Sex-dependent placental mQTL provide insight into the prenatal origins of childhood-onset traits and conditions

William Casazza,[1,2] Amy M. Inkster,[3,4] Giulia F. Del Gobbo,[3,4,5] Victor Yuan,[3,4] Fabien Delahaye,[6,7] Carmen Marsit,[8] Yongjin P. Park,[9,10] Wendy P. Robinson,[3,4] Sara Mostafavi,[1,11] Jessica K Dennis[*1,2,3]

Affiliations

[1]Centre for Molecular Medicine and Therapeutics, BC Children's Hospital, Vancouver, BC, Canada

[2]Bioinformatics Graduate Program, University of British Columbia, Vancouver, British Columbia, Canada

[3]Department of Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada

[4]BC Children's Hospital Research Institute, Vancouver, British Columbia, Canada

[5]Children's Hospital of Eastern Ontario Research Institute, University of Ottawa, Ottawa, Ontario, Canada

[6]Institut Pasteur de Lille, Lille, Hauts-de-France, France

[7]Albert Einstein College of Medicine, The Bronx, NY, USA

[8]Rollins School of Public Health, Emory University, Atlanta, GA, USA

[9]Department of Statistics, University of British Columbia, Vancouver, British Columbia, Canada

[10]Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, British Columbia, Canada

[11]Paul Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA

***Corresponding author:**

Jessica K Dennis, Msc, Phd

Investigator, BC Children's Hospital Research Institute

Assistant Professor, Faculty of Medicine, Department of Medical Genetics
The University of British Columbia
950 West 28th Avenue, Vancouver BC V5Z 4H4 Canada
Phone 604 875 2000 x7871

Summary:
Molecular quantitative trait loci (QTL) allow us to understand the biology captured in genome-wide association studies (GWAS). The placenta guides fetal development, and has notable molecular sex-differences (e.g. in DNA methylation). We therefore hypothesized that placental mQTL explain variation in childhood onset traits, likely to a different degree in males vs females. We analyzed 442 term male and female placenta from two studies and found 49,252 methylation (CpG) sites with methylation QTL (mQTL) and 2,489 CpG sites with sex-dependent mQTL. All mQTL were enriched in regions active in neonatal tissues that typically affect gene expression. All mQTL were enriched for growth- and immune-related traits, but male- and female-specific mQTL were more enriched than cross-sex mQTL. mQTL colocalized with trait loci at 322 CpG sites, with 77 (21%) specific to males or females. Overall, mQTL specific to male and female placenta capture otherwise overlooked variation in childhood traits.
Keywords: Developmental origins of health and disease, methylation quantitative trait loci, placenta, sex differences, complex trait analysis, placenta

Introduction

GWAS findings hold valuable clues about trait mechanisms.[1,2] Deciphering these clues, however, requires additional data on gene regulation, as over 90% of SNPs identified in GWAS lie in gene regulatory regions, as opposed to in the protein-coding gene region itself.[3,4] Molecular quantitative trait locus (molQTL) analysis is a powerful strategy to interpret the gene regulatory functions of GWAS SNPs. Under the umbrella of molQTL, expression quantitative trait loci (eQTL) are the most widely studied, and relative to SNPs matched by minor allele frequency (MAF), they are enriched for GWAS loci.[5] However, only 43% of eQTL share the same causal variant (i.e. colocalize) with a GWAS locus,[5] and up to 77% of eQTL in linkage disequilibrium (LD) with a trait-associated SNP are shared across more than one tissue.[6] As a result, the majority of GWAS loci have either no known effects on expression, or their relationship with

traits is clouded by their broad effects on gene expression across tissues.[5,7] Moreover, few new eQTL are being discovered in the most widely used eQTL resource, GTEx, even as sample sizes exceed 600,[7] suggesting that eQTL discovery using post-mortem adult tissues is limited.

Therefore, to advance the functional interpretation of GWAS SNPs, we must extend molQTL discovery across molecular traits, tissues, and contexts (e.g., sex). In this study, we identify DNA methylation QTL (mQTL) in placenta, including mQTL that differ between males and females. DNAm is an attractive molecular trait for functional interpretation of GWAS results because it can provide insights on the precise molecular mechanism (i.e., biochemical modification to the DNA sequence) by which GWAS SNPs associate with traits and conditions.[8] In addition, variation in DNAm is genetically influenced. In blood, the tissue in which DNAm is most widely studied, 21% of the variation of DNAm is explained by additive genetic variation in cis (i.e. via mQTL).[9] Importantly, mQTL provide information on gene regulation beyond what is provided by eQTL. For example, mQTL cover roughly twice as many genes as eQTL in blood,[5,9,10] and in one study measuring both mQTL and eQTL in blood in the same participants, only 7% of mQTL with eQTL effects actually mediated the association with gene expression, suggesting that the DNAm was regulating a distinct molecular process.[10]

The placenta is of the same genetic make up as the fetus and it is one of the first organs to form during gestation. Throughout pregnancy it is responsible for the exchange of oxygen, nutrients, and hormones between mother and fetus. Despite its central role in human development, however, placenta is under-studied and is not represented in large-scale molQTL resources like GTEx. Several studies have previously investigated placental molQTL and their relationship with complex traits and conditions.[11–16] In the largest mQTL analysis to date, Delahaye et al.[12] analyzed 303 placental samples from the The Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) study and characterized a small number (N = 4,342) of strongly associated (i.e., passing a stringent permutation test and quality threshold) mQTL, which were found to overlap two type 2 diabetes loci. Tekola-Ayele et al.[13] analyzed the same samples with a similar approach and using colocalization analysis, implicated four genes with placental DNAm and gene expression that shared genetic loci with birth weight. While these analyses of the NICHD study demonstrated the relevance of placental molecular traits to postnatal outcomes, the stringent thresholds used to map mQTL means that many mQTL remain unmapped, especially in larger sample sizes. In addition, both Delahaye and

Tekola-Ayele analyzed at most two GWAS traits, and the relationship of placental mQTL to multiple postnatal outcomes has yet to be investigated. The Rhode Island Child Health Study (RICHS)[15,17] is another large study that has collected molecular data from 149 placental samples, but mQTL have not been mapped in the RICHS study, and overall, a more comprehensive study is needed to better understand how placental mQTL affect genome-wide risk of traits.

Additionally, neither the NICHD nor RICHS studies have investigated sex differences in the genetic regulation of placental molecular traits. Oliva et al. recently analyzed sex differences in eQTL across 44 GTEx tissues and found that sex-dependent eQTL (i.e. cross-tissue mQTL with a genotype by sex interaction) were remarkably tissue specific: of 369 sex-dependent eQTL found in at least one tissue, only one was shared by two tissues.[18] Moreover, they found 74 eQTL that colocalized with GWAS loci in either males or females, 24 of which showed no evidence of colocalization in eQTL computed in all subjects. These results suggest that sex-dependent molQTL yield functional interpretations of GWAS loci beyond what is provided by cross-sex analysis. Analyzing sex-dependent molQTL could be especially important in placenta, since sex is strongly associated with placental molecular traits such as DNAm, even after accounting for cell type proportions,[19,20] which can bias sex-dependent molQTL analysis.[18]

In this work we identify placental mQTL with a shared effect in males and females (i.e. cross-sex mQTL) and mQTL that are modified by sex (i.e. sex-dependent mQTL) by meta-analyzing data from 411 term placentas from the NICHD study and RICHS. We then compare cross-sex and sex-dependent mQTL with regards to genomic location and gene expression patterns. Finally we quantify the relevance of cross-sex and sex-dependent mQTL to childhood-onset traits and conditions using stratified LD score regression (S-LDSC)[21–23] and colocalization.

Results

Sex-dependent placental mQTL are distinct from cross-sex placental mQTL

We meta-analyzed mQTL across the NICHD and RICHS studies (Table S1), using MeCS[24] to account for bias induced by correlation in mQTL effects between studies. Meta-analyzed mQTL effects (or genotype by sex interaction effects in the sex-dependent analysis) were called at a Bonferroni corrected $p < 0.05$. We defined 6 sets of mQTL (Figure 1,

Table 1) which we use in all downstream analyses: (i) cross-sex, or those mQTL with an effect independent of sex (i.e. mQTL that are typically mapped); (ii) sex-dependent, which have an effect modified by sex; (iii) male-stratified, which have an effect in male samples; (iv) female-stratified, which have an effect in female samples; (v) male-specific,which have an effect in males that also differs from its effect in females (intersection of sets (ii) and (iii)); and (vi) female-specific, which have an effect in females that differs from its effect in males (intersection of sets (ii) and (iv)).

We found 49,252 CpGs with a cross-sex mQTL, 351 CpGs with a male-specific mQTL, and 255 CpGs with a a female-specific mQTL. Of the 351 CpG sites with a male-specific mQTL, 185 (53%) also had a cross-sex mQTL. Of the 255 CpG sites with a female-specific mQTL, 153 (60%) also had a cross-sex mQTL. 75 CpG sites had both male- and female-specific effects (see Figure 1 for an example of these types of mQTL effects), of which 74 (99%) also had a cross-sex mQTL (Figures S1A,B, and Figures S1C,D for a comparison with male- and female-stratified mQTL).

In analyzing the distribution of mQTL across chromosomes, we observed that the proportion of CpG sites with a cross-sex vs. sex-dependent mQTL differed for 10 of 23 chromosomes (Table S2, Figure 2A, chromosomes with a proportion test[25–27] adjusted $p < 0.05$ marked with an asterisk). The proportion of male- vs. female-specific mQTL was consistent across chromosomes (Figure 2B, for male- and female-stratified differences see Tables S2).

We further investigated the difference in male- vs female-specific beta values (i.e. strength of SNP-CpG association, for a similar analysis using allelic Fold-Change[28] instead of beta values, see STAR Methods). To benchmark, across adult tissues in GTEx V8 with sex-dependent eQTL, the majority of SNP-eGene associations were in the same direction in both sexes but had differences in effect size.[18] In our placental meta-analysis, the average effect in male-specific mQTL was smaller than the effect observed in female-specific mQTL (mean beta 0.021 in males, 0.035 in females, Wilcoxon rank-sum test two-sided $p < 0.026$, Figure 2C, Table S3). This difference was larger when we consider the absolute value of the effect size (mean absolute beta 0.067 in males vs 0.079 in females, Wilcoxon rank-sum test two-sided $p < 2.3e-11$, Table S3, Figure S1C). This effect size difference did not appear to be driven solely by differences in mQTL effect sizes on the X chromosome: female-specific mQTL had a larger mean signed effect size than male-specific mQTL in 6 autosomes, and a larger mean absolute

effect size in 9 autosomes (Wilcoxon rank-sum test one-sided, Bonferroni-Holm corrected $p < 0.05$, Figure 2D, Table S3). Effects were highly correlated between male- and female-specific mQTL (Spearman's $\rho = 0.87$, Figure S1H). Thus, despite representing mQTL specific to one sex, male- and female-specific mQTL tended to have the same direction of effect.

As a point of comparison, we calculated the replication (π_1) of placental mQTL across mQTL in two other perinatal tissues, umbilical cord blood and fetal brain, identified in independent datasets (Figure S1E,F; STAR Methods).[29,30] We observed a relatively large proportion of placental mQTL overlapping cord blood mQTL (π_1=0.76) and and even larger proportion of placental mQTL overlapping fetal brain mQTL (π_1=0.84). Effect sizes of placental mQTL correlated poorly with those in cord blood (Spearman's ρ=-0.31) but largely shared the same direction of effect in fetal brain (Spearman's ρ=0.65), which suggests a higher degree of similarity between mQTL in placenta vs in fetal brain.

Placental mQTL are enriched in regions controlling gene expression and primarily occur at CpG sites with intermediate levels of DNAm

In order to establish whether sex-dependent mQTL occur in distinct genomic regions compared to cross-sex mQTL, we applied GARFIELD[31] to the minimum mQTL p value for each SNP (STAR Methods). Briefly, GARFIELD tests whether a given set of SNPs associated with a particular phenotype are enriched in a set of genomic regions, accounting for LD between SNPs and redundancy in annotations. In other words, GARFIELD will not overestimate enrichment due to many correlated mQTL existing in the same location, and will exclude annotations with similar enrichment across all mQTL. As expected, the enrichment of each mQTL set in each annotation was typically larger for the set with a larger number of mQTL, with cross-sex having the most mQTL, followed by male-stratified, female-stratified, then sex-dependent mQTL (Table 1). For each set of annotations, male- and female-stratified mQTL had the same ordering of odds ratios as in cross-sex mQTL, albeit with smaller estimates. Note that male- and female-specific mQTL were excluded here, as GARFIELD uses only a single p value for each SNP and therefore could not handle the main and interaction term p values that define the male- and female-specific analyses. Across ChromHMM annotations in human cell lines (Figure 3A), we observed that cross-sex placental mQTL were more enriched in transcription start sites (TSS) than chromatin states (TSS odds ratio (OR) of 2.68 vs a mean OR=1.60 in the remaining selected states), which

was not observed when looking at sex-dependent mQTL (TSS OR=1.22 vs a mean OR=1.47 across other annotations). Additionally we noted the enrichment of these placental mQTL sets in histone modifications averaged across several tissues in ENCODE[31,32] (Figure 3B). For cross-sex mQTL, the highest enrichment was in H3K9 acetylation (mean OR=2.50), followed by H3K4 trimethylation (mean OR=2.41), both of which are indicative of active gene promoters.[33] For sex-dependent mQTL, the highest enrichment was in H4K20 monomethylation (OR=1.69), which is associated with transcription activation.[34] Together, these results suggest minor, but potentially meaningful, differences in the molecular mechanisms active at sex-dependent vs cross-sex mQTL.

Next, we investigated the mean DNAm of CpG sites with at least one cross-sex, sex-dependent, male-specific, or female-specific mQTL, as well as the distribution of these sites relative to CpG islands (CGIs) and gene regions as mapped Illumina's HumanMethylation450k array annotation (STAR Methods). Both cross-sex and sex-dependent mQTL were primarily associated with CpG sites with intermediate DNAm (with 72% and 71% of CpG sites respectively having a weighed mean beta between 0.2 and 0.8). These proportions are much larger than what is observed in blood (40% of sites with intermediate DNAm)[9] and is expected given that the placenta has a high proportion of CpG sites with an intermediate amount of DNAm, as has been previously reported (Figure 3E).[35] As for the remaining CpG sites, 16% of CpG sites with cross-sex or sex-dependent mQTL had high methylation (beta > 0.8), with the remaining 12% and 13% of CpG sites respectively having low methylation (beta < 0.2). Within mQTL with male-specific effects, 19% had low methylation, 60% had intermediate methylation, and 21% had high methylation. For mQTL with female-specific effects 14% had low methylation, 55% had intermediate methylation and 31% had high methylation (Figure 3E).

In follow-up, we looked at the weighted-mean DNAm of these CpG sites by position relative to CGIs and gene regions (Figures 3C-E). CGIs are where the majority of DNAm occurs, and the location of a given CpG site relative to the CGI is associated with the particular function of that CpG. For example, high DNAm in CpG shores (within 2kb of a CpG island) and shelves (within 2-4 kb of a CpG island) is associated with higher nearby gene expression.[36]

We observed similar CGI distributions across all four mQTL sets. The majority of highly methylated (beta > 0.8) CpG sites were in open sea regions (51% of cross-sex, 52% of sex-dependent, 57% of male-specific and 52% of female-specific), followed by regions within

shores or shelves (36% of cross-sex, 36% of sex-dependent, 35% of male-specific, and 38% of female-specific), and the remaining CpG sites within CGIs themselves (13% of cross-sex, 12% of sex-dependent, 10% of male-specific, and 10% of female-specific) (Figure 3C). We complemented the analysis of position relative to CGIs with an analysis of position relative to gene bodies. As in the CGI analysis, we observed similar patterns across all four mQTL sets, with highly methylated CpG sites located in gene bodies (51% of cross-sex, 51% of sex-dependent, 56% of male-specific, and 53% of female-specific), which typically indicates active transcription of those genes. Intermediately methylated sites were evenly split between gene bodies and intergenic regions (37% and 33% for both cross-sex mQTL and sex-dependent mQTL, 37% and 30% for male-specific mQTL, and 38% and 30% for female-specific mQTL respectively). Finally, lowly methylated sites were primarily within 1500bp of the transcription start site of genes (39% in cross-sex, 39% in sex-dependent, 41% of male-specific, and 43% of female-specific), indicating active transcription (Figure 3D).

Sex-dependent and cross-sex mQTL show similar enrichment in genomic regions active across tissues

We again turned to GARFIELD to investigate if mQTL associated with these CpG sites could have any tissue-specific activity. We considered DNAse1 hypersensitivity (DHS) sites in fetal tissues (Figure 4A), as well as in adult tissues and cell lines (Figure 4B). For both cross-sex and sex-dependent mQTL, enrichment was highest in fetal membrane (OR=2.82 vs OR=1.58 respectively), followed by fetal placenta DHS sites (OR=2.74 vs OR=1.43 respectively). For adult tissues and cell lines, cross-sex mQTL were most enriched in myometrium (uterine wall tissue, OR=2.71), bone (OR=2.68), liver (OR=2.64), and colon (OR=2.64). Slight differences were found in sex-dependent mQTL, with enrichment being highest in bone (OR=1.56), blastula (an early stage of embryonic development, OR=1.51), colon (OR=1.49) and liver (OR=1.47). Thus despite showing weaker enrichment overal, sex-dependent mQTL show similar patterns of activity to cross-sex mQTL across tissues.

Given the generally higher enrichment of cross-sex mQTL in the previous analysis, we chose to check whether these CpG were in genes specifically expressed in adult tissues. Cross-sex mQTL were spread across over half of the genes annotated to CpG sites in the 450k array (12,746 of roughly 20,000[37]), in contrast to the much smaller set of genes annotated to

CpGs with a sex-dependent, male- or female-specific mQTL (of 976 genes annotated to a sex-dependent mQTL, 201 were annotated to CpG sites with male-specific mQTL and 153 were annotated to a CpG site with a female-specific mQTL, Table 1, Figure S1). Next, we asked whether each gene set was likely to be expressed in all 54 GTEx tissues (i.e. genes with a median transcripts per million (TPM) >= 20 in all tissues, comprising 460 genes total, STAR Methods). Genes annotated to CpGs with a cross-sex mQTL were enriched in genes expressed across all tissues (overlap of 170, adjusted hypergeometric test $p < 1.09e-5$, Table SI). Meanwhile, sex-dependent and female-specific mQTL genes showed no enrichment (2 genes, $p < 0.26$ and 2 genes, $p < 0.71$ respectively), and no male-specific genes were expressed across all tissues (0 genes, $p < 0.13$). Thus it appears that genes annotated to CpG sites with a sex-dependent mQTL are less likely to be expressed across all 54 GTEx V8 tissues than genes annotated to CpG sites with a cross-sex mQTL.

Male- and female-specific placental mQTL show higher enrichment than cross-sex placental mQTL in immune-related and growth-related traits

In this study we were particularly interested in how placental mQTL specific to males or females shed light on sex differences in the prevalence and manifestation of traits that are measured during childhood (STAR Methods). We included GWAS of 18 childhood traits (Table SI; Figure 5),[38,39] as well as a GWAS of maternal pre-eclampsia (i.e., the fetal genetic effect on maternal pre-eclampsia risk,[40] which is a placentally-mediated condition. We then used S-LDSC to estimate the proportion of SNP-heritability of these 19 complex traits that was explained by our placental mQTL sets. We grouped childhood traits into 3 categories: neuropsychiatric (N=6), immune-related (N=3), and growth-related (N=8). None of the neuropsychiatric trait GWAS were enriched for our mQTL sets, either individually, or when all 6 traits were meta-analyzed (Table SII for all S-LDSC results). Several immune- and growth-related traits, on the other hand, were enriched for our placental mQTL sets, both individually, and in meta-analysis within groups.

We found an enrichment (FDR < 0.05) of cross-sex placental mQTL in childhood onset asthma, type 1 diabetes, birth weight, and growth at the start of puberty (Figure 6A). We also observed an enrichment for birth length, and growth later in puberty at a within-trait FDR < 0.05. The τ^* coefficient for cross-sex mQTL in birth weight was also significant ($\tau^* = 4.397*10^{-3}$,

within-trait FDR < 0.05, Figure S2A), suggesting that enrichment was not explained by overlap of the cross-sex mQTL with other functional annotation categories in the genome. In line with the hypothesis that mQTL with sex-dependent effects may better explain heritability for certain conditions, we observed an even greater enrichment of male- and female-specific mQTL (FDR < 0.05) compared to cross-sex mQTL for birth weight and growth at the start of puberty, and at a within-trait FDR < 0.05 we observed enrichment in childhood BMI.

In meta-analysis, female-specific mQTL were more enriched than cross-sex mQTL for immune related traits (1.75, SE=0.249 in female-specific mQTL vs 1.29, SE=0.06 in cross-sex mQTL) (Figure 6B). Additionally, both male- and female-specific mQTL proved to be more enriched in growth related traits when compared to either cross-sex, mQTL (1.66, SE = 0.167 in male-specific mQTL, 1.78, SE=0.137 in female-specific mQTL, and 1.20, SE= 0.08 in cross-sex mQTL). For meta-analyzed τ^* estimates, see Table S4, Figures S2B,E,G. Thus overall, we observed that enrichment for placental mQTL was higher within growth and immune related traits, with both male- and female-specific mQTL showing particularly high enrichment in these categories, even when compared to male- or female-stratified mQTL enrichments.

In sensitivity analyses, we also examined sex-stratified h^2 estimates and the proportion of sex-stratified h^2_{SNP} explained by our mQTL sets. Since the sex-stratified GWAS sample sizes were smaller, fewer traits passed our quality control checks (STAR Methods): 4 neuropsychiatric traits (2 with both sexes), 2 immune-related traits (1 with both sexes), and 3 growth-related traits (2 with both sexes) (Figure S2C). Additional traits are in Table SI.

When looking at sex-stratified GWAS of neuropsychiatric traits, we observed a single small enrichment of male-stratified mQTL in SCZ (1.16, p < 0.00542, SE = 0.121, Figure S2D). Additionally the normalized S-LDSC coefficient, (τ^*),[23] for this enrichment was significant at a within trait FDR < 0.05 ($\tau^*=1.23*10^{-3}$, Figure S2F). Additionally male- and female-specific mQTL were enriched in sex-stratified GWAS for growth at the start of puberty at an FDR < 0.05 (Figure S2D). It is worth noting that traits enriched for male- and female-specific mQTL all showed a larger enrichment for that trait than either cross-sex, male- or female-stratified mQTL. In other words, the per-SNP heritability explained by male- and female-stratified mQTL was lower than when we restricted those annotations to CpG sites having an mQTL that interacted with sex. Alternatively, the τ^* coefficient for placental mQTL in immune and growth related GWAS was only significant at a within trait FDR < 0.05 for cross-sex mQTL in birth weight (τ^*=

$4.34*10^{-3}$, Figure S2F), which may suggest that despite covering many SNPs that contribute to the heritability of each trait with significant enrichment, these annotations likely also contain a number of SNPs that do not inform trait risk.

Placental mQTL colocalize primarily within growth and immune related traits , with additional CpG sites colocalizing at male- and female-specific placental mQTL

We subsequently applied colocalization analyses to assess whether any of the mQTL we identified explained signals in the GWAS of 18 childhood traits and maternal pre-eclampsia. We conducted this analysis for cross-sex and sex-stratified mQTL with GWAS loci within a 150kb window centered on each CpG site. Colocalization was conducted on p values using *coloc*[41] (STAR Methods). We found a considerable number of colocalized CpG sites in several of the childhood traits with moderate enrichment in S-LDSC, such as type 1 diabetes, child onset asthma, and birth weight (Figure 6C, Table SIII). We also observed a high number of colocalized sites for SCZ, which notably has a large number of GWAS loci.[42]

Each of these four traits showed colocalization in male- and female-stratified mQTL that was not detected in cross-sex mQTL, which we deem male- and female-specific in this analysis (Table SIII). It is worth noting that this definition depends on both whether mQTL colocalize with GWAS loci, and whether a particular mQTL in a given category passed a threshold for significance (for examples of colocalized CpG sites, see Figure S3). For example, 128 CpG sites (annotated to 70 genes) whose cross-sex mQTL colocalized with child onset asthma loci. Of these, 57 CpG sites (annotated to 36 genes) only colocalized in cross-sex mQTL, and 53 CpG sites (annotated to 23 genes) additionally colocalized in both male- and female-stratified mQTL, 13 CpG sites (annotated to 8 genes) colocalized in male-stratified mQTL and 5 CpG sites (annotated to 3 genes) colocalized in female-stratified mQTL. In Figure 6C, we label all of these CpG sites as Cross-sex (STAR Methods). An additional 45 colocalized CpG sites were female-specific (13 genes), 19 colocalized CpG sites were male-specific (9 genes), and 4 CpG were both male- and female-specific (3 genes). We observe similar patterns for the remaining 3 traits with the most CpG sites having colocalized mQTL (Figure 6C, Table SIII).

To get a picture of differences in sex-specific vs cross-sex colocalization, we chose to summarize genes colocalized with childhood onset asthma loci (54 cross-sex, 5 male-specific,

and 10 female-specific, 69 total). Notably, childhood onset asthma shows a distinct pattern of onset in males vs females.[43,44] A considerable number of the genes found were located within the HLA region[44,45] (STAR Methods) in mQTL with both cross-sex and sex-specific effects. Notably, in cross-sex, genes from the HLA region were in the minority (14 of 54, 26%, in childhood onset asthma). Using Enrichr,[46–48] we tested for the enrichment of the remaining genes found outside of the HLA region in these traits at an adjusted p value < 0.05, and a minimum overlap of 3 genes, across the three main genome ontology annotations. The remaining cross-sex colocalized CpG sites were most enriched for the regulation of type I interferon gamma production (adjusted p < 4.8e-4, GO:0032479, overlapping *TRAF3*, *IRF1*, *STAT6*, *POLR3H*, and *NFKB1*). For female-specific CpG sites 5 of 10 genes were in the HLA region, with 3 of the remaining 5 genes being part of the major histocompatibility complex class III (*BAT1*, *BAT2*, and *BAT3*).[49] 3 of 5 genes for male-specific CpG sites were in the HLA region, with the gene *NEU4*, which is related to metabolism, and predicted gene *TSBP1* (*C6ORF10*) which encodes a protein most highly expressed in the testis.[50]

Discussion

In this study, we conducted a meta-analysis of mQTL across two term placental studies (NICHD and RICHS), and a comprehensive sex-dependent mQTL analysis. We demonstrated that sex-dependent mQTL are largely located at distinct CpG sites from cross-sex mQTL, and are annotated to genes expressed in different tissues than cross-sex mQTL. In contrast to blood mQTL, which are mainly lowly or highly methylated,[9] placental mQTL tended to be associated with intermediately methylated CpG sites. Although the biological function of intermediate methylation in placenta is likely different than in adult tissues, sites with intermediate methylation in adult tissues are likely involved in regulating genes with a context-dependent function.[51] We showed that cross-sex mQTL were enriched in immune- and growth-related trait SNP-based heritability, and that male- and female-specific mQTL were more enriched than cross-sex mQTL across traits in these categories. Perhaps most importantly, we found several CpG sites that only colocalized in a male- or female-specific manner. Thus, sex-dependent mQTL appear to capture some underlying aspect of heritable trait risk that is not being captured by cross-sex placental mQTL alone.

Thus, there is further need to test whether placental mQTL specifically, from term placenta or otherwise, are likely involved in the basis of complex traits. Notably, past eQTL analyses,[12,15,16] and integrative molecular QTL analyses in placenta collected primarily from preterm births,[14,52] have made meaningful progress in outlining the scope of traits likely to be influenced by the genetic regulation of placental gene expression. We and others have found as many as 2,745 CpG sites across 582 genes associated with placental sex, and these genes are primarily related to immune function and growth-factor signalling.[19,20] Studies have further shown that placental eQTL colocalize with a combination of childhood anthropometric raits such as BMI and birth weight,[16,52] and potentially immune related traits such as asthma and type 2 diabetes, all of which have evidence of sex-differential genetic risk.[53] Despite some evidence of enrichment in neuropsychiatric traits like schizophrenia, these effects appear to be much larger across immune and growth-related traits for which colocalization with placental eQTL has already been found.[15,52] Our study corroborates and builds on these findings by showing that mQTL with sex-specific effects are more enriched for these effects.

Building on this discussion, placental morphology, possibly through its effect on birth size, has been shown to associate risk for conditions onset throughout life course.[54–56] Interestingly, asthma is one trait that has been shown to associate with both placental morphology and small body size (including birth length and weight).[54] Importantly, many of the loci in this study showing colocalization with placental mQTL were within hits in the HLA region, which frequently shows association with childhood onset asthma and allergic diseases.[43,57] This includes genes containing SNPs most associated with increased risk for asthma hospitalizations in adults,[58] such as *HLA-DQB1* (colocalized in cross-sex mQTL) and *HLA-DRB6* (colocalized in female-specific mQTL), where genetic risk in *HLA-DRB6* was shown to be mediated through gene expression in lung tissue. Thus the additional functional annotation done here was able to further flesh out the molecular function of top disease variants.

There are several limitations to this study. First, our view of tissue-specificificy for placental mQTL is limited, as efforts are still underway to characterize mQTL across tissues . However, it is worth noting that the similarity of cross-sex placental mQTL to mQTL in other perinatal tissues (π_1=0.76 in umbilical cord blood and π_1=0.84 fetal brain, see Results) appears higher than the similarity between placental eQTL and adult tissues from GTEx (ranging from π_1=0.32 in cerebral hemisphere to π_1=0.69 in fibroblasts).[12] Second, it is important here to make

the distinction between sex-stratified analyses in this work and analyses that look at sex-specific effects (i.e. those mQTL that have a significant genotype-sex interaction). Sex-dependent mQTL are less numerous than cross-sex mQTL due to the increased power required to detect interaction effects.[59] As a result, in all of our experiments comparing sex-dependent mQTL (and by extension male- and female-specific mQTL) to cross-sex mQTL, sex-dependent mQTL had larger standard errors (e.g. in enrichment estimates) in comparison to cross-sex mQTL, which hampered statistical inference. As we accrue larger samples sizes in molecular studies of placenta and other tissues, so too will our ability to identify robust differences between the enrichment of male- and female-specific mQTL in complex traits. Notably, meta-analyzing the enrichment of sex-specific mQTL within the 3 trait categories considered here (neuropsychiatric, immune-related, and growth-related) yields estimates with a smaller standard error than in individual traits, allowing for an stronger comparison with meta-analyzed enrichment estimates from cross-sex mQTL. Moreover, we note that sex-stratified mQTL, which are less affected by this loss in power, and their associated CpG sites appeared to largely have the same patterns of enrichment across each experiment when compared to cross-sex mQTL (Figures 3A-B,E; S2A,B,D-G). Thus there is a clear distinction between mQTL effects detected in one sex (as in the male- and female-stratified analyses) and mQTL which have an interaction with sex (as in the male-and female-specific analyses).

Similarly, we were limited in our ability to test male- and female-specific mQTL for enrichment in GWAS of their respective sex. Although sex-stratified GWAS results are increasingly being shared,[60] these GWAS are underpowered. Only the GWAS of ADHD, ASD, SCZ, hayfever or eczema age of onset, pubertal growth start, and total pubertal growth were sufficiently powered to meet our criteria for inclusion in the enrichment analyses in both sexes. For these traits, we found higher enrichments of male- and female-specific mQTL (Figure S2D,E) as compared to enrichments found for sex-pooled GWAS, which emphasizes the need for large sex-stratified GWAS. Lastly, LDSC methods have mainly been validated in reference panels matching the genetic ancestry of GWAS cohorts.[21–23] Here we made use of the 1000 Genomes European reference panel and elected to assess effects in European GWAS summary statistics, thus ignoring effects that may arise in other populations. This likely has a non-trivial effect on our results, as we apply these methods to an ancestrally diverse cohort. In particular, the ancestry of the NICHD cohort is considerably more ancestrally diverse as compared to the

RICHS cohort (72 of 303 subjects in NICHD vs 99 of 149 subjects in RICHS classified as EUR via GRAF-pop, Table S5). Notably mixed ancestry has been shown to impact in the colocalization of eQTL and GWAS signal.[61] As our understanding of molecular traits and genetic associations in people of diverse ancestries grows, we will develop a better view of the biological mechanisms that underlie complex traits and conditions.

Overall, this study demonstrates that the genetic regulation of placental DNAm is at least partially sex-dependent. Sex-dependent signal occurs at distinct functional genomic regions, and appears to be distinguishable from cross-sex genetic effects on placental DNAm. Placental mQTL explain a significant proportion of the heritability of conditions relating to immune function and childhood growth, and both male- and female-specific mQTL are more enriched than cross-sex mQTL across these conditions, despite being far fewer in number. Trends we observed in the enrichment of placental mQTL translated to higher instances of colocalization between mQTL and GWAS loci, and including male- and female-specific colocalization allowed us to detect 235 CpG sites (annotated to 77 genes) that did not show sufficient evidence of colocalization in cross-sex mQTL. Taken together, our findings demonstrate that careful consideration of sex in mQTL analyses has the potential to provide additional information about the basis of complex traits, particularly when the tissue, molecular features, and traits being queried present differently in males vs females.

Acknowledgements

This project was supported by a British Columbia Children's Hospital Research Institute Catalyst Grant. JDK is supported by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation, and is a Michael Smith Health Research BC Scholar.
phs001586 : Rhode Island Child Health Study (RICHS): This work was supported by the National Institutes of Health [NIH-NIMH R01MH094609, NIH-NIEHS R01ES022223, NIH-NIEHS PO1ES022832, NIH-NIEHS R24ES028507, NIH-NIEHS R21ES028226, and NIH-NIEHS R01ES025145]. phs001717: This research was supported by the Intramural Research Program of the Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health (Contract Numbers: HHSN275200800013C;HHSN27500006;HHSN275200800003I;HHSN275200800014C;HHSN275200800012C;HHSN275200800028C; HHSN275201000009C). We also want to thank the

Genomics Shared Resource of Roswell Park Cancer Institute, supported by National Cancer Institute (NCI) grant P30CA016056.
This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of British Columbia (https://doi.org/10.14288/SOCKEYE).

Author Contributions

Conceptualization, W.C., J.K.D. and S.M.; Methodology, W.C.,Y.P, and J.K.D.; Software, W.C. A.I., J.D. V.Y.; Validation W.C.,A.I., J.D. V.Y.; Formal analysis, W.C.; Resources, J.K.D. and S.M.; Data Curation, W.C; Writing - Original Draft, W.C.; Writing - Review and Editing, J.K.D.; Visualization, W.C.; Supervision, J.K.D. and S.M.

Declaration of Interests

The authors declare no competing interests.

Figures

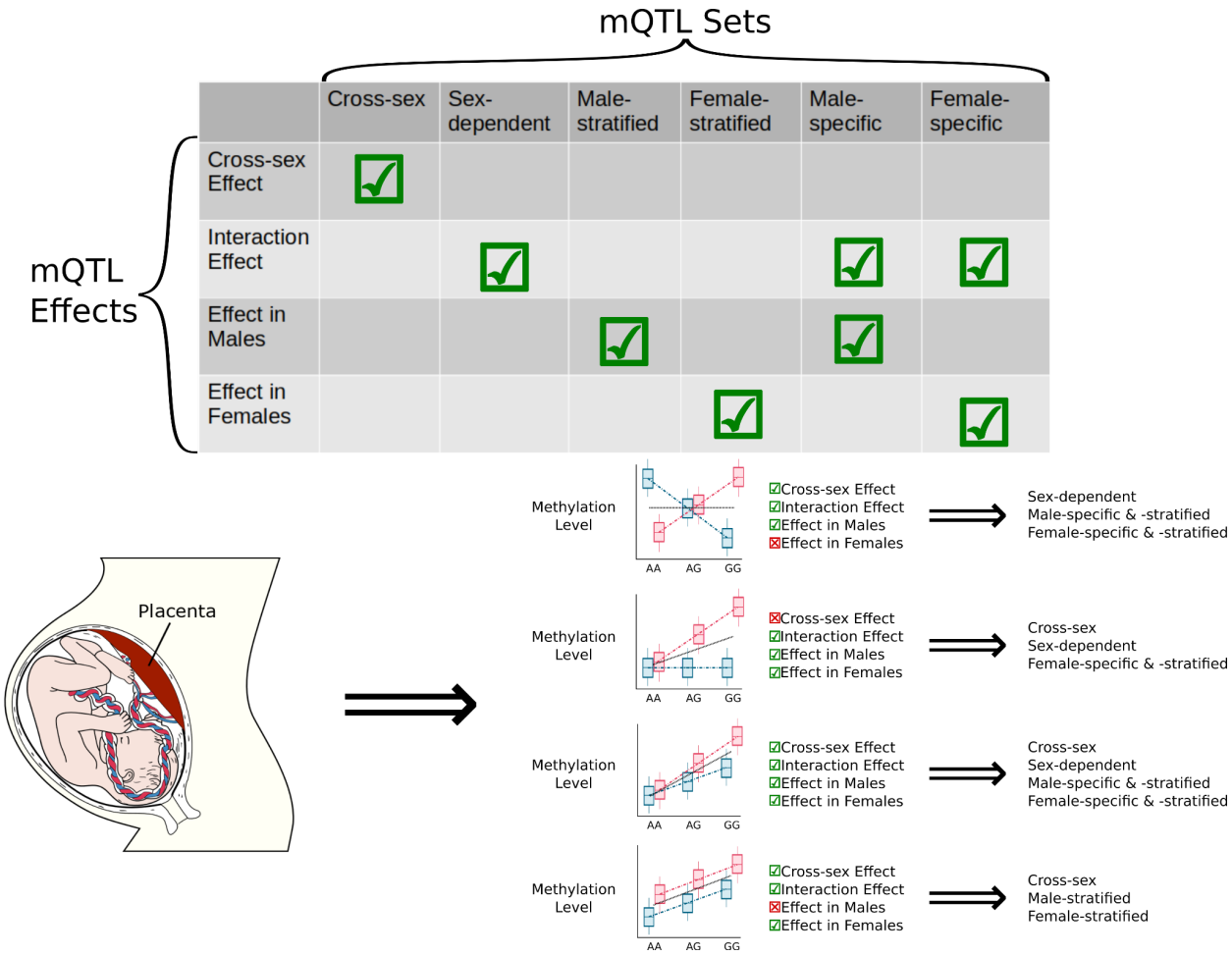

Figure 1. Defining categories of placental mQTL. An overview of how different categories of placental mQTL are classified in this work. In brief, 6 mQTL categories are defined by their effect in a male- or female-stratified analysis, their effect in a cross-sex mQTL analysis with all samples, and by whether they have a genotype by sex interaction effect in an analysis with all samples. These categories are (i) cross-sex mQTL, which have an effect independent of sex, and which match the definition of mQTL called in most analyses; (ii) sex-dependent mQTL, which have an effect that differs between male and female samples, (iii) male-stratified mQTL, which have an effect in males, (iv) female-stratified mQTL, which have an effect in females, (v) male-specific mQTL, which have an effect in males that also differs from its effect in females, and (vi) female-specific mQTL, which have an effect in females that also differs from its effect in males.

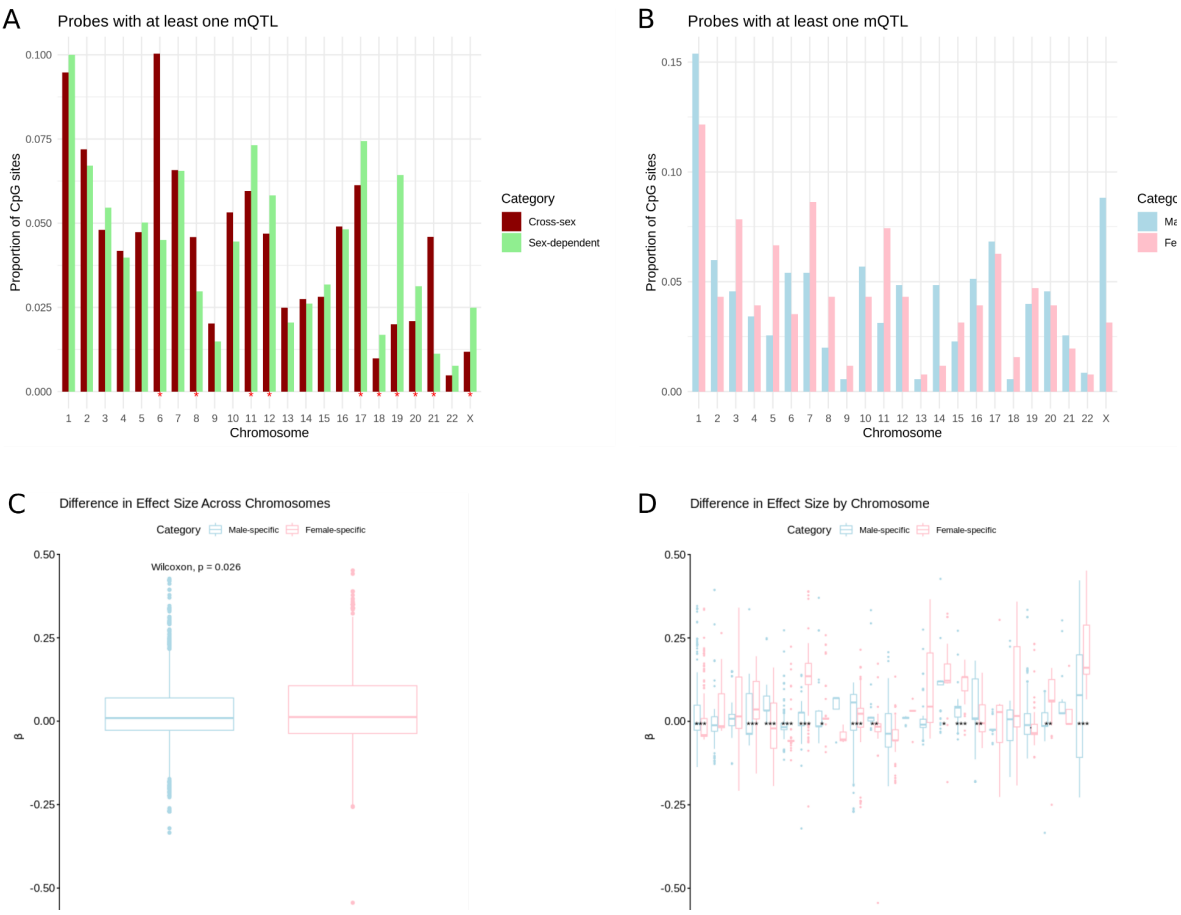

Figure 2. mQTL detected across chromosomes. 4 placental mQTL effects were called and meta-analyzed across 2 term placenta cohorts (STAR Methods) resulting in sets of mQTL with cross-sex, sex-dependent, male- and female-specific effects that were significant at a Bonferroni corrected $p < 0.05$. A) The proportion CpG sites with at least one cross-sex vs sex-dependent mQTL by chromosome. Chromosomes with a significant difference in proportion are marked with an asterisk (FDR < 0.05, proportion test). B) The proportion of CpG sites with at least one male- vs female-specific mQTL called per chromosome. No chromosomes in this analysis had a significant difference in proportion of CpG sites with at least one mQTL (i.e. meeting a FDR < 0.05, proportion test). C) The difference in mean effect size in male- vs female-specific mQTL across all chromosomes, with female-specific effects being significantly larger on average (wilcoxon rank-sum test p value = 0.026). D) Difference in mean effect size for male- and female-specific mQTL for each chromosome with wilcoxon rank-sum test p values after Bonferroni-Holm correction denoted as follows: .: p <=0.1; *: p <= 0.05, **: p <= 0.01;***: p <= 0.001.

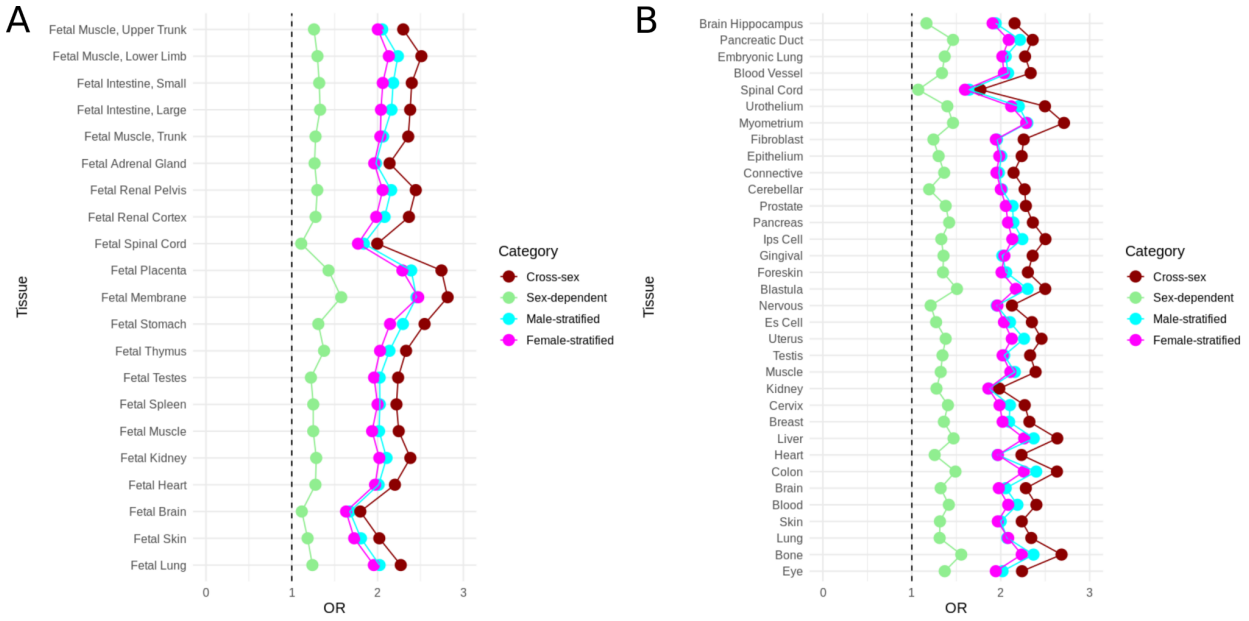

Figure 3. Functional role of placental mQTL. GARFIELD analysis was conducted with respect to the minimum mQTL p value for each SNP (STAR Methods), and the enrichment of each mQTL category is displayed here with respect to A) the average enrichment of each mQTL set across human stem cell lines in chromatin states, B) the average enrichment of mQTL SNPs in chromatin marks averaged across tissues. The proportion of CpG sites with at least one mQTL (STAR Methods for description of each mQTL category) are displayed here versus their inverse-variance weighted mean DNAm in NICHD and RICHS subjects used in mQTL analysis. We enumerate proportions C) within, outside of, upstream of or downstream of a CpG island and D) within a gene region. In E) we plot the density of CpG sites corresponding to mQTL of four categories, in addition to sites without mQTL in any of these categories.

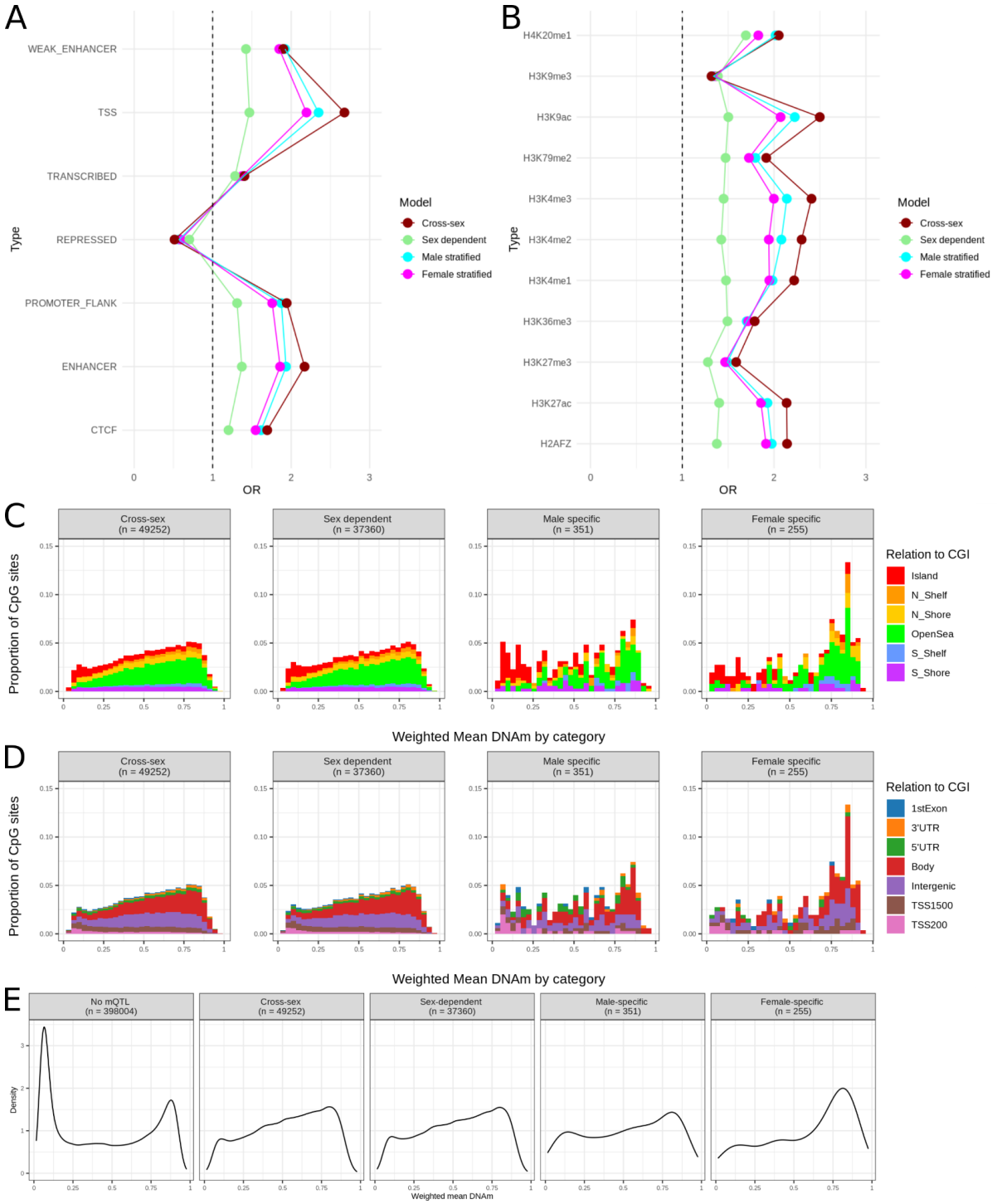

Figure 4. Enrichment of placental mQTL across ENCODE Tissues. We estimated GARFIELD enrichment for mQTL in DNAse hypersensitivity sites in A) fetal and B) adult tissues from ENCODE. Estimates are averaged over samples from the same tissue and stem cell lines (STAR Methods).

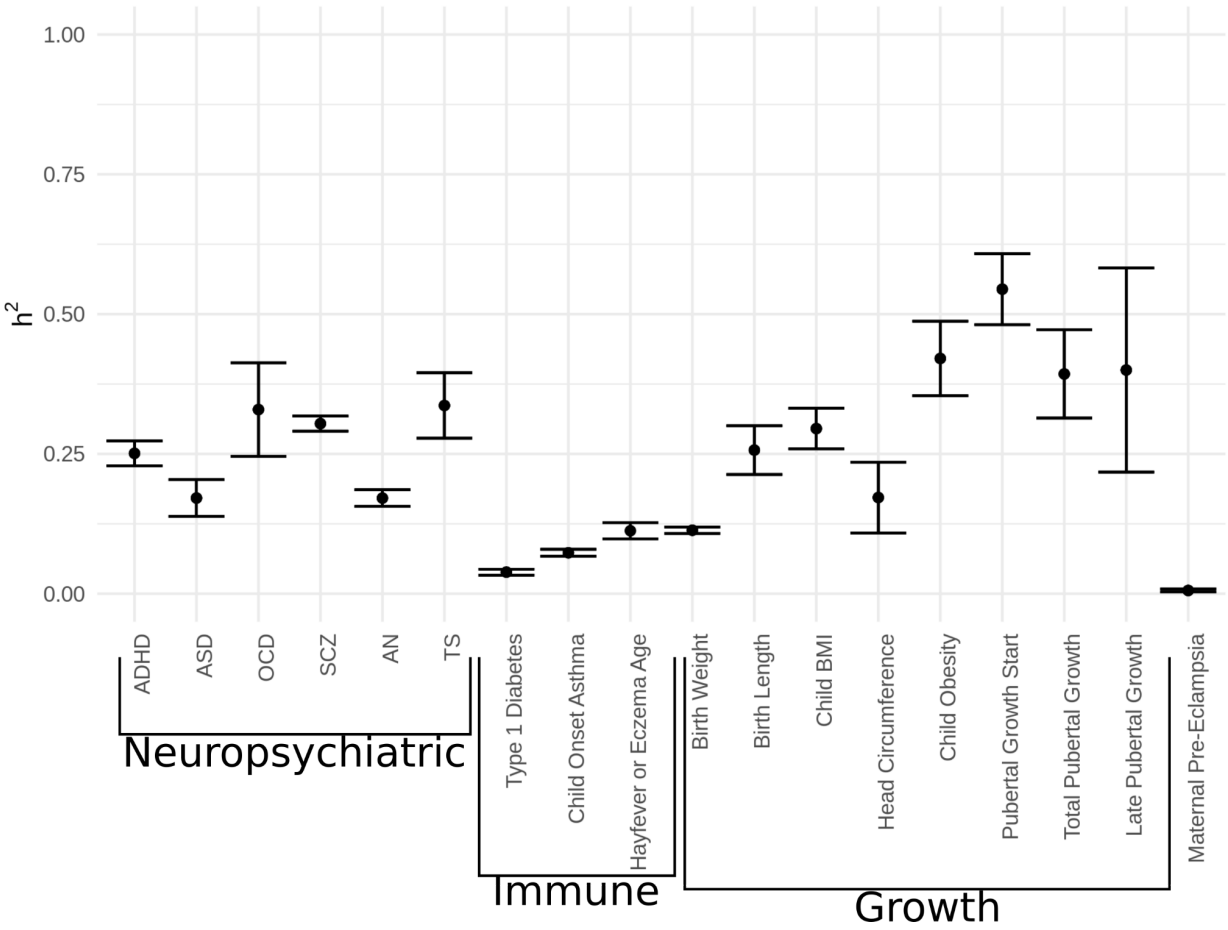

Figure 5. Heritability of traits for traits analyzed via LD score regression. SNP heritability (h2) were estimated by LD score regression for traits in Table SI meeting our criteria (STAR Methods). Traits that passed QC are plotted here for reference and labeled with their set of related traits. Estimates above are all derived from cross-sex GWAS summary statistics.

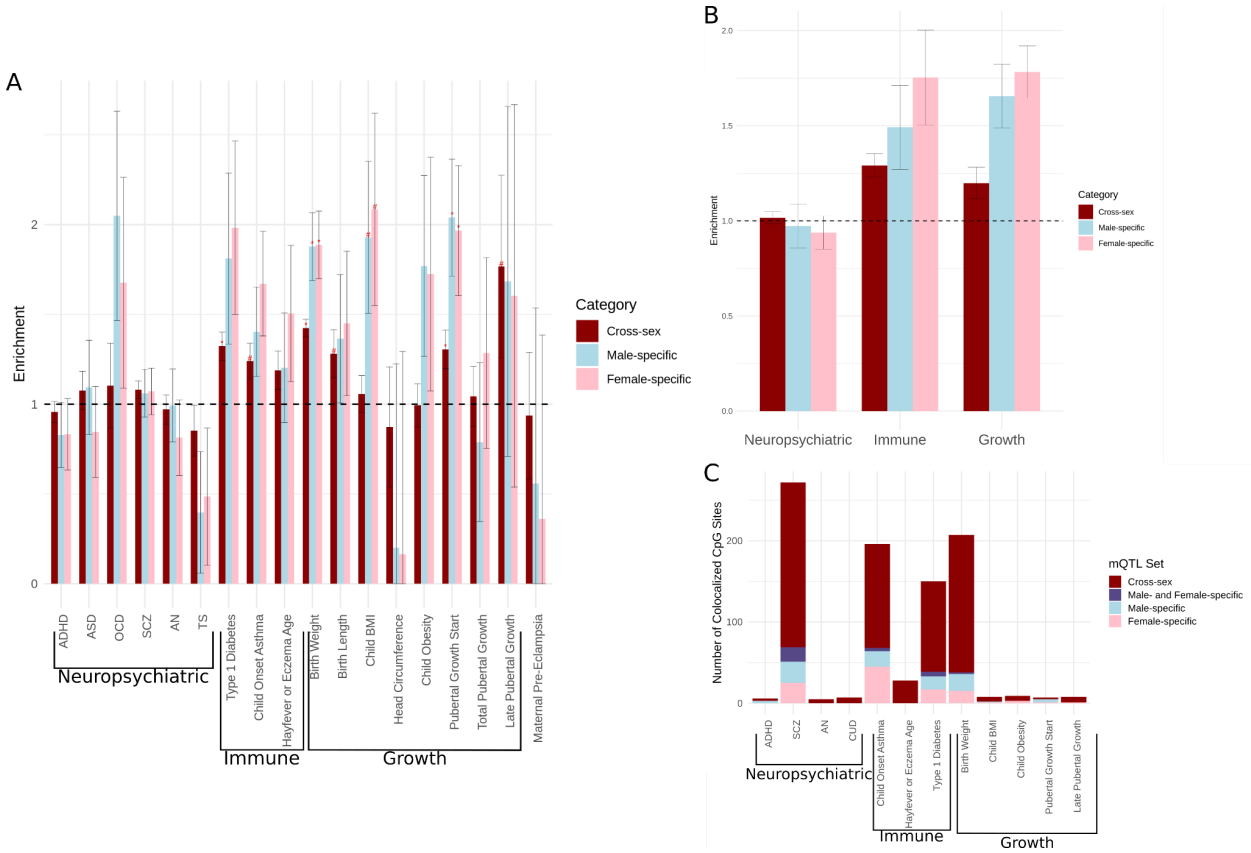

Figure 6. Enrichment of placental mQTL in 19 complex traits. The enrichment of placental mQTL in genetic associations with 19 complex traits using several methods. A) Stratified linkage disequilibrium score regression (LDSC) conducted on sets of placental mQTL accounting for the effects of 97 standard baseline regulatory effects from a variety of sources. Enrichments with a nominal p value < 0.05 are marked with #, whereas enrichments significant at an FDR < 0.05, accounting for each mQTL annotation and GWAS assessed, are marked with an asterisk. B) Estimates from a random effect meta-analysis of LDSC enrichment conducted across three categories: neuropsychiatric, immune, and growth related. C) Enumerating CpG sites with an mQTL which colocalizes (shares the same locus) with genetic trait associations at a posterior probability (H4) > 0.9.

Tables

Table 1. Counts of mQTL, associated CpG sites and corresponding genes in each mQTL set. CpG sites with at least one mQTL were included, and corresponding genes were identified by the default Illumina annotation.

mQTL Set	No. mQTL	No. CpG Sites	No. Corresponding Genes
Cross-sex	1,699,445	49,252	12,746
Sex-dependent	27,706	2,489	976
Male-specific	2,876	351	201
Female-specific	1,981	255	153
Male-stratified	865,986	31,384	9,507
Female-stratified	578,481	25,180	8,264

STAR Methods

Data description

We obtained placental genotype and DNAm data from two different studies, both of which sampled fetal side placental tissue from pregnancies lasting >=37 weeks. Discovery sample: The National Institute of Child Health and Human Development (NICHD) study[12] included 303 subjects (151 male, 152 female) of diverse reported maternal ethnicity (Table S1), the vast majority of which (96%) had no reported pre-eclampsia. Birth weight and length are not publicly available. All samples were genotyped using the Illumina HumanOmni2.5 Beadchip, and DNAm

was measured using the Illumina HumanMethylation450k array (dbGaP accession phs001717.v1.p1).

Replication sample: The Rhode Island Child Health Study (RICHS)[15,17] included 149 subjects (74 male and 75 female) of primarily non-hispanic white reported ethnicity. The samples were selected for being either large or small for gestational age (either > 90% or < 10% on the Fenton growth chart[62]). The majority of subjects (55.1%) are in the average range (i.e. in between 10% and 90%)(Table S1). No serious prenatal complications or congenital or chromosomal abnormalities were present in this sample. Samples were genotyped on the Illumina Expanded Multi-Ethnic Genotyping Array (Mega-EX) (dbGaP accession phs001586.v1.p1). DNAm was measured on the IlluminaHumanMethylation450k array (GEO accession GSE75248). DNAm and SNP genotype data were linked by a common ID file shared by Carmen J. Marsit.

Genotype data processing and quality control

Genotyping data from NICHD and RICHS were processed separately, but were subject to the same quality control (QC) and processing protocol, which was based on the RICOPILI pipeline.[63] Using plink 1.9,[64] we removed SNPs that were strand ambiguous, had a call rate <0.05, or MAF<0.01. We removed individuals with a mismatch between recorded and genotyped sex, SNP missingness >0.02, and with excess heterozygosity (Fhet>0.2). SNPs on the X chromosome were kept, using the default 0/2 encoding in plink for SNPs on the X chromosome in males.

We used the GRAFpop algorithm[65] on autosomal SNPs to assign individuals to one of 7 ancestry groups: EUR (European), AFR (African), AFR_AM (African American), LAT_AM_1 (Latin American 1), LAT_AM_2 (Latin American 2), EAS (East Asian), PAC (Asian Pacific Islander), and SAS (South Asian), which we used to help decide the number of ancestry PCs to include in downstream analyses. Genetic ancestry assignments agreed well with maternal self-reported ethnicity (Table S5). Next, to define ancestry PCs, we first pruned SNPs based on LD in a pairwise manner using plink (running --indep-pairwise 200 100 0.2 twice), snps with an MAF <0.05, and those in highly recombinant regions (the HLA region, chr6:28,477,797-33,448,354 in GRCh37, and regions of long range LD[66]). Then samples were filtered to ensure within each ancestry group identity by descent did not exceed an $|F_{ST}| > 0.2$

between samples, and 10 ancestry PCs were computed using FastPCA.[67] From these 10, we ultimately selected the first five to include in all downstream analyses on the basis of their ability to separate distinct populations of subjects (defined by self-reported race or GRAF-pop estimated ancestry) in each study (Figures S4A-D).

Before imputing SNP genotypes, we removed variants that were associated with genotype batch (Bonferroni corrected $p < 0.05$) and variants with a Hardy-Weinberg equilibrium $p < 1e\text{-}10$. Variants were then aligned and imputed to the 1000 genomes phase 3 reference cosmopolitan reference panel[68] using the Michigan imputation server. Variants were filtered to have an imputation quality $R2 > 0.3$ and an ancestry-specific MAF within 0.1 of their corresponding 1000 genomes phase 3 group. Last, imputed SNPs associated with genotype batch were removed. The NICHD study had 275 samples with 14,110,641 SNPs remaining for analysis, whereas RICHS had 136 samples with 6,139,984 SNPs remaining for analysis.

DNA methylation array processing and quality control

Illumina HumanMethylation450k array data from the NICHD study and RICHS were processed separately using the bioconductor program *minfi.*[69] Data from RICHS were first normalized for dye bias (*preprocessNoob, minfi*[69]) since the data were provided in raw array IDAT form as opposed to called beta values.

First the pattern of methylation across all probes was checked to confirm that it followed a hemi-methylated pattern characteristic of the placenta[35] (Figure 3E). Next, we removed individuals with sex mismatches, i.e., individuals with sex chromosome DNAm values that did not cluster with samples in their reported sex (2 samples total, neither of which were flagged as sex-mismatched during genotyping QC). We then removed probes with a detection $P > 0.01$, and probes that either failed or were missing, in > 20% of samples. Next we removed SNP probes (annotated by Illumina and identified by others[70]), Y chromosome probes, and non-specific cross-hybridizing probes.[71] Lastly, the data were quantile normalized using *preprocessQuantile* from the R package *minfi.*[69] After implementing these QC steps, the NICHD and RICHS studies had 447,232 and 446,976 probes remaining for analysis.

Mapping cross-sex, sex-dependent, male-stratified, and female-stratified mQTL

We conducted four different analyses to detect four different cis-mQTL effects: a cross-sex analysis, an interaction analysis (in which the effect of SNP on DNAm differed by sex, as captured by a SNP*sex interaction term), an analysis of males only, and an analysis of females only (Figure 1). Each analysis was conducted separately in NICHD and RICHS.

We used matrixEQTL[72] for mQTL mapping of methylation beta values and imputed genotype dosages. Each SNP within a 75 kb upstream or downstream of each CpG site was regressed onto DNAm, accounting for gestational age, sex, methylation array batch, 5 genetic ancestry PCs, and 9 DNAm PCs. As interaction effects (i.e. sex-dependent effects) are harder to detect, and require a mQTL effect, we elected to exclude trans-mQTL, or SNPs > 75kb from an associated CpG site, from this analysis.[59,73,74] We chose the 75 kb window as the majority of cis-mQTL in other tissues have been shown to be contained within 50kb of a CpG site.[12,75] The number of DNAm PCs to include was determined based on the number of mQTL declared on chromosome 21 at a Bonferroni corrected $p < 0.05$, varying the number of PCs included while keeping other covariates fixed and selecting the number of PCs for which the number of mQTL declared did not improve (Figures S4E,F).[76,77]

Since sex-dependent molQTL detection can be biased by sex differences in cell type proportions,[18] as an additional sanity check, we ensured that the top 10 DNAm PCs captured estimated placental cell type proportions (Figures S5C,D). Placental cell type proportions were estimated using the data available on cell-type specific methylation in placenta in the R package *PlaNET*,[78] and the Houseman algorithm in *minfi*.[69] Reassuringly, we found that cell-type proportion was most correlated with the top 5 DNAm PCs in both studies. We also checked if there was a difference in estimated cell-type proportion between males and females in each study (Figures S5A,B). We observed a sex difference in average estimated stromal cell proportion in the NICHD study, but these would have been accounted for by the DNAm PCs.

Before meta-analyzing results from the NICHD and RICHS samples, we tested cross-sex and sex-dependent mQTL identified in NICHD for replication in RICHS using the π_1 statistic,[79] which is a better measure of comparability than the raw proportion of overlapping sites because it accounts for between-study differences in the number of tested SNP-CpG pairs, which can artificially inflate or deflate estimates of overlap. For the π_1 analysis, we relied on cross-sex and sex-dependent mQTL called in each study at an FDR < 0.05 because calling mQTL at a

Bonferroni corrected $p < 0.05$ resulted in unstable estimates of π_1. We identified cross-sex mQTL with linear models testing for association between genotype and DNAm, and associations were declared statistically significant at a liberal threshold (FDR < 0.05). Sex-dependent mQTL were those with a statistically significant (FDR < 0.05) SNP*sex interaction term. In the NICHD study, we found 275,806 CpGs with at least one mQTL and 60537 CpGs with at least one mQTL that interacted with sex. As expected, mQTL previously called in the NICHD study by Delahaye et al.[12] were all called with our model, even when we used a more stringent threshold (Bonferroni corrected $p < 0.05$), which resulted in 60,320 CpG sites with at least one significant mQTL. In RICHS, we found 129,361 CpGs with at least one mQTL and 19,530 CpGs with at least one mQTL that differed by sex. The π_1 estimate for the cross-sex mQTL was 0.74, which is the proportion of mQTL in the NICHD study that have similar effects in the RICHS study (computed over 2,691,024 SNP-CpG pairs available in both studies). Sex-dependent mQTL showed modest replication ($\pi_1 = 0.28$, computed over 80,363 SNP-CpG pairs available in both studies), which is in line with values previously reported for sex-dependent eQTL in GTEx tissues ($\pi_1 = 0.28$ in breast tissue, π_1 ranging from 0-0.12 in the remaining 43 GTEx V8 tissues analyzed by Oliva et al.).[18]

Reassured by the above result, we then meta-analyzed across the NICHD and RICHS samples for each of the four analyses using MeCS software,[24] which was designed to account for the highly correlated nature of molecular cis-QTL effects between independent studies. Within each analysis, we retained only statistically significant (Bonferroni corrected $p < 0.05$) mQTL, or mQTL with a statistically significant SNP*sex interaction effect in the interaction analysis. From these four analyses, we defined six sets of mQTL: (i) cross-sex; (ii) sex-dependent; (iii) male-stratified; (iv) female-stratified; (v) male-specific; and (vi) female-specific. The (v) male- and (vi) female-specific mQTL were those with significant stratified and interaction effects (i.e. the intersection of sex-dependent and male- or female-stratified mQTL). The reported effect size for each male- and female–specific mQTL was taken from male- and female-stratified mQTL. In counting mQTL, associated CpG sites, and their targeted genes, SNPs with a MAF < 0.05 were excluded in order to ensure mQTL were less likely to be a result of outliers in DNAm.

As a point of comparison for our meta-analyzed beta values, allelic fold-change (aFC)[28] was computed for each effect in males and females to sex-stratified analysis was conducted for sbmQTL called at an FDR < 0.05 in NICHD. Correlation between male and female cis-mQTL

allelic fold-change in sex-biased mQTL in NICHD cohort was moderate (Spearman's rank correlation $\rho = 0.16$), with the allelic fold change in female-specific mQTL (aFC = 0.09) being significantly smaller on average than in males (aFC = 0.14) (Wilcoxon rank sum test, $p < 4.3e-4$). In fact, the average absolute aFC was larger in males than females(0.32 in males, 0.29 in females, Wilcoxon rank sum test $p < 3.0e-7$), and upon excluding the X chromosome this difference remained only in absolute aFC (0.31 in males, 0.29 in females wilcoxon rank sum test $P < 5.1e-6$), with no significant difference in signed aFC between sexes (Wilcoxon rank sum test $p < 0.22$). The difference in aFC observed between males and females for mQTL on the X chromosome was large on average (0.51 vs 0.33 absolute aFC in males versus females Wilcoxon rank sum $P < 4.0e-7$, 0.32 vs 0.10 signed aFC in males vs females Wilcoxon rank sum test $P < 3.09e-38$), with effects showing a larger correlation between males and females (Spearman's rank correlation $\rho = 0.37$).

Comparing cross-sex placental mQTL to other perinatal tissues

As a main focus of this study is outlining the developmental origins of complex traits, we compared placental mQTL to mQTL from other tissues collected before and after birth. We accessed mQTL summary statistics for umbilical cord blood mQTL from the accessible resource for integrative epigenomic studies (ARIES; N=711; available via http://www.mqtldb.org/), originally from the Avon longitudinal study of parents and children,[29] and fetal brain mQTL summary statistics from the human developmental biology resource[80] (HDBR; N=173; available via https://epigenetics.essex.ac.uk/mQTL/). The ARIES study was comprised of 51% male samples, and in mQTL mapping authors accounted for sex, 10 ancestry PCs, batch, and cell-type proportion estimates for white blood cell counts. mQTL in the ARIES study were called at a $P < 1 * 10^{-14}$ for all pairwise associations between 8,074,398 imputed SNPs at an MAF > 0.05 and 395,625 CpG sites passing their QC. This is more stringent than our bonferroni corrected alpha < 0.05 computed for associations computed in cis, which corresponds to a threshold of $P < 3.77 * 10^{-10}$. Meanwhile, the HDBR study in homogenized fetal brain tissue consisted of 54% male samples ranging from 8-24.1 weeks post conception. When mapping mQTL, they accounted for sex, age, and 2 genotyping PCs. Additionally, they had a more stringent methylation QC and included only directly genotyped SNPs at MAF > 0.05, resulting in 430,304 SNPs and 314,554 CpG sites tested for all pairwise associations. mQTL were called at a Bonferroni-corrected

threshold of $P < 3.69 * 10^{-13}$. Notably, the HDBR study excluded the sex chromosomes from their mQTL analysis.

GARFIELD analysis

We used GARFIELD[31] to quantify the enrichment of mQTL in different genomic regions across tissues, accounting for linkage disequilibrium (LD) and redundant annotations (i.e. annotations with similar enrichment to one another across mQTL). We used the default genomic regions provided by the software developers. GARFIELD was run for all 6 mQTL sets by mapping each mQTL to its minimum p value for association with DNAm regardless of the probe with which it was associated. We assessed the enrichment of mQTL significant at a threshold of $1 * 10^{-9}$ (accounting for the roughly 100 million SNP-CpG pairs tested in cross-sex, sex-dependent, male-stratified, or female-stratified mQTL analysis). To simplify our results, we report the mean odds ratio over annotations that come from the same tissue and regulatory mark.

Gene set enrichment

CpG sites were mapped to genes using the default Illumina encoding published with the 450k Array. Gene set enrichment analyses were performed using FUMA[81] GENE2FUNCTION, excluding genes that fall within the major histocompatibility complex (MHC region). Genes were selected for each set based on whether or not a CpG site with at least one significant mQTL was annotated to them. Using FUMA we conducted a series of hypergeometric tests, for which we treated each mQTL-based gene set as selected from the universe of all genes on the 450K array annotation. Enrichments featured here include genes differentially expressed up or down in one of 54 GTEx v8 tissues relative to all remaining tissues.

In addition to these sets, we included enrichment based on median expression of genes across GTEx v8 tissues that are publicly available for download. Whereas FUMA is concerned with whether genes are over-represented in genes differentially regulated in a single tissue, this tested whether genes corresponding to each mQTL set were enriched in genes expressed across all 54 GTEx v8 tissues. We declared a gene to be ubiquitously expressed based on whether or not its median TPM was greater than either 5, 10, 15, or 20 in all tissues. Enrichment was assessed using a hypergeometric test, again with all 450K array genes matching reported genes from GTEx as a background set.

Linkage disequilibrium score regression analyses

We estimated GWAS trait heritability and genetic correlations using LDSC software[21,22,82,83] (https://github.com/bulik/ldsc). We partitioned heritability by placental mQTL sets using S-LDSC. S-LDSC is most powerful when the annotation (in this case, mQTL sets) cover at least 1% of the genome. To fit this criterion, and to maximize enrichment in our mQTL sets, we re-defined the placental mQTL by applying a previously developed method.[23] Briefly, for each CpG site with at least one mQTL at an FDR < 0.05, we fine-mapped meta-analyzed mQTL associations with a nominal $p < 0.05$ using CAVIAR[23] The output from CAVIAR is a set of SNPs with a 95% likelihood of containing the SNP causal to changes in DNAm (i.e. 95% credible set). As fine-mapping for interaction effects is difficult to interpret, we did not consider sex-dependent mQTL as its own category for this analysis. Instead for CpG sites with at least one male-stratified or male-specific mQTL, as well as female-stratified or female-stratified mQTL, we ran fine-mapping with respect to male- and female-stratified mQTL p values respectively. We chose not to use the maximum causal posterior probability (max CPP) measure for each SNP for our annotations as the effective sum of our male- and female-specific annotations was small, less than 50,000 compared to the over 1 million per annotation that we achieved when using the 95% credible set (i.e. assigning 95% credible set SNPs a value of 1, Figures S5F,G). For an estimate of the number of CpG sites that may be affected by SNPs in cis, we computed narrow-sense heritability using GCTA[84–86] assessing all SNPs within 75kb of a CpG site with respect to its M-value, accounting for all covariates included in cis-mQTL analyses. This ensures that CpG sites we consider in our annotation are more likely to be explained by some number of additive genetic effects. Figure S5E shows the distribution of h^2 values for CpG sites passing with a heritability $p < 0.10$, which was roughly 164,434 CpG sites.

Once we had defined the mQTL sets for use in S-LDSC, the stratified LD scores for our annotations (i.e., mQTL sets) were built using all subjects available in the 1000 Genomes Phase III EUR population, again using the author's software, their publicly available baseline v2.2 LD scores (consisting of 97 annotations), and excluding SNPs within the MHC region of the human genome.[21,22] The EUR population was chosen as opposed to a transancestral cohort since LDSC

has only been evaluated on EUR and EAS cohorts with ancestrally matched summary statistics, with little documentation of the effects of admixture,[21–23,82] and the majority of GWAS used in our analyses were in European populations (Availability and criteria for GWAS summary statistics). For age of hayfever or eczema diagnosis, which was computed both with respect to raw age and normalized age, we present raw age as it has a higher heritability z-score as determined via LD-score regression (Figure 5,S2C, Table SI).

We report enrichment as the proportion of trait heritability explained by the SNPs in each mQTL set, divided by the proportion of SNPs included in the mQTL set. We also quantify enrichment using τ^* (Table SI), which is defined previously as:[82]

$$\tau^* = \frac{\tau\, s.d.(c)}{h_g^2/M}$$

Where h_g^2 is the SNP heritability of the trait, s.d.(c) is the standard deviation of the annotation, τ is the coefficient of that annotation, and M is the number of variants used to compute heritability. Thus, tau can be interpreted as the standardized average contribution of SNPs in an annotation to the total SNP-heritability of a trait. An advantage of τ^* over enrichment is that it quantifies effects that are specific to that annotation, i.e., after taking into account the overlap between the annotation of interest, and other annotations in the model. An elevated tau value suggests that enrichment is not explained by overlap with other annotations in the model. Negative τ^* values indicate that the annotation on its own reduces h^2_{SNP}, on average.

S-LDSC was run separately for each of the 6 mQTL sets, accounting for the 97 baseline annotations. Following S-LDSC analysis in single traits, we meta-analyzed enrichment and τ^* estimates across trait groups using the *rmeta*[87] package. Estimates from trait categories with sex-stratified GWAS summary statistics available were analyzed with mQTL annotations matching the sex of their component GWAS (GWAS labelled _female or _male in Table S4). As in Hormozdiari et al, 2018,[23] we note that childhood BMI and childhood obesity have a genetic correlation of > 0.9, leaving us with 20 independent traits. τ^* and its standard error from meta-analysis were used to generate a normally distributed Z score for computing meta-analyzed p values. No such procedure is currently defined for computing significance from meta-analyzed S-LDSC enrichment values, so significance is not computed.

Colocalization of placental mQTL and GWAS loci

We performed a colocalization analysis to characterize the extent to which the proportion of heritability attributed to placental mQTL corresponded to shared genetic loci between mQTL and GWAS signals. We focused on all sets of SNPs within 75kb of CpG sites with a mQTL associated at a $p < 5e-8$, that were also associated with a GWAS trait at $p < 5e-8$, and we computed colocalization using the *coloc* R package.[41] Notably, unlike our protocol in S-LDSC, we included the HLA region in colocalization analysis given that this analysis is not dependent on genome-wide patterns of LD, and the HLA regions are of particular interest to the immune-related traits analyzed. All genes considered to be within the HLA region were selected from chr6:28,477,797-33,448,354 in genome build hg19, and were gathered from their corresponding table on the UCSC genome browser at http://genome.ucsc.edu.[88] Colocalization was computed with cross-sex GWAS summary statistics. We defined colocalization as having a probability of Hypothesis 4 > 0.9, which corresponds to the likelihood of both the level of DNAm and the trait sharing a single causal variant in that region.

In plotting these results (Figure 6C), we wanted to highlight the colocalizations that were found only through sex-stratified analyses. Therefore, all colocalizations identified in the cross-sex analysis were labeled as such, even if they were also identified in the male- and female-stratified analyses. A small number of colocalizations (N=30 CpG sites annotated to N=4 genes) were identified in both male- and female-stratified analyses, but not in cross-sex analyses. We label these "Male- and Female-specific" colocalizations. Finally, we plot colocalizations that were found in the male- and female-stratified analyses only. We include additional GWAS in this analysis that did not meet their SNP-heritability z-score[21,22] threshold, but still have a median age of onset < 25,[38] which is roughly considered to be the end of brain maturation and adolescence[39] (Availability and criteria for GWAS summary statistics).

Availability and criteria for GWAS summary statistics

For neuropsychiatric traits, we made the choice to only include traits which have a median age of onset < 25,[38] which is roughly considered to be the end of brain maturation and adolescence.[39]

Summary statistics were formatted for all LD score analyses using the munge_sumstats.py script made available by the LDSC authors. This includes the following quality control measures: only indel SNPs are kept, strand ambiguous SNPs are excluded, duplicate SNPs are excluded, SNPs have an imputation INFO score > 0.9, a MAF > 0.01, $0 < p <= 1$, SNPs with a number of samples < 90th quantile divided by 1.5 are excluded, and the median Z score of all SNPs are within 0.1 of its expected null value (0 for signed statistics, 1 for an odds ratio). This largely matches what has been done in previous publications dealing with sex-stratified summary statistics, with the exception of differences MAF threshold used across different studies[60,89]

Genome-wide summary statistics for all cross-sex traits in the Psychiatric Genetics Consortium (PGC), and for those related to developmental traits, were downloaded in April 2021 and are publicly available (Table SI). Sex-stratified GWAS from PGC were obtained from Martin et al, 2021,[60] and they are described in their study. Summary statistics for preeclampsia were requested separately via an InterPreGen access request. In S-LDSC and genetic correlation analyses, we imposed a filter of a GWAS SNP heritability z-score[21,22] > 1.644854, corresponding to a trait heritability estimate that was significant at a nominal $p < 0.05$.

Large Supplemental Table Legends

Table SI. Information and accession for GWAS considered in this study. All GWAS considered in this study, along with various metrics for their inclusion and the status of their inclusion in stratified linkage disequilibrium score (S-LDSC) analysis.

Table SII. Stratified LD score regression results for all mQTL sets. Stratified LD score regression was performed with respect to each sex-aware mQTL set and GWAS passing our inclusion criteria (as outlined in STAR Methods, Table SI). This table includes enrichment estimates and tau star coefficients for cross-sex, male- and female-specific, and male- and female-stratified mQTL.

Table SIII. Counts of CpG sites and targeted genes colocalized with GWAS loci across cross-sex, male- and female-stratified placental mQTL. Colocalization analysis was

performed between cross-sex, male-stratified, and female-stratified mQTL p values and genome wide significant GWAS loci using *coloc* (STAR Methods). Colocalized CpG sites (p (H4) > 0.9) with at least one mQTL were labeled with their target gene from Illumina's HumanMethylation450k Array annotation, and then enumerated for each mQTL set, with overlap in colocalization between groups with respect to CpG sites and their target gene is counted here.

References

1. King EA, Davis JW, Degner JF. Are drug targets with genetic support twice as likely to be approved? Revised estimates of the impact of genetic support for drug mechanisms on the probability of drug approval. PLoS Genet. 2019 Dec 12;15(12):e1008489.
2. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015 Aug;47(8):856–60.
3. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. Science. 2012 Sep 7;337(6099):1190–5.
4. Schork AJ, Thompson WK, Pham P, Torkamani A, Roddey JC, Sullivan PF, et al. All SNPs Are Not Created Equal: Genome-Wide Association Studies Reveal a Consistent Pattern of Enrichment among Functionally Annotated SNPs. PLoS Genet. 2013 Apr 25;9(4):e1003449.
5. THE GTEX CONSORTIUM. The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 2020 Sep 11;369(6509):1318–30.
6. Gamazon ER, Segrè AV, van de Bunt M, Wen X, Xi HS, Hormozdiari F, et al. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. Nat Genet. 2018 Jul;50(7):956–67.
7. Umans BD, Battle A, Gilad Y. Where Are the Disease-Associated eQTLs? Trends in Genetics [Internet]. 2020 Sep 7 [cited 2020 Nov 18]; Available from: http://www.sciencedirect.com/science/article/pii/S0168952520302092
8. Do C, Shearer A, Suzuki M, Terry MB, Gelernter J, Greally JM, et al. Genetic–epigenetic interactions in cis: a major focus in the post-GWAS era. Genome Biol [Internet]. 2017 Jun 19 [cited 2020 Nov 23];18. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5477265/
9. Min JL, Hemani G, Hannon E, Dekkers KF, Castillo-Fernandez J, Luijk R, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. Nat Genet. 2021 Sep;53(9):1311–21.
10. Pierce BL, Tong L, Argos M, Demanelis K, Jasmine F, Rakibuz-Zaman M, et al. Co-occurring expression and methylation QTLs allow detection of common causal variants and shared biological mechanisms. Nature Communications [Internet]. 2018 [cited 2022 Jun 6];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5824840/
11. Kikas T, Rull K, Beaumont RN, Freathy RM, Laan M. The Effect of Genetic Variation on the Placental Transcriptome in Humans. Front Genet. 2019 Jun 11;10:550.
12. Delahaye F, Do C, Kong Y, Ashkar R, Salas M, Tycko B, et al. Genetic variants influence on the placenta regulatory landscape. PLOS Genetics. 2018 Nov 19;14(11):e1007785.

13. Tekola-Ayele F, Zeng X, Chatterjee S, Ouidir M, Lesseur C, Hao K, et al. Placental multi-omics integration identifies candidate functional genes for birthweight. Nat Commun. 2022 May 2;13(1):2384.
14. Bhattacharya A, Li Y, Love MI. MOSTWAS: Multi-Omic Strategies for Transcriptome-Wide Association Studies. PLOS Genetics. 2021 Mar 8;17(3):e1009398.
15. Peng S, Deyssenroth MA, Di Narzo AF, Lambertini L, Marsit CJ, Chen J, et al. Expression quantitative trait loci (eQTLs) in human placentas suggest developmental origins of complex diseases. Hum Mol Genet. 2017 Sep 1;26(17):3432–41.
16. Peng S, Deyssenroth MA, Narzo AFD, Cheng H, Zhang Z, Lambertini L, et al. Genetic regulation of the placental transcriptome underlies birth weight and risk of childhood obesity. PLOS Genetics. 2018 Dec 31;14(12):e1007799.
17. Appleton AA, Murphy MA, Koestler DC, Lesseur C, Paquette AG, Padbury JF, et al. Prenatal Programming of Infant Neurobehaviour in a Healthy Population. Paediatr Perinat Epidemiol. 2016 Jul;30(4):367–75.
18. Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, Wucher V, Gewirtz ADH, Cotter DJ, et al. The impact of sex on gene expression across human tissues. Science [Internet]. 2020 Sep 11 [cited 2020 Sep 15];369(6509). Available from: https://science.sciencemag.org/content/369/6509/eaba3066
19. Martin E, Smeester L, Bommarito PA, Grace MR, Boggess K, Kuban K, et al. Sexual epigenetic dimorphism in the human placenta: implications for susceptibility during the prenatal period. Epigenomics. 2017 Mar;9(3):267–78.
20. Inkster AM, Yuan V, Konwar C, Matthews AM, Brown CJ, Robinson WP. A cross-cohort analysis of autosomal DNA methylation sex differences in the term placenta. Biology of Sex Differences. 2021 May 27;12(1):38.
21. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. Nature Genetics. 2015 Mar 2;47(3):291–5.
22. Finucane HK, Bulik-Sullivan B, Gusev A, Trynka G, Reshef Y, Loh PR, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nature genetics. 2015 Nov;47(11):1228–35.
23. Hormozdiari F, Gazal S, van de Geijn B, Finucane HK, Ju CJT, Loh PR, et al. Leveraging molecular quantitative trait loci to understand the genetic architecture of diseases and complex traits. Nature Genetics. 2018 Jul;50(7):1041–7.
24. Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. Nature Communications. 2018 Jun 11;9(1):2282.
25. Newcombe RG. Two-sided confidence intervals for the single proportion: comparison of seven methods. Statistics in Medicine. 1998;17(8):857–72.
26. Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Statistics in Medicine. 1998;17(8):873–90.
27. Wilson EB. Probable Inference, the Law of Succession, and Statistical Inference. Journal of the American Statistical Association. 1927 Jun 1;22(158):209–12.
28. Mohammadi P, Castel SE, Brown AA, Lappalainen T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. Genome Res. 2017 Nov;27(11):1872–84.
29. Gaunt TR, Shihab HA, Hemani G, Min JL, Woodward G, Lyttleton O, et al. Systematic

identification of genetic influences on methylation across the human life course. Genome Biology. 2016 Mar 31;17(1):61.
30. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation QTLs in the developing brain and their enrichment in schizophrenia risk loci. Nat Neurosci. 2016 Jan;19(1):48–54.
31. Iotchkova V, Ritchie GRS, Geihs M, Morganella S, Min JL, Walter K, et al. GARFIELD classifies disease-relevant genomic features through integration of functional annotations with association signals. Nat Genet. 2019 Feb 1;51(2):343–53.
32. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): data portal update. Nucleic Acids Res. 2018 04;46(D1):D794–801.
33. Karmodiya K, Krebs AR, Oulad-Abdelghani M, Kimura H, Tora L. H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. BMC Genomics. 2012 Aug 24;13(1):424.
34. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet. 2008 Jul;40(7):897–903.
35. Konwar C, Del Gobbo G, Yuan V, Robinson WP. Considerations when processing and interpreting genomics data of the placenta. Placenta. 2019 Sep 1;84:57–62.
36. Edgar R, Tan PPC, Portales-Casamar E, Pavlidis P. Meta-analysis of human methylomes reveals stably methylated sequences surrounding CpG islands associated with high gene expression. Epigenetics & Chromatin. 2014 Oct 23;7(1):28.
37. Infinium HumanMethylation450 BeadChip. :4.
38. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. Mol Psychiatry. 2021 Jun 2;1–15.
39. Arain M, Haque M, Johal L, Mathur P, Nel W, Rais A, et al. Maturation of the adolescent brain. Neuropsychiatr Dis Treat. 2013;9:449–61.
40. Steinthorsdottir V, McGinnis R, Williams NO, Stefansdottir L, Thorleifsson G, Shooter S, et al. Genetic predisposition to hypertension is associated with preeclampsia in European and Central Asian women. Nat Commun. 2020 Nov 25;11(1):5976.
41. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 2014 May 15;10(5):e1004383.
42. Pardiñas AF, Holmans P, Pocklington AJ, Escott-Price V, Ripke S, Carrera N, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nature Genetics. 2018 Mar;50(3):381–9.
43. Ferreira MAR, Mathur R, Vonk JM, Szwajda A, Brumpton B, Granell R, et al. Genetic Architectures of Childhood- and Adult-Onset Asthma Are Partly Distinct. Am J Hum Genet. 2019 Apr 4;104(4):665–84.
44. Pividori M, Schoettler N, Nicolae DL, Ober C, Im HK. Shared and Distinct Genetic Risk Factors for Childhood Onset and Adult Onset Asthma: Genome- and Transcriptome-wide Studies. Lancet Respir Med. 2019 Jun;7(6):509–22.
45. Redondo MJ, Steck AK, Pugliese A. Genetics of type 1 diabetes. Pediatr Diabetes. 2018 May;19(3):346–53.

46. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(Web Server issue):W90–7.
47. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. Current Protocols. 2021;1(3):e90.
48. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 Apr 15;14:128.
49. Gruen JR, Weissman SM. Human MHC class III and IV genes and disease associations. Frontiers in Bioscience-Landmark. 2001 Aug 1;6(3):960–72.
50. Safran M, Rosen N, Twik M, BarShir R, Stein TI, Dahary D, et al. The GeneCards Suite. In: Abugessaisa I, Kasukawa T, editors. Practical Guide to Life Science Databases [Internet]. Singapore: Springer; 2021 [cited 2022 May 3]. p. 27–56. Available from: https://doi.org/10.1007/978-981-16-5812-9_2
51. Elliott G, Hong C, Xing X, Zhou X, Li D, Coarfa C, et al. Intermediate DNA methylation is a conserved signature of genome regulation. Nat Commun. 2015 Feb 18;6(1):6363.
52. Bhattacharya A, Freedman AN, Avula V, Harris R, Liu W, Pan C, et al. Placental genomics mediates genetic associations with complex health traits and disease. Nat Commun. 2022 Feb 4;13(1):706.
53. Randall JC, Winkler TW, Kutalik Z, Berndt SI, Jackson AU, Monda KL, et al. Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. PLOS Genetics. 2013 Jun 6;9(6):e1003500.
54. Barker DJP, Thornburg KL. Placental programming of chronic diseases, cancer and lifespan: A review. Placenta. 2013 Oct 1;34(10):841–5.
55. Barker DJP, Thornburg KL. The Obstetric Origins of Health for a Lifetime. Clinical Obstetrics & Gynecology. 2013 Sep;56(3):511–9.
56. Del Gobbo GF, Konwar C, Robinson WP. The significance of the placental genome and methylome in fetal and maternal health. Hum Genet. 2020 Sep 1;139(9):1183–96.
57. Schoettler N, Rodríguez E, Weidinger S, Ober C. Advances in Asthma and Allergic Disease Genetics – Is Bigger Always Better? J Allergy Clin Immunol. 2019 Dec;144(6):1495–506.
58. Yan Q, Forno E, Herrera-Luis E, Pino-Yanes M, Yang G, Oh S, et al. A genome-wide association study of asthma hospitalizations in adults. Journal of Allergy and Clinical Immunology. 2021 Mar 1;147(3):933–40.
59. Zhang P, Lewinger JP, Conti D, Morrison JL, Gauderman WJ. Detecting Gene-Environment Interactions for a Quantitative Trait in a Genome-Wide Association Study. Genet Epidemiol. 2016 Jul;40(5):394–403.
60. Martin J, Khramtsova EA, Goleva SB, Blokland GAM, Traglia M, Walters RK, et al. Examining Sex-Differentiated Genetic Effects Across Neuropsychiatric and Behavioral Traits. Biological Psychiatry. 2021 Jun 15;89(12):1127–37.
61. Gay NR, Gloudemans M, Antonio ML, Abell NS, Balliu B, Park Y, et al. Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. Genome Biology. 2020 Sep 11;21(1):233.
62. Fenton TR, Kim JH. A systematic review and meta-analysis to revise the Fenton growth chart for preterm infants. BMC Pediatrics. 2013 Apr 20;13(1):59.
63. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al.

RICOPILI: Rapid Imputation for COnsortias PIpeLIne. Bioinformatics. 2020 Feb 1;36(3):930–3.
64. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience [Internet]. 2015 Dec 1 [cited 2020 Nov 3];4(1). Available from: http://academic.oup.com/gigascience/article/4/1/s13742-015-0047-8/2707533
65. Jin Y, Schaffer AA, Feolo M, Holmes JB, Kattman BL. GRAF-pop: A Fast Distance-Based Method To Infer Subject Ancestry from Multiple Genotype Datasets Without Principal Components Analysis. G3 (Bethesda). 2019 May 31;9(8):2447–61.
66. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc. 2010 Sep;5(9):1564–73.
67. Galinsky KJ, Bhatia G, Loh PR, Georgiev S, Mukherjee S, Patterson NJ, et al. Fast Principal-Component Analysis Reveals Convergent Evolution of ADH1B in Europe and East Asia. Am J Hum Genet. 2016 Mar 3;98(3):456–72.
68. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015 Oct;526(7571):68–74.
69. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA Methylation microarrays. Bioinformatics. 2014;30(10):1363–9.
70. Chen Y an, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics. 2013 Feb 1;8(2):203–9.
71. Price EM, Robinson WP. Adjusting for Batch Effects in DNA Methylation Microarray Data, a Lesson Learned. Frontiers in Genetics [Internet]. 2018 [cited 2020 May 4];9. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5864890/
72. Andrey A. Shabalin. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012 May 15;28(10):1353–8.
73. Bookman EB, McAllister K, Gillanders E, Wanke K, Balshaw D, Rutter J, et al. Gene-Environment Interplay in Common Complex Diseases: Forging an Integrative Model—Recommendations From an NIH Workshop. Genet Epidemiol. 2011 May;35(4):217–25.
74. Assary E, Vincent JP, Keers R, Pluess M. Gene-environment interaction and psychiatric disorders: Review and future directions. Seminars in Cell & Developmental Biology. 2018 May 1;77:133–43.
75. Do C, Lang CF, Lin J, Darbary H, Krupska I, Gaba A, et al. Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. Am J Hum Genet. 2016 May 5;98(5):934–55.
76. Mostafavi S, Battle A, Zhu X, Urban AE, Levinson D, Montgomery SB, et al. Normalizing RNA-Sequencing Data by Modeling Hidden Covariates with Prior Knowledge. PLoS One [Internet]. 2013 Jul 18 [cited 2020 Mar 3];8(7). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3715474/
77. Ng B, White CC, Klein HU, Sieberts SK, McCabe C, Patrick E, et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. Nature Neuroscience. 2017 Sep 4;20(10):1418–26.
78. Yuan V, Hui D, Yin Y, Peñaherrera M, Beristain A, Robinson W. Cell-specific

Characterization of the Placental Methylome [Internet]. In Review; 2020 Oct [cited 2020 Nov 18]. Available from: https://www.researchsquare.com/article/rs-38223/v3
79. Storey JD, Tibshirani R. Statistical significance for genomewide studies. PNAS. 2003 Aug 5;100(16):9440–5.
80. Hannon E, Spiers H, Viana J, Pidsley R, Burrage J, Murphy TM, et al. Methylation quantitative trait loci in the developing brain and their enrichment in schizophrenia-associated genomic regions. Nat Neurosci. 2016 Jan;19(1):48–54.
81. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nature Communications. 2017 Nov 28;8(1):1826.
82. Gazal S, Finucane HK, Furlotte NA, Loh PR, Palamara PF, Liu X, et al. Linkage disequilibrium–dependent architecture of human complex traits shows action of negative selection. Nat Genet. 2017 Oct;49(10):1421–7.
83. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015 Nov;47(11):1236–41.
84. Yang J, Lee SH, Goddard ME, Visscher PM. Genome-Wide Complex Trait Analysis (GCTA): Methods, Data Analyses, and Interpretations. In Humana Press, Totowa, NJ; 2013 [cited 2019 Aug 20]. p. 215–36. Available from: http://link.springer.com/10.1007/978-1-62703-447-0_9
85. Cebamanos L, Gray A, Stewart I, Tenesa A. Regional heritability advanced complex trait analysis for GPU and traditional parallel architectures. Bioinformatics. 2014 Apr 15;30(8):1177–9.
86. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, et al. A gene-based association method for mapping traits using reference transcriptome data. Nature Genetics. 2015 Sep 10;47(9):1091–8.
87. Lumley T. rmeta: Meta-Analysis [Internet]. 2018. Available from: https://CRAN.R-project.org/package=rmeta
88. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, et al. The Human Genome Browser at UCSC. Genome Res. 2002 Jan 6;12(6):996–1006.
89. Bernabeu E, Canela-Xandri O, Rawlik K, Talenti A, Prendergast J, Tenesa A. Sex differences in genetic architecture in the UK Biobank. Nat Genet. 2021 Sep;53(9):1283–9.
90. Hormozdiari F, Kostem E, Kang EY, Pasaniuc B, Eskin E. Identifying Causal Variants at Loci with Multiple Signals of Association. Genetics. 2014 Oct;198(2):497–508.
91. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Available from: https://www.R-project.org/
92. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the tidyverse. Journal of Open Source Software. 2019;4(43):1686.
93. Dowle M, Srinivasan A. data.table: Extension of `data.frame` [Internet]. 2021. Available from: https://CRAN.R-project.org/package=data.table
94. Triche TJ, Weisenberger DJ, Berg DVD, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. Nucleic Acids Research. 2013;41(7):e90.
95. Wei T, Simko V. R package “corrplot”: Visualization of a Correlation Matrix [Internet]. 2021. Available from: https://github.com/taiyun/corrplot