Abundant Parent-of-origin Effect eQTL in Humans: The Framingham Heart Study

Yongtao Guan and Daniel Levy
National Heart, Lung, and Blood Institute

June 5, 2024

Abstract

Parent-of-origin effect (POE) is a phenomenon whereby an allele's effect on a phenotype depends both on its allelic identity and parent from whom the allele is inherited, as exemplified by the polar overdominance in the ovine callypyge locus and the human obesity *DLK1* locus. Systematic studies of POE of expression quantitative trait loci (eQTL) are lacking. In this study we use trios among participants in the Framingham Heart Study to examine to what extend POE exists for gene expression of whole blood using whole genome sequencing and RNA sequencing. For each gene and the SNPs in cis, we performed eQTL analysis using genotype, paternal, maternal, and joint models, where the genotype model enforces the identical effect sizes on paternal and maternal alleles, and the joint model allows them to have different effect sizes. We compared models using Bayes factors to identify paternal, maternal, and opposing eQTL, where paternal and maternal effects have opposite directions. The resultant variants are collectively called POE eQTL. The highlights of our study include: 1) There are more than $2,000$ genes harbor POE eQTL and majority POE eQTL are not in the vicinity of known imprinted genes; 2) Among 180 genes harboring opposing eQTL, 99 harbor exclusively opposing eQTL, and 58 of the 99 are phosphoprotein coding genes, reflecting significant enrichment; 3) Paternal eQTL are enriched with GWAS hits, and genes harboring paternal eQTL are enriched with drug targets. Our study demonstrates the abundance of POE in gene expression, illustrates the complexity of gene expression regulation, and provides a resource that is complementary to existing resources such as GTEx. We revisited two previous POE findings in light of our POE results. A SNP residing in *KCNQ1* that is maternally associated with diabetes is a maternal eQTL of *CDKN1C*, not *KCNQ1*. A SNP residing in *DLK1* that showed paternal polar overdominance for human obesity is a maternal eQTL of *MEG3*, offering an explanation for the baseline risk of homozygous samples through association between *MEG3* expression and obesity. Finally, we advised caution on conducting Mendelian randomization using gene expression as the exposure.

1 Introduction

Parent-of-origin effect (POE) is a phenomenon whereby an allele's effect on a phenotype not only depends on its allelic identity, but also on the parent from whom the allele is inherited (DeChiara et al., 1991; Cockett et al., 1996). POE can be driven by factors such as sex bias in transmission of genetic variants (Tomé et al., 2011) and maternal genetic effects that influences the environment in which the offspring develops (Hager et al., 2008). The main driver, however, is genomic imprinting, which is a phenomenon where one of the two alleles at a locus is functionally silenced by methylation. Importantly, which parental copy of the allele being silenced is highly consistent for that gene across samples. Imprinting appears to play an important role in embryonic and placental development and social behavior (Reik and Walter, 2001; Garfield et al., 2011). Genes underpinning syndromic disorders such as Prader-Willi, Beckwtih-Weidemann, and Angelman show POE (Lawson et al., 2013). The same is also true for common disease phenotypes, such as breast cancer, diabetes, and cardiovascular disease (Kong et al., 2009; Hanson et al., 2013; Mozaffari et al., 2019).

In genetic association studies, particularly genome-wide association studies (GWAS), the absence of POE is usually assumed implicitly. This assumption is largely out of convenience, as in most GWAS, the parent-of-origin of an allele is not directly observable, and not inferable due to lack of first and second degree relatives in the sample. Exceptions include an Icelandic cohort, with their extensively documented pedigrees and the abundance of closely related samples (Kong et al., 2009), a Hutterite pedigree (Mozaffari et al., 2019), the Framingham Heart Study with three-generations of participants (Kannel et al., 1979), and the biobank datasets, for which parent-of-origin can be inferred using distant relatives (Hofmeister et al., 2022). Another contributing factor to the implicit assumption may be the presumed scarcity of documented POE in humans. Although POE is believed to be associated with a wide range of complex traits and diseases (Lawson et al., 2013), significant associations reported in GWAS setting are of limited quantity. For example, in a Hutterite study, Mozaffari et al. (2019) examined 21 phenotypes and produced 18 significant associations with POE, fewer than one association per phenotype.

Mouse studies demonstrated that, despite a limited number of imprinted genes, most phenotypes display POE, and loci that show POE were not enriched for known imprinted genes. Using reciprocal F1 crosses, Mott et al. (2014) showed that non-imprinted genes can generate POE by interactions with imprinted loci. Focusing on metabolic traits and using a mouse population at different levels of intercrossing, Macias-Velasco et al. (2022) identified

a network comprised of three imprinted and six non-imprinted genes that show POE. In humans, however, loci with POE in relation to phenotypes are either near imprinted regions by design (Kong et al., 2009; Hanson et al., 2013), or near regions with characteristic of imprinting from a genome-wide scan (Mozaffari et al., 2019). To what extend POE contributes to phenotypic variation in humans and how widespread the loci with POE in the human genome remains elusive.

Using whole genome sequencing and RNA sequencing data from the Framingham Heart Study, and taking advantage of its relative abundance of trios, we set to investigate to what extent and degree gene expressions is affected by parent-of-origin. There exist multiple studies on POE on gene expression (Zhabotynsky et al., 2019; Jadhav et al., 2019; Deng et al., 2020) or POE of methylation and gene expression jointly (Zink et al., 2018), but these studies are either underpowered due to small sizes or focus on analysis of allele specific gene expression instead of eQTL. A systematic study of POE eQTL is lacking and our study aims to fill this void.

2 Results

2.1 Overview of data processing

We identified 1477 trios with whole genome sequencing (WGS) data and children within the trios having RNAseq data. After routine QC of genotype data, for each trio, we first phased the child based on rules of Mendelian inheritance, which resolved all non triple heterozygous genotypes. The genotype data has negligible Mendelian error rate, and see details in Methods on how to handle them.

We then marked the remaining triple heterozygous genotypes as missing, which produced phased paternal and maternal haplotypes for each child in the trio. Next we imputed missing alleles into their paternal and maternal haplotype background, and the parental haplotype with a higher imputed allele dosage was assigned the reference allele, the other haplotype was assigned the alternative allele. Simulation studies showed this mask and impute phasing to be highly accurate (Supplementary Table S1). Thus at each SNP, we have a genotype vector, a paternal vector, and a maternal vector.

To process RNAseq data, we started from Transcript Per Million (TPM) values, and selected $16,824$ genes that have $< 5\%$ missing values (i.e., 0 TMP). We then corrected for the GC content bias using an approach that is similar to the one used by EDAseq (Risso et al., 2011). The corrected gene expression values were then quantile normalized.

2.2 Bayes factors as evidence for association

For genetic association we used a linear mixed model, fitted by a novel and efficient method that is designed for analyzing multiomics datasets (Guan and Levy, 2024a). The genetic relatedness matrix (twice of kinship matrix) used for the linear mixed model was estimated by Kindred (Guan and Levy, 2024c). The covariates included age, sex, body mass index, and white blood cell composition. This study focuses on genetic association between gene expression and genetic variants in cis, defined as within 1Mb from transcription start site.

We elected to use Bayes factors (Stephens and Balding, 2009) instead of p-values as the evidence of association for the following considerations. First, at many SNPs, paternal and maternal alleles have different allele frequencies, presumably due to sampling variation. P-values are not comparable unless they have the same power, while the same power requires equal allele frequencies. On the other hand Bayes factors are comparable because they take into account power through prior specifications. Second, Bayes factors panelize genetic variants with small allele frequencies to reduce false positives (Guan and Stephens, 2011; Zhou and Guan, 2018). This is useful as we observed extremely low paternal or maternal allele frequencies. Third, Bayes factors are also convenient in comparing different alternative models, such as paternal effects versus genotype effect, a setting that is difficult for p-values.

Gene	TSS	SNP	$lgBF_g$	$lgBF_1$	$lgBF_0$	$lgBF_j$	β_1	β_0	P
MEG3	100.78	rs12881545	35.7	-0.7	86.3	85.7	1.0	-21.2	54.0
PPIEL	39.53	rs79473113	47.6	84.6	0.5	85.8	-20.5	-2.7	38.3
NDN	23.69	rs3743340	36.5	68.5	-0.2	67.7	18.6	0.8	33.2
PEG10	94.66	rs7801134	2.6	13.3	-0.1	13.6	-8.2	2.4	12.6
ZNF331	53.52	rs1284523	6.3	-0.9	15.9	14.9	-0.5	8.8	10.3
ZNF890P	5.12	rs112844843	125.8	94.7	31.1	133.9	23.2	13.6	10.1
FAM50B	3.85	rs111515624	3.5	11.5	-0.5	10.9	-7.4	0.6	8.6
NECAB3	33.66	rs2626556	-0.7	1.2	4.6	5.5	-2.9	4.9	7.8
SNURF	24.95	rs4906936	5.1	12.1	-0.9	11.2	7.7	-0.1	7.6
LSM7	2.32	rs393651	-1.0	3.5	1.6	4.9	4.4	-3.3	7.5

Table 1: Example of eQTL statistics. $lgBF_x$ columns are $\log_{10} BF_x$. Column β_j's are effect estimates in joint analysis used to compute BF_j. P column contains $-\log_{10}$ p-value with the null hypothesis $\beta_0 = \beta_1$ and alternative hypothesis $\beta_0 \neq \beta_1$. Transcript start site (TSS) is in Mb, and the coordinate is from HG38. Bona fide imprinted genes are colored in blue.

For each gene and its cis-SNP, we computed four Bayes factors: genotype Bayes factor (BF_g), paternal Bayes factor (BF_1), maternal Bayes factor (BF_0), and joint Bayes factors

(BF_j). When computing joint Bayes factor, we also computed paternal and maternal effect sizes and attached a p-value to test whether they differ significantly. Examples of test statistics are provided in Table 1. BF_g, BF_0 and BF_1 are one degree of freedom tests, but BF_j is a two degree of freedom test. Thus to achieve $BF_j > BF_g$, the paternal and maternal effect sizes have to differ significantly to compensate for the extra degree of freedom. Examples in Table 1 are chosen because their paternal and maternal effects differ significantly, and we have $BF_j >> BF_g$. Another useful observation is that BF_1 is correlated with $|\beta_1|$, and BF_0 is correlated with $|\beta_0|$. But BF_1 and BF_0 are oblivious to the signs of β_1 and β_0, and signs are informative and interesting in our context.

We used $\log_{10} BF = 4$ as the significance threshold. If the prior odds for a cis-eQTL is 1 out of 1000, this threshold gives the posterior probability of association (PPA) of 0.91. If the prior odds for a cis-eQTL is 1 out of 100, this threshold gives the PPA of 0.99. $1-PPA$ is the Bayesian counterpart of the local false discovery rate (c.f. Soloff et al., 2024).

2.3 Overview from sentinel eQTL of joint analysis

Table 1 contains 10 genes and test statistics of their sentinel eQTL. A sentinel eQTL for a gene is defined as the eQTL with largest BF_j for that gene. A gene with significant eQTL is called eGene. These sentinel eQTL are either $BF_1 >> BF_0 \approx 1$ such as *NDN*, or $BF_0 >> BF_1 \approx 1$ such as *MEG3*, or $BF_j >> BF_g \approx 1$ such as *NECAB3*. In Table 1 the sentinel eQTL were ordered according to column P, which measures the significance of differences between paternal and maternal effects. Reassuringly, six out of top ten sentinel eQTL with most significant P are from bona fide imprinted genes (colored in blue) according to `http://geneimprint.com` .

Figure 1 plots paternal effect vs maternal effect for all sentinel eQTL whose $\log_{10} BF_j > 3$. The color of dots corresponding to degree of difference between paternal and maternal effects. The marked sentinel eQTL include known imprinted genes such as maternally expressed *ZNF331* and *MEG3*, and paternally expressed *FAM50B, GNAS, SGCE, NDN, SNURF, SNRPN,* and *PEG10*. Imprinted genes are located along the x and y-axes. At least two genes appear to be novel imprinted genes: *PPIEL* and *CCR9*. Gene *ZNF890P* provides an example of paternal and maternal effects being in the same direction, but their sizes differ significantly. Finally, *NECAB3* and *LSM7* are two examples of paternal and maternal effects that are in opposite directions.

This approach based on the sentinel eQTL suggests that 1) the combined effect of paternal and maternal allele can be across all 360 degrees; 2) when paternal and maternal

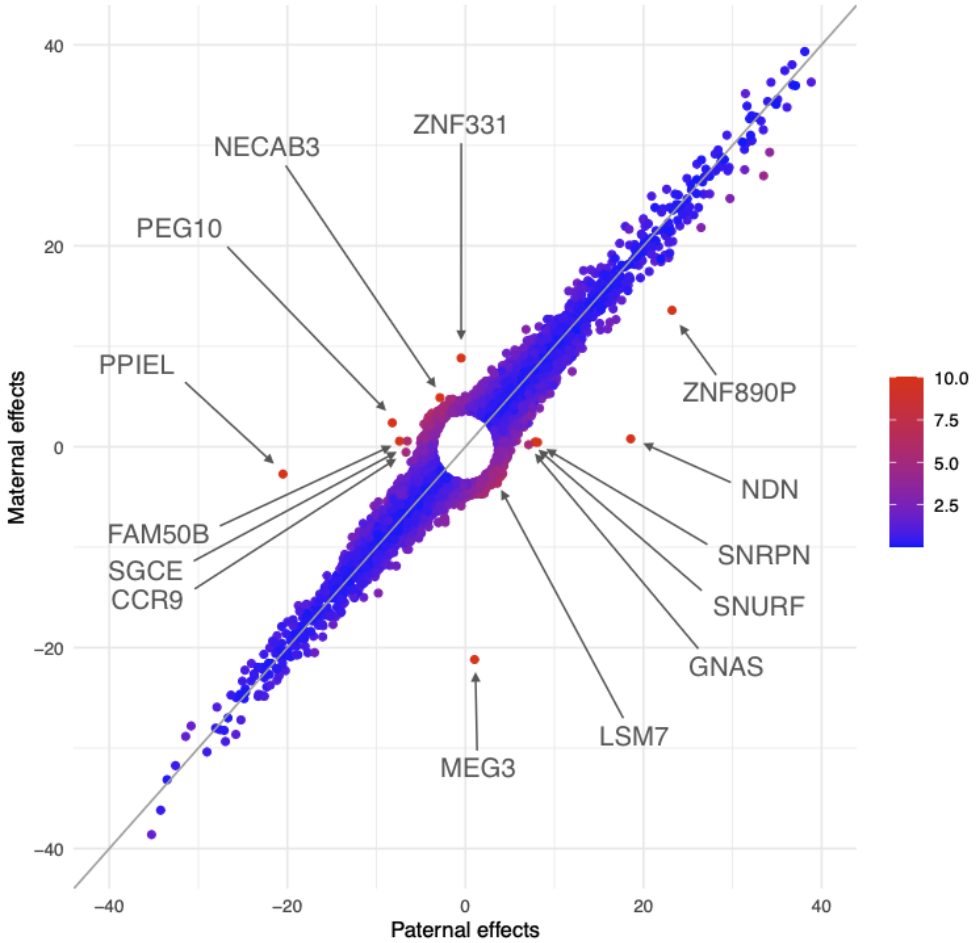

Figure 1: Comparison of paternal effects and maternal effects of sentinel eQTL in joint analysis. Each point is the most prominent eQTL with $\log_{10} BF_j > 3$ of gene. The paternal effect is on x-axis and maternal effect y-axis. The coloring reflects significance of differentials between paternal effects and maternal effects, with red more significant than blue.

effects are in opposite directions, both effects sizes are modest compared to when they are in the same direction; and 3) there are quite a few sentinel eQTL whose paternal and maternal effects show opposite directions.

2.4 *NECAB3*: opposite paternal and maternal effects

The existence of eGenes whose eQTL show opposite paternal and maternal effects is intriguing. We focused on the example of *NECAB3* to examine further details. Figure 2 left panel shows that all eQTL of *NECAB3* have opposite paternal and maternal effects. Consequently, the joint Bayes factors BF_j's are much larger than their corresponding genotype Bayes factors BF_g's (right bottom).

We selected as an example SNP rs4911348, which is different from the sentinel eQTL shown in Figure 1, and looked into details of its association with the gene expression of *NECAB3* (phenotype). Three panels of boxplots (right top) show that genotype has no association with the phenotype, and both paternal alleles and maternal alleles are associated with the phenotype, but in opposite directions. Opposite paternal and maternal effects were also observed in Hutterite POE study, where ten associations of opposite effects across nine traits were reported (Mozaffari et al., 2019).

2.5 POE eQTL are abundant

Next we looked beyond the sentinel eQTL and examined all significant eQTL. We noted that there are $15,893$ eQTL from $14,733$ SNPs and $1,824$ eGenes that are insignificant for the SNP test, but significant for the paternal, or maternal, or joint test. That is, $\log_{10} BF_g < 4$, but either $\log_{10} BF_1 > 4$, or $\log_{10} BF_0 > 4$, or $\log_{10} BF_j > 4$. In other words, these eQTL cannot be detected without analyzing POE, which implies that POE can contribute to recover missing heritability.

With reference Figure 1, we proposed criteria to define subsets of eQTL. The first subset eQTL locate along the x-axis, mimicking sentinel eQTL of *NDN* and *FAM50B*; the second subset locate along the y-axis, mimicking sentinel eQTL of *ZNF331* and *MEG3*; the third subset locate along the secondary diagonal, mimicking sentinel eQTL of *NECAB3* and *LSM7*; and the fourth subset follows (blues dots) along the diagonal line. Note Figure 1 only involves BF_j, our criteria to define gene sets also takes into account of BF_1, BF_0, and BF_g (details in Methods).

The first set of eQTL have significant paternal effects but insignificant maternal effects. We identified $15,576$ such *paternal eQTL* (set S_P) from $14,372$ SNPs associated with $1,188$

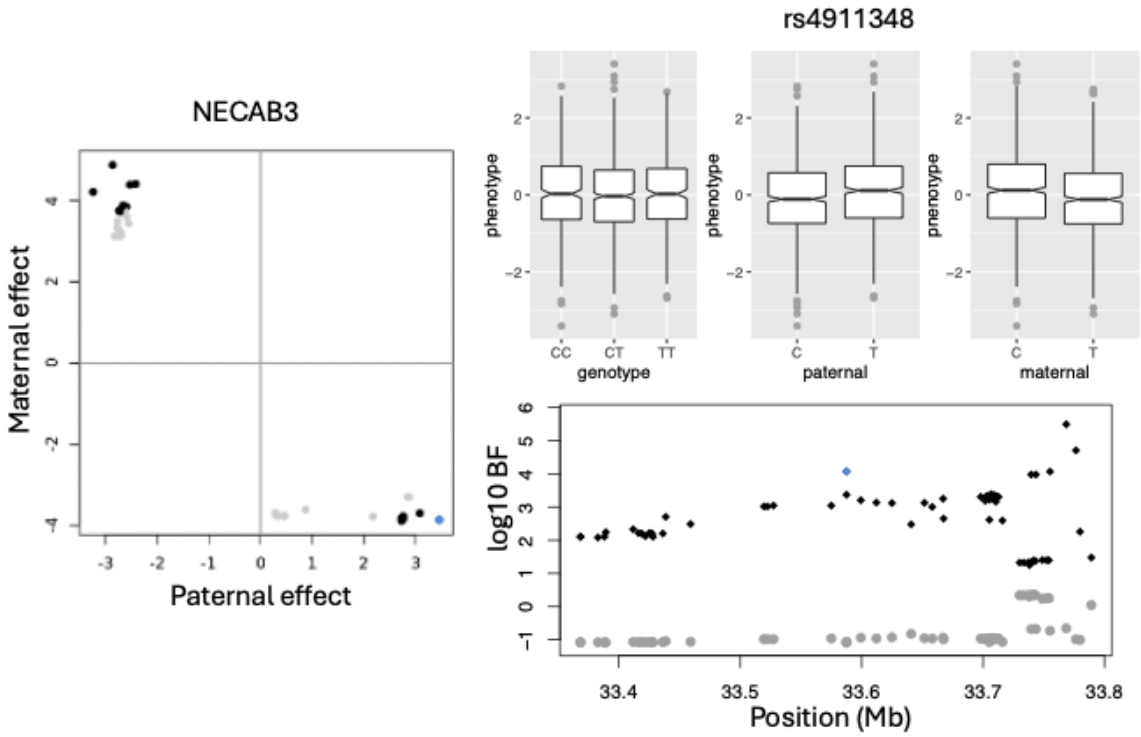

Figure 2: An example of gene that whose paternal and maternal eQTL have opposite effect sizes. On the left is paternal vs maternal effect sizes normalized by their corresponding standard deviation. The effect sizes were estimated by joint analysis. The non-significant eQTL are colored in gray and significant eQTL colored in black. An eQTL marked in blue was chosen to show boxplots with gene expression (phenotype) of NECAB3. The three panels of boxplots are for genotypes, paternal alleles, and maternal alleles. For this SNP rs4911348, the genotypes has no association with phenotype, but both paternal alleles and maternal alleles are associated with the phenotype, and the paternal and maternal effects are in opposite direction. Right panel bottom show Manhattan plot of cis-eQTL of the gene NECAB3. The genotype Bayes factor were colored in gray and joint test Bayes factor were colored in black. SNP rs4911348 was highlighted in blue.

eGenes. The second set of eQTL have significant maternal effects but insignificant paternal effects. We identified $14,783$ such *maternal eQTL* (set S_M) from $13,293$ SNPs associated with $1,209$ eGenes. We refer to these paternal and maternal eQTL as *imprinting eQTL*. The third set of eQTL have opposite paternal and maternal effects, such that joint Bayes factors are much larger than genotype Bayes factor $BF_j >> BF_g$. We identified 688 such *opposing eQTL* (set S_O) from 485 SNPs that are associated with 180 eGenes. Imprinting eQTL and opposing eQTL are referred to as *POE eQTL*.

The fourth set of eQTL require that both paternal and maternal alleles alone are associated with an expression phenotype, and the two effects are in the same direction. For these eQTL $BF_g > BF_j$ because similar effect sizes favor one degree of freedom test. We identified $884,119$ such *genotype eQTL* from $577,701$ SNPs and 4940 eGenes. The percent of eGenes with genotype eQTL is on par, but smaller than GTEx study (Consortium et al., 2020), presumably because $\log_{10} BF_g > 4$ is a more stringent threshold for eQTL analysis.

2.6 Paternal eQTL are enriched with GWAS hits

Using ANNOVAR (Wang et al., 2010), we annotated all four sets of eQTL (Table 2). We made the following observations: 1) The median distance to transcription start site (TSS) for opposing eQTL is 673.2Kb, much larger than other types of eQTL, and majority of opposing eQTL S_O are located in introns of nearby genes. 2) The percent of GWAS hits for SNPs in S_O is 0.022, significantly lower than that of SNPs in genotype eQTL S_G's 0.054 (test for proportion $P = 0.001$). This is reasonable as GWAS mainly use genotype test and are agnostic to opposing eQTL. 3) The precent of GWAS hits for SNPs in paternal eQTL S_P is higher than that of SNPs in S_G (test for proportion $P = 9 \times 10^{-11}$), and the percent of GWAS hits among SNPs in maternal eQTL S_M has no significant difference to that of SNPs in S_G (test for proportion $P = 0.051$).

Number of eQTL per SNP (r) measures degree of pleiotropy of a set of SNPs. SNPs in S_P and SNPs in S_M have similar pleiotropy $r \approx 1.1$, SNPs in S_G have the largest $r = 1.53$, and SNPs in S_O display $r = 1.41$. In other words, SNPs are more exclusive for imprinting eQTL, less exclusive for genotype eQTL, and somewhere in between for opposing eQTL. Table 2 (bottom) compares how the degree of pleiotropy affects their annotation for SNPs in S_G, where set P_k contains SNPs in S_G that are eQTL of at least k eGenes. The most interesting observation is that the percent of GWAS hits increases with rising SNP pleiotropy. This feature is only observed in genotype eQTL.

Set	eQTL	eGene	Len	D_{TSS}	Genic	Intron	Exon	ncRNA	SNP	GWAS
S_O	688	180	28.1	**673.2**	0.260	**0.663**	0.006	0.026	485	**0.023**
S_P	15,576	1188	23.3	122.9	0.362	0.490	0.016	0.080	14,372	**0.066**
S_M	14,783	1209	23.8	122.5	0.365	0.487	0.016	0.076	13,293	0.050
S_G	884,119	4940	24.6	84.1	0.383	0.449	0.017	0.091	577,701	0.054
P_1	884,119	4940	24.6	84.1	0.383	0.449	0.017	0.091	577,701	0.054
P_2	471,354	2855	17.7	96.8	0.400	0.403	0.019	0.112	164,936	0.071
P_3	269,546	1536	13.8	114.4	0.440	0.349	0.020	0.129	64,032	0.088
P_4	172,871	858	11.0	134.2	0.488	0.295	0.017	0.143	31,807	0.097
P_5	126,107	510	8.6	143.8	0.508	0.284	0.016	0.140	20,116	0.110
P_6	94,367	383	7.7	146.7	0.520	0.284	0.015	0.135	13,768	0.117
P_7	46,925	272	6.1	131.8	0.593	0.234	0.015	0.106	5,861	0.133
P_8	26,695	178	3.7	116.6	0.619	0.221	0.010	0.099	2,971	0.144
P_9	12,783	124	2.3	121.8	0.483	0.355	0.011	0.111	1,232	0.198

Table 2: Top tabular: Annotation of eQTL. S_O is a set of opposing eQTL, S_P is a set of paternal eQTL, S_M is a set of maternal eQTL, and S_G is a set of genotype eQTL. The column eQTL contain counts of eQTL in each set, and column SNP contains counts of distinct SNPs of eQTL in the set. The column *eGene* is the number of genes associated with eQTL in the set. The column of Len is the median length in Kb of the eGenes. (The pattern is the same with the mean length.) The column D_{TSS} contains median distance in Kb to transcription start site. The column GWAS contains percent of GWAS hits among SNPs, with GWAS p-value threshold of 5×10^{-8}. Bottom tabular: A SNP set P_k contain SNPs in S_G that are eQTL of at least k eGenes. (Note $P_1 = S_G$.) Percent of GWAS hits for SNP set P_k increases with k.

2.7 POE eGenes are abundant

There are $1,188$ eGenes harboring paternal eQTL and $1,209$ eGenes harboring maternal eQTL. Their union contains $2,139$ eGenes harbor imprinting eQTL, and their intersection contains 258 eGenes harboring both paternal and maternal eQTL. There are $4,940$ eGenes harboring genotype eQTL, and $1,867$ eGenes harboring both genotype eQTL and imprinting eQTL. Supplementary Figure S1 showed an example of an eGene harboring both maternal and paternal eQTL: *GZMH*, and an example of an eGene harboring both genotype eQTL and imprinting eQTL: *DSE*. The most prominent feature is that paternal eQTL and maternal eQTL cluster in separate genomic regions, and genotype eQTL and imprinting eQTL also cluster in separate genomic region.

2.8 eGenes that harbor exclusively imprinting eQTL

There are 129 eGenes (set G_1) that harbor exclusively paternal eQTL, and 139 eGenes (set G_0) that harbor exclusively maternal eQTL. Naturally, G_1 contain many bona fide imprinted genes with paternal expression such as *FAM50B, NDN, SGCE, SNRPN, SNURF* and *PEG10*. Remaining genes in G_1 are candidates for imprinted genes with paternal expression, for example *CCR9* and *NCOA2*. G_0 contain bona fide imprinted genes with maternal expression such as *GRB10, ZNF331* and *MEG3*. The remaining in G_0 are candidates for imprinted genes with maternal expression, for example *COA8* and *ZNF888*.

The Supplementary Material provides a full list of the candidate imprinted genes from the eQTL analysis. Note however, that not all imprinted genes have eQTL, therefore this list is incomplete. Moreover, a gene on the list is considered behaving like an imprinted gene, but may not be a bona fide imprinted gene. For example, a gene that has no cis-eQTL on its own, may "acquires" a cis-eQTL through gene-gene interaction with a putative imprinted gene such that the cis-eQTL is in fact its trans-eQTL.

We used G_1 and G_0 as proxies for paternal and maternal imprinted genes, and used cytoBands these genes occupies as proxies for imprinted regions to examine to what extend paternal and maternal eQTL locate in the same cytoBands as these proxy imprinted genes. There are 811 autosome cytoBands in humans, 765 of them larger than 1 Mb, and the median size is 3.2 Mb (Cheung et al., 2001). The 129 proxy paternal imprinted genes occupy 89 cytoBands (set C_{P_I}), while $15,576$ paternal eQTL occupy 445 cytoBands (set C_{P_e}), among them $9,947$ (or 64%) were not in the cytoBands set C_{P_I}. On the other hand, the 139 proxy maternal imprinted genes occupy 96 cytoBands (set C_{M_I}), and $14,783$ maternal eQTL occupy 465 cytoBands (set C_{M_e}), among them $7,987$ (or 54%) were not in

the cytoBand set C_{M_I}. Therefore, a majority of imprinting eQTL are not in the vicinity of the proxy imprinted regions. Interestingly, C_{P_I} and C_{M_I} share only 14 cytoBands, 8% of their union of 171 cytoBands. As a comparison, C_{P_e} and C_{M_e} share 349 cytoBands, 62% of their union of 561 cytoBands.

2.9 eGenes that harbor exclusively opposing eQTL

There are 99 eGenes (set G_2) that harbor only opposing eQTL (Supplementary). Among them 58 encode phosphoproteins (Supplementary), a significant enrichment (FDR $= 5.5 \times 10^{-4}$) according to DAVID, a web server for functional enrichment analysis (Sherman et al., 2022). A phosphoprotein is a protein that is posttranslationally modified by the attachment of either a single phosphate group, or a complex molecule such as 5'-phospho-DNA, through a phosphate group. Because phosphorylation often serves as an on/off switch, targeting phosphoproteins or the enzymes regulating them can restore normal cell signaling in diseases. In addition, drugs can modulate the functions of these proteins, either inhibiting or enhancing their activities. This makes phosphoprotein attractive drug targets. Indeed, according to a therapeutic target database (Zhou et al., 2024), among 58 phosphoproteins genes that harbor exclusively opposing eQTL, there are three successful drug targets *ITPR1, ITPR2*, and *REL*, and three additional clinical trial targets *CD46, NOTCH3,* and *XPO1*.

Other notable phosphoproteins genes include *CDKN2A*, *PHIP*, and *LEP*, among others. Where *CDKN2A* produces two major proteins p16(*INK4*), which is a cyclin-depnednet kinase inhibitor, and p14(*ARF*), which binds the p53-stabilizing protein MDM2 (Serrano et al., 2000); *PHIP* is associated with Chung-Jansen syndrome, featuring behavioral problems, intellectual disability, obesity, and dysmorphia (Webster et al., 2016; Jansen et al., 2018); and *LEP* encodes leptin, a protein that plays a critical role in the regulation of body weight. Leptin is secreted by white adipocytes, and it binds to the leptin receptor in the brain, which in turn inhibits appetite and promotes energy expenditure (Maffei et al., 1995; Weigle et al., 1997).

2.10 Paternal eGenes are enriched with drug targets

We used a therapeutic target database (Zhou et al., 2024) to examine the enrichment with drug targets of the four sets of eGenes, grouped by their eQTL. Table 3 contains the enrichment of the four gene sets in either successful targets, clinical trial targets, or combined targets. The significant enrichment that (following Bonferroni correction of 12

tests) is highlighted. Both paternal eGenes and genotype eGenes are enriched in clinical trial targets and combined targets, but paternal eGenes are significantly more enriched than genotype eGenes (one-sided test for proportion $P = 0.035$). The enrichment of drug targets for paternal eGenes echoes the enrichment of GWAS hits for paternal eQTL (Table 2). This asymmetry between paternal and maternal eGenes perhaps finds its root in the conflicting interests of two sets of genes in relation to transfer of nutrients from the mother to her offspring (Moore and Haig, 1991). In maze, gene expression of the hybrid is regulated exclusively by the paternally transmitted alleles (Swanson-Wagner et al., 2009).

	Set	eGenes	nTarget	$\hat{f}$	f_0	$P(\hat{f} > f_o)$
Successful target	G_O	180	7	0.0389	0.0220	0.098
	G_P	1188	38	0.0320	0.0220	0.012
	G_M	1209	23	0.0190	0.0220	0.728
	G_G	4940	124	0.0251	0.0220	0.075
Clinical trial target	G_O	180	7	0.0389	0.0421	0.512
	G_P	1188	69	**0.0581**	0.0421	**0.004**
	G_M	1209	67	0.0554	0.0421	0.013
	G_G	4940	254	**0.0514**	0.0421	**0.0006**
Combined target	G_O	180	14	0.0778	0.0641	0.275
	G_P	1188	107	**0.0901**	0.0641	**0.0002**
	G_M	1209	90	0.0744	0.0641	0.079
	G_G	4940	374	**0.0757**	0.0641	**0.0005**

Table 3: Enrichment of drug target genes. G_O contains eQTL with both paternal and maternal effects but they are in opposite direction. G_P contains eQTL of paternal effects. G_M contains eQTL of maternal effects. G_G contain eQTL that have both paternal and maternal effects and they are in the same direction. Column nGene are sizes of gene sets, Column nTarget are counts of successful drug targets. The last column contains one-sided test-of-proportion p-values. The highlighted p-values are significant after Bonferroni correction of 12 tests.

3 Discussion

We performed parent-of-origin analysis of eQTL in whole blood gene expression using 1477 trios from the Framingham Heart Study. Some highlights of our study include: 1) There are more than 2,000 genes harboring POE eQTL and majority POE eQTL are not in the vicinity of known imprinted genes; 2) Among 180 genes harboring opposing eQTL, 99 harbor exclusively opposing eQTL, and 58 of the 99 are phosphoprotein coding genes,

reflecting significant enrichment; 3) Paternal eQTL are enriched with GWAS hits, and genes harboring paternal eQTL are enriched with drug targets.

Our study demonstrated abundance of POE in gene expression, illustrated additional complexity of gene expression regulation, and provided a resource that is complementary to existing resources such as GTEx. We conclude by revisiting two previous findings in light of our work, and advising caution on conducting Mendelian randomization using gene expression as the exposure.

3.1 T2D and maternal allele of rs2237892

SNP rs2237892 in the last intron of *KCNQ1* was found in association with type 2 diabetes (T2D) in a Japanese cohort (Yasuda et al., 2008), and the association was later confirmed in a Chinese cohort (Liu et al., 2009). Three other SNPs (rs2283228, rs2237895, and rs2237897) in the same last intron of *KCNQ1* were found in association with T2D in another Japanese cohort (Unoki et al., 2008). Both rs2237895 and rs2237897 were replicated in a Singaporean cohort, a Danish cohort (Unoki et al., 2008), and a Chinese cohort (Liu et al., 2009). In these studies, *KCNQ1* was identified as T2D candidate gene.

Based on these results from the genotype test, Kong et al. (2009) demonstrated that SNP rs2237892 was maternally associated with T2D in an Iceland cohort. This maternal association was replicated with a larger odds ratio in a Pima Amerindian cohort (Hanson et al., 2013). According to data from GTEx, none of these SNPs are eQTL of *KCNQ1* and *CDKN1C* in any tissue, consistent with the fact that these are imprinted genes and GTEx study is oblivious to parent of origin. Our eQTL analysis shows that SNPs rs2237892, rs2283228, and rs2237897 are maternal eQTL of *CDKN1C*, a down-stream neighbor of *KCNQ1*, and none of these SNPs are eQTL of *KCNQ1*.

We therefore suggest that *CDKN1C*, instead of *KCNQ1*, is a T2D candidate gene, for the following reasons: 1) These GWAS SNPs connect T2D and *CDKN1C* quantitatively, but connect T2D and *KCNQ1* only geographically; 2) A study suggests that *CDKN1C* mutations may represent a novel monogenic form of diabetes (Kerns et al., 2014); 3) A boy carrying a frameshift mutation in *CDKN1C* was diabetic from week 29 (Berland et al., 2022); and 4) Targeted demethylation at the CDKN1C/p57 locus induces human β cell replication, while the loss of insulin-secreting β cell is characteristic among T1D and T2D (Ou et al., 2019).

3.2 Polar overdominance and rs1802710

Overdominance is a pattern of inheritance such that both homozygous AA and BB have the same baseline trait value, while heterozygous AB has a higher trait value. Polar overdominance (Cockett et al., 1996) introduces asymmetry into overdominance to separates AB into two groups according to parent-of-origin of A (or B), such that AB with paternal A (or B) has a high trait value, while AB with maternal A (or B) has the baseline trait value.

Wermter et al. (2008) studied trios of extremely obese offspring and identified rs1802710 in exon 5 of *DLK1*, homologous to the ovine callipyge locus (Cockett et al., 1996), whose allelic transmission pattern was consistent with polar overdominance: frequent transmission of the paternal C allele to obese children, but the relative risk for carriers of the homozygous CC genotype was not increased compared to the reference TT genotype.

Our eQTL results show that *DLK1* has no significant eQTL, but the maternal C allele of rs1802710 reduces expression of *MEG3*, 53Kb downstream of *DLK1*. Both *MEG3* and *DLK1* are imprinted with *MEG3* maternally expressed and *DLK1* paternally expressed. The expression of *MEG3* was shown to be significantly higher in the obese group (Daneshmoghadam et al., 2021). Therefore the maternal C allele is associated with reduced risks of obesity, which balances out the increased risk conferred by paternal C allele, thus offering an explanation for the baseline risk of homozygous CC samples in (Wermter et al., 2008). Since rs1802710 is not an eQTL of any gene in any tissue according to GTEx data, our study is critical to link rs1802710 with *MEG3*, albeit not in the adipose tissue.

3.3 Mendelian randomization

Mendelian randomization (MR) refers to the random allocation of alleles at the time of gamete formation. Observational epidemiology studies use MR to infer the causal effect of an exposure on an phenotype (Smith and Ebrahim, 2003), as if the random allocation of alleles is comparable to a randomized clinical trial. An important assumption for MR, among many other important assumptions, is that the phenotype conferred by a specific genetic variant is homogeneous in the population, and exchangeable between paternally and maternally inherited alleles. POE is a direct violation of this assumption (Bochud et al., 2008).

There is a growing interest in using gene expression as an exposure to perform MR across various traits (Porcu et al., 2019; van der Graaf et al., 2020). If an exposure has eQTL with POE, ignoring the parent-of-origin will bias prediction of a subset of samples.

If the bias is severe it will change the ranking of the exposure. Consequently, it either compromises the power of the MR or leads to a false conclusion of a causal relationship between the exposure and the phenotype.

Our study demonstrated the abundance of POE in gene expressions, and therefore advised caution when conducting MR using gene expression as the exposure. Our suggestion is checking the list of POE SNPs and eGenes we provided in the Supplementary Material and excluding those that show POE towards the exposure. The same suggestion is also applicable to computing polygenic risk scores, and imputing gene expression such as PrediXcan (Gamazon et al., 2015) and a Bayesian method motivated by it (Qi et al., 2018).

4 Methods

4.1 Genotype data

We first used pedigree and availability of RNAseq data to narrow the samples down to 2955 individuals and identified all bi-allelic SNPs following a TOPMed protocol (Taliun et al., 2021). We then used Kindred to infer pairwise kinship for each chromosome, and computed mean and standard deviation (SD) of the kinship. A sample pair that is parent-offspring can be distinguished from full sibs by the SD (Supplementary Figure S2). Trios are consisted of sample A, B, and C such that $\phi(A, B) \approx 0$ and $\phi(A, C) \approx \phi(B, C) \approx 0.25$ where $\phi(\cdot, \cdot)$ is the kinship. In the end, we identified 1477 trios that have whole genome sequencing data and whose children have RNAseq data. In total, we had $7,752,281$ autosomal bi-allelic SNPs with minor allele frequency > 0.01 (not corrected for relatedness). Across all 1477 trios, the Mendelian error is negligible 4.76 per $100,000$ SNPs per trio. To handle Mendelian errors, we assumed that the child was correctly genotyped, and errors occurred in the parents, but if the child is heterozygous, we assumed this to be triple heterozygous SNP. On average 1477 trios had 65 triple heterozygous genotypes per SNP. The triple heterozygous SNPs cannot be phased by Mendelian inheritance, but can be phased based on a linkage disequilibrium model (below).

4.2 Inference of parent of origin

For each trio, we first phased all markers that are not triple heterozygous using rules of Mendelian inheritance, where for the child in each trio we obtained paternal and maternal haplotypes with triple heterozygous markers whose phase were unresolved. We marked

those triple heterozygous markers as missing. For markers that showed Mendelian incompatibility, we assumed that the child's genotype was correct and marked it as missing if it was heterozygous. Using a hidden Markov model that was designed to model haplotype variation (Guan, 2014), we imputed missing markers in each haplotypes to obtain dosage estimates (between 0 and 1). Then for each marker that was marked as missing, we compared imputed dosages between paternal and maternal haplotypes, and assigned the large dosage as 1 and small dosage as 0. We thus obtained paternal and maternal haplotypes with no missing data. We simulated trios using haplotypes from the 1000 Genomes project (Auton et al., 2015), and investigated the accuracy of the phasing by the mask and imputation approach described above. The phasing was highly accurate, the error rate was less than 1 out of 200 triple heterozygous SNPs (Supplementary Table S1).

4.3 RNAseq data

The RNAseq data of 1477 children was obtained in three batches with 1381 in batch 1; 14 in batch 2, and 82 in batch 3. Our primary goal was to correct for GC content bias. To this end, we first removed genes whose proportion of 0 TPM values was greater than 5%, which retained $16,969$ genes out of total 58103, a majority of which are coding genes. Among those, $16,824$ genes have GC content. To correct for GC content bias, we took an approach used by EDAseq to regress out the GC content from $\log(1+TMP)$, but instead of using local linear regression as documented in EDAseq, we used local quadratic regression implemented in loess in R. Specifically, let $y = \log(1 + TMP)$ be a gene expression and g be the GC content, we fit $ly = loess(y, g)$ and compute $\hat{y} = median(y) + ly\$residuals$. We then quantile normalized $\hat{y}$ by $qqnorm(\hat{y}, plot.it = F)\$y$ (in R) separately for each gene.

4.4 eQTL analysis

We used IDUL to perform eQTL analysis. IDUL fits linear mixed models to achieve exact optimal. It was specifically developed to analyze multi-omics data to achieve high efficiency by reusing the intermediate computations (Guan and Levy, 2024a). Previous analysis showed that for the Framingham Heart Study participants, the genomic inflation was well controlled using a linear mixed model with kinship matrix computed by Kindred (Guan and Levy, 2024c). In this study, we also controlled age, sex, BMI, and white blood cell compositions in peripheral blood. We now briefly describe how Bayes factors

were computed. Consider a model

$$\begin{aligned} \mathbf{y} &= \mathbf{Wa} + \mathbf{x}\beta + \mathbf{Zu} + \mathbf{e} \\ \mathbf{u} &\sim \mathrm{MVN}_n(0, \tau^{-1}\ \eta\mathbf{K}) \\ \mathbf{e} &\sim \mathrm{MVN}_n(0, \tau^{-1}\ \mathbf{I}_n) \end{aligned} \tag{1}$$

where $\mathbf{W}$ contains conventional covariates such as age and sex, including a column of 1, $\mathbf{x}$ contains genetic variant(s) to be tested for association, $\mathbf{u}$ is the random effect with $\mathbf{Z}$ as its loading matrix and twice of kinship $\mathbf{K}$ as its covariance (both $\mathbf{Z}$ and $\mathbf{K}$ are known), MVN_n denotes an n-dimensional multivariate normal distribution, $\mathbf{I}_n$ is n-dimensional identity matrix. Denote $\mathbf{X} = (\mathbf{W}, \mathbf{x})$ and $\mathbf{b} = (\mathbf{a}, \beta)$, then $\mathbf{Xb}$ is the fixed effect, and we assume $\mathbf{X}$ has a full rank c. In genetic association studies, the random effect $\mathbf{Zu}$ is a nuisance term that absorbs part of the phenotype $\mathbf{y}$ that is attributable to population stratification and relatedness. The maximum likelihood estimate (MLE) of η can be efficiently obtained (Guan and Levy, 2024a), plug $\hat{\eta}$ back into (1), and specify the following conjugate prior

$$\begin{aligned} \mathbf{a} &\sim N(0, \tau^{-1}V_{\mathbf{a}}) \\ \beta &\sim N(0, \tau^{-1}\sigma^2) \\ \tau &\sim \Gamma(\kappa_1/2, \kappa_2/2) \end{aligned} \tag{2}$$

and let $V_{\mathbf{a}} \to \infty$, $\kappa_1 \to 0$ and $\kappa_2 \to 0$, Bayes factor can be evaluated efficiently in a closed form. In this study we used $\sigma = 0.5$, following prior work of Bayes factors for linear models (Servin and Stephens, 2007). Complete details on computing Bayes factors for linear mixed model can be found in (Guan and Levy, 2024b).

4.5 P-value for difference between paternal maternal effects

For joint analysis, we have

$$\mathbf{y} = \mathbf{Wa} + \mathbf{x}_1\beta_1 + \mathbf{x}_2\beta_2 + \mathbf{Zu} + \mathbf{e} \tag{3}$$

with the prior for β_1 and β_2 as

$$\begin{aligned} \beta_1 &\sim N(0, \tau^{-1}\sigma^2) \\ \beta_2 &\sim N(0, \tau^{-1}\sigma^2) \end{aligned} \tag{4}$$

The Bayes factor can be similarly computed in a closed form. We obtain the posterior estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$, their variances s_1^2 and s_2^2 and covariance s_{12}. To test null hypothesis

$\hat{\beta}_1 = \hat{\beta}_2$, we compute test statistics $t = \frac{(\hat{\beta}_1 - \hat{\beta}_2)^2}{s_1^2 + s_2^2 - 2s_{12}}$. Since under the null t following χ^2 distribution with 1 degree of freedom, we can compute a p-value.

4.6 Threshold for eQTL sets and gene sets

For paternal eQTL set S_P, we required $\log_{10} BF_1 > 4$ and $\log_{10} BF_0 < \theta$; For maternal eQTL set S_M, we required $\log_{10} BF_0 > 4$ and $\log_{10} BF_1 < \theta$; For opposing eQTL set S_O, we required $\log_{10} BF_j - \log_{10} BF_g > 4$ and $\log_{10} BF_1 > \theta$ and $\log_{10} BF_0 > \theta$, and $\beta_1 * \beta_0 < 0$; For genotype eQTL set S_G, we required $\log_{10} BF_g > 4$ and $\log_{10} BF_1 > \theta$ and $\log_{10} BF_0 > \theta$, and $\beta_1 * \beta_0 > 0$. It's easy to see for a particular type of eQTL, different θ produced nested eQTL sets. We tried $\theta = 0, \log_{10} 2$, and $\log_{10} 3$ to obtain eQTL sets (Supplementary Table S2), from which we obtained gene sets harboring those eQTL G_P, G_M, G_A and G_G, from which we obtained genes harboring exclusively paternal eQTL $G_1 = G_P \setminus (G_M \cup G_A \cup G_G)$, genes harboring exclusive maternal eQTL $G_0 = G_M \setminus (G_P \cup G_A \cup G_G)$, and genes harboring exclusively opposing eQTL $G_0 = G_A \setminus (G_P \cup G_M \cup G_G)$. We compared known imprinted genes with paternal expression with G_1, and known imprinted genes with maternal expression with G_0. $\theta = \log_{10} 2$ produced minimum G_1 and G_0 that contain known imprinted genes.

References

Auton, A., G. R. Abecasis, D. M. Altshuler, R. M. Durbin, D. R. Bentley, A. Chakravarti, A. G. Clark, and et al. (2015). A global reference for human genetic variation. *Nature 526*(7571), 68–74.

Berland, S., B. I. Haukanes, P. B. Juliusson, and G. Houge (2022, February). Deep exploration of a CDKN1C mutation causing a mixture of Beckwith-Wiedemann and IMAGe syndromes revealed a novel transcript associated with developmental delay. *J. Med. Genet. 59*(2), 155–164.

Bochud, M., A. Chiolero, R. C. Elston, and F. Paccaud (2008, April). A cautionary note on the use of mendelian randomization to infer causation in observational epidemiology. *Int. J. Epidemiol. 37*(2), 414–6; author reply 416–7.

Cheung, V. G., N. Nowak, W. Jang, I. R. Kirsch, S. Zhao, X. N. Chen, T. S. Furey, U. J. Kim, W. L. Kuo, M. Olivier, J. Conroy, A. Kasprzyk, H. Massa, R. Yonescu, S. Sait, C. Thoreen, A. Snijders, E. Lemyre, J. A. Bailey, A. Bruzel, W. D. Burrill, S. M. Clegg, S. Collins, P. Dhami, C. Friedman, C. S. Han, S. Herrick, J. Lee, A. H. Ligon, S. Lowry, M. Morley, S. Narasimhan, K. Osoegawa, Z. Peng, I. Plajzer-Frick, B. J. Quade, D. Scott, K. Sirotkin, A. A. Thorpe, J. W. Gray, J. Hudson, D. Pinkel, T. Ried, L. Rowen, G. L.

Shen-Ong, R. L. Strausberg, E. Birney, D. F. Callen, J. F. Cheng, D. R. Cox, N. A. Doggett, N. P. Carter, E. E. Eichler, D. Haussler, J. R. Korenberg, C. C. Morton, D. Albertson, G. Schuler, P. J. de Jong, B. J. Trask, and BAC Resource Consortium (2001, February). Integration of cytogenetic landmarks into the draft sequence of the human genome. *Nature 409*(6822), 953–958.

Cockett, N. E., S. P. Jackson, T. L. Shay, F. Farnir, S. Berghmans, G. D. Snowder, D. M. Nielsen, and M. Georges (1996, July). Polar overdominance at the ovine callipyge locus. *Science 273*(5272), 236–238.

Consortium, T. G., F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. de Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, M. J. Gloudemans, R. Guigó, A. R. Hame, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, H. K. Im, B. Jo, S. Kasela, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, Y. Liang, S. Mangul, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, A. B. Nobel, M. Oliva, Y. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, P. A. Branton, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P. Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, F. Jasmine, L. Jiang, R. Kaul, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, B. L. Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, H. Tang, M. Wang, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, and S. Volpi (2020). The gtex consortium atlas of genetic regulatory effects across human tissues. *Science 369*(6509), 1318–1330.

Daneshmoghadam, J., A. Omidifar, N. Akbari Dilmaghani, Z. Karimi, S. Emamgholipour, and M. Shanaki (2021, May). The gene expression of long non-coding RNAs (lncRNAs): MEG3 and H19 in adipose tissues from obese women and its association with insulin resistance and obesity indices. *J. Clin. Lab. Anal. 35*(5), e23741.

DeChiara, T. M., E. J. Robertson, and A. Efstratiadis (1991, February). Parental imprinting of the mouse insulin-like growth factor II gene. *Cell 64*(4), 849–859.

Deng, S., J. Hardin, C. I. Amos, and F. Xiao (2020, August). Joint modeling of eQTLs and parent-of-origin effects using an orthogonal framework with RNA-seq data. *Hum. Genet. 139*(8), 1107–1117.

Gamazon, E. R., H. E. Wheeler, K. P. Shah, S. V. Mozaffari, K. Aquino-Michaels, R. J. Carroll, A. E. Eyler, J. C. Denny, GTEx Consortium, D. L. Nicolae, N. J. Cox, and H. K. Im (2015, September). A gene-based association method for mapping traits using reference transcriptome data. *Nat. Genet. 47*(9), 1091–1098.

Garfield, A. S., M. Cowley, F. M. Smith, K. Moorwood, J. E. Stewart-Cox, K. Gilroy, S. Baker, J. Xia, J. W. Dalley, L. D. Hurst, L. S. Wilkinson, A. R. Isles, and A. Ward (2011, January). Distinct physiological and behavioural functions for parental alleles of imprinted grb10. *Nature 469*(7331), 534–538.

Guan, Y. (2014, March). Detecting structure of haplotypes and local ancestry. *Genetics 196*(3), 625–642.

Guan, Y. and D. Levy (2024a). Asymptotically exact fit for linear mixed model. *bioRxiv*.

Guan, Y. and D. Levy (2024b). Bayes factor for linear mixed model in genetic association studies. *bioRxiv*.

Guan, Y. and D. Levy (2024c, 02). Estimation of inbreeding and kinship coefficients via latent identity-by-descent states. *Bioinformatics 40*(2), btae082.

Guan, Y. and M. Stephens (2011). Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *The Annals of Applied Statistics 5*(3), 1780 – 1815.

Hager, R., J. M. Cheverud, and J. B. Wolf (2008, March). Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting. *Genetics 178*(3), 1755–1762.

Hanson, R. L., T. Guo, Y. L. Muller, J. Fleming, W. C. Knowler, S. Kobes, C. Bogardus, and L. J. Baier (2013, August). Strong parent-of-origin effects in the association of KCNQ1 variants with type 2 diabetes in american indians. *Diabetes 62*(8), 2984–2991.

Hofmeister, R. J., S. Rubinacci, D. M. Ribeiro, A. Buil, Z. Kutalik, and O. Delaneau (2022, November). Parent-of-Origin inference for biobanks. *Nat. Commun. 13*(1), 6668.

Jadhav, B., R. Monajemi, K. K. Gagalova, D. Ho, H. H. M. Draisma, M. A. van de Wiel, L. Franke, B. T. Heijmans, J. van Meurs, R. Jansen, GoNL Consortium, BIOS Consortium, P. A. C. 't Hoen, A. J. Sharp, and S. M. Kiełbasa (2019, June). RNA-Seq in 296 phased trios provides a high-resolution map of genomic imprinting. *BMC Biol. 17*(1), 50.

Jansen, S., A. Hoischen, B. P. Coe, G. L. Carvill, H. Van Esch, D. G. M. Bosch, U. A. Andersen, C. Baker, M. Bauters, R. A. Bernier, B. W. van Bon, H. L. Claahsen-van der Grinten, J. Gecz, C. Gilissen, L. Grillo, A. Hackett, T. Kleefstra, D. Koolen, M. Kvarnung, M. J. Larsen, C. Marcelis, F. McKenzie, M.-L. Monin, C. Nava, J. H. Schuurs-Hoeijmakers, R. Pfundt, M. Steehouwer, S. J. C. Stevens, C. T. Stumpel, F. Vansenne, M. Vinci, M. van de Vorst, P. d. Vries, K. Witherspoon, J. A. Veltman, H. G. Brunner, H. C. Mefford, C. Romano, L. E. L. M. Vissers, E. E. Eichler, and B. B. A. de Vries (2018, January). A genotype-first approach identifies an intellectual disability-overweight syndrome caused by PHIP haploinsufficiency. *Eur. J. Hum. Genet. 26*(1), 54–63.

Kannel, W. B., M. Feinleib, P. M. McNamara, R. J. Garrison, and W. P. Castelli (1979, September). An investigation of coronary heart disease in families. the framingham offspring study. *Am. J. Epidemiol. 110*(3), 281–290.

Kerns, S. L., J. Guevara-Aguirre, S. Andrew, J. Geng, C. Guevara, M. Guevara-Aguirre, M. Guo, C. Oddoux, Y. Shen, A. Zurita, R. G. Rosenfeld, H. Ostrer, V. Hwa, and A. Dauber (2014, October). A novel variant in CDKN1C is associated with intrauterine growth restriction, short stature, and early-adulthood-onset diabetes. *J. Clin. Endocrinol. Metab. 99*(10), E2117–22.

Kong, A., V. Steinthorsdottir, G. Masson, G. Thorleifsson, P. Sulem, S. Besenbacher, A. Jonasdottir, A. Sigurdsson, K. T. Kristinsson, A. Jonasdottir, M. L. Frigge, A. Gylfason, P. I. Olason, S. A. Gudjonsson, S. Sverrisson, S. N. Stacey, B. Sigurgeirsson, K. R. Benediktsdottir, H. Sigurdsson, T. Jonsson, R. Benediktsson, J. H. Olafsson, O. T. Johannsson, A. B. Hreidarsson, G. Sigurdsson, DIAGRAM Consortium, A. C. Ferguson-Smith, D. F. Gudbjartsson, U. Thorsteinsdottir, and K. Stefansson (2009, December). Parental origin of sequence variants associated with complex diseases. *Nature 462*(7275), 868–874.

Lawson, H. A., J. M. Cheverud, and J. B. Wolf (2013, September). Genomic imprinting and parent-of-origin effects on complex traits. *Nat. Rev. Genet. 14*(9), 609–617.

Liu, Y., D. Z. Zhou, D. Zhang, Z. Chen, T. Zhao, Z. Zhang, M. Ning, X. Hu, Y. F. Yang, Z. F. Zhang, L. Yu, L. He, and H. Xu (2009, July). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes in the population of mainland china. *Diabetologia 52*(7), 1315–1321.

Macias-Velasco, J. F., C. L. St Pierre, J. P. Wayhart, L. Yin, L. Spears, M. A. Miranda, C. Carson, K. Funai, J. M. Cheverud, C. F. Semenkovich, and H. A. Lawson (2022, March). Parent-of-origin effects propagate through networks to shape metabolic traits. *Elife 11*.

Maffei, M., H. Fei, G. H. Lee, C. Dani, P. Leroy, Y. Zhang, R. Proenca, R. Negrel, G. Ailhaud, and J. M. Friedman (1995, July). Increased expression in adipocytes of ob RNA in mice with lesions of the hypothalamus and with mutations at the db locus. *Proc. Natl. Acad. Sci. U. S. A. 92*(15), 6957–6960.

Moore, T. and D. Haig (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in Genetics 7*(2), 45–49.

Mott, R., W. Yuan, P. Kaisaki, X. Gan, J. Cleak, A. Edwards, A. Baud, and J. Flint (2014, January). The architecture of parent-of-origin effects in mice. *Cell 156*(1-2), 332–342.

Mozaffari, S. V., J. M. DeCara, S. J. Shah, C. Sidore, E. Fiorillo, F. Cucca, R. M. Lang, D. L. Nicolae, and C. Ober (2019, January). Parent-of-origin effects on quantitative phenotypes in a large hutterite pedigree. *Commun. Biol. 2*(1), 28.

Ou, K., M. Yu, N. G. Moss, Y. J. Wang, A. W. Wang, S. C. Nguyen, C. Jiang, E. Feleke, V. Kameswaran, E. F. Joyce, A. Naji, B. Glaser, D. Avrahami, and K. H. Kaestner (2019, 1). Targeted demethylation at the cdkn1c/p57 locus induces human beta cell replication. *The Journal of Clinical Investigation 129*(1), 209–214.

Porcu, E., S. Rüeger, K. Lepik, eQTLGen Consortium, BIOS Consortium, F. A. Santoni, A. Reymond, and Z. Kutalik (2019, July). Mendelian randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *Nat. Commun. 10*(1), 3300.

Qi, Z., S. Xie, R. Chen, H. A. Aisa, G. C. Hon, and Y. Guan (2018). Tissue-specific gene expression prediction associates vitiligo with suox through an active enhancer. *bioRxiv*.

Reik, W. and J. Walter (2001, January). Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet. 2*(1), 21–32.

Risso, D., K. Schwartz, G. Sherlock, and S. Dudoit (2011, December). GC-content normalization for RNA-Seq data. *BMC Bioinformatics 12*(1), 480.

Serrano, J., S. U. Goebel, P. L. Peghini, I. A. Lubensky, F. Gibril, and R. T. Jensen (2000, November). Alterations in the p16INK4a/CDKN2A tumor suppressor gene in gastrinomas. *J. Clin. Endocrinol. Metab. 85*(11), 4146–4156.

Servin, B. and M. Stephens (2007, 07). Imputation-based analysis of association studies: Candidate regions and quantitative traits. *PLOS Genetics 3*(7), 1–13.

Sherman, B. T., M. Hao, J. Qiu, X. Jiao, M. W. Baseler, H. C. Lane, T. Imamichi, and W. Chang (2022, July). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res. 50*(W1), W216–W221.

Smith, G. D. and S. Ebrahim (2003, February). 'mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol. 32*(1), 1–22.

Soloff, J. A., D. Xiang, and W. Fithian (2024). The edge of discovery: Controlling the local false discovery rate at the margin. *The Annals of Statistics 52*(2), 580 – 601.

Stephens, M. and D. J. Balding (2009, October). Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet. 10*(10), 681–690.

Swanson-Wagner, R. A., R. DeCook, Y. Jia, T. Bancroft, T. Ji, X. Zhao, D. Nettleton, and P. S. Schnable (2009, November). Paternal dominance of trans-eQTL influences gene expression patterns in maize hybrids. *Science 326*(5956), 1118–1120.

Taliun, D., D. N. Harris, M. D. Kessler, J. Carlson, Z. A. Szpiech, R. Torres, and et al. (2021, February). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature 590*(7845), 290–299.

Tomé, S., G. B. Panigrahi, A. López Castel, L. Foiry, D. W. Melton, G. Gourdon, and C. E. Pearson (2011, June). Maternal germline-specific effect of DNA ligase I on CTG/CAG instability. *Hum. Mol. Genet. 20*(11), 2131–2143.

Unoki, H., A. Takahashi, T. Kawaguchi, K. Hara, M. Horikoshi, G. Andersen, D. P. K. Ng, J. Holmkvist, K. Borch-Johnsen, T. Jørgensen, A. Sandbaek, T. Lauritzen, T. Hansen, S. Nurbaya, T. Tsunoda, M. Kubo, T. Babazono, H. Hirose, M. Hayashi, Y. Iwamoto, A. Kashiwagi, K. Kaku, R. Kawamori, E. S. Tai, O. Pedersen, N. Kamatani, T. Kadowaki, R. Kikkawa, Y. Nakamura, and S. Maeda (2008, September). SNPs in KCNQ1 are associated with susceptibility to type 2 diabetes in east asian and european populations. *Nat. Genet. 40*(9), 1098–1102.

van der Graaf, A., A. Claringbould, A. Rimbert, BIOS Consortium, H.-J. Westra, Y. Li, C. Wijmenga, and S. Sanna (2020, October). Mendelian randomization while jointly modeling cis genetics identifies causal relationships between gene expression and lipids. *Nat. Commun. 11*(1), 4930.

Wang, K., M. Li, and H. Hakonarson (2010, September). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res. 38*(16), e164.

Webster, E., M. T. Cho, N. Alexander, S. Desai, S. Naidu, M. R. Bekheirnia, A. Lewis, K. Retterer, J. Juusola, and W. K. Chung (2016, November). De novo PHIP-predicted deleterious variants are associated with developmental delay, intellectual disability, obesity, and dysmorphic features. *Cold Spring Harb. Mol. Case Stud. 2*(6), a001172.

Weigle, D. S., S. L. Ganter, J. L. Kuijper, D. L. Leonetti, E. J. Boyko, and W. Y. Fujimoto (1997, February). Effect of regional fat distribution and Prader-Willi syndrome on plasma leptin levels. *J. Clin. Endocrinol. Metab. 82*(2), 566–570.

Wermter, A.-K., A. Scherag, D. Meyre, K. Reichwald, E. Durand, T. T. Nguyen, K. Koberwitz, P. Lichtner, T. Meitinger, H. Schäfer, A. Hinney, P. Froguel, J. Hebebrand, and G. Brönner (2008, September). Preferential reciprocal transfer of paternal/maternal DLK1 alleles to obese children: first evidence of polar overdominance in humans. *Eur. J. Hum. Genet. 16*(9), 1126–1134.

Yasuda, K., K. Miyake, Y. Horikawa, K. Hara, H. Osawa, H. Furuta, Y. Hirota, H. Mori, A. Jonsson, Y. Sato, K. Yamagata, Y. Hinokio, H.-Y. Wang, T. Tanahashi, N. Nakamura, Y. Oka, N. Iwasaki, Y. Iwamoto, Y. Yamada, Y. Seino, H. Maegawa, A. Kashiwagi, J. Takeda, E. Maeda, H. D. Shin, Y. M. Cho, K. S. Park, H. K. Lee, M. C. Y. Ng, R. C. W. Ma, W.-Y. So, J. C. N. Chan, V. Lyssenko, T. Tuomi, P. Nilsson, L. Groop, N. Kamatani, A. Sekine, Y. Nakamura, K. Yamamoto, T. Yoshida, K. Tokunaga, M. Itakura, H. Makino, K. Nanjo, T. Kadowaki, and M. Kasuga (2008, September). Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet. 40*(9), 1092–1097.

Zhabotynsky, V., K. Inoue, T. Magnuson, J. Mauro Calabrese, and W. Sun (2019, September). A statistical method for joint estimation of cis-eQTLs and parent-of-origin effects under family trio design. *Biometrics 75*(3), 864–874.

Zhou, Q. and Y. Guan (2018). On the null distribution of bayes factors in linear regression. *Journal of the American Statistical Association 113*(523), 1362–1371.

Zhou, Y., Y. Zhang, D. Zhao, X. Yu, X. Shen, Y. Zhou, S. Wang, Y. Qiu, Y. Chen, and F. Zhu (2024, January). TTD: Therapeutic target database describing target druggability information. *Nucleic Acids Res. 52*(D1), D1465–D1477.

Zink, F., D. N. Magnusdottir, O. T. Magnusson, N. J. Walker, T. J. Morris, A. Sigurdsson, G. H. Halldorsson, S. A. Gudjonsson, P. Melsted, H. Ingimundardottir, S. Kristmundsdottir, K. F. Alexandersson, A. Helgadottir, J. Gudmundsson, T. Rafnar, I. Jonsdottir, H. Holm, G. I. Eyjolfsson, O. Sigurdardottir, I. Olafsson, G. Masson, D. F. Gudbjartsson, U. Thorsteinsdottir, B. V. Halldorsson, S. N. Stacey, and K. Stefansson (2018, November). Insights into imprinting from parent-of-origin phased methylomes and transcriptomes. *Nat. Genet. 50*(11), 1542–1552.