myriad RBM; Austin, Texas) tubes containing proprietary TruCulture media. DNA was extracted using AllPrep DNA/RNA Mini Kits (Qiagen). DNA samples were bisulfite converted and hybridized to the Illumina HumanMethylation 450K array at the University of Chicago Functional Genomics Center. Samples were processed using default parameters using the R package minfi [7], normalized using SWAN (subset within-array normalization [68]) and quantile normalized similar to previous methylation studies [76]. Probes were removed if: (1) mapped non-uniquely to a bisulfite-converted genome; (2) mapped to sex chromosomes; (3) had a probe detection p-value >0.01 in at least 25% of samples; and (4) contained common

SNPs within the probe sequence, as previously described[11]. Principal components analysis (PCA) was used to identify significant technical covariates, and the ComBat function [48] within the R package sva [58] was used to correct for chip effect. Analyses of DNA methylation levels were conducted using beta values, which were converted from M-values using the lumi R package [31].

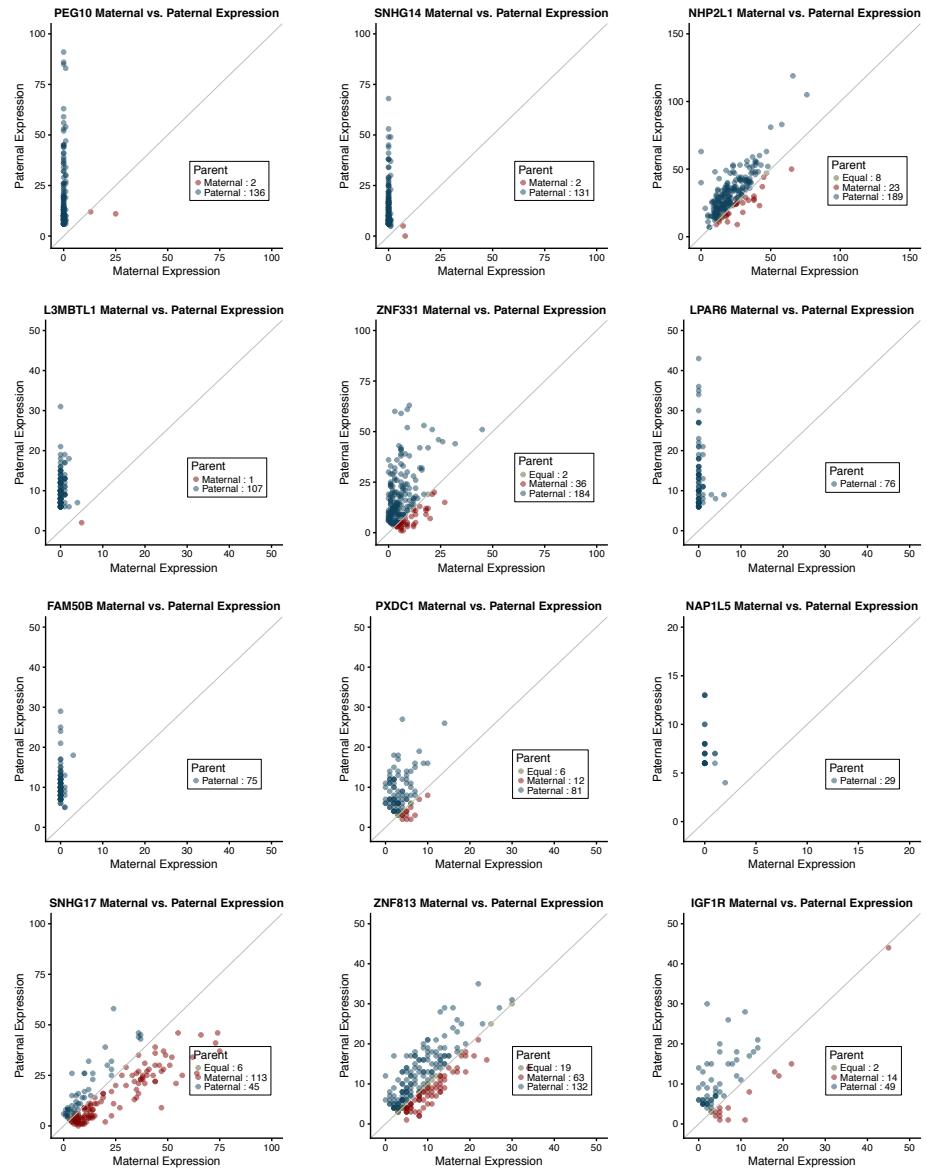


Figure 3.4: Plots of maternal and paternal expression for remaining genes with parent of origin asymmetry. Maternal gene expression along x-axis and paternal gene expression along y-axis.

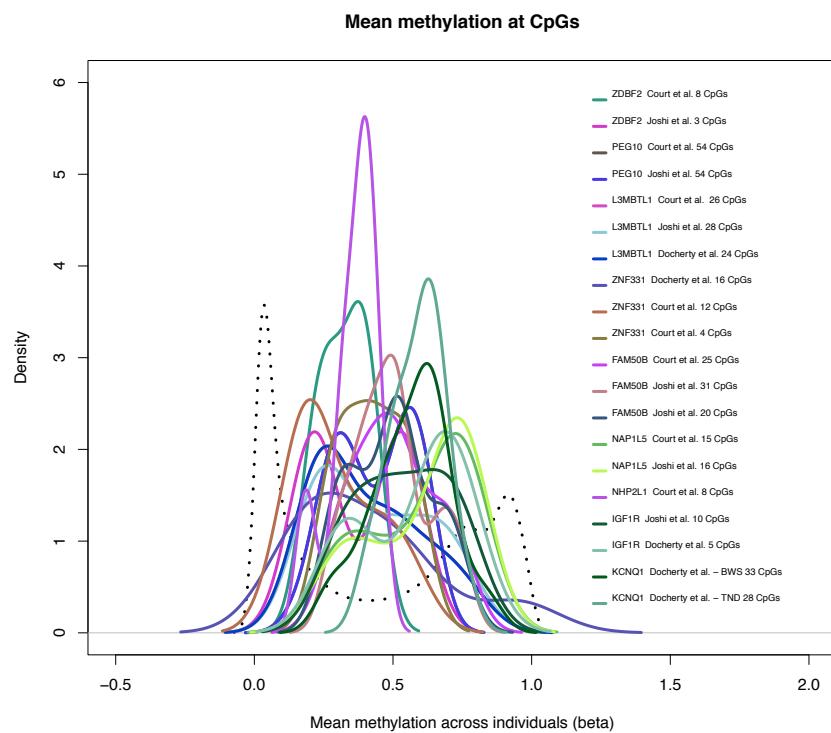


Figure 3.5: Density plot for DMRs for all imprinted genes. DNA methylation density plots for all imprinted genes with methylation levels in Joshi *et al.*[49] and Court *et al.*[24]

| Gene | Number of Individuals | Gene Name | Imprinted Status | Observed Expression Pattern in LCLs | Pattern consistent with known expression? |
|-----------------|-----------------------|----------------------|------------------|-------------------------------------|---|
| ENSG00000182636 | 22 | <i>NDN</i> | Known | Paternal | Consistent |
| ENSG00000259261 | 11 | <i>IGHV4OR15-8</i> | | Maternal | |
| ENSG00000128739 | 9 | <i>SNRPN</i> | Known | Paternal | Consistent |
| ENSG00000185044 | 9 | <i>RP11-435B5.4</i> | | Maternal | |
| ENSG00000089876 | 6 | <i>DHX32</i> | | Paternal | |
| ENSG00000262333 | 5 | <i>HNRNPA1P16</i> | | Paternal | |
| ENSG00000196378 | 5 | <i>ZNF34</i> | | Maternal | |
| ENSG00000198744 | 5 | <i>RP5-857K21.11</i> | | Maternal | |
| ENSG00000233757 | 5 | <i>AC092835.2</i> | | Maternal | |
| ENSG00000272933 | 5 | <i>RP11-47A8.5</i> | | Maternal | |
| ENSG00000154611 | 4 | <i>PSMA8</i> | | Paternal | |
| ENSG00000228109 | 4 | <i>MFI2-AS1</i> | | Paternal | |
| ENSG00000103269 | 4 | <i>RHBDL1</i> | | Maternal | |
| ENSG00000129654 | 4 | <i>FOXJ1</i> | | Maternal | |
| ENSG00000129757 | 3 | <i>CDKN1C</i> | Known | Paternal | Inconsistent |
| ENSG00000137936 | 3 | <i>BCAR3</i> | | Paternal | |
| ENSG00000148926 | 3 | <i>ADM</i> | | Paternal | |
| ENSG00000204482 | 3 | <i>LST1</i> | | Paternal | |
| ENSG00000211669 | 3 | <i>IGLV3-10</i> | | Paternal | |
| ENSG00000223509 | 3 | <i>RP11-632K20.7</i> | | Paternal | |
| ENSG00000260442 | 3 | <i>RP11-22P6.3</i> | | Paternal | |
| ENSG00000270441 | 3 | <i>RP11-694I15.7</i> | | Paternal | |
| ENSG00000160828 | 3 | <i>STAG3L2</i> | | Maternal | |
| ENSG00000165886 | 3 | <i>UBTD1</i> | | Maternal | |
| ENSG00000181284 | 3 | <i>TMEM102</i> | | Maternal | |
| ENSG00000214269 | 3 | <i>LGMNP1</i> | | Maternal | |
| ENSG00000258561 | 3 | <i>RP11-72M17.1</i> | | Maternal | |
| ENSG00000075089 | 2 | <i>ACTR6</i> | | Paternal | |
| ENSG00000105518 | 2 | <i>TMEM205</i> | | Paternal | |
| ENSG00000115457 | 2 | <i>IGFBP2</i> | | Paternal | |
| ENSG00000137821 | 2 | <i>LRRK49</i> | | Paternal | |
| ENSG00000158517 | 2 | <i>NCF1</i> | | Paternal | |
| ENSG00000183604 | 2 | <i>RP11-347C12.2</i> | | Paternal | |
| ENSG00000211637 | 2 | <i>IGLV4-69</i> | | Paternal | |
| ENSG00000211940 | 2 | <i>IGHV3-9</i> | | Paternal | |

Table 3.3: **Genes with only Maternal/ only Paternal gene expression.** Genes expressed only from one parent in number of individuals listed in column 2 (Number of Individuals). Imprinting status and pattern of gene expression in LCLs listed in columns 4 and 5.

| Gene | Number of Individuals | Gene Name | Imprinted Status | Observed Expression Pattern in LCLs | Pattern consistent with known expression? |
|-----------------|-----------------------|----------------------|------------------|-------------------------------------|---|
| ENSG00000233426 | 2 | <i>EIF3FP3</i> | | Paternal | |
| ENSG00000240041 | 2 | <i>IGHJ4</i> | | Paternal | |
| ENSG00000240731 | 2 | <i>RP5-890O3.9</i> | | Paternal | |
| ENSG00000272145 | 2 | <i>NFYC-AS1</i> | | Paternal | |
| ENSG0000025156 | 2 | <i>HSF2</i> | | Maternal | |
| ENSG00000108298 | 2 | <i>RPL19</i> | | Maternal | |
| ENSG00000133216 | 2 | <i>EPHB2</i> | | Maternal | |
| ENSG00000133328 | 2 | <i>HRASLS2</i> | | Maternal | |
| ENSG00000134864 | 2 | <i>GGACT</i> | | Maternal | |
| ENSG00000158481 | 2 | <i>CD1C</i> | | Maternal | |
| ENSG00000169019 | 2 | <i>COMMD8</i> | | Maternal | |
| ENSG00000175701 | 2 | <i>LINC00116</i> | | Maternal | |
| ENSG00000196465 | 2 | <i>MYL6B</i> | | Maternal | |
| ENSG00000198155 | 2 | <i>ZNF876P</i> | | Maternal | |
| ENSG00000215030 | 2 | <i>RPL13P12</i> | | Maternal | |
| ENSG00000232640 | 2 | <i>RP1-266L20.2</i> | | Maternal | |
| ENSG00000233493 | 2 | <i>TMEM238</i> | | Maternal | |
| ENSG00000235400 | 2 | <i>RP4-641G12.4</i> | | Maternal | |
| ENSG00000240652 | 2 | <i>RP11-832N8.1</i> | | Maternal | |
| ENSG00000243364 | 2 | <i>EFNA4</i> | | Maternal | |
| ENSG00000255135 | 2 | <i>RP11-111M22.3</i> | | Maternal | |
| ENSG00000267152 | 2 | <i>CTD-2528L19.6</i> | | Maternal | |
| ENSG00000033122 | 1 | <i>LRRC7</i> | | Paternal | |
| ENSG00000096080 | 1 | <i>MRPS18A</i> | | Paternal | |
| ENSG00000100442 | 1 | <i>FKBP3</i> | | Paternal | |
| ENSG00000100632 | 1 | <i>ERH</i> | | Paternal | |
| ENSG00000109083 | 1 | <i>IFT20</i> | | Paternal | |
| ENSG00000111875 | 1 | <i>ASF1A</i> | | Paternal | |
| ENSG00000116819 | 1 | <i>TFAP2E</i> | | Paternal | |
| ENSG00000121089 | 1 | <i>NACA3P</i> | | Paternal | |
| ENSG00000122218 | 1 | <i>COPA</i> | | Paternal | |
| ENSG00000128011 | 1 | <i>LRFN1</i> | | Paternal | |
| ENSG00000129673 | 1 | <i>AANAT</i> | | Paternal | |
| ENSG00000140459 | 1 | <i>CYP11A1</i> | | Paternal | |
| ENSG00000148187 | 1 | <i>MRRF</i> | | Paternal | |

Table 3.4: **Genes with only Maternal/ only Paternal gene expression (Continued - 1).** Genes expressed only from one parent in number of individuals listed in column 2 (Number of Individuals). Imprinting status and pattern of gene expression in LCLs listed in columns 4 and 5.

| Gene | Number of Individuals | Gene Name | Imprinted Status | Observed Expression Pattern in LCLs | Pattern consistent with known expression? |
|-----------------|-----------------------|-----------------------|------------------|-------------------------------------|---|
| ENSG00000150456 | 1 | <i>N6AMT2</i> | | Paternal | |
| ENSG00000151366 | 1 | <i>NDUFC2</i> | | Paternal | |
| ENSG00000154640 | 1 | <i>BTG3</i> | | Paternal | |
| ENSG00000158716 | 1 | <i>DUSP23</i> | | Paternal | |
| ENSG00000158806 | 1 | <i>NPM2</i> | | Paternal | |
| ENSG00000163634 | 1 | <i>THOC7</i> | | Paternal | |
| ENSG00000165121 | 1 | <i>RP11-213G2.3</i> | | Paternal | |
| ENSG00000167286 | 1 | <i>CD3D</i> | | Paternal | |
| ENSG00000173715 | 1 | <i>C11orf80</i> | | Paternal | |
| ENSG00000173762 | 1 | <i>CD7</i> | | Paternal | |
| ENSG00000175550 | 1 | <i>DRAP1</i> | | Paternal | |
| ENSG00000179603 | 1 | <i>GRM8</i> | | Paternal | |
| ENSG00000181038 | 1 | <i>METTL23</i> | | Paternal | |
| ENSG00000181852 | 1 | <i>RNF41</i> | | Paternal | |
| ENSG00000183506 | 1 | <i>PI4KAP2</i> | | Paternal | |
| ENSG00000197568 | 1 | <i>HHLA3</i> | | Paternal | |
| ENSG00000198356 | 1 | <i>ASNA1</i> | | Paternal | |
| ENSG00000204472 | 1 | <i>AIF1</i> | | Paternal | |
| ENSG00000211594 | 1 | <i>IGKJ4</i> | | Paternal | |
| ENSG00000211595 | 1 | <i>IGKJ3</i> | | Paternal | |
| ENSG00000211965 | 1 | <i>IGHV3-49</i> | | Paternal | |
| ENSG00000215548 | 1 | <i>RP11-764K9.4</i> | | Paternal | |
| ENSG00000225329 | 1 | <i>RP11-325F22.5</i> | | Paternal | |
| ENSG00000226121 | 1 | <i>AHCTF1P1</i> | | Paternal | |
| ENSG00000233912 | 1 | <i>AC026202.3</i> | | Paternal | |
| ENSG00000239819 | 1 | <i>IGKV1D-8</i> | | Paternal | |
| ENSG00000239830 | 1 | <i>RPS4XP22</i> | | Paternal | |
| ENSG00000243312 | 1 | <i>RP11-397E7.1</i> | | Paternal | |
| ENSG00000244055 | 1 | <i>AC007566.10</i> | | Paternal | |
| ENSG00000253998 | 1 | <i>IGKV2-29</i> | | Paternal | |
| ENSG00000257261 | 1 | <i>RP11-96H19.1</i> | | Paternal | |
| ENSG00000259699 | 1 | <i>HMGB1P8</i> | | Paternal | |
| ENSG00000260219 | 1 | <i>RP11-347C12.10</i> | | Paternal | |
| ENSG00000260655 | 1 | <i>CTA-250D10.23</i> | | Paternal | |
| ENSG00000264473 | 1 | <i>hsa-mir-4538</i> | | Paternal | |

Table 3.5: **Genes with only Maternal/ only Paternal gene expression (Continued - 2).** Genes expressed only from one parent in number of individuals listed in column 2 (Number of Individuals). Imprinting status and pattern of gene expression in LCLs listed in columns 4 and 5.

| Gene | Number of Individuals | Gene Name | Imprinted Status | Observed Expression Pattern in LCLs | Pattern consistent with known expression? |
|-----------------|-----------------------|----------------------|------------------|-------------------------------------|---|
| ENSG00000268568 | 1 | <i>AC007228.9</i> | | Paternal | |
| ENSG00000106211 | 1 | <i>HSPB1</i> | | Maternal | |
| ENSG00000118514 | 1 | <i>ALDH8A1</i> | | Maternal | |
| ENSG00000126709 | 1 | <i>IFI6</i> | | Maternal | |
| ENSG00000131773 | 1 | <i>KHDRBS3</i> | | Maternal | |
| ENSG00000135914 | 1 | <i>HTR2B</i> | | Maternal | |
| ENSG00000136104 | 1 | <i>RNASEH2B</i> | | Maternal | |
| ENSG00000136463 | 1 | <i>TACO1</i> | | Maternal | |
| ENSG00000148444 | 1 | <i>COMMD3</i> | | Maternal | |
| ENSG00000156873 | 1 | <i>PHKG2</i> | | Maternal | |
| ENSG00000163249 | 1 | <i>CCNYL1</i> | | Maternal | |
| ENSG00000164794 | 1 | <i>KCNV1</i> | | Maternal | |
| ENSG00000172586 | 1 | <i>CHCHD1</i> | | Maternal | |
| ENSG00000174871 | 1 | <i>CNIH2</i> | | Maternal | |
| ENSG00000178922 | 1 | <i>HYI</i> | | Maternal | |
| ENSG00000183426 | 1 | <i>NPIPA1</i> | | Maternal | |
| ENSG00000185885 | 1 | <i>IFITM1</i> | Predicted | Maternal | Consistent |
| ENSG00000197279 | 1 | <i>ZNF165</i> | | Maternal | |
| ENSG00000199753 | 1 | <i>SNORD104</i> | | Maternal | |
| ENSG00000215302 | 1 | <i>CTD-3092A11.1</i> | | Maternal | |
| ENSG00000226085 | 1 | <i>UQCRFS1P1</i> | | Maternal | |
| ENSG00000227053 | 1 | <i>RP11-395B7.4</i> | | Maternal | |
| ENSG00000232573 | 1 | <i>RPL3P4</i> | | Maternal | |
| ENSG00000237973 | 1 | <i>hsa-mir-6723</i> | | Maternal | |
| ENSG00000240356 | 1 | <i>RPL23AP7</i> | | Maternal | |
| ENSG00000240449 | 1 | <i>RP4-584D14.5</i> | | Maternal | |
| ENSG00000253485 | 1 | <i>PCDHGA5</i> | | Maternal | |
| ENSG00000254681 | 1 | <i>PKD1P5</i> | | Maternal | |
| ENSG00000254887 | 1 | <i>CTC-378H22.1</i> | | Maternal | |
| ENSG00000261504 | 1 | <i>RP11-317P15.4</i> | | Maternal | |
| ENSG00000262691 | 1 | <i>CTC-277H1.7</i> | | Maternal | |
| ENSG00000266208 | 1 | <i>CTD-2267D19.3</i> | | Maternal | |
| ENSG00000268030 | 1 | <i>AC005253.2</i> | | Maternal | |
| ENSG00000272468 | 1 | <i>RP1-86C11.7</i> | | Maternal | |

Table 3.6: **Genes with only Maternal/ only Paternal gene expression (Continued - 3).** Genes expressed only from one parent in number of individuals listed in column 2 (Number of Individuals). Imprinting status and pattern of gene expression in LCLs listed in columns 4 and 5.

CHAPTER 4

PARENT OF ORIGIN EFFECTS ON GENE EXPRESSION

4.1 Abstract

In this chapter, I explore the impact of parental origin of genetic variation on gene expression. We are interested in identifying any variants that are eQTLs but differ in direction by the parent the variant was inherited from. We performed opposite effect eQTL (oeQTL) and *cis* maternal and paternal eQTL (mat-eQTL, pat-eQTL) using lymphoblastoid cell line (LCL) gene expression in 306 Hutterites. We did not find any variants that have opposite effects by parental origin on gene expression with either of these two approaches. We also used a χ^2 test to search for parent specific effects on reciprocal heterozygotes using parent specific gene expression and identified SNPs that have modest parent of origin eQTL effects that need to be investigated further.

4.2 Introduction

Imprinted genes have one allele silenced in a parent of origin specific manner. In humans, approximately 150 imprinted loci have been identified, many of which play important roles in development and growth [35, 87, 14]. Dysregulation of imprinted genes or regions can cause diseases that show parent of origin effects, such as Prader-Willi or Angelman syndrome, among others [87]. Dysregulation of imprinted genes can be caused by large deletions but also by single variant mutations. Imprinted regions have also been associated with complex traits, such as height and age of menarche [14, 115], as well as common diseases such as obesity and some cancers [87]. We know that SNPs associated with traits are more likely to be eQTLs [77], and here we explore if different parentally inherited alleles can have different impacts on gene expression and can be identified as parent of origin eQTLs with potential impacts on traits. We are not the first to look for parent of origin effects on gene expression:

Garg et al. used gene expression in LCLs from HapMap trios to identify thirty imprinting eQTLs with parent of origin specific effects on expression of which two were known imprinted genes [38]. Garg et al. looked for impact of parent of origin specific effects on gene expression [38], but none have looked for eQTLs on parent specific expression coming from the same haplotype as the parentally inherited allele.

Using RNA-seq and allele specific expression (ASE) we can map genes to parental haplotypes that will inform us of gene expression from parental chromosomes. With parentally mapped gene expression data, we can ask if genetic variation on the parental haplotype can influence gene expression from the same haplotype. We are the first to look for 1) parental genetic variation that can have opposite effects on gene expression, as well as 2) maternal or paternally inherited genetic variation that could affect parental gene expression on the same chromosome.

We use methods to detect opposite parent of origin effects on total expression, as well as parent-specific expression in the Hutterites, a founder population of European descent, for which we have phased genotype data [61]. We use RNA-seq from LCLs to map transcripts to parental haplotypes and use the parental gene expression to look for variation in *cis* that would effect gene expression. Our study is likely underpowered to identify any opposite effect parent of origin eQTLs but we do identify a few parent of origin eQTLs where parentally inherited variation affects parent specific gene expression from the same haplotype. There is no known biological mechanism as to why parent of origin eQTLs could exist outside of imprinted loci.

4.3 Results

For each of 306 individuals, the total number of transcripts at each gene was assigned as maternally inherited, paternally inherited, or unknown parent of origin. The last group included transcripts without heterozygote SNPs or SNPs without parent of origin information.

Transcripts were assigned to the parentally inherited categories using SNPs in the reads and matching alleles to either the known maternally or paternally inherited alleles. All the genes analyzed had some transcripts of unknown origin (average 97.8%, range 8.3-100%). For each gene we assigned parental origin to an average of 1.8% of transcripts (range: 0-34.7%), and for each individual we assigned parental origin to an average of 1.4% of transcripts (range: 0-1.7%). On average, about 40 SNPs per gene were used to assign the transcripts of a gene to a parent (range 1-1839 SNPs).

4.3.1 Opposite Parent of Origin eQTL (oeQTL)

Our original oeQTL identified three significant opposite effect associations but these associations were driven by one individual's genotype. The significant associations are shown in Figures 4.1. Once we subset SNPs on having at least three individuals in each of three genotype groups, we did not find any significant results (Bonferroni corrected p-value).

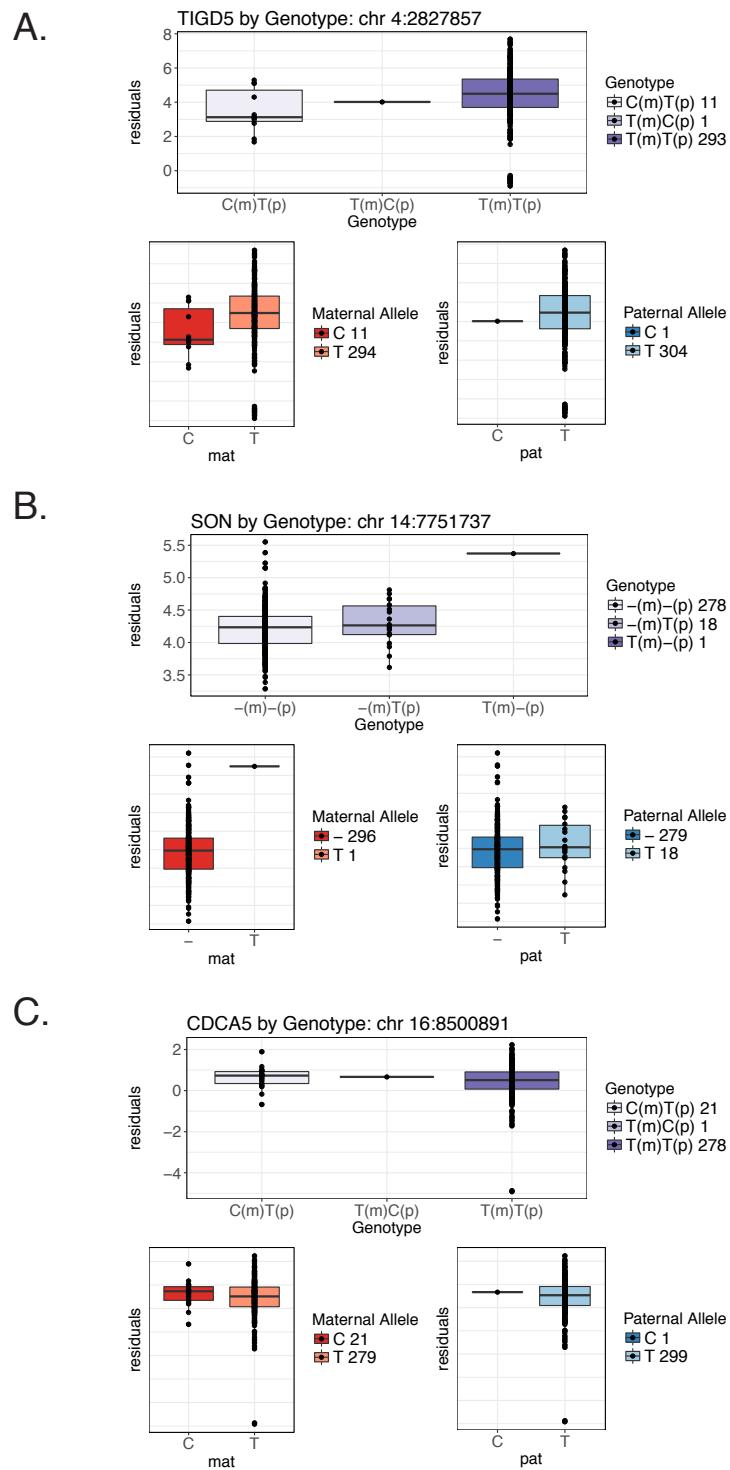


Figure 4.1: **Opposite effect eQTLs driven by one individual's genotype.** The three most significant opposite effect eQTLs for genes A) *TIGD5* (p-value 3.1e-09), B) *SON* (p-value 4.3e-09) and C) *CDCA5* (p-value 3.5e-09). The parent of origin eQTL is driven by one heterozygous individual for each gene.

4.3.2 Single Parent eQTL (mat-eQTL, pat-eQTL)

We performed the mat-eQTL and pat-eQTL analysis, using parent of origin normalized expression. We normalized the parental gene expression data using library sizes from the total gene expression(see Methods). However, the data was sparse and zeros drove most of the analysis. The significant maternal and paternal associations were driven by zeros in the data shown in Figures 4.2 and 4.3.

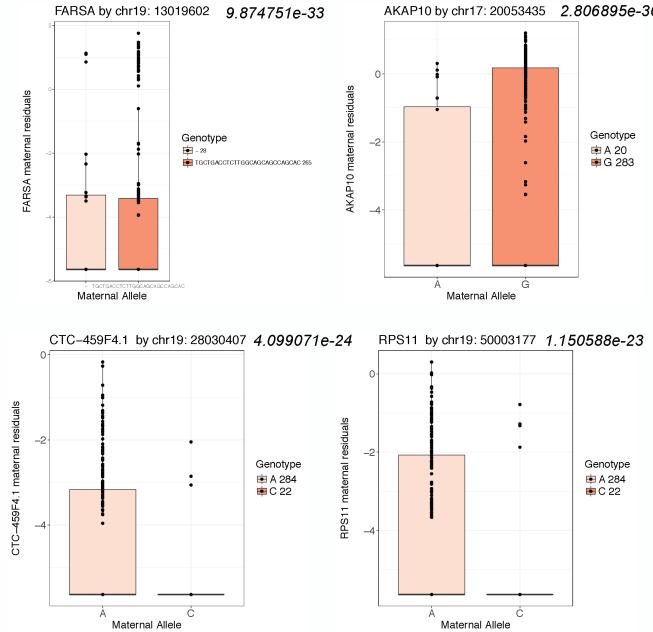


Figure 4.2: Maternal eQTL Associations Driven by Null Values. The four most significant maternal eQTL associations. Most of the individuals have no value of expression for these genes, we see most of the genes have a median that corresponds to a value of zero after normalization.

To address this problem, we redid the same analysis using only informative reads, removing zeros that were due to absence of heterozygous SNPs in the gene (see Methods for more detail). There were 7,398,096 SNP-gene pairs we could compare across both single parent eQTLs. For eQTLs significant in both, (60,549 SNP gene pairs), the effect sizes were all

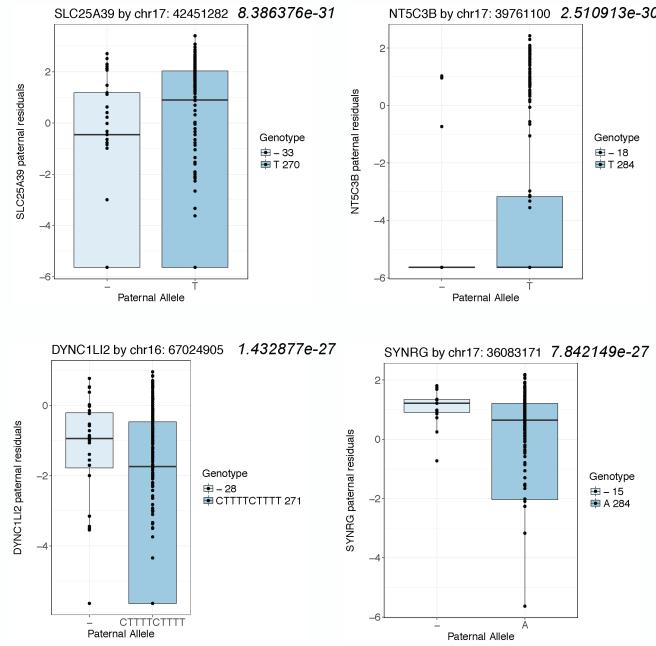


Figure 4.3: Paternal eQTL Associations Driven by Null Values. The four most significant paternal eQTL associations. Most of the individuals have no value of expression for these genes, we see most of the genes have a median that corresponds to a value of zero after normalization.

in the same direction: no SNPs had opposite effects on their corresponding parental gene expression (Figure 4.4). The imbalance of positive and negative effect sizes in Figure 4.4 is likely due in large part to the sparsity of the data, where most individuals have an expression value of zero and any individuals with some measure of expression drive the effect size to be positive.

We compared SNP gene pairs that were significant (Bonferroni) in one parent, and not significant ($p > 0.05$) in the other parent. 7,712 SNP-gene pairs were maternally significant and not paternally significant and 10,815 paternal significant associations were not maternally significant. An example of each is shown in Figure 4.5 where the maternally inherited A allele at 20:20036897 in Figure 4.5A is associated with increased maternal expression, but at least half of the individuals with the paternally inherited A allele at the same SNP have

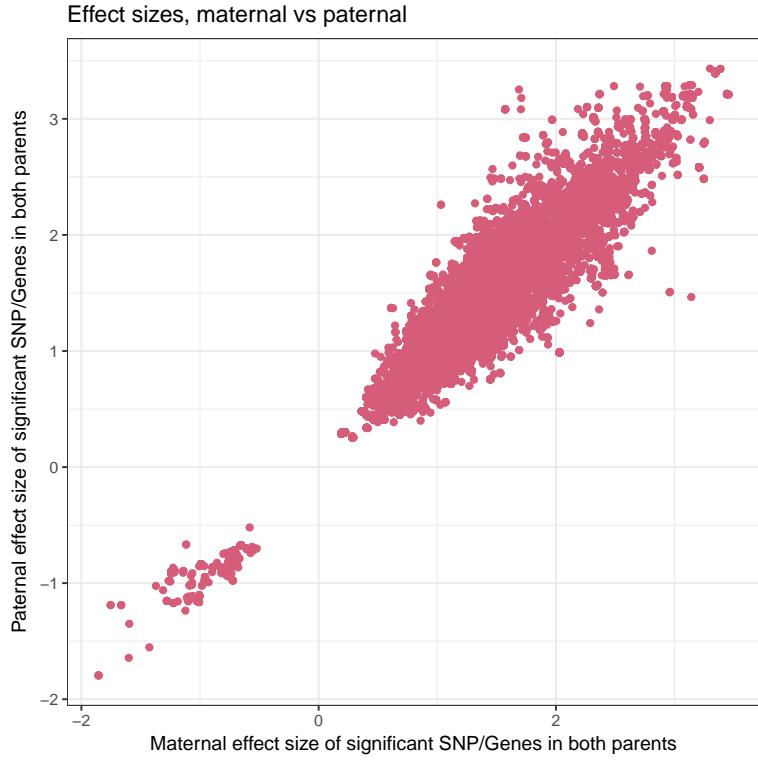


Figure 4.4: **Similar effect sizes across mat-eQTL and pat-eQTL.**

no expression from the paternal haplotype. In Figure 4.5 B, the A allele at 9:134850002 is associated with increased paternal expression when inherited from the father but not when inherited from the mother.

4.3.3 Parent of Origin (PO) - ASE Test

To detect parent of origin effects on expression using a different approach, we did a PO ASE test (see Methods) using parental gene expression count data. We identified 56,800 significant results using a Bonferroni corrected p-value . The top ten significant genes with their most significant SNPs are included in Table 4.1. The top four genes with their most significant SNPs shown in Figures 4.6, 4.7, 4.8, and 4.9. Of the 10 most significant genes, 5 are imprinted (*SNHG14*, *ZDBF2*, *PEG10*, *L3MBTL1*, *FAM50B*.)

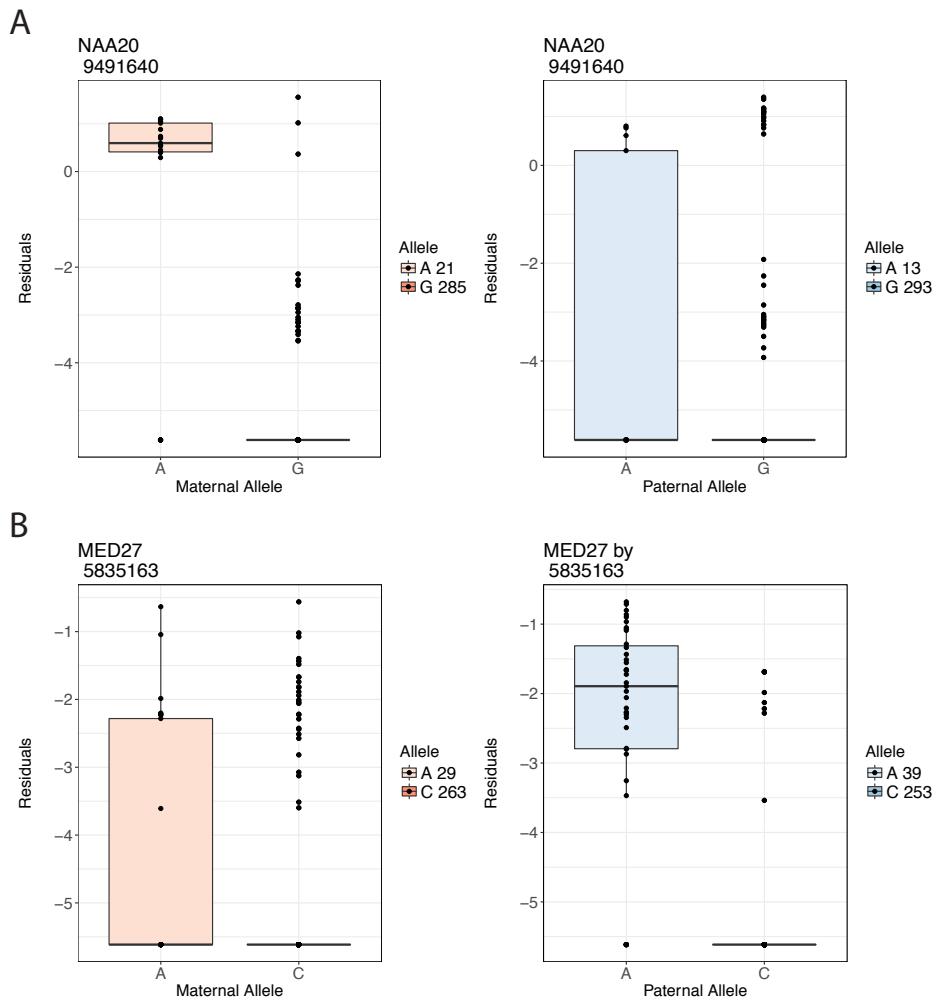


Figure 4.5: Parent specific eQTLs significant in one parent and not the other.
 (A) Variant on chromosome 20:20036897 has a significant mat-eQTL association (left) with gene *NAA20* ($p\text{-value} = 2.32\text{e-}27$) where the pat-eQTL (right) from the same variant is not significant ($p\text{-value} = 2.18\text{e-}02$). In contrast, (B) Variant on chromosome 9:134850002 has a significant pat-eQTL association (right) with gene *MED27* ($p\text{-value} = 4.7\text{e-}32$) and mat-eQTL is not significant ($p\text{-value} = 2.75\text{e-}03$) (left).

| Chr | bp | Gene | Maternal Reads (Ref/Alt) | Paternal Reads (Ref/Alt) | p-value |
|-----|---------|------------------|--------------------------|--------------------------|-----------|
| 15 | 8034579 | <i>SNHG14</i> * | 15/11 | 587/983 | 0 |
| 2 | 1459666 | <i>ZDBF2</i> * | 34/49 | 638/1190 | 0 |
| 5 | 3342675 | <i>ERAP2</i> | 3948/2445 | 4790/7067 | 0 |
| 7 | 4671649 | <i>PEG10</i> * | 42/4 | 635/1073 | 0 |
| 8 | 5271050 | <i>PABPC1</i> | 600/2564 | 21/494 | 0 |
| 20 | 9548868 | <i>L3MBTL1</i> * | 13/18 | 527/699 | 9.88e-248 |
| 14 | 8007882 | <i>IGHG1</i> | 794/156 | 1711/918 | 5.15e-191 |
| 6 | 3666034 | <i>FAM50B</i> * | 0/6 | 481/362 | 1.58e-180 |
| 22 | 9799426 | <i>IGLV2-5</i> | 240/499 | 23/1 | 6.56e-145 |
| 6 | 3758995 | <i>BTN3A2</i> | 3609/1049 | 3328/3330 | 8.40e-144 |

Table 4.1: **Top Ten Significant Genes from PO-ASE Test.** Five of the top ten significant genes from the PO-ASE test are imprinted genes (* represents imprinted genes). P-value of 0 corresponds to a really small p-value (<10e-248)

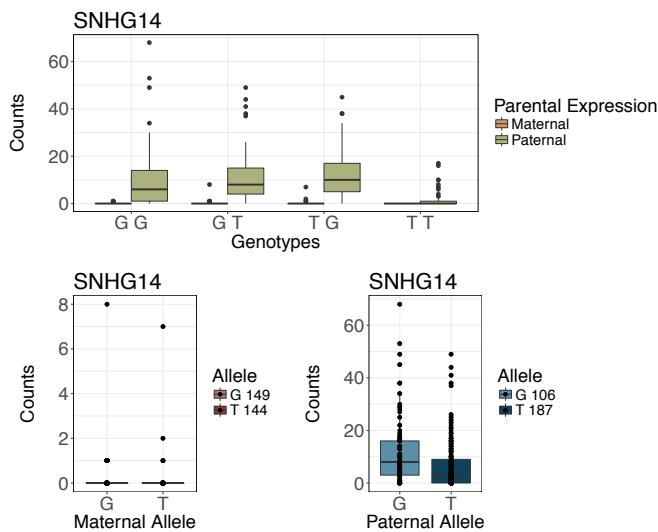


Figure 4.6: **Significant PO-ASE association with gene *SNHG14*.** For imprinted gene *SNHG14* we see variation in expression based on the allele and the parent it was inherited from. The paternally inherited T and G allele have different expression levels. Due to the imprinted status of the gene, we also see consistently more paternal expression from both alleles than maternal expression.

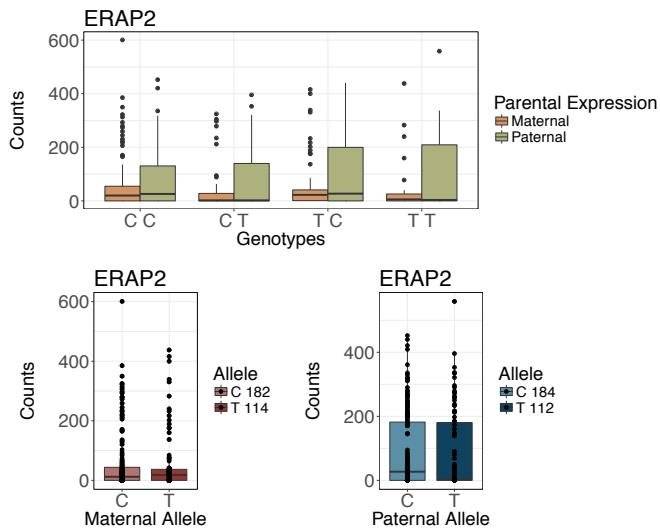


Figure 4.7: **Significant PO-ASE association with gene *ERAP2*.** For gene *ERAP2* Alleles C and T have different expression levels based on which parent the allele was inherited from. We see more expression form the paternal G allele than the paternal A allele in the reciprocal heterozygotes. Similar to imprinted genes, we see more paternal expression and less maternal expression across both alleles.

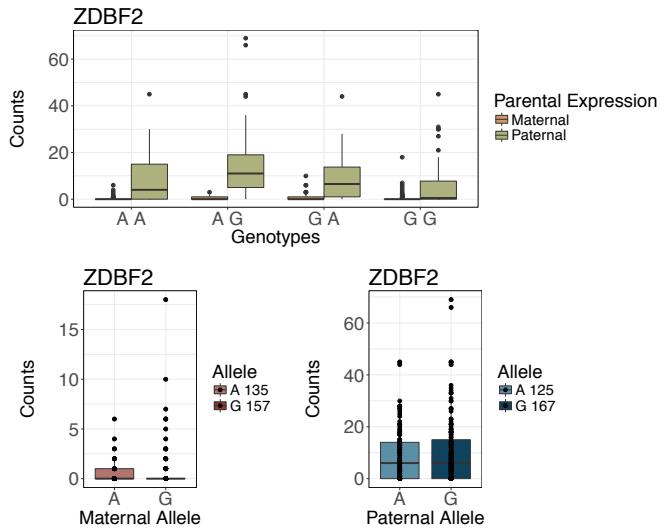


Figure 4.8: **Significant PO-ASE association with gene *ZDBF2*.** In *ZDBF2*, a maternally imprinted gene, there is consistently less maternal expression but still different in expression among parentally inherited expression, as well as by parentally inherited allele where there is more expression from the parental haplotype with the paternal G allele than the paternal A allele in the two reciprocal heterozygotes.

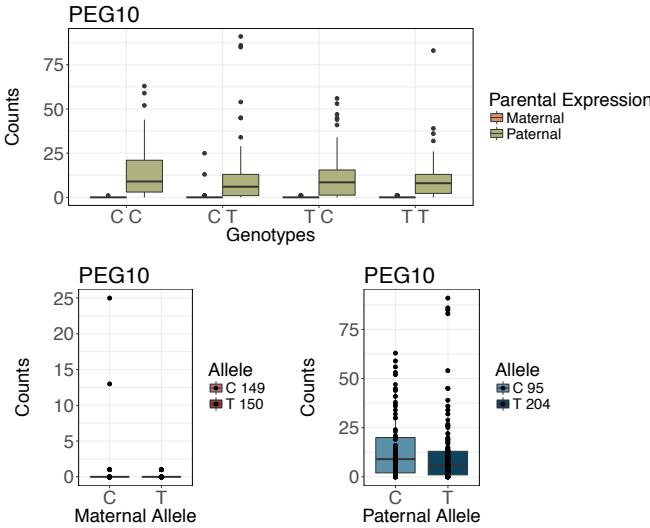


Figure 4.9: Significant PO-ASE association with gene *PEG10*. For maternally imprinted gene *PEG10* we see significant differences between maternal and paternal expression as well as paternally expression in the reciprocal heterozygotes. There is consistently less maternal expression due to the imprinting status of the gene and then more expression from the haplotype with the paternal expression with the C allele compared to the T allele in the reciprocal heterozygotes.

4.3.4 Modified ASE Test on Symmetrically Expressed Genes

We were not surprised to find that the most significant genes from the PO-ASE test were imprinted genes since one haplotype of imprinted genes is silenced, and the other haplotype will be expressed contributing to the significant difference in parental gene expression. We are searching for genes that have different parental expression based on a parentally inherited allele in *cis*. Imprinted genes, as those we identified in Figures 4.6, 4.8, and 4.9, are consistently not expressed from the imprinted haplotype, irrespective of which allele was inherited. We wanted to identify a parent of origin allele effect on expression.

To exclude imprinted genes in the analysis we only tested genes that did not have significant asymmetrical expression (see Methods). We were searching for genes that were more symmetrically expressed from each parental haplotype to identify if a parentally inherited allele can affect expression from the same haplotype in *cis*. We identified 15,340 significant

SNPs (not pruned for LD) with 518 genes (using Bonferroni p-value 5.99e-09). The two most significant genes from the PO-ASE test and with symmetric expression are plotted with their most significant SNPs in Figures 4.10 and 4.11. The top ten genes and their most significant SNPs are listed in Table 4.2.

| Chr | bp | Gene | Maternal reads (Ref/Alt) | Paternal reads (Ref/Alt) | p-value |
|-----|---------|---------|--------------------------|--------------------------|------------|
| 1 | 404610 | SEC22B | 148/101 | 513/478 | 1.11e-98 |
| 3 | 2222154 | ZMAT3 | 1320/1650 | 257/2085 | 3.53e-83 |
| 8 | 5425272 | PARP10 | 128/334 | 324/1018 | 1.06e-71 |
| 12 | 7253839 | OAS3 | 4465/939 | 2069/4327 | 1.06e-65 |
| 11 | 6368707 | IRF7 | 620/2463 | 2079/167 | 2.94e-6520 |
| 14 | 8008568 | IGHV2-5 | 291/295 | 22/135 | 1.80e-62 |
| 17 | 8884867 | CCDC137 | 1449/865 | 173/1819 | 1.45e-57 |
| 14 | 7958818 | ITPK1 | 387/686 | 862/1056 | 2.48e-57 |
| 17 | 8865144 | SEPT9 | 560/2595 | 2594/1305 | 7.50e-50 |
| 17 | 8730038 | SLFN5 | 2528/3084 | 1678/2539 | 1.27e-48 |

Table 4.2: Top Ten Significant Genes from PO-ASE Test after Filtering Asymmetrically Expressed Genes. Filtering on symmetrical expression removed imprinted genes from being the most significant genes from PO-ASE test.

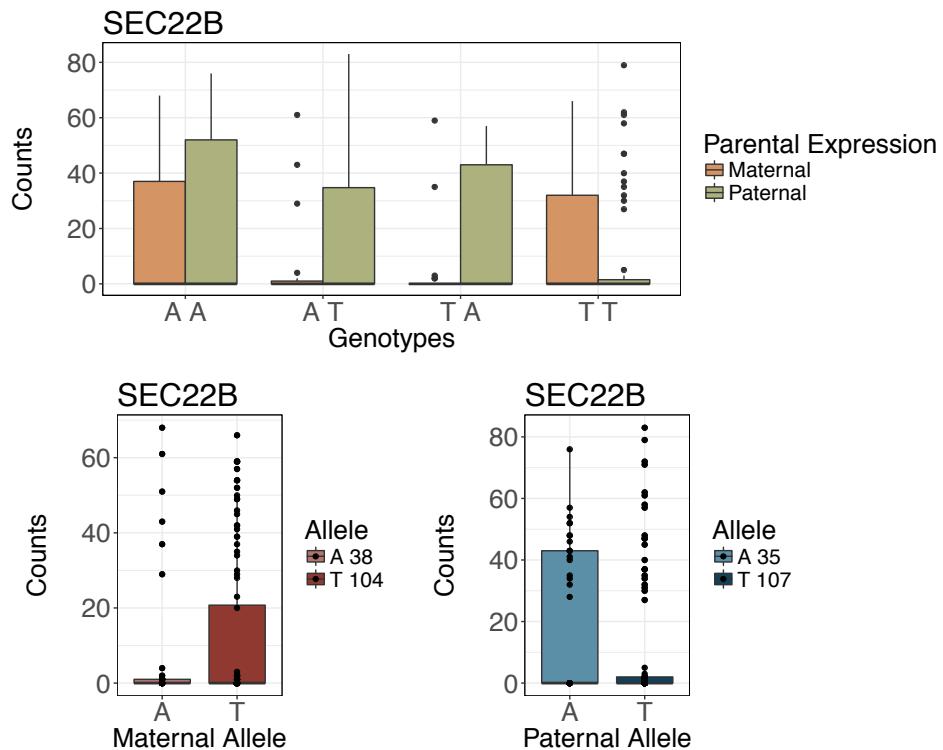


Figure 4.10: **Significant Association from PO-ASE test with *SEC22B*.** The most significant SNP (1:404610) with the most significant gene *SEC22B*. In the top plot with all four genotypes we see expression from the paternal allele in all four genotypes except TT. Maternal expression is only seen in the homozygotes. In the plots by parental allele below, it is clear that the maternal T allele is associated with increased maternal expression compared to the maternal A allele at this SNP. We see the opposite effect with paternal allele where the paternal A allele is associated with increased paternal expression and the paternal T allele with decreased expression. Based on the parental origin of the allele, the expression from the haplotype is different.

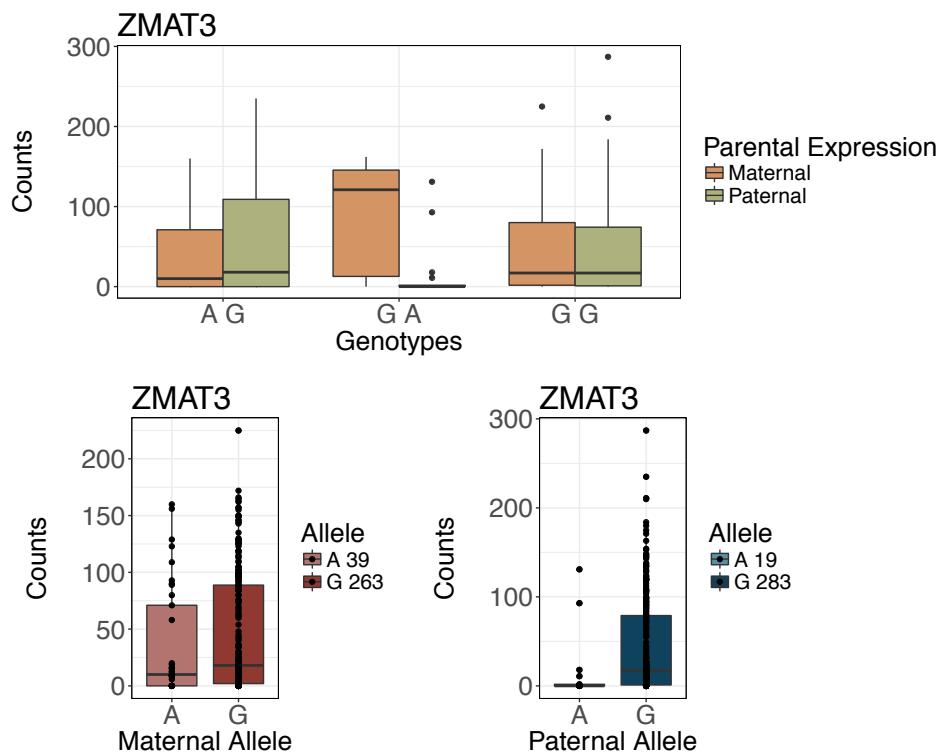


Figure 4.11: **Significant Association from PO-ASE test with *ZMAT3*.** The most significant SNP (3:2222154) with the second most significant gene *ZMAT3*. In the top plot with all three genotypes we see expression from the paternal allele in all four genotypes except GA. Maternal expression in all genotypes. In the plots by parental allele below, it is clear that the maternal alleles are associated with maternal expression, but the paternal A allele, in contrast to the paternal G allele, is not associated with paternal expression.

4.4 Discussion

Previous studies using parental alleles and gene expression have identified imprinted genes and genetic variation that affects quantitative traits [115, 12, 14, 38], but none to our knowledge have looked at how genetic variation can impact parent of origin expression from the same haplotype as the genetic variation.

Here we used parental specific gene expression in 306 Hutterites to characterize genetic effects on parental expression. We first performed a parent of origin opposite effect eQTL (oeQTL) using total gene expression. We then did a maternal and a paternal eQTL study of maternal and paternal gene expression (mat-eQTL, pat-eQTL), respectively. Finally, we tested for parent of origin effects among reciprocal heterozygotes.

Our opposite effect model has been successful in identifying opposite effects of parentally inherited variants on quantitative traits in the Hutterites but we were not able to find any with gene expression in LCLs. These could be due to a number of reasons, including sample size and the tissue studied. LCLs are transformed cell types and the transformation could alter imprinting mechanisms. We also performed a *cis* mat-eQTL and pat-eQTL study. This test identified known significant eQTLs that showed up in both the mat-eQTL and pat-eQTL results since the effect does not depend on the parent of origin. None of the variants compared across the mat-eQTL and pat-eQTL showed opposite effects by parent of origin. We were not able to find any maternal or paternally only effects on gene expression without modifying our test.

We found that most of the negative results were driven by sparsity in the data. Null expression values for gene expression could be due to two factors: 1) no heterozygous/parent of origin SNPs in the gene such that homozygous reads could not be assigned to a parent, or 2) there are heterozygous SNPs in the genes but there are no reads. Genes without any heterozygous SNPs for an individual were considered missing and not included in the analysis. We maintained the values for those genes for the individuals with at least one

heterozygous site in a gene. Although this resulted in different numbers of individuals and genes to be tested, and provided a more conservative and informative data set, we still did not find any significant opposite effects on gene expression.

Finally, we performed a PO- ASE test among reciprocal heterozygotes to identify effects of parental variation on gene expression. The missing gene expression (i.e. uninformative) for some individuals decreased the numbers of reciprocal heterozygotes we could test for each gene.

These few results from the PO-ASE test and the opposite effect eQTL could be due to many limitations of our study. Although we were able to determine the parent of origin for many transcripts in the Hutterites, we could not assign every RNA sequencing read to a parent due to lack of heterozygous sites or missing parent of origin information for alleles. Missing parental gene expression resulted in very sparse data. Second, we conducted these studies in LCLs, and therefore would miss effects in other tissues or developmental time points. Additionally, our models to test for parent of origin eQTL effects are effective but could be much improved, such as to model over-dispersion in gene expression.

In summary, we did not identify any genetic variation with opposite parental effects on either parentally mapped gene expression or total gene expression. We did identify SNPs with parent of origin eQTLs even though our data are noisy and underpowered. Deeper sequencing and better modeling could potentially identify more genetic variation that impacts parental gene expression if such a biological mechanism exists. We expect these possible parent of origin eQTLs from the PO-ASE test to represent variation that can impact imprinting or other gene silencing mechanisms by parent of origin. For example, for the *ZMAT3* gene in Figure 4.11 we see expression from all haplotypes except from the paternal A allele. It could suggest that the paternal G allele is disrupting the mechanism that turns off paternal expression at this gene or that the paternal A allele activates a mechanism to silence the gene expression from the paternal haplotype. If more evidence for such effects are identified,

it could lead to new approaches to understanding genetic variation and its impact on gene expression and, ultimately, disease and human health.

4.5 Methods

4.5.1 Genotypes and Sample Information

LCL RNA-seq transcripts for 306 individuals were mapped to parental haplotypes as in Chapter 3. We used the measures of total as well as maternal and paternal expression in this study. We used multiple approaches to characterize parent of origin effects on gene expression. To be conservative, we used 306 Hutterite individuals for which we have parental genotypes and tested SNPs for which we have at least three individuals in at least three of four parent of origin genotype classes (such that we have at least three individuals in at least one heterozygote category and one heterozygote individual will not drive our analysis). We used QCed SNPs with MAF >5%.

4.5.2 RNA-seq QC

Multiple approaches required different QC methods. For the total gene expression, we used normalized gene expression. First, we removed lowly expressed genes with a log count per million (cpm) greater than 1 in at least 20 individuals. The R/Bioconductor package edgeR was used to convert the RNA-seq counts to log2 TMM-normalized CPM values[93, 92]. Technical covariates correlated with gene expression Principal Components were regressed out (RIN, DNA concentration, RNA concentration, Flowcell/Lane).

4.5.3 Parent of Origin Expression QC

Maternal gene expression was used as both counts and as normalized gene expression. Maternal gene expression counts were used directly from STAR gene count output[29] subsetted

on genes included in the total gene expression analysis. Normalized maternal expression was calculated similar to total gene expression using edgeR and converting RNA-seq counts to log₂ TMM normalized CPM values using normalization factors (library sizes) from the total gene expression (maternal gene expression too sparse on its own). Same method was used to get paternal gene expression counts and normalized paternal gene expression.

4.5.4 Informative Genes

To separate informative parental gene expression from uninformative parental gene expression I compiled all of the heterozygous SNPs for each individual for each gene that was expressed in LCLs. If a gene for an individual did not have any heterozygous parent of origin SNPs (i.e. informative SNPs), the gene was considered missing (converted to NA for downstream analysis). If there was at least one heterozygous parent of origin SNP in the corresponding gene, the gene expression value was not altered, since zero expression for that gene for that parent could be informative. This nulled different numbers of genes for different individuals (Figure 4.12).

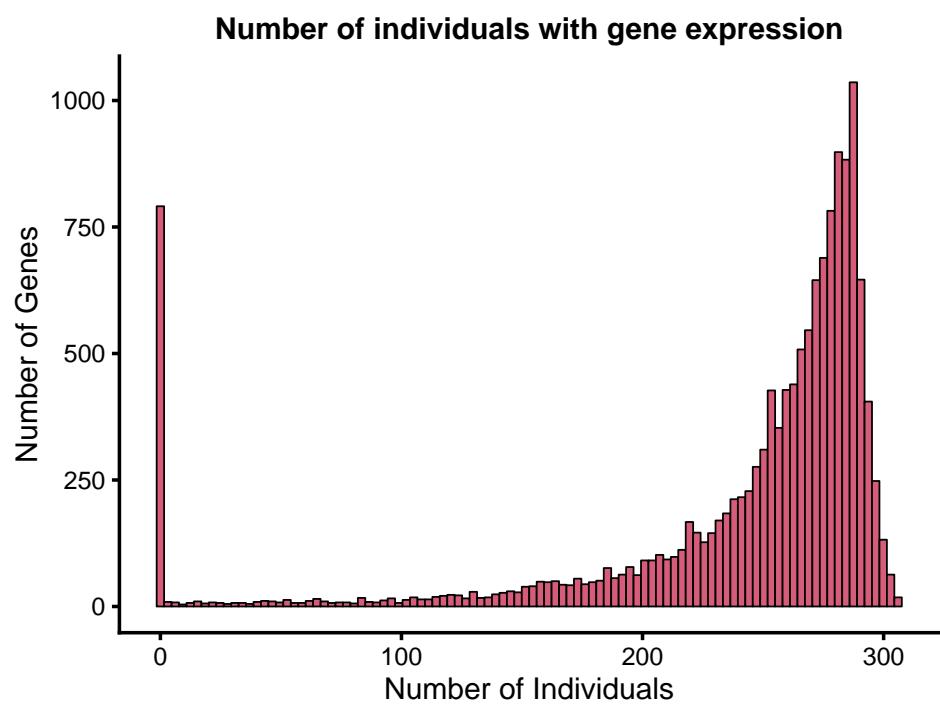


Figure 4.12: Number of Individuals with Gene Expression.

4.5.5 Opposite Parent of Origin eQTL

We used the same method outlined in Chapter 2 to detect if SNPs had opposite effects on total gene expression by parental origin. We tested SNPs in *cis*, defined as +/-250kb from the TSS of the gene. The model as implemented in GEMMA[114] where we used sex and age as covariates and corrected for relatedness (see Methods from Chapter 2).

4.5.6 Single Parent eQTL

To use the parent of origin expression, we performed a *cis* eQTL testing for specific parental effects on the same parental gene expression as follows, where n is the number of individuals, Y_P and Y_M are an $n \times 1$ vector of quantitative traits corresponding to paternal and maternal expression. W is an $n \times c$ matrix of covariates (fixed effects) including intercept 1. α is a $c \times 1$ vector of covariate coefficients. X_M is an $n \times 1$ vector of maternal alleles, and X_P an $n \times 1$ vector of paternal alleles. β_M and β_P are the effect sizes of maternal and paternal alleles, respectively. g is a vector of genetic effects with $g \sim N(0, A(\sigma_g)^2)$ where A is the genetic relatedness matrix; ϵ is a vector of non-genetic effects with $\epsilon \sim N(0, I(\sigma_e)^2)$.

$$Y_M = W\alpha + X_M\beta_M + g + \epsilon \quad (4.1)$$

$$Y_P = W\alpha + X_P\beta_P + g + \epsilon \quad (4.2)$$

We defined *cis* as +/- 250kb from the TSS of the gene. The model was implemented in GEMMA[114] where we used sex and age as covariates and corrected for relatedness.

4.5.7 PO ASE Test

We used a simple χ^2 test on the reciprocal heterozygotes on their corresponding allelic expression using maternal and paternal count data corresponding to haplotype specific expression.

Layout of the matrix for the test is in Table 4.3.

| Genotype | A allele (ref) expression | G allele (alt) expression |
|------------|---------------------------|---------------------------|
| A_{MG_P} | 148 (maternal expression) | 478 (paternal expression) |
| G_{MA_P} | 513 (paternal expression) | 101 (maternal expression) |

Table 4.3: **Setup for PO-ASE test.** For reciprocal genotypes we use the maternal and paternal gene expression counts corresponding to opposite alleles. The chi-squared test here will determine if there is an allele and parental effect.

4.5.8 Not Asymmetrically Expressed Genes

We used a binomial test to determine which genes had more maternal expression than paternal expression. We summed paternal and maternal gene expression across all individuals for a gene and used the sum of maternal and paternal gene expression in a binomial test to determine if the ratio of maternal to the sum of maternal and paternal expression is only 1/2. We kept genes not significant (Bonferroni, $p > 3.5\text{e-}06$) for the Modified ASE Test.

CHAPTER 5

CONCLUSION

As we look past the "low hanging fruit" of GWAS, we turn to other biological mechanisms where genetics can influence our traits, including studies of rare variation[45, 60], gene by environment interactions, and parent of origin effects, among others. The Hutterites are an ideal population to look for parent of origin effects on quantitative disease related traits and gene expression because of their common environmental exposures and similar genetic background[111, 1, 80]. With the availability of RNA-seq data, novel methods for imputing and phasing data at a population scale[61], and extensive phenotyping of the Hutterites, we are able to start investigating such questions related to parent of origin effects.

5.1 A novel method to detect opposite effects of parentally inherited variants on cardiovascular disease and asthma associated traits

In Chapter 2, I describe our study to detect genetic variation than can impact cardiovascular disease and asthma associated traits when inherited from one parent and not the other. In order to maximize our data, we developed a novel method that can be used to detect parentally inherited alleles with opposite effects on quantitative traits, depending on which parent they were inherited from. These associations would be missed in traditional GWAS. Previous studies have explored parent of origin effects but mostly using trios [38, 4, 44] and on binary disease status [55, 4]. Only a few studies have searched for parent of origin effects on quantitative traits, starting with height [14, 115].

Using our method, we identify parent of origin effects (POEs) on 11 phenotypes in the Hutterites, most of which are risk factors for cardiovascular disease. Most of the loci we identified have features of imprinted regions and many of the variants are associated with

expression of nearby genes. Most of the phenotypes are also associated with metabolic traits consistent with the parent conflict hypothesis of imprinted genes first put forward by Haig [41, 13, 85]. The idea suggests that parental interests may be in conflict such that paternal alleles favor growth of the fetus at the expense of the mother while maternal alleles favor restricting resources to the fetus to ensure preservation of her own nutritional needs. We show that POEs, which can be opposite in direction, are relatively common in humans, are possible imprinted regions, and have potentially important clinical effects.

It is necessary to replicate the parent of origin effects we identified in a different population to verify these effects exist. The associations did not replicate in the Sardinia population, although there were few suggestively significant. Additionally a power analysis will need to be done to confirm our method would pick up genetic variants of the effect size and minor allele frequency that we identified.

It is important to further investigate these regions. One of the regions with a POE identified in our study with *LINC01081* has been studied in detail by another group showing that a parent of origin effect exists at this region [102]. Others have shown POEs at known imprinted regions that affect height[14, 115] that we were not able to replicate in the Hutterites likely due to our smaller sample size. It would be the next step to show whether or not the regions we identify with POEs lie in or near previously unidentified imprinted regions.

5.2 Identifying two novel imprinted genes in known imprinted regions using parent of origin gene expression

In Chapter 3, I identified two new imprinted genes using parent of origin and allele specific expression. Using a novel method as a variation on WASP [104], I mapped RNA-seq reads to parental haplotypes using SNPs in the reads and parent of origin information of SNPs. We identified known imprinted genes and two novel imprinted genes, *PXDC1* and *PWAR6*,

among our genes with asymmetrical parent of origin gene expression. We validated the patterns of gene expression using RNA-seq from peripheral blood leukocytes (PBLs). To validate the imprinted genes further, we used DNA methylation levels in the PBLs to confirm imprinting control regions (ICRs) previously defined[49, 24]. Our two new imprinted genes lie in known imprinted regions with known ICRs providing more evidence for their imprinting status.

This is the largest pedigree based genome-wide scan for imprinted regions to date with 306 Hutterite individuals. We also provide a new way of mapping reads to parent of origin haplotypes and identify two new possible imprinted genes.

We would still need to further validate *PXDC1* and *PWAR6* as imprinted genes in a different population and possibly in additional cell types. Further characterization of these loci are still required. We identified these genes as imprinted but had overall very few parentally mapped reads. This would need to be replicated with many more parentally mapped reads. It was somewhat surprising they were not previously discovered as imprinted genes since they are in known and somewhat well characterized imprinted regions. It is possible they are tissue and/or developmentally specific as has been shown for other imprinted genes [12]. It is also possible that, although imprinted, variation at these imprinted genes does not affect disease, contrary to other imprinted genes and how they have been previously identified.

5.3 Can genetic variation by parent of origin influence gene expression on the same haplotype?

In Chapter 4, I explored if parentally inherited genetic variation can affect gene expression as well as haplotype specific gene expression (maternal and paternal specific expression). Using the method described in Chapter 2 to detect opposite effect associations, we did not identify any SNPs that had opposite effects on total gene expression depending on which parent the allele as inherited from. To get at the same opposite effects in a different way, we

performed a maternal eQTL (mat-eQTL) with maternal alleles on maternal expression as well as a paternal eQTL (pat-eQTL) with paternal alleles on paternal expression. Across all the SNP-gene pairs analyzed we did not identify any effects that were opposite in direction.

To determine if any alleles had a specific effect on gene expression when inherited from one parent or the other, we used a parent of origin allele specific expression (PO-ASE) test and identified alleles that when inherited from one parent, and not the other, had a different effect on gene expression from the same haplotype as the allele.

While our model detected genes that look like they could have a PO allele specific effect, there is still room for improvement. We use gene expression counts but our model is not accounting for overdispersion, which needs to be accounted for when looking at gene expression. We also looked for PO-ASE effects with very few parentally mapped reads. For standard eQTL studies we require at least 10 million reads, whereas, here, we have on average 1.8% of 10 million reads mapped to a parental haplotype resulting in, on average, about half of 1.8 million reads to use for an eQTL study. Our study is likely, therefore, underpowered to identify SNPs that have opposite effects on gene expression depending on which parent it was inherited from. More reads are required to perform this study and find parent specific effects.

Even with these parent of origin eQTLs identified in Chapter 4, it is not clear what biological mechanism could lead to such an effect. It is possible that the parentally inherited allele that contributes to gene expression differently from the rest could disrupt normal expression by inducing silencing of the gene in a chromosome specific manner. It is also possible that the opposite allele - from the same parent - results in aberrant over expression of the allele. Both of these would result in a parent of origin effect on gene expression but why or how these would occur is unknown.

5.4 Future Directions

With this large pedigree data we are able to answer some questions about parent of origin effects on gene expression and quantitative traits, but it barely scratches the surface of biological mechanisms that we don't yet have a pipeline for or the ideal dataset to detect such effects, in contrast to genome wide association studies. From Chapter 2, our method to detect opposite parent of origin effects has been progress in this direction. The method needs to be tested on multiple datasets to confirm the results are accurate. There would also need to be a power analysis done to confirm that most of the associations are not false positives. Although these opposite effect variants don't show an effect on gene expression in our study or others [14], we still need to understand the mechanism by which these variants are acting. Better characterization of the expression of genes nearby these variants needs to be done, across different tissues, to get at whether these effects are a result of imprinting. Methylation and chromatin in these regions could be studied once the gene or genes contributing to the parent of origin phenotype are identified. Characterization of allele specific interactions using chromatin conformation capture at heterozygous SNPs could also help unravel the different interactions among parental alleles if one is not imprinted and both genes are actively expressed. Ultimately, functional characterization of these SNPs remains necessary to understand the mechanism behind these opposite parental effects and the traits they impact.

In Chapter 3 we identify known imprinted genes in lymphoblastoid cell lines as well as two novel imprinted genes. Using a larger dataset, with accurate parent of origin calls, more parent of origin calls, more sequencing reads, and across different tissues is necessary to confirm these novel imprinted genes as well as confirm more known imprinted genes. Imprinted genes have been shown to be tissue specific [12] and thus performing a similar study in more tissues and samples could increase the resolution on imprinted genes and their mechanisms of gene expression.

Although we were not able to do it in Chapter 4, it remains necessary to identify variation that can impact imprinted genes or impact gene expression by parent of origin. Similar to diseases that result from mutations in imprinted genes, it is possible genetic variation that affects gene expression in a parent of origin manner can affect diseases or disease associated traits.

In this dissertation, I characterize parent of origin effects of some variants, and their small effects on quantitative disease associated traits, but there still remains a lot to be discovered and better understood for us to have a better understanding of the genome and its impact on human health and disease. These parent of origin effects, although small, likely do contribute to heritability of traits as more researchers, including us, have identified genetic variation that can impact traits by parent of origin [14, 115, 38, 55?].

5.5 Concluding remarks

In this thesis, I contribute to the growing knowledge of additional biological mechanisms, specifically, parent of origin effects, that can contribute to phenotypes and disease risk missed in standard GWAS. In all the chapters of this dissertation we are able to leverage the parent of origin information for alleles in the Hutterites and use them to uncover novel effects of genetic variation on gene expression and quantitative traits. In Chapter 2 I developed a new method to test for opposite effects of parentally inherited genetic variation on quantitative traits and we find 11 parent of origin effects, including maternal, paternal, and opposite parental effects. In Chapter 3, I mapped RNA-seq to parental haplotypes using parent of origin of alleles and uncover known and new imprinted genes (*PXDC1* and *PWAR6*), which show similar patterns of expression in PBLs and have well characterized ICRs in the region. In Chapter 4, I explored how parent of origin genetic variation could influence gene expression. We did not find any opposite effects on gene expression but identified alleles that when inherited from one parent showed altered gene expression patterns.

Throughout this research, I investigated the basis for parent of origin effects for which we have not had the methodology and or the appropriate dataset I have today to explore such a phenomenon. The novelty of these data and the questions I asked has allowed for the expansion of this research from one original aim to a full dissertation covering 3 separate projects and future directions such as measuring parent of origin heritability on traits and gene expression. The research in this dissertation provides methods and a basis for studying parent of origin effects and other less well characterized effects that could contribute to heritability of traits and disease risk.

References

- [1] Mark Abney, Mary Sara McPeek, and Carole Ober. Broad and Narrow Heritabilities of Quantitative Traits in a Founder Population. *The American Journal of Human Genetics*, 68(5):1302–1307, May 2001.
- [2] Lara K Abramowitz and Marisa S Bartolomei. Genomic imprinting: recognition and marking of imprinted loci. *Current opinion in genetics & development*, 22(2):72–78, April 2012.
- [3] Ana Paula Abreu, Andrew Dauber, Delanie B Macedo, Sekoni D Noel, Vinicius N Brito, John C Gill, Priscilla Cukier, Iain R Thompson, Victor M Navarro, Priscila C Gagliardi, Tânia Rodrigues, Cristiane Kochi, Carlos Alberto Longui, Dominique Beckers, Francis de Zegher, Luciana R Montenegro, Berenice B Mendonca, Rona S Carroll, Joel N Hirschhorn, Ana Claudia Latronico, and Ursula B Kaiser. Central precocious puberty caused by mutations in the imprinted gene MKRN3. *New England Journal of Medicine*, 368(26):2467–2475, June 2013.
- [4] Holly F Ainsworth, Jennifer Unwin, Deborah L Jamison, and Heather J Cordell. Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genetic Epidemiology*, 35(1):19–45, December 2010.
- [5] Hala Al Adhami, Brendan Evano, Anne Le Digarcher, Charlotte Gueydan, Emeric Dubois, Hugues Parrinello, Christelle Dantec, Tristan Bouschet, Annie Varrault, and Laurent Journot. A systems-level approach to parental genomic imprinting: the imprinted gene network includes extracellular matrix genes and regulates cell cycle exit and differentiation. *Genome Research*, 25(3):353–367, March 2015.
- [6] S Anders, P T Pyl, and W Huber. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2):166–169, January 2015.
- [7] Martin J Aryee, Andrew E Jaffe, Hector Corrada-Bravo, Christine Ladd-Acosta, Andrew P Feinberg, Kasper D Hansen, and Rafael A Irizarry. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*, 30(10):1363–1369, May 2014.
- [8] Tomas Babak. Identification of imprinted loci by transcriptome sequencing. *Methods in molecular biology (Clifton, N.J.)*, 925:79–88, 2012.
- [9] Tomas Babak, Brian DeVeale, Christopher Armour, Christopher Raymond, Michele A Cleary, Derek van der Kooy, Jason M Johnson, and Lee P Lim. Global Survey of Genomic Imprinting by Transcriptome Sequencing. *Current Biology*, 18(22):1735–1741, November 2008.

- [10] Tomas Babak, Brian DeVeale, Emily K Tsang, Yiqi Zhou, Xin Li, Kevin S Smith, Kim R Kukurba, Rui Zhang, Jin Billy Li, Derek van der Kooy, Stephen B Montgomery, and Hunter B Fraser. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature genetics*, 47(5):544–549, May 2015.
- [11] Nicholas E Banovich, Xun Lan, Graham McVicker, Bryce van de Geijn, Jacob F Degner, John D Blischak, Julien Roux, Jonathan K Pritchard, and Yoav Gilad. Methylation QTLs are associated with coordinated changes in transcription factor binding, histone modifications, and gene expression levels. *PLoS genetics*, 10(9):e1004663, September 2014.
- [12] Yael Baran, Meena Subramaniam, Anne Biton, Taru Tukiainen, Emily K Tsang, Manuel A Rivas, Matti Pirinen, Maria Gutierrez-Arcelus, Kevin S Smith, Kim R Kukurba, Rui Zhang, Celeste Eng, Dara G Torgerson, Cydney Urbanek, GTEx Consortium, Jin Billy Li, Jose R Rodriguez-Santana, Esteban G Burchard, Max A Seibold, Daniel G MacArthur, Stephen B Montgomery, Noah A Zaitlen, and Tuuli Lappalainen. The landscape of genomic imprinting across diverse adult human tissues. *Genome Research*, 25(7):927–936, July 2015.
- [13] D P Barlow and M S Bartolomei. Genomic Imprinting in Mammals. *Cold Spring Harbor Perspectives in Biology*, 6(2):a018382–a018382, February 2014.
- [14] Stefania Benonisdottir, Asmundur Oddsson, Agnar Helgason, Ragnar P Kristjansson, Gardar Sveinbjornsson, Arna Oskarsdottir, Gudmar Thorleifsson, Olafur B Davidsdottir, Gudny A Arnadottir, Gerald Sulem, Brynjar O Jansson, Hilma Holm, Kristjan F Alexandersson, Laufey Tryggvadottir, G Bragi Walters, Sigurjon A Gudjonsson, Lucas D Ward, Jon K Sigurdsson, Paul D Iordache, Michael L Frigge, Thorunn Rafnar, Augustine Kong, Gisli Masson, Hannes Helgason, Unnur Thorsteinsdottir, Daniel F Gudbjartsson, Patrick Sulem, and Kari Stefansson. Epigenetic and genetic components of height regulation. *Nature Communications*, 7:13490, November 2016.
- [15] Jeremie Boucher, Marika Charalambous, Kim Zarse, Marcelo A Mori, Andre Kleinridders, Michael Ristow, Anne C Ferguson-Smith, and C Ronald Kahn. Insulin and insulin-like growth factor 1 receptors are required for normal expression of imprinted genes. *Proceedings of the National Academy of Sciences of the United States of America*, 111(40):14512–14517, October 2014.
- [16] C M Brideau, K E Eilertson, J A Hagarman, C D Bustamante, and P D Soloway. Successful Computational Prediction of Novel Imprinted Genes from Epigenomic Features. *Molecular and Cellular Biology*, 30(13):3357–3370, June 2010.
- [17] Karin Buiting. Prader-Willi syndrome and Angelman syndrome. *American journal of medical genetics. Part C, Seminars in medical genetics*, 154C(3):365–376, August 2010.

- [18] D Canoy, V Beral, A Balkwill, F L Wright, M E Kroll, G K Reeves, J Green, B J Cairns, and for the Million Women Study Collaborators*. Age at Menarche and Risks of Coronary Heart and Other Vascular Diseases in a Large UK Cohort. *Circulation*, 131(3):237–244, January 2015.
- [19] B M Cattanach and M Kirk. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, 315(6019):496–498, June 1985.
- [20] H S Chandra and S W Brown. Chromosome imprinting and the mammalian X chromosome. *Nature*, 253(5488):165–168, January 1975.
- [21] Andrew Chess. Monoallelic Gene Expression in Mammals. *Annual Review of Genetics*, 50(1):317–327, November 2016.
- [22] Sanaa Choufani, Cheryl Shuman, and Rosanna Weksberg. Beckwith-Wiedemann syndrome. *American journal of medical genetics. Part C, Seminars in medical genetics*, 154C(3):343–354, August 2010.
- [23] Trees-Juen Chuang, Yu-Hsiang Tseng, Chia-Ying Chen, and Yi-Da Wang. Assessment of imprinting- and genetic variation-dependent monoallelic expression using reciprocal allele descendants between human family trios. *7(1):1–12*, July 2017.
- [24] Franck Court, Chiharu Tayama, Valeria Romanelli, Alex Martin-Trujillo, Isabel Iglesias-Platas, Kohji Okamura, Naoko Sugahara, Carlos Simón, Harry Moore, Julie V Harness, Hans Keirstead, Jose Vicente Sanchez-Mut, Eisuke Kaneki, Pablo Lapunzina, Hidenobu Soejima, Norio Wake, Manel Esteller, Tsutomu Ogata, Kenichiro Hata, Kazuhiko Nakabayashi, and David Monk. Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. *Genome Research*, 24(4):554–569, April 2014.
- [25] H V Crouse. The Controlling Element in Sex Chromosome Behavior in *Sciara*. *Genetics*, 45(10):1429–1443, October 1960.
- [26] Darren A Cusanovich, Minal Caliskan, Christine Billstrand, Katelyn Michelini, Claudia Chavarria, Sherryl De Leon, Amy Mitrano, Noah Lewellyn, Jack A Elias, Geoffrey L Chupp, Roberto M Lang, Sanjiv J Shah, Jeanne M Decara, Yoav Gilad, and Carole Ober. Integrated analyses of gene expression and genetic association studies in a founder population. *Human Molecular Genetics*, 25(10):2104–2112, May 2016.
- [27] Caroline Daelemans, Matthew E Ritchie, Guillaume Smits, Sayeda Abu-Amero, Ian M Sudbery, Matthew S Forrest, Susana Campino, Taane G Clark, Philip Stanier, Dominic Kwiatkowski, Panos Deloukas, Emmanouil T Dermitzakis, Simon Tavaré, Gudrun E Moore, and Ian Dunham. High-throughput analysis of candidate imprinted genes and allele-specific gene expression in the human term placenta. *BMC genetics*, 11(1):25, April 2010.

- [28] Florence Demenais, Patricia Margaritte-Jeannin, Kathleen C Barnes, William O C Cookson, Janine Altmüller, Wei Ang, R Graham Barr, Terri H Beaty, Allan B Becker, John Beilby, Hans Bisgaard, Unnur Steina Bjornsdottir, Eugene Bleeker, Klaus Bønnelykke, Dorret I Boomsma, Emmanuelle Bouzigon, Christopher E Brightling, Myriam Brossard, Guy G Brusselle, Esteban Burchard, Kristin M Burkart, Andrew Bush, Moira Chan-Yeung, Kian Fan Chung, Alexessander Couto Alves, John A Curtin, Adnan Custovic, Denise Daley, Johan C de Jongste, Blanca E Del-Rio-Navarro, Kathleen M Donohue, Liesbeth Duijts, Celeste Eng, Johan G Eriksson, Martin Farrall, Yuliya Fedorova, Bjarke Feenstra, Manuel A Ferreira, Australian Asthma Genetics Consortium (AAGC) collaborators, Maxim B Freidin, Zofia Gajdos, Jim Gauderman, Ulrike Gehring, Frank Geller, Jon Genuneit, Sina A Gharib, Frank Gilliland, Raquel Granell, Penelope E Graves, Daniel F Gudbjartsson, Tari Haahtela, Susan R Heckbert, Dick Heederik, Joachim Heinrich, Markku Heliövaara, John Henderson, Blanca E Himes, Hiroshi Hirose, Joel N Hirschhorn, Albert Hofman, Patrick Holt, Jouke Hottinga, Thomas J Hudson, Jennie Hui, Medea Imboden, Vladimir Ivanov, Vincent W V Jaddoe, Alan James, Christer Janson, Marjo-Riitta Jarvelin, Deborah Jarvis, Graham Jones, Ingileif Jonsdottir, Pekka Jousilahti, Michael Kabesch, Mika Kähönen, David B Kantor, Alexandra S Karunas, Elza Khusnutdinova, Gerard H Koppelman, Anita L Kozyrskyj, Eskil Kreiner, Michiaki Kubo, Rajesh Kumar, Ashish Kumar, Mikko Kuokkanen, Lies Lahousse, Tarja Laitinen, Catherine Laprise, Mark Lathrop, Susanne Lau, Young-Ae Lee, Terho Lehtimäki, Sébastien Letort, Albert M Levin, Guo Li, Liming Liang, Laura R Loehr, Stephanie J London, Daan W Loth, Ani Manichaikul, Ingo Marenholz, Fernando J Martinez, Melanie C Matheson, Rasika A Mathias, Kenji Matsumoto, Hamdi Mbarek, Wendy L McArdle, Mads Melbye, Erik Melén, Deborah Meyers, Sven Michel, Hamida Mohamdi, Arthur W Musk, Rachel A Myers, Maartje A E Nieuwenhuis, Emiko Noguchi, George T O'Connor, Ludmila M Ogorodova, Cameron D Palmer, Aarno Palotie, Julie E Park, Craig E Pennell, Göran Pershagen, Alexey Polonikov, Dirkje S Postma, Nicole Probst-Hensch, Valery P Puzyrev, Benjamin A Raby, Olli T Raitakari, Adaikalavan Ramasamy, Stephen S Rich, Colin F Robertson, Isabelle Romieu, Muhammad T Salam, Veikko Salomaa, Vivi Schlünssen, Robert Scott, Polina A Selivanova, Torben Sigsgaard, Angela Simpson, Valérie Siroux, Lewis J Smith, Maria Solodilova, Marie Standl, Kari Stefansson, David P Strachan, Bruno H Stricker, Atsushi Takahashi, Philip J Thompson, Gudmar Thorleifsson, Unnur Thorsteinsdottir, Carla M T Tiesler, Dara G Torgerson, Tatsuhiko Tsunoda, André G Uitterlinden, Ralf J P van der Valk, Amaury Vaysse, Sailaja Vedantam, Andrea von Berg, Erika von Mutius, Judith M Vonk, Johannes Waage, Nick J Wareham, Scott T Weiss, Wendy B White, Magnus Wickman, Elisabeth Widén, Gonnieke Willemse, L Keoki Williams, Inge M Wouters, James J Yang, Jing Hua Zhao, Miriam F Moffatt, Carole Ober, and Dan L Nicolae. Multiancestry association study identifies new asthma risk loci that colocalize with immune-cell enhancer marks. *Nature genetics*, 50(1):42–53, January 2018.

- [29] Alexander Dobin and Thomas R Gingeras. *Mapping RNA-seq Reads with STAR*,

volume 29 of *Mapping RNA-seq Reads with STAR*. John Wiley & Sons, Inc., Hoboken, NJ, USA, August 2002.

- [30] Louise E Docherty, Faisal I Rezwan, Rebecca L Poole, Hannah Jagoe, Hannah Lake, Gabrielle A Lockett, Hasan Arshad, David I Wilson, John W Holloway, I Karen Temple, and Deborah J G Mackay. Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. *Journal of medical genetics*, 51(4):229–238, April 2014.
- [31] Pan Du, Warren A Kibbe, and Simon M Lin. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24(13):1547–1548, July 2008.
- [32] Thomas Eggermann. Russell-Silver syndrome. *American journal of medical genetics. Part C, Seminars in medical genetics*, 154C(3):355–364, August 2010.
- [33] Evan E Eichler, Jonathan Flint, Greg Gibson, Augustine Kong, Suzanne M Leal, Jason H Moore, and Joseph H Nadeau. Missing heritability and strategies for finding the underlying causes of complex disease. *Nature reviews. Genetics*, 11(6):446–450, June 2010.
- [34] Francesca Marta Elli, Agnes Linglart, Intza Garin, Luisa de Sanctis, Paolo Bordogna, Virginie Grybek, Arrate Pereda, Federica Giachero, Elisa Verrua, Patrick Hanna, Giovanna Mantovani, and Guiomar Perez de Nanclares. The Prevalence of GNAS Deficiency-Related Diseases in a Large Cohort of Patients Characterized by the EuroPHP Network. *The Journal of clinical endocrinology and metabolism*, 101(10):3657–3668, October 2016.
- [35] J Greg Falls, David J Pulford, Andrew A Wylie, and Randy L Jirtle. Genomic Imprinting: Implications for Human Disease. *The American Journal of Pathology*, 154(3):635–647, March 1999.
- [36] Michael D Fountain, Emmelien Aten, Megan T Cho, Jane Juusola, Magdalena A Walkiewicz, Joseph W Ray, Fan Xia, Yaping Yang, Brett H Graham, Carlos A Bacino, Lorraine Potocki, Arie van Haeringen, Claudia A L Ruivenkamp, MD, Pedro Mancias, Hope Northrup, Mary K Kukolich, Marjan M Weiss, Conny M A van Ravenswaaij-Arts, Inge B Mathijssen, Sebastien Levesque, Naomi Meeks, Jill A Rosenfeld, Danielle Lemke, Ada Hamosh, Suzanne K Lewis, Simone Race, Laura L Stewart, Beverly Hay, Andrea M Lewis, Rita L Guerreiro, Jose T Bras, Marcia P Martins, Gerarda Derkx-Lubsen, Els Peeters, Connie Stumpel, Sander Stegmann, Levinus A Bok, Gijs W E Santen, and Christian P Schaaf. The phenotypic spectrum of Schaaf-Yang syndrome: 18 new affected individuals from 14 families. *Genetics in medicine : official journal of the American College of Medical Genetics*, 19(1):45–52, January 2017.
- [37] A Gabory, M A Ripoche, A Le Digarcher, F Watrin, A Ziyyat, T Forne, H Jammes, J F X Ainscough, M A Surani, L Journot, and L Dandolo. H19 acts as a trans regulator

- of the imprinted gene network controlling growth in mice. *Development*, 136(20):3413–3421, September 2009.
- [38] Paras Garg, Christelle Borel, and Andrew J Sharp. Detection of Parent-of-Origin Specific Expression Quantitative Trait Loci by Cis-Association Analysis of Gene Expression in Trios. *PLoS ONE*, 7(8):e41695, August 2012.
- [39] Greg Gibson. Rare and common variants: twenty arguments. *Nature reviews. Genetics*, 13(2):135–145, January 2012.
- [40] GTEx Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, 348(6235):648–660, May 2015.
- [41] David Haig. The Kinship Theory of Genomic Imprinting. *Annual Review of Ecology and Systematics*, 31(1):9–32, November 2000.
- [42] I Hatada and T Mukai. Genomic imprinting of p57KIP2, a cyclin-dependent kinase inhibitor, in mouse. *Nature genetics*, 11(2):204–206, October 1995.
- [43] Ann-Kathrin Hoffmann, Xenia Naj, and Stefan Linder. Daam1 is a regulator of filopodia formation and phagocytic uptake of Borrelia burgdorferi by primary human macrophages. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, 28(7):3075–3089, July 2014.
- [44] Richard Howey and Heather J Cordell. PREMIM and EMIM: tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. *BMC Bioinformatics*, 13(1):149, June 2012.
- [45] Catherine Igartua, Sahar V Mozaffari, Dan L Nicolae, and Carole Ober. Rare non-coding variants are associated with plasma lipid traits in a founder population. *Scientific reports*, 7(1):16415, November 2017.
- [46] Yiannis Ioannides, Kemi Lokulo-Sodipe, Deborah J G Mackay, Justin H Davies, and I Karen Temple. Temple syndrome: improving the recognition of an underdiagnosed chromosome 14 imprinting disorder: an analysis of 51 published cases. *Journal of medical genetics*, 51(8):495–501, August 2014.
- [47] D R Johnson. Hairpin-tail: a case of post-reductional gene action in the mouse egg. *Genetics*, 76(4):795–805, April 1974.
- [48] W Evan Johnson, Cheng Li, and Ariel Rabinovic. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics (Oxford, England)*, 8(1):118–127, January 2007.
- [49] Ricky S Joshi, Paras Garg, Noah Zaitlen, Tuuli Lappalainen, Corey T Watson, Nidha Azam, Daniel Ho, Xin Li, Stylianos E Antonarakis, Han G Brunner, Karin Buiting, Sau Wai Cheung, Bradford Coffee, Thomas Eggermann, David Francis, Joep P

Geraedts, Giorgio Gimelli, Samuel G Jacobson, Cedric Le Caignec, Nicole de Leeuw, Thomas Liehr, Deborah J Mackay, Stephen B Montgomery, Alistair T Pagnamenta, Peter Papenhausen, David O Robinson, Claudia Ruivenkamp, Charles Schwartz, Bernhard Steiner, David A Stevenson, Urvashi Surti, Thomas Wassink, and Andrew J Sharp. DNA Methylation Profiling of Uniparental Disomy Subjects Provides a Map of Parental Epigenetic Bias in the Human Genome. *American journal of human genetics*, 99(3):555–566, September 2016.

- [50] Goo Jun, Matthew Flickinger, Kurt N Hetrick, Jane M Romm, Kimberly F Doheny, Goncalo R Abecasis, Michael Boehnke, and Hyun Min Kang. Detecting and Estimating Contamination of Human DNA Samples in Sequencing and Array-Based Genotype Data. *The American Journal of Human Genetics*, 91(5):839–848, November 2012.
- [51] Masayo Kagami, Kenji Kurosawa, Osamu Miyazaki, Fumitoshi Ishino, Kentaro Matsuoka, and Tsutomu Ogata. Comprehensive clinical studies in 34 patients with molecularly defined UPD(14)pat and related conditions (Kagami-Ogata syndrome). *European Journal of Human Genetics*, 23(11):1488–1498, November 2015.
- [52] Masayo Kagami, Keiko Matsubara, Kazuhiko Nakabayashi, Akie Nakamura, Shinichiro Sano, MD, Kohji Okamura, PhD1, Kenichiro Hata, Maki Fukami, Tsutomu Ogata, and PhD2. Genome-wide multilocus imprinting disturbance analysis in Temple syndrome and Kagami-Ogata syndrome. *Genetics in medicine : official journal of the American College of Medical Genetics*, 19(4):476–482, April 2017.
- [53] Jennifer M Kalish, Connie Jiang, and Marisa S Bartolomei. Epigenetics and imprinting in human disease. *The International journal of developmental biology*, 58(2-4):291–298, 2014.
- [54] Lihua Kang, Jingnan Sun, Xue Wen, Jiuwei Cui, Guanjun Wang, Andrew R Hoffman, Ji-Fan Hu, and Wei Li. Aberrant allele-switch imprinting of a novel IGF1R intragenic antisense non-coding RNA in breast cancers. *European journal of cancer (Oxford, England : 1990)*, 51(2):260–270, January 2015.
- [55] Augustine Kong, Valgerdur Steinthorsdottir, Gisli Masson, Gudmar Thorleifsson, Patrick Sulem, Soren Besenbacher, Aslaug Jonasdottir, Asgeir Sigurdsson, Kari Th Kristinsson, Adalbjorg Jonasdottir, Michael L Frigge, Arnaldur Gylfason, Pall I Olason, Sigurjon A Gudjonsson, Sverrir Svartisson, Simon N Stacey, Bardur Sigurgeirsson, Kristrun R Benediktsdottir, Helgi Sigurdsson, Thorvaldur Jonsson, Rafn Benediktsdottir, Jon H Olafsson, Oskar Th Johannsson, Astradur B Hreidarsson, Gunnar Sigurdsson, DIAGRAM Consortium, Anne C Ferguson-Smith, Daniel F Gudbjartsson, Unnur Thorsteinsdottir, and Kari Stefansson. Parental origin of sequence variants associated with complex diseases. *Nature*, 462(7275):868–874, December 2009.
- [56] Charles A Laurin, Gabriel Cuellar-Partida, Gibran Hemani, George Davey Smith, Jian Yang, and David M Evans. Partitioning Phenotypic Variance Due To Parent-Of-Origin Effects Using Genomic Relatedness Matrices. *bioRxiv*, pages 1–41, May 2017.

- [57] Heather A Lawson, James M Cheverud, and Jason B Wolf. Genomic imprinting and parent-of-origin effects on complex traits. *Nature reviews. Genetics*, 14(9):609–617, August 2013.
- [58] Jeffrey T Leek, W Evan Johnson, Hilary S Parker, Andrew E Jaffe, and John D Storey. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*, 28(6):882–883, March 2012.
- [59] Juan Li, Anthony J Bench, George S Vassiliou, Nasios Fourouclas, Anne C Ferguson-Smith, and Anthony R Green. Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. *Proceedings of the National Academy of Sciences*, 101(19):7341–7346, May 2004.
- [60] Xin Li, Yungil Kim, Emily K Tsang, Joe R Davis, Farhan N Damani, Colby Chiang, Gaelen T Hess, Zachary Zappala, Benjamin J Strober, Alexandra J Scott, Amy Li, Andrea Ganna, Michael C Bassik, Jason D Merker, GTEx Consortium, Laboratory, Data Analysis &Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, NIH/NHGRI, NIH/NIMH, NIH/NIDA, Biospecimen Collection Source Site—NDRI, Biospecimen Collection Source Site—RPCI, Biospecimen Core Resource—VARI, Brain Bank Repository—University of Miami Brain Endowment Bank, Leidos Biomedical—Project Management, ELSI Study, Genome Browser Data Integration &Visualization—EBI, Genome Browser Data Integration &Visualization—UCSC Genomics Institute, University of California Santa Cruz, Ira M Hall, Alexis Battle, and Stephen B Montgomery. The impact of rare variation on gene expression across tissues. *Nature*, 550(7675):239–243, October 2017.
- [61] Oren E Livne, Lide Han, Gorka Alkorta-Aranburu, William Wentworth-Sheilds, Mark Abney, Carole Ober, and Dan L Nicolae. PRIMAL: Fast and Accurate Pedigree-based Imputation from Sequence Data in a Founder Population. *PLOS Computational Biology*, 11(3):e1004139, March 2015.
- [62] Isabel Lokody. Gene expression: Consequences of parent-of-origin effects. *Nature reviews. Genetics*, 15(3):145–145, January 2014.
- [63] Philippe P Luedi, Fred S Dietrich, Jennifer R Weidman, Jason M Bosko, Randy L Jirtle, and Alexander J Hartemink. Computational and experimental identification of novel human imprinted genes. *Genome Research*, 17(12):1723–1730, December 2007.
- [64] Weiwei Luo, Zi Zhao Lieu, Ed Manser, Alexander D Bershadsky, and Michael P Sheetz. Formin DAAM1 Organizes Actin Filaments in the Cytoplasmic Nodal Actin Network. *PLoS ONE*, 11(10):e0163915–22, October 2016.
- [65] Mary F Lyon and Sohaila Rastan. Parental source of chromosome imprinting and its relevance for X chromosome inactivation. *Differentiation*, 26(1):63–67, June 1984.

- [66] D J G Mackay, S E Boonen, J Clayton-Smith, J Goodship, J M D Hahnemann, S G Kant, P R Njølstad, N H Robin, D O Robinson, R Siebert, J P H Shield, H E White, and I K Temple. A maternal hypomethylation syndrome presenting as transient neonatal diabetes mellitus. *Human Genetics*, 120(2):262–269, September 2006.
- [67] Deborah J G Mackay and I Karen Temple. Human imprinting disorders: Principles, practice, problems and progress. *European journal of medical genetics*, 60(11):618–626, November 2017.
- [68] Jovana Maksimovic, Lavinia Gordon, and Alicia Oshlack. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome biology*, 13(6):R44, June 2012.
- [69] Giovanna Mantovani, Anna Spada, and Francesca Marta Elli. Pseudohypoparathyroidism and Gs α -cAMP-linked disorders: current view and open issues. *Nature Reviews Endocrinology*, 12(6):347–356, June 2016.
- [70] Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17(1):10, August 2011.
- [71] S Matsuoka, J S Thompson, M C Edwards, J M Bartletta, P Grundy, L M Kalikin, J W Harper, S J Elledge, and A P Feinberg. Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, p57KIP2, on chromosome 11p15. *Proceedings of the National Academy of Sciences*, 93(7):3026–3030, April 1996.
- [72] James McGrath and Davor Solter. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell*, 37(1):179–183, May 1984.
- [73] Richard Mott, Wei Yuan, Pamela Kaisaki, Xiangchao Gan, James Cleak, Andrew Edwards, Amelie Baud, and Jonathan Flint. The Architecture of Parent-of-Origin Effects in Mice. *Cell*, 156(1-2):332–342, January 2014.
- [74] Surabhi Mulchandani, Elizabeth J Bhoj, Minjie Luo, Nina Powell-Hamilton, Kim Jenny, Karen W Gripp, Miriam Elbracht, Thomas Eggermann, Claire L S Turner, I Karen Temple, Deborah J G Mackay, Holly Dubbs, David A Stevenson, Leah Slatton, Elaine H Zackai, Nancy B Spinner, Ian D Krantz, and Laura K Conlin. Maternal uniparental disomy of chromosome 20: a novel imprinting disorder of growth failure. *Genetics in medicine : official journal of the American College of Medical Genetics*, 18(4):309–315, April 2016.
- [75] R D Nicholls, JHM Knoll, M G Butler, and S Karam Nature. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *nature.com*, 1989.
- [76] Jessie Nicodemus-Johnson, Rachel A Myers, Noburu J Sakabe, Debora R Sobreira, Douglas K Hogarth, Edward T Naureckas, Anne I Sperling, Julian Solway, Steven R White, Marcelo A Nobrega, Dan L Nicolae, Yoav Gilad, and Carole Ober. DNA

methylation in lung cells is associated with asthma endotypes and genetic risk. *JCI insight*, 1(20):e90151, December 2016.

- [77] Dan L Nicolae, Eric Gamazon, Wei Zhang, Shiwei Duan, M Eileen Dolan, and Nancy J Cox. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS genetics*, 6(4):e1000888, April 2010.
- [78] Mats I Nilsson, Aliyah A Nissar, Dhuha Al-Sajee, Mark A Tarnopolsky, Gianni Parise, Boleslav Lach, Dieter O Fürst, Peter F M van der Ven, Rudolf A Kley, and Thomas J Hawke. Xin Is a Marker of Skeletal Muscle Damage Severity in Myopathies. *The American Journal of Pathology*, 183(6):1703–1709, December 2013.
- [79] Carole Ober. Asthma Genetics in the Post-GWAS Era. *Annals of the American Thoracic Society*, 13 Suppl 1:S85–90, March 2016.
- [80] Carole Ober, Mark Abney, and Mary Sara McPeek. The Genetic Dissection of Complex Traits in a Founder Population. *The American Journal of Human Genetics*, 69(5):1068–1079, November 2001.
- [81] Tsutomu Ogata and Masayo Kagami. Kagami-Ogata syndrome: a clinically recognizable upd(14)pat and related disorder affecting the chromosome 14q32.2 imprinted region. *Journal of human genetics*, 61(2):87–94, February 2016.
- [82] Hiroaki Okae, Hitoshi Hiura, Yuichiro Nishida, Ryo Funayama, Satoshi Tanaka, Hatusune Chiba, Nobuo Yaegashi, Keiko Nakayama, Hiroyuki Sasaki, and Takahiro Arima. Re-investigation and RNA sequencing-based identification of genes with placenta-specific imprinted expression. *Human Molecular Genetics*, 21(3):548–558, October 2011.
- [83] Nuala A O’Leary, Mathew W Wright, J Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S Joardar, Vamsi K Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M McGarvey, Michael R Murphy, Kathleen O’Neill, Shashikant Pujar, Sanjida H Rangwala, Daniel Rausch, Lillian D Riddick, Conrad Schoch, Andrei Shkeda, Susan S Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E Tully, Anjana R Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D Murphy, and Kim D Pruitt. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44(D1):D733–D745, January 2016.
- [84] Paolo Pariante, Raffaele Dotolo, Massimo Venditti, Diana Ferrara, Aldo Donizetti, Francesco Aniello, and Sergio Minucci. First Evidence of DAAM1 Localization During

the Post-Natal Development of Rat Testis and in Mammalian Sperm. *Journal of cellular physiology*, 231(10):2172–2184, October 2016.

- [85] Manus M Patten, Michael Cowley, Rebecca J Oakey, and Robert Feil. Regulatory links between imprinted genes: evolutionary predictions and consequences. *Proceedings of the Royal Society B: Biological Sciences*, 283(1824):20152760, February 2016.
- [86] John Rb Perry, Felix Day, Cathy E Elks, Patrick Sulem, Deborah J Thompson, Teresa Ferreira, Chunyan He, Daniel I Chasman, Tõnu Esko, Gudmar Thorleifsson, Eva Albrecht, Wei Q Ang, Tanguy Corre, Diana L Cousminer, Bjarke Feenstra, Nora Franceschini, Andrea Ganna, Andrew D Johnson, Sanela Kjellqvist, Kathryn L Lunetta, George McMahon, Ilja M Nolte, Lavinia Paternoster, Eleonora Porcu, Albert V Smith, Lisette Stolk, Alexander Teumer, Natalia Tšernikova, Emmi Tikkанen, Sheila Ulivi, Erin K Wagner, Najaf Amin, Laura J Bierut, Enda M Byrne, Jouke-Jan Hottenga, Daniel L Koller, Massimo Mangino, Tune H Pers, Laura M Yerges-Armstrong, Jing Hua Zhao, Irene L Andrulis, Hoda Anton-Culver, Femke Atsma, Stefania Bandinelli, Matthias W Beckmann, Javier Benitez, Carl Blomqvist, Stig E Bojesen, Manjeet K Bolla, Bernardo Bonanni, Hiltrud Brauch, Hermann Brenner, Julie E Buring, Jenny Chang-Claude, Stephen Chanock, Jinhui Chen, Georgia Chenevix-Trench, J Margriet Collée, Fergus J Couch, David Couper, Andrea D Covello, Angela Cox, Kamila Czene, Adamo Pio D'adamo, George Davey Smith, Immaculata De Vivo, Ellen W Demerath, Joe Dennis, Peter Devilee, Aida K Dieffenbach, Alison M Dunning, Gudny Eiriksdottir, Johan G Eriksson, Peter A Fasching, Luigi Ferrucci, Dieter Flesch-Janys, Henrik Flyger, Tatiana Foroud, Lude Franke, Melissa E Garcia, Montserrat García-Closas, Frank Geller, Eco Ej de Geus, Graham G Giles, Daniel F Gudbjartsson, Vilmundur Gudnason, Pascal Guénel, Suiqun Guo, Per Hall, Ute Hamann, Robin Haring, Catharina A Hartman, Andrew C Heath, Albert Hofman, Maartje J Hooning, John L Hopper, Frank B Hu, David J Hunter, David Karasik, Douglas P Kiel, Julia A Knight, Veli-Matti Kosma, Zoltán Kutalik, Sandra Lai, Diether Lambrechts, Annika Lindblom, Reedik Mägi, Patrik K Magnusson, Arto Mannermaa, Nicholas G Martin, Gisli Masson, Patrick F McArdle, Wendy L McArdle, Mads Melbye, Kyriaki Michailidou, Evelin Mihailov, Lili Milani, Roger L Milne, Heli Nevanlinna, Patrick Neven, Ellen A Nohr, Albertine J Oldehinkel, Ben A Oostra, Aarno Palotie, Munro Peacock, Nancy L Pedersen, Paolo Peterlongo, Julian Peto, Paul Dp Pharoah, Dirkje S Postma, Anneli Pouta, Katri Pylkäs, Paolo Radice, Susan Ring, Fernando Rivadeneira, Antonietta Robino, Lynda M Rose, Anja Rudolph, Veikko Salomaa, Serena Sanna, David Schlessinger, Marjanka K Schmidt, Mellissa C Southey, Ulla Sovio, Meir J Stampfer, Doris Stöckl, Anna M Storniolo, Nicholas J Timpson, Jonathan Tyrer, Jenny A Visser, Peter Vollenweider, Henry Völzke, Gerard Waeber, Melanie Waldenberger, Henri Wallaschofski, Qin Wang, Gonke Willemsen, Robert Winqvist, Bruce Hr Wolfenbuttel, Margaret J Wright, Australian Ovarian Cancer Study, GENICA Network, kConFab, LifeLines Cohort Study, InterAct Consortium, Early Growth Genetics (EGG) Consortium, Dorret I Boomsma, Michael J Econs, Kay-Tee Khaw, Ruth Jf Loos, Mark I McCarthy, Grant W Montgomery, John P Rice, Elizabeth A Streeten, Unnur Thorsteins-

dottir, Cornelia M van Duijn, Behrooz Z Alizadeh, Sven Bergmann, Eric Boerwinkle, Heather A Boyd, Laura Crisponi, Paolo Gasparini, Christian Gieger, Tamara B Harris, Erik Ingelsson, Marjo-Riitta Jarvelin, Peter Kraft, Debbie Lawlor, Andres Metspalu, Craig E Pennell, Paul M Ridker, Harold Snieder, Thorkild Ia Sørensen, Tim D Spector, David P Strachan, André G Uitterlinden, Nicholas J Wareham, Elisabeth Widén, Marek Zygmunt, Anna Murray, Douglas F Easton, Kari Stefansson, Joanne M Murabito, and Ken K Ong. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature*, 514(7520):92–97, October 2014.

- [87] Jo Peters. The role of genomic imprinting in biology and disease: an expanding view. *Nature reviews. Genetics*, 15(8):517–530, August 2014.
- [88] Nuno D Pires and Ueli Grossniklaus. Different yet similar: evolution of imprinting in flowering plants and mammals. *F1000Prime Reports*, 6, August 2014.
- [89] C Proudhon and D Bourc’his. Identification and resolution of artifacts in the interpretation of imprinted gene expression. *Briefings in Functional Genomics*, 9(5-6):374–384, January 2011.
- [90] Wolf Reik. Genomic imprinting and genetic disorders in man. *Trends in Genetics*, 5:332–336, January 1989.
- [91] DR Johnson Genetics Research and 1974. Further observations on the hairpin-tail (Thp) mutation in the mouse. *cambridge.org*, 24(02):207, April 2009.
- [92] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, January 2010.
- [93] Mark D Robinson and Alicia Oshlack. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome biology*, 11(3):R25, 2010.
- [94] Federico A Santoni, Georgios Stamoulis, Marco Garieri, Emilie Falconnet, Pascale Ribaux, Christelle Borel, and Stylianos E Antonarakis. Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. *American journal of human genetics*, 100(3):444–453, March 2017.
- [95] C Sapienza. Genome imprinting and dominance modification. *Annals of the New York Academy of Sciences*, 564:24–38, 1989.
- [96] Bertrand Servin and Matthew Stephens. Imputation-Based Analysis of Association Studies: Candidate Regions and Quantitative Traits. *PLoS genetics*, 3(7):e114, 2007.
- [97] Michelle M Stein, Cara L Hrusch, Justyna Gozdz, Catherine Igartua, Vadim Pivniouk, Sean E Murray, Julie G Ledford, Mauricius Marques dos Santos, Rebecca L Anderson, Nervana Metwali, Julia W Neilson, Raina M Maier, Jack A Gilbert, Mark Holbreich, Peter S Thorne, Fernando D Martinez, Erika von Mutius, Donata Vercelli, Carole Ober,

and Anne I Sperling. Innate Immunity and Asthma Risk in Amish and Hutterite Farm Children. *New England Journal of Medicine*, 375(5):411–421, August 2016.

- [98] Liora Z Strichman-Almashanu, Michael Bustin, and David Landsman. Retroposed copies of the HMG genes: a window to genome dynamics. *Genome Research*, 13(5):800–812, May 2003.
- [99] Pernilla Stridh, Sabrina Ruhrmann, Petra Bergman, Mélanie Thessén Hedreul, Sevasti Flytzani, Amennai Daniel Beyeen, Alan Gillett, Nina Krivosija, Johan Öckinger, Anne C Ferguson-Smith, and Maja Jagodic. Parent-of-Origin Effects Implicate Epigenetic Regulation of Experimental Autoimmune Encephalomyelitis and Identify Imprinted Dlk1 as a Novel Risk Gene. *PLoS genetics*, 10(3):e1004265, March 2014.
- [100] Jingnan Sun, Wei Li, Yunpeng Sun, Dehai Yu, Xue Wen, Hong Wang, Jiuwei Cui, Guanjun Wang, Andrew R Hoffman, and Ji-Fan Hu. A novel antisense long noncoding RNA within the IGF1R gene locus is imprinted in hematopoietic malignancies. *Nucleic acids research*, 42(15):9588–9601, September 2014.
- [101] M A H Surani, S C Barton, and M L Norris. Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature*, 308(5959):548–550, April 1984.
- [102] Przemyslaw Szafranski, Tomasz Gambin, Avinash V Dharmadhikari, Kadir Caner Akdemir, Shalini N Jhangiani, Jennifer Schuette, Nihal Godiwala, Svetlana A Yatsenko, Jessica Sebastian, Suneeta Madan-Khetarpal, Urvashi Surti, Rosanna G Abellar, David A Bateman, Ashley L Wilson, Melinda H Markham, Jill Slamon, Fernando Santos-Simarro, María Palomares, Julián Nevado, Pablo Lapunzina, Brian Hon-Yin Chung, Wai-Lap Wong, Yoyo Wing Yiu Chu, Gary Tsz Kin Mok, Eitan Kerem, Joel Reiter, Namasivayam Ambalavanan, Scott A Anderson, David R Kelly, Joseph Shieh, Taryn C Rosenthal, Kristin Scheible, Laurie Steiner, M Anwar Iqbal, Margaret L McKinnon, Sara Jane Hamilton, Kamilla Schlade-Bartusiak, Dawn English, Glenda Hendson, Elizabeth R Roeder, Thomas S DeNapoli, Rebecca Okashah Littlejohn, Daynna J Wolff, Carol L Wagner, Alison Yeung, David Francis, Elizabeth K Fiorino, Morris Edelman, Joyce Fox, Denise A Hayes, Sandra Janssens, Elfride De Baere, Björn Menten, Anne Loccuifier, Lieve Vanwalleghem, Philippe Moerman, Yves Szajnajer, Amy S Lay, Jennifer L Kussmann, Jasneek Chawla, Diane J Payton, Gael E Phillips, Erwin Brosens, Dick Tibboel, Annelies de Klein, Isabelle Maystadt, Richard Fisher, Neil Sebire, Alison Male, Maya Chopra, Jason Pinner, Girvan Malcolm, Gregory Peters, Susan Arbuckle, Melissa Lees, Zoe Mead, Oliver Quarrell, Richard Sayers, Martina Owens, Charles Shaw-Smith, Janet Lioy, Eileen McKay, Nicole de Leeuw, Ilse Feenstra, Liesbeth Spruijt, Frances Elmslie, Timothy Thiruchelvam, Carlos A Bacino, Claire Langston, James R Lupski, Partha Sen, Edwina Popek, and Paweł Stankiewicz. Pathogenetics of alveolar capillary dysplasia with misalignment of pulmonary veins. *Human Genetics*, 135(5):569–586, April 2016.

- [103] N Takagi and M Sasaki. Preferential inactivation of the paternally derived X chromosome in the extraembryonic membranes of the mouse. *Nature*, 256(5519):640–642, August 1975.
- [104] Bryce van de Geijn, Graham McVicker, Yoav Gilad, and Jonathan K Pritchard. WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods*, 12(11):1061–1063, November 2015.
- [105] Annie Varrault, Charlotte Gueydan, Annie Delalbre, Anja Bellmann, Souheir Housami, Cindy Aknin, Dany Severac, Laetitia Chotard, Malik Kahli, Anne Le Digarcher, Paul Pavlidis, and Laurent Journot. Zac1 Regulates an Imprinted Gene Network Critically Involved in the Control of Embryonic Growth. *Developmental Cell*, 11(5):711–722, November 2006.
- [106] Emma L Wakeling, Frédéric Brioude, Oluwakemi Lokulo-Sodipe, Susan M O’Connell, Jennifer Salem, Jet Bliek, Ana P M Canton, Krystyna H Chrzanowska, Justin H Davies, Renuka P Dias, Béatrice Dubern, Miriam Elbracht, Eloise Giabicani, Adda Grimberg, Karen Grønskov, Anita C S Hokken-Koelega, Alexander A Jorge, Masayo Kagami, Agnes Linglart, Mohamad Maghnie, Klaus Mohnike, David Monk, Gundrun E Moore, Philip G Murray, Tsutomu Ogata, Isabelle Oliver Petit, Silvia Russo, Edith Said, Meropi Toumba, Zeynep Tümer, Gerhard Binder, Thomas Eggermann, Madeleine D Harbison, I Karen Temple, Deborah J G Mackay, and Irène Netchine. Diagnosis and management of Silver-Russell syndrome: first international consensus statement. In *Nature reviews. Endocrinology*, pages 105–124. North West Thames Regional Genetics Service, London North West Healthcare NHS Trust, Watford Road, Harrow HA1 3UJ, UK., Nature Publishing Group, February 2017.
- [107] Qinchuan Wang, Jenny Li-Chun Lin, Albert J Erives, Cheng-I Lin, and Jim Jung-Ching Lin. New insights into the roles of Xin repeat-containing proteins in cardiac development, function, and disease. *International review of cell and molecular biology*, 310:89–128, 2014.
- [108] Y Wang, J R O’Connell, P F McArdle, J B Wade, S E Dorff, S J Shah, X Shi, L Pan, E Rampersaud, H Shen, J D Kim, A R Subramanya, N I Steinle, A Parsa, C C Ober, P A Welling, A Chakravarti, A B Weder, R S Cooper, B D Mitchell, A R Shuldiner, and Y P C Chang. Whole-genome association study identifies STK39 as a hypertension susceptibility gene. *Proceedings of the National Academy of Sciences*, 106(1):226–231, January 2009.
- [109] C R Weinberg, A J Wilcox, and R T Lie. A Log-Linear Approach to Case-Parent-Triad Data: Assessing Effects of Disease Genes That Act Either Directly or through Maternal Effects and That May Be Subject to Parental Imprinting. *The American Journal of Human Genetics*, 62(4):969–978, April 1998.
- [110] Clarice R Weinberg. Methods for Detection of Parent-of-Origin Effects in Genetic

Studies of Case-Parents Triads. *The American Journal of Human Genetics*, 65(1):229–235, July 1999.

- [111] Lauren A Weiss, Mark Abney, Edwin H Cook Jr., and Carole Ober. Sex-Specific Genetic Architecture of Whole Blood Serotonin Levels. *The American Journal of Human Genetics*, 76(1):33–41, January 2005.
- [112] Wellcome Trust Case Control Consortium, Peter Donnelly, Willem H Ouwehand, Nick Craddock, Nilesh J Samani, John A Todd, Mark I McCarthy, Jeffrey C Barrett, Andrew P Morris, Lon R Cardon, Chris C A Spencer, Matthew A Brown, Panos Deloukas, Martin Farrall, Alistair S Hall, Andrew T Hattersley, Christopher G Mathew, and Miles Parkes. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, June 2007.
- [113] Noah Zaitlen, Peter Kraft, Nick Patterson, Bogdan Pasaniuc, Gaurav Bhatia, Samuela Pollack, and Alkes L Price. Using Extended Genealogy to Estimate Components of Heritability for 23 Quantitative and Dichotomous Traits. *PLoS genetics*, 9(5):e1003520, May 2013.
- [114] Xiang Zhou and Matthew Stephens. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7):821–824, June 2012.
- [115] Magdalena Zoledziewska, Carlo Sidore, Charleston W K Chiang, Serena Sanna, Antonella Mulas, Maristella Steri, Fabio Busonero, Joseph H Marcus, Michele Marongiu, Andrea Maschio, Diego Ortega Del Vecchyo, Matteo Floris, Antonella Meloni, Alessandro Delitala, Maria Pina Concas, Federico Murgia, Ginevra Biino, Simona Vaccargiu, Ramaiah Nagaraja, Kirk E Lohmueller, UK10K Consortium, Nicholas J Timpson, Nicole Soranzo, Ioanna Tachmazidou, George Dedoussis, Eleftheria Zeggini, Understanding Society Scientific Group, Sergio Uzzau, Chris Jones, Robert Lyons, Andrea Angius, Goncalo R Abecasis, John Novembre, David Schlessinger, and Francesco Cucca. Height-reducing variants and selection for short stature in Sardinia. *Nature genetics*, 47(11):1352–1356, November 2015.
- [116] O Zuk, E Hechter, SR Sunyaev Proceedings of the, and 2012. The mystery of missing heritability: Genetic interactions create phantom heritability. *National Acad Sciences*, 2012.