Passive *r*GE or Developmental Gene-Environment Cascade? An Investigation of the Role of Xenobiotic Metabolism Genes in the Association Between Smoke Exposure During Pregnancy and Child Birth Weight

Kristine Marceau[1,2] · Rohan H. C. Palmer[1,3] · Jenae M. Neiderhiser[4] · Taylor F. Smith[5] · John E. McGeary[1,3] · Valerie S. Knopik[1,3]

Abstract There is considerable evidence that smoke exposure during pregnancy (SDP) environmentally influences birth weight after controlling for genetic influences and maternal characteristics. However, maternal smoking during pregnancy—the behavior that leads to smoke exposure during pregnancy—is also genetically-influenced, indicating the potential role of passive gene-environment correlation. An alternative to passive gene-SDP correlation is a cascading effect whereby maternal and child genetic influences are causally linked to prenatal exposures, which then have an 'environmental' effect on the development of the child's biology and behavior. We describe and demonstrate a conceptual framework for disentangling passive *r*GE from this cascading GE effect using a systems-based polygenic scoring approach comprised of genes shown to be important in the xenobiotic (substances foreign to the body) metabolism pathway. Data were drawn from 5044 families from the Avon Longitudinal Study of Parents and Children with information on maternal SDP, birth weight, and genetic polymorphisms in the xenobiotic pathway. Within a k-fold cross-validation approach ($k = 5$), we created weighted maternal and child polygenic scores using 18 polymorphisms from 10 genes that have been implicated in the xenobiotic metabolism pathway. Mothers and children shared variation in xenobiotic metabolism genes. Amongst mothers who smoked during pregnancy, neither maternal nor child xenobiotic metabolism polygenic scores were associated with a higher likelihood of smoke exposure during pregnancy, or the severity of smoke exposure during pregnancy (and therefore, neither proposed mechanism was supported), or with child birth weight. SDP was consistently associated with lower child birth weight controlling for the polygenic scores, maternal educational attainment, social class, psychiatric problems, and age. Limitations of the study design and the potential of the framework using other designs are discussed.

Keywords ALSPAC · Passive gene-environment correlation · Birth weight · Smoking during pregnancy · Polygenic

✉ Kristine Marceau
Kristine_Marceau@Brown.edu

[1] Division of Behavioral Genetics, Department of Psychiatry, Rhode Island Hospital, Coro West Suite 204, 1 Hoppin St, Providence, RI 02903, USA

[2] Center for Alcohol and Addiction Studies, Brown University, Providence, RI, USA

[3] Department of Psychiatry and Human Behavior, Brown University, Providence, RI, USA

[4] Psychology Department, The Pennsylvania State University, State College, PA, USA

[5] Department of Psychology and Child Development, California Polytechnic State University, San Luis Obispo, CA, USA

Introduction

Smoke exposure during pregnancy (generally assessed via maternal smoking during pregnancy) has been extensively studied as an environmental teratogen which can adversely affect child physical and behavioral development (e.g., Gaysina et al. 2013; Herrmann et al. 2008; Knopik 2009; Kramer 1987; Milberger et al. 1996; Stone et al. 2014; Ward et al. 2007). In recent years this association, particularly with behavioral outcomes, has been challenged by researchers using genetically-informed approaches that

have shown reductions in the strength of associations when controlling for genetic influences (e.g., Agrawal et al. 2008; D'Onofrio et al. 2003; Gaysina et al. 2013; Knopik 2009; Rice et al. 2009; Thapar et al. 2003). Importantly, associations of SDP with neurobiological and behavioral outcomes later in development may be mediated by low birth weight (Agrawal et al. 2010). Generally, SDP is associated with about a 5 % reduction in birth weight (on average ~ 150 g Kramer 1987; Kramer et al. 2001), following a dose response pattern, with more SDP resulting in lower birth weight and with SDP later in pregnancy exaggerating the effect. Low birth weight has been associated with increased morbidity and mortality (McIntire et al. 1999), and the link between low birth weight and mortality appears stable (e.g., from 1997 to 2002) despite advances in obstetric and neonatal health care (Fanaroff et al. 2007). Here, we examine competing hypotheses for how SDP is associated with low birth weight, as understanding potential causes and possible mechanisms of low birth weight is imperative for improving mental and physical health outcomes as well as reducing infant mortality.

Smoke exposure during pregnancy may cause low birth weight due to the environmental (intrauterine) impact of nicotine and toxins in cigarette smoke on placental and fetal development (e.g., as evidenced by animal models; Ernst et al. 2001). However, if specific maternal and child genetic variation or sets of genetic variants are associated with SDP exposure, and mothers pass those variants (or sets of variants) in addition to exposure to SDP to the child, passive gene-environment correlation (*r*GE) may explain at least part of this association. Much of the early research examined zero-order associations without considering possible confounding factors. More recently, genetically-informed studies have examined whether SDP actually exerts an environmental influence on child development beyond the contributions of genetics and other maternal characteristics. There is considerable evidence that SDP does environmentally influence birth weight, although the association may be attenuated after controlling for genetic influences and other maternal characteristics (e.g., maternal age, education, SES, etc.; Agrawal et al. 2010; D'Onofrio et al. 2003; Juárez and Merlo 2013; Knopik et al. 2015b; Rice et al. 2009). However, several quantitative genetic study designs (e.g., children of twins) have also shown that SDP is under genetic influence, indicating a *potential* role of passive *r*GE (Agrawal et al. 2008; D'Onofrio et al. 2003). That is, maternal genes may confound the association of SDP and birth weight via the likelihood of mothers smoking, severity of mothers' smoking, or via differential metabolism of smoking byproducts that may influence birth weight. In further support of the passive *r*GE pathway is the finding that controlling for family influences reduces the SDP-birth weight association, suggesting that this relationship is partially confounded by maternal characteristics or genetics (Juárez and Merlo 2013). In sum, existing evidence is unclear as to whether and to what extent the SDP-birth weight association may be a direct environmental influence or confounded by passive *r*GE (e.g., can be accounted for by genetic factors).

Whether SDP has a causal environmental effect on low birth weight or whether the association arises via passive *r*GE has important implications. For example, in general, if SDP has a causal, environmental effect on low birth weight, then SDP is a plausible target for interventions aiming to reduce negative sequelae related to low birth weight (e.g., later ADHD problems, infant mortality and morbidity). However, if SDP is spuriously associated with low birth weight, through a passive *r*GE, then intervention efforts could be better spent on other plausibly causal targets during pregnancy (e.g., other teratogens, dietary choices). Thus, in this study we present a new conceptual framework for understanding whether SDP has a causal environmental effect on low birth weight or whether the association arises via passive *r*GE. This work complements the existing literature using specialized genetically-informed designs and attempts to further elucidate the mechanisms of the association between SDP and birth weight.

Mechanisms of genetic and SDP influences on child outcomes

The possibility of passive *r*GE is particularly important to investigate. As noted above, passive *r*GE occurs when genetic variation that the parent and child share influences both the environment provided by the parent and the outcome of interest in the child. That is, passive *r*GE is a non-causal explanation such that genetic influences shared by the mother and child, which are also associated with the exposure (SDP), explain the association of SDP and birth weight. Thus, the association of interest (SDP and birth weight) would be spurious, and instead SDP would be a variable that represents genetic risk for birth weight. Thus, if passive *r*GE is operating, it would be unlikely to uncover a mechanism by which SDP influences birth weight because of the confounding effect of genetic transmission; a direct environmental mechanism would not drive the association in this case. In contrast, genetic influences may only partly confound the association; passive *r*GE may explain a part of the association leaving a smaller role for direct environmental mechanisms than would be concluded from non-genetically informed studies. Of the various quantitative genetic study designs, thus far, only children-of-twins designs have been able to show that maternal genetic influences are associated with her smoking during

pregnancy (e.g., D'Onofrio et al. 2003). These findings provide evidence of possible (not definitive) passive *r*GE in the association of SDP and child outcomes.

An alternative to passive gene-SDP correlation is a cascading effect whereby genetic influences (some of which the child is likely to inherit) are causally linked to prenatal exposures, which then have an 'environmental' effect on the development of the child's biology and behavior. Importantly, this is a plausibly causal mechanism, which includes genetic and environmental influences unfolding in a temporal process through biologically based mechanisms. It is possible to disentangle passive *r*GE from the cascading effect described above using molecular genetic data. The criteria for concluding passive *r*GE and a cascading effect are presented in Fig. 1. For both passive *r*GE and the cascade, two conditions must be met: (a) mother and child must share the genetic variation of interest, and (b) maternal genetic variation must be related to SDP. For passive *r*GE to be occurring, (c) that same genetic variation in the child must also be related to SDP. However, this is not a necessary condition for a developmental cascade, because there could be a teratogenic effect of SDP independent from an association of child genetic variation with SDP. A significant correlation of child genetic variation and SDP does not necessarily rule out the cascade effect because the association may only partly account for relations of SDP and birth weight. That is, there may still be environmental teratogenic effects of SDP on birth weight that are either independent, or potentially moderated by child genetic variation. The key difference needed to understand whether passive *r*GE is occurring or whether there is a potentially causal cascade including both genetic and environment influences is found in (d) the nature of the association of SDP with the child outcome, in this case birth weight. If SDP is related to the outcome at the zero-level only, but not after controlling for genetic influences, then passive *r*GE is occurring. This is because a key concept of *r*GE is that it describes genetic mediation of environment-outcome associations (Plomin 2014). That is, if genetics fully mediate the association, and therefore the association is explained by passive *r*GE, then the association would need to disappear when the relevant genes are added to the model. However, if SDP is related to the outcome even after controlling for genetic influences (particularly Gc [child genetic influences], but also Gm [maternal genetic influences], in Fig. 1), then there is stronger evidence for a developmental GE cascade. That is, the genetic confound is insufficient to fully explain the association of SDP and birth weight, and so there could be a biologically based mechanism by which SDP exposure leads to lower birth weight.

The utility of molecular genetic data

As noted above, thus far, quantitative genetic designs have been used to infer (in the case of children-of-twins studies) or control for (in the case of adoption designs) passive *r*GE. It is important to confirm findings from quantitative genetic designs using other strategies in order to ensure that findings do not arise from sampling techniques. In particular, it is important to use measured genes (Plomin 2014) to determine whether maternal genetic influences are associated with the amount of smoke exposure during pregnancy experienced by the developing fetus. Examining measured genes may better elucidate specific biologically-based mechanisms by which SDP or potential passive gene-prenatal environment correlation contributes to child development than quantitative genetic designs.

Molecular genetic data allows for better specificity than is possible using latent factors to assess genetic influences

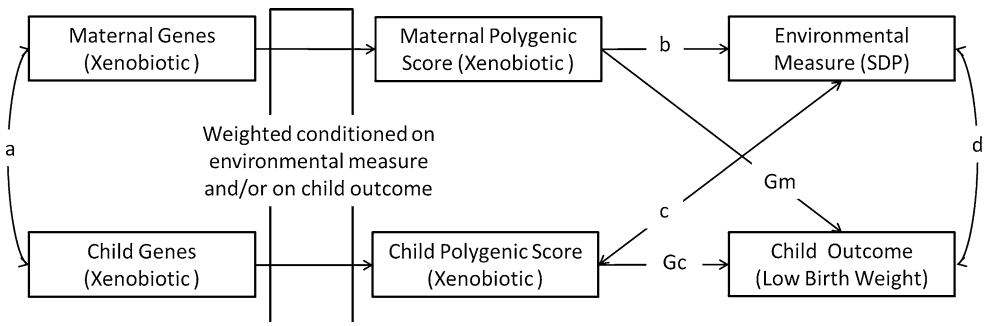

Fig. 1 Conceptual Model. For both passive *r*GE and a GE cascade, *a* (mother and child share genetic variation) and *b* (maternal genetic variation is related to the environmental measure) must be present. The presence of *a* and *b* are equivalent to what has typically been called passive *r*GE or possible passive *r*GE. A more conservative test of passive *r*GE also requires *c* (child genetic variation shared with maternal genetic variation correlated with the environmental measure) to be present. If *d* (environmental measure associated with outcome) is significant only at baseline, and not when also accounting for the contributions of *Gm* (maternal genetic influences) and *Gc* (child genetic influences), and *a*, *b*, and *c* are true, this is strong evidence of passive *r*GE specific to the biological mechanism in question. If *d* is significant after controlling for *Gm* and *Gc*, and *a* and *b* are true, then the evidence is more supportive of a GE cascade

as in quantitative genetic studies. One commonly used approach for creating and examining the polygenicity of behavioral phenotypes involves a two-step approach. First, a discovery sample is used to identify the gene variants most highly correlated with the outcome of interest. Then, a polygenic risk score is computed and tested in an independent sample using those most relevant genes identified in the discovery sample (e.g., Purcell et al. 2009; Salvatore et al. 2014). However, when specific biological mechanisms are of interest, theoretically and biologically-based sets of genes can be grouped in order to index the genetic variation underlying the specific biological pathway or mechanism of interest (rather than empirically identifying the most highly associated gene variants across the entire genome). Thus far, systems-based genetic analyses have been most frequently implemented using computed polygenic scores comprised of a set of genes carefully chosen to characterize a specific biological mechanism (e.g., Derringer et al. 2012; Juhasz et al. 2014). However, other methods besides the computation of polygenic scores can be used (e.g., Genome-Wide Complex Trait Analysis (GCTA); Yang et al. 2011). Polygenic scores may be particularly useful in the absence of complete genome-wide data and in the case of smaller sample sizes, as there is a single regression test of the polygenic score on the outcome after the score is created, whereas GCTA requires an extremely large sample in order to obtain enough genetic variation across unrelated individuals to fit the biometric models. The present study utilizes a systems-based polygenic scoring approach for investigating the possibility of a specific, identifiable passive *r*GE underlying the association of maternal smoking during pregnancy and child birth weight.

Xenobiotic metabolism genes

As an example of one plausible biological mechanism that may represent passive *r*GE or a cascading effect of genes and SDP for birth weight, we consider the role of xenobiotic metabolism genes. Maternal xenobiotic metabolism genes help regulate the metabolism of teratogens. This includes the metabolism of nicotine, but also many other teratogens (e.g., pesticides, alcohol, other toxins). Thus, maternal xenobiotic metabolism genes putatively have an effect on the amount of nicotine (and other teratogens) that will be available to cross the placenta and affect fetal development. Further, maternal xenobiotic metabolism genes are likely to also have an effect on the extent to which mothers smoke, and their ability or desire to quit or continue smoking during pregnancy. For example, some genes implicated in the xenobiotic metabolism pathway (e.g., CYP2A6) have been linked with smoking behavior (Thorgeirsson et al. 2010), and some (e.g., in the alcohol dehydrogenase family) have been selected for inclusion on a chip designed to index genes associated with nicotine dependence and other smoking phenotypes (http://biorealmresearch.com/smokescreen/). Thus, in the case of genes in the xenobiotic metabolism pathways, we would expect maternal genetic variation to be associated with the extent to which children are exposed to maternal smoking during pregnancy (e.g., path b in the conceptual model, Fig. 1). Mothers are also likely to pass these genes related to drug metabolism to the child, as mothers and children share 50 % of their genes on average (e.g., path a in the conceptual model, Fig. 1). Thus, child xenobiotic metabolism genes make up the 'third line of defense' (after the placenta; Blumenfeld et al. 2009) in the metabolism of nicotine and will further have an effect on the amount of nicotine and other smoke byproducts available to affect fetal development[1] (e.g., path c in Fig. 1).

Passive rGE

If maternal and child xenobiotic metabolism genes are associated with SDP exposure (paths b and c in the conceptual model, Fig. 1, respectively) and mothers pass the genes in addition to the SDP exposure to the child (path a in the conceptual model, Fig. 1), then passive *r*GE may explain the association of SDP and birth weight. In the literature, passive *r*GE is clearly defined as non-causal. Thus far, because passive *r*GE has primarily been explored using quantitative genetic designs, it is often considered in broad terms. Here, we extend the broader concept of passive *r*GE to consider a specific example of a narrower, biologically informed passive *r*GE. A spurious association of SDP and birthweight, when xenobiotic metabolism genes are implicated (e.g., there is a direct influence of xenobiotic metabolism genes on birth weight rendering the SDP-birthweight association spurious), is an example of a specific passive *r*GE. That is, the SDP-birthweight association would not be a direct environmental influence but rather explained by the action of xenobiotic metabolism genes on birthweight even when mothers do not smoke during pregnancy. Thus, for this specific passive *r*GE to occur, xenobiotic metabolism genes must have a direct influence on birth weight.

There is some evidence of this direct association in candidate gene studies (e.g., Infante-Rivard et al. 2006; Nukui et al. 2004) although recent GWAS studies of birth weight have not found significant hits among xenobiotic

[1] Xenobiotic metabolism genes may also be expressed in placental tissue, which represents the second line of defense for the developing child (Blumenfeld et al. Blumenfeld et al. 2009). However, here we focus on maternal and child genetic variation only, as epigenetic information from the placenta is needed to thoroughly investigate this mechanism which is out of the scope of the current manuscript.

metabolism genes specifically (although it is unclear how well xenobiotic genes were covered in the arrays used, or whether there was sufficient exposure to manifest effects; Freathy et al. 2010; Horikoshi et al. 2013). The mechanism of action of xenobiotic metabolism genes does allow for xenobiotic metabolism genes to have a direct influence on birth weight. For example, genes in the glutathione S-transferase family each play an important role in detoxification by coding for enzymes that catalyze the conjugation of reduced glutathione with a variety of hydrophobic and electrophilic compounds. Deletions in the glutathione S-transferase mu 1 (GSTM1) gene and from the glutathione S-transferase theta 1 (GSTT1) gene (and especially both risk alleles together) have been shown to be associated with fetal growth restriction in infants exposed to organochlorine pesticides (Sharma et al. 2012), and the GSTT1 deletion has been associated with small-for-gestational age, especially among youth exposed to smoking during pregnancy (Infante-Rivard et al. 2006). Consuming more iron during pregnancy was associated with higher birth weight in infants without the GSTM1 deletion, even adjusting for a host of maternal and fetal covariates, including urinary cotinine levels (Hur et al. 2013). Thus, genes in the glutathione S-transferase family are mechanistically linked to birth weight through the ways in which exposures to multiple teratogens are metabolized in both the mother and child. These findings provide evidence that genes implicated in the xenobiotic pathway may have direct effects on birth weight, and further that their effects on birth weight can be independent of smoking during pregnancy as a specific teratogenic exposure.

In terms of passive *r*GE, SDP then may be an environmental influence that is associated with xenobiotic metabolism genes in the mother (e.g., associated with maternal smoking behavior and her inability to quit smoking during pregnancy), but not actually exert an influence on birth weight if the direct association of these genes in the fetus on birth weight is driven by the metabolism and level of exposure of other teratogens experienced by the mother and fetus (e.g., pesticides, dietary choices). These other exposures may or may not be associated with SDP and are infrequently measured and controlled in studies of SDP and birth weight. Thus, it is possible that xenobiotic metabolism genes have a more broad direct effect on birth weight, which could render the more specific association of SDP and birth weight spurious and therefore fall under the mechanism of a specific passive *r*GE.

Developmental gene-environment cascade

Alternatively, the xenobiotic metabolism genes may launch a cascade influencing (1) mothers' ability or inability to quit, subsequently exposing the child to SDP, (2) mother's and child's ability to metabolize nicotine and other smoke byproducts and therefore modifying the extent of the child's exposure to SDP, and (3) subsequently the association of SDP to birth weight. As noted above, the key difference (if paths a, b, and c are present) would be whether the association of SDP and birth weight (path d in the conceptual model, Fig. 1) is significant after controlling for the influence of maternal (path Gm in the conceptual model, Fig. 1) and child (path Gc in the conceptual model, Fig. 1) xenobiotic metabolism genes. This gene-environment cascade mechanism is supported by findings that the association of SDP and birth weight remains robust even when controlling for genetic influences and other maternal characteristics (Agrawal et al. 2010; D'Onofrio et al. 2003; Juárez & Merlo 2013; Knopik et al. 2015b; Rice et al. 2009).

Present study

The goal of the current study is to test whether there is a passive gene-SDP correlation, or whether SDP-birth weight associations represent a developmental gene-environment cascade using a large cohort of mothers and their children. We use a conservative systems-based approach in order to ascertain whether genes from the maternal xenobiotic pathway are associated with her smoking behavior during pregnancy. We also test specifically for passive gene-SDP associations by testing whether mothers and children share xenobiotic metabolizing gene variation, and whether child xenobiotic metabolizing genes are also associated with SDP at the zero-order level, and after controlling for the polygenic score indexing xenobiotic metabolism gene variation and other maternal characteristics. Based on previous literature, we expect that (a) mother and child will share at least 50 % (expected for mother- to- child transmission) of xenobiotic metabolizing gene variation, and (b) that maternal gene variation will be related to SDP. We were unable to hypothesize whether (c) child xenobiotic metabolizing gene variation would be related to SDP. However, we do expect that (d) associations of SDP and birth weight would be present, even after controlling for genetic influences. Because of hypothesis (d), overall we expected results to be consistent with a cascade model rather than passive *r*GE.

Method

Participants

Data were drawn from the Avon Longitudinal Study of Parents and Children (ALSPAC; Boyd et al. 2013). ALSPAC recruited 14,541 pregnant women resident in

Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992. The total sample represented 15,458 fetuses; 14,775 were live births and 14,701 were alive at 1 year of age. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local (Rhode Island Hospital) Research Ethics Committees. We used only the portion of the live births for which there was information on maternal smoking during pregnancy (N = 11,133), and birth weight (N = 13,901), and had the relevant polymorphisms (see below) for mothers (N = 7553) and children (N = 6754). For multiple births (N = 406), we randomly chose only one twin for inclusion in the analysis. This resulted in an analytic sample of 5044 families. Please see Boyd et al. (2013) for additional sample and recruitment details.

Measures

Please note that the study website contains details of all the data that is available through a fully searchable data dictionary (http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/).

Xenobiotic metabolism genes

We used 18 polymorphisms from 10 genes that have been implicated in the xenobiotic metabolism pathway (see Table 1 for details). These included rs4986782 and rs4987076 from NAT1, and rs1799930, rs1799931, and rs1801280 from NAT2 within the N-acetyltransferase family; rs1695 from GSTP1 and deletions from GSTM1 and GSTT1 within the glutathione S-transferase family; rs284779 from ADH7, rs975833, rs1229966, and rs2866151 from ADH1A, rs2066701 and rs4147536 from ADH1B, and rs3762894, rs4148884, rs4699714 from ADH4 within the alcohol dehydrogenase family; rs28399433 from CY2A6_9 within cytochrome P450, family 2, subfamily A. See Table 1 for more information on the gene families, functions, as well as specific genes and polymorphisms included here.

Smoking during pregnancy (SDP)

The quantity of cigarettes smoked on average per day across the first 3 months of pregnancy was assessed via self-report when mothers were 18 weeks pregnant. The current number of cigarettes smoked per day was also assessed via self-report when mothers were 32 weeks pregnant. These variables were used to create a SDP severity score theoretically consistent with research showing dose–response patterns of exposure to nicotine and low birth weight (Ernst et al. 2001). The severity score was built on the following assumptions: (1) continuing to smoke later in pregnancy represents a higher likelihood of risk than successfully quitting in or shortly after the first trimester; (2) smoking later in pregnancy imparts greater risk than smoking earlier in pregnancy (in the rare instances where mothers begin to smoke after the first trimester; e.g., Dwyer et al. 2009; Hebel et al. 1988); (3) smoking less than a half pack per day, smoking between a half and whole pack per day, and smoking more than a pack per day represent qualitatively different levels of risk (McNeil 1995). As such, the severity score had 7 levels (see Knopik et al. 2015a also in this Special Issue):

0 = no smoking during pregnancy in either the first trimester or later in pregnancy (N = 8036)
1 = 1-10 cigarettes per day in the first trimester, no smoking later in pregnancy (N = 473)
2 = 11-20 cigarettes per day in the first trimester, no smoking later in pregnancy (N = 45)
3 = 21 + cigarettes per day in the first trimester, no smoking later in pregnancy (N = 33)
4 = any smoking later in pregnancy but not during the first trimester (N = 169)
5 = 1-10 cigarettes per day later in pregnancy and any smoking in the first trimester (N = 1289)
6 = 11-20 cigarettes per day later in pregnancy and any smoking in the first trimester (N = 457)
7 = 21 + cigarettes per day later in pregnancy and any smoking in the first trimester (N = 343)

Birth weight

Birth weight was assessed from obstetric data, recorded by the ALSPAC measurers, and via birth notification. We used the ALSPAC preferred birth weight (detailed notation available on the study website; http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/). Briefly, if all birth weights from each data source were identical, that was the preferred birth weight. In cases where the disagreement across the different assessments was >100 g birth weight was set to missing. In cases where the disagreement across the different assessments was <100 g, the lower birth weight was used.

Covariates

We included the following covariates: child sex and maternal educational attainment, social class, psychiatric problems, and age. Maternal educational attainment, social class, psychiatric problems, and age were assessed when mothers were 32 weeks pregnant. Social class is a standard variable derived by the ALSPAC team comprised of occupation information (e.g., occupation, industry, managerial status). 'Psychiatric problems' is an indicator

Table 1 Xenobiotic metabolism genes included in the present study

Family	Function	Gene	rs#
N-acetyltransferase family	Codes an enzyme that is involved in the metabolism of drugs by catalyzing the transfer of an acetyl group during the drug metabolism process	N-acetyltransferase 1 (NAT1)	rs4986782
			rs4987076
	Codes an enzyme that both activates and deactivates the compounds catalyzed by the enzyme encoded by NAT1 as well as carcinogens, and governs the speed of transferring acetyl groups during drug metabolism (which is associated with drug toxicity)	N-acetyltransferase 2 (NAT2)	rs1799930
			rs1799931
			rs1801280
Glutathione S-transferase family	Plays important role in detoxification by coding for enzymes that catalyze the conjugation of reduced glutathione with a variety of hydrophobic and electrophilic compounds	Glutathione S-transferase pi 1 (GSTP1)	rs1695
		Glutathione S-transferase mu 1 (GSTM1)	Deletion
		Glutathione S-transferase theta 1 (GSTT1)	Deletion
Alcohol dehydrogenase family	Plays important role in the metabolism of a wide variety of drugs, including alcohol	Alcohol dehydrogenase 7 (ADH7)	rs284779 intron variant
		Alcohol dehydrogenase 1A (ADH1A)	rs975833 intron variant
			rs1229966 intron variant 2 KB upstream
			rs2866151
		Alcohol dehydrogenase 1B (ADH1B)	s2066701 intron variant
			rs4147536 intron variant
		Alcohol dehydrogenase 4 (ADH4)	rs3762894 intron variant 2 KB upstream
			rs4148884 intron variant 2 KB upstream
			rs4699714 intron variant
Cytochrome P450, family 2, subfamily A	Encodes cytochrome P450 proteins, which catalyze many reactions involved in drug metabolism	Cytochrome P450, family 2, subfamily A, polypeptide 6 (CYP2A6_9)	rs28399433

variable denoting the absence/presence of any of the following: drug addiction, alcoholism, schizophrenia, anorexia nervosa, severe depression, and "other psychiatric problem" either recently or in the past, assessed during pregnancy. Detailed notation for each covariate is available on the study website; http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/.

Analytic strategy

In order to test for criterion (a) we examined whether mothers and children were concordant for the number of minor alleles for each polymorphism. This was operationalized as the percentage of genes for which mothers and children were concordant for the number of minor alleles for each polymorphism. We then averaged the concordance (e.g., percentage of the number of minor alleles shared) across all 18 polymorphisms in order to obtain an average percentage of minor alleles shared within the portion of the xenobiotic pathway sampled here.

To create polygenic scores comprised of xenobiotic metabolism genes, we used a k-fold cross validation approach using a p threshold of 1, therefore including all genes in the xenobiotic pathway regardless of significance

of the individual SNP. First, we split the participants randomly into one of five (i.e., k = 5) folds. Then, inside of a primary loop

1. We selected four folds as the discovery set and one fold as the test set.

   a.
   Embedded in a second loop, we

   i. mean centered each polymorphism,
   ii. identified monomorphic polymorphisms, and
   iii. ran a series of baseline regressions that assessed the main effect of each individual SNP on the outcome (described below) in the discovery sample (comprised of 80 % of the sample). Any monomorphic polymorphisms identified in the specific training set were given a beta-coefficient of 0 and *p* value of 1.

      A. Maternal polymorphisms predicting SDP (zero-inflated Poisson regression)
      B. Child polymorphisms predicting SDP (zero-inflated Poisson regression)
      C. Child polymorphisms predicting birth weight (linear regression)
      D. Maternal polymorphisms predicting birth weight (linear regression)

   iv. From this series of baseline regressions (one per polymorphism), we saved the coefficients. This loop ran 18 times, equal to the number of xenobiotic pathway polymorphisms available in the data.

2. Next, (outside of the second loop but inside the primary loop) we created a matrix of the genotypes in the test sample after centering the genotypes in the test matrix, and gave any missing values a value of 0 (equal to the test sample average number of minor alleles for that polymorphism). This is a mean imputation of missing polymorphism information in the test sample.
3. Four polymorphisms in the alcohol dehydrogenase family were in LD: rs1229966 with rs975833, rs2066701, and (negatively) rs2866151; and rs975833 with rs2066701 for both mothers and children. Therefore, we also pruned for LD ($R^2 > .70$). For each pair of SNPs in LD, we kept the coefficient and *p*-value of the polymorphisms more highly associated with the outcome, and set the coefficient to zero for the polymorphisms more weakly associated with the outcome so that it would not contribute to the polygenic score.
4. Then, we multiplied the test matrix by the polymorphism coefficients to create our polygenic scores. Specifically, the test matrix is 18 columns (for the 18 polymorphisms) by N rows (1 per individual). It gets multiplied by a vector that is 18 values long (the 18 polymorphism weights from the training set). First, each polymorphism weight is multiplied by each individual's corresponding polymorphism, and then all the resulting values in that row (e.g., for that individual) is summed. This results in a single value for that individual's weighted polygenic score.
5. Finally, we conducted a series of hypothesis-testing regressions (described below). This primary loop was repeated (k) times, so that each fold was the test set once.

Thus, at the end of the primary loop (k) we had two vectors for each regression analysis. There was a list of 5 (one for each loop (k), corresponding to each unique test sample) coefficients from each regression assessing the effect of the polygenic score on each outcome, and a list of 5 *p*-values for those coefficients. Consistency across these results indicates a stable effect.

Within this framework, several regression analyses were conducted. For models of smoking during pregnancy (SDP), we used zero-inflated Poisson regressions to account for the zero-inflated nature of SDP. For models of child birth weight, linear regressions were used.

To assess criterion (b) relation of maternal genes and SDP:

1. Maternal polygenic score conditioned on SDP (e.g., consisting of weighted coefficients from the models where individual maternal polymorphisms predicted SDP; A above), predicting SDP.
2. Maternal polygenic score conditioned on SDP (e.g., the same score as in regression 1) and a series of potential confounding variables (child sex, maternal educational attainment, social class, psychiatric problems, and age) added as covariates, predicting SDP.

To assess criterion (c) relation of child genes and SDP:

3. Child polygenic score conditioned on SDP (e.g., consisting of weighted coefficients from the models where individual child polymorphisms predicted SDP; B above) predicting SDP.
4. Child polygenic score conditioned on SDP (e.g., the same score as in regression 3) and the potential confounding variables predicting SDP.

To assess the baseline relation of genetic influence on birth weight:

5. Child polygenic score conditioned on birth weight (e.g., consisting of weighted coefficients from the models where individual child polymorphisms predicted birth weight; C above) predicting child birth weight

To assess criterion (d) relation of SDP and birth weight accounting for child genetic influences:

6. Child polygenic score conditioned on birth weight (e.g., the same score as in regression 5), SDP, and the potential confounding variables, predicting child birth weight.

We also examined the contribution of maternal genes on birth weight for completeness:

7. Maternal polygenic score conditioned on birth weight (e.g., consisting of weighted coefficients from the models where individual maternal polymorphisms predicted birth weight; D above), predicting child birth weight
8. Maternal polygenic score conditioned on birth weight (e.g., the same score as in regression 7), SDP, and the potential confounding variables predicting SDP predicting child birth weight.

Results

We first assessed the percentage of polymorphisms for which mothers and children were concordant for the number of minor alleles in each of the 18 xenobiotic metabolism genes. Across the 18 polymorphisms, mother and child concordance was 57 % (statistically significantly above the 50 % expected for mother- to- child transmission). Thus, the criterion (a) is met.

Results from the regression analyses within the K-fold validation approach are summarized in Table 2. Neither maternal nor child xenobiotic metabolism polygenic scores were associated with a higher likelihood of smoking during pregnancy, or to the severity of smoking during pregnancy if SDP was endorsed. Thus, criteria (b) and (c) were not met. SDP was associated with child birth weight. The zero-order association in the full sample indicated that SDP was associated with lower child birth weight ($r = -.15$, unstandardized $\beta = -37.30$, $p < .0001$). We examined this association within the cross-validation approach in order to test the association more conservatively. Even controlling for the polygenic score (both maternal and child) and maternal educational attainment, social class, psychiatric problems, and age predicted child birth weight, SDP was consistently (e.g., in each of the 5 folds) associated with lower birth weight. Neither the maternal nor child xenobiotic metabolism polygenic score was directly associated with child birth weight. Because criterion (b) and (c) were not met, the data cannot support the passive *r*GE or developmental GE cascade mechanism. We can only conclude that SDP is consistently associated with birth weight above and beyond other modeled maternal characteristics and the influence of the polygenic contribution of xenobiotic metabolism genes sampled here.

Discussion

We presented a theoretical method for disentangling causal from non-causal joint effects of genetic and environmental influences using molecular genetic data as a way to corroborate findings from twin and family studies. Empirically, we corroborated a very well-characterized association of prenatal smoking exposure and low child birth weight in a very large sample using a conservative test—a k-fold cross-validation approach. The association held consistently across folds even when controlling for other maternal characteristics and a polygenic score representing polymorphisms implicated in the xenobiotic metabolism pathway. Our data did not meet the initial criteria needed to separate these mechanisms. Nonetheless, we believe that our conceptual framework will be useful for future studies harnessing molecular genetic data to test findings from quantitative genetic designs. Corroboration across study types and across quantitative and molecular genetic study designs is imperative, as each sample design comes with its own specific assumptions and limitations.

Our findings potentially suggest that xenobiotic metabolism genes are not likely contributing to the overlap in genetic and SDP influences found in previous children-of-twins and child-based twin studies. However, that conclusion is tempered by the fact that the xenobiotic- or drug-metabolizing pathways are highly complex, and we had very limited coverage of the system with the 18 polymorphisms included in this study. We hope to explore these questions using a more thorough examination including a better sample of polymorphisms in the xenobiotic metabolism pathway genes in the future. The limitation of our insufficient coverage of a complex system is compounded by the complexity of gene products, and that polymorphisms inherently measure gene structure, which is only partly responsible for gene function. We summed the effects of the polymorphisms into the weighted polygenic score. A risk is that some polymorphisms have slightly positive and others slightly negative effects on SDP. Thus, when summed, some polymorphisms may wash the effects of other out, and we are more likely to have a polygenic

Table 2 Regression Results

	Fold 1 β	Fold 2 β	Fold 3 β	Fold 4 β	Fold 5 β
Predicting the likelihood of not smoking during pregnancy (Inflated Zeros)					
Maternal XMGV					
Baseline	1.17	−0.56	0.68	0.08	−0.12
Controlled	2.29	−0.49	1.22	0.95	0.03
Child XMGV					
Baseline	−1.15	−0.96	−0.09	−0.96	−0.31
Controlled	−2.61*	−0.55	−0.19	−1.13	−0.23
Predicting SDP severity (Count, if SDP is present)					
Maternal XMGV					
Baseline	−0.08	−0.02	−0.22	1.57	−0.10
Controlled	−0.15	−0.54	−1.16	1.65	0.09
Child XMGV					
Baseline	0.11	−0.50	−0.10	0.15	−0.42
Controlled	−0.35	−0.47	−0.16	0.88	−0.68
Predicting Birth Weight					
Main effects Only					
Maternal XMGV	0.66	0.23	0.88	0.55	0.58
Child XMGV	0.03	0.63	0.56	0.39	−.31
SDP	−35.23*	−51.46*	−46.27*	−19.54*	−35.31*
Controlled					
Maternal XMGV	0.98	−0.14	−0.28	0.93	0.71
Child XMGV	−0.06	0.54	0.34	0.15	−0.40
SDP controlling for covariates and maternal XMGV	−21.46*	−44.82*	−57.17*	−18.24*	−39.46*
SDP controlling for covariates and child XMGV	−22.51*	−44.84*	−57.16*	−19.27*	−39.59*

Covariates include child sex, maternal educational attainment, social class, psychiatric problems, and age

XMGV xenobiotic metabolism gene variants

* $p < .001$

score with null effects.[2] Whereas this does limit our likelihood of finding a clear polygenic signal related to the outcome, we believe that this most closely resembles the underlying biology. Indeed, an individual likely has many genes imparting risk and many others that act protectively against the teratogenic effects of SDP, or of mothers' ability to quit smoking. This may in part explain the null findings for direct effects of xenobiotic metabolism genes on SDP exposure and birth weight in this study. The field examining polygenicity is very rapidly advancing, and as our polygenic scoring methods improve, it will be worth revisiting the questions examined here to determine whether methodological limitations led to the null findings for polygenic effects of xenobiotic metabolism genes on SDP. Alternatively, xenobiotic metabolism pathway genes may not be the relevant biological pathway for genetic influences on SDP, or may be one of several pathways acting together.

[2] We also constructed scores using only the polymorphisms positively associated with conditioning variable in one score and the polymorphisms negatively associated with the conditioning variable in a separate score, and there were no differences in results

Conceptual framework for passive *r*GE and developmental genetic-environmental cascade

The most important contribution of this study is the theoretical framework for disentangling these important mechanisms. We proposed a way to use molecular genetic data to disentangle the inherently non-causal mechanism (e.g., confounded with family background) of passive *r*GE from a potentially causal (e.g., teratogenic) mechanism whereby genetic and environmental influences unfold temporally through biologically based mechanisms. Given a set of underlying (fully testable) assumptions (Fig. 1a) that mothers and children share the genetic variation of interest, (Fig. 1b) that those maternal genetic variations are related to the environmental exposure, and (Fig. 1c) that child genetic variations are also related to the environmental exposure (especially for passive *r*GE), one key

parameter (Fig. 1d) differentiates the two mechanisms. If the association of the exposure and outcome is present initially, but disappears when the polygenic score is added to the model, then this is evidence of a specific passive *r*GE. However, if the association is present and persists when the maternal (Fig. 1, Gm) and child (Fig. 1, Gc) polygenic scores are added to the model, then this is evidence of a developmental GE cascade.

Thus far, quantitative genetic designs have been able to investigate (e.g., children-of-twins) or control for (e.g., adoption, in vitro designs) passive *r*GE. A limitation of quantitative genetic designs is that influences are necessarily non-specific. Using polygenic scores (or other novel methods) comprised of theoretically relevant gene sets can help hone in on whether genetic and environmental influences work together for specific biologically based mechanism important for child development, or to understand specific passive *r*GEs. Systems-based polygenic approaches—when the system is adequately characterized—should explain more variance in the phenotype than broader molecular genetic approaches because a clear effect (e.g., with less noise) related to the mechanism of interest may be observed. However, this clear effect can only explain the amount of the variance in the phenotype that the specific mechanism under consideration actually plays (not necessarily expected to be anywhere close to 100 %). It may become more difficult to find genetic influences that play a meaningful role in the mechanism of interest when the mechanism is insufficiently covered (e.g., only few relevant polymorphisms from only few relevant genes are included) in the polygenic score. Further, including phenotypes that are closer to the process (e.g., conditioning the polygenic score on a measure of efficiency of drug metabolism) may increase the likelihood of observing genetic influences that play a meaningful role in the process of interest by cutting down the number of levels the genetic influences need to operate across in order to influence the phenotype. Thus, conditioning the polygenic score on phenotypes as close to the biological mechanism of interest and ensuring good coverage of both polymorphisms in each gene and relevant genes in the biological system will increase the utility of this approach in future work.

Other limitations and future directions

In addition to the limitations already discussed, there are a number of other considerations important to keep in mind. Including maternal and child genetic variation can help elucidate causality in quantitative genetic designs (e.g., the extended children of twins model; Narusyte et al. 2008). However, as we move toward specificity of genetic (e.g., do not account for 100 % of the genetic variation) and environmental influences to examine specific mechanisms of interest (e.g., the role of xenobiotic metabolism genes and SDP exposure for low birth weight), we no longer can infer causality (e.g., using the conceptual framework presented here).

This is due to a number of limitations. First, it is probable that related and confounding pathways exist, but are not explicitly modeled, as is common in any study of human behavior. The concepts of equifinality (the likelihood that the same outcome is a downstream effect of multiple possible temporally preceding influences) and multifinality (the likelihood that a single developmental influence can result in multiple outcomes through multiple pathways; Cicchetti and Rogosch 1996) lead to predictions that multiple biological pathways are likely to be implicated in multiple outcomes. Specificity is very important to understand these mechanisms, but is necessarily limited because of how many important factors are necessarily excluded from the model. Second, excluded genes are highly likely to be associated with the genes included, which diminishes our confidence that (especially small) gene sets are causally linked to the exposure or outcome of interest. Third, SDP was assessed via self-report and therefore may be subject to error, for example, due to non-disclosure. Future studies may do well to confirm SDP exposure with cotinine levels, although self-report and retrospective report have been shown to be valid in the literature (e.g., Knopik et al. 2015b; Pickett et al. 2009; Reich et al. 2003).

Further, there is a problem of correlated residual error in the genetic influence whenever multiple genes are examined together. This problem varies with the methods used to examine joint effects of multiple genes. The problem of correlated residual error in genotypes is attenuated considerably by using genetic relatedness matrixes in GCTA, but is generally unaccounted for when using polygenic risk scores (thus far). Therefore, while the framework presented here has the potential to disentangle specific passive *r*GEs from plausibly causal biological mechanisms, we stress that this framework does not infer causality. Recent developments in using genome-wide data have allowed for explicit tests of passive *r*GE using an extended Genome-wide Complex Trait Analysis (GCTA; Yang et al. 2011) approach incorporating both maternal and child genome-wide data (m-GCTA; Eaves et al. 2014). A strength of m-GCTA is that it maps quite well onto quantitative genetic study designs, using similar variance decomposition strategies and examining broad-sense additive heritability. This approach has great potential to disentangle causal environmental effects from those confounded by maternal and child genes, and passive *r*GE in order to corroborate findings from quantitative genetic models.

We continue to hope that examining variation from multiple genes together will result in a stronger signal to examine gene-environment interplay, and believe that taking theoretically derived, systems-based approaches will further augment our power for understanding when and how genetic variation influences behavior in conjunction with the environment. In terms of the present study, more work is needed to understand whether the xenobiotic metabolism pathway is actually unrelated to the smoking during pregnancy- birth weight association because of the limitations of the data used here. Further, we plan to vet the current conceptual framework using simulated data in order to further provide a proof of concept for the utility of polygenic scores related to specific pathways to discover when passive *r*GE plays a role versus when a GE cascade may influence child development. As convergence across multiple study designs with different limitations provides the strongest evidence for gene-environment interplay, we encourage the use of multiple ways of examining polygenicity and the continued use of both quantitative and molecular genetic approaches.

Acknowledgments We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant Ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. We sincerely thank Lindon Eaves for his thoughtful feedback on the model and conceptual approach. This publication is the work of the authors and K Marceau, R Palmer, JM Neiderhiser, T Smith, & VS Knopik will serve as guarantors for the contents of this paper. Authors were supported by the following sources: T32MH019927 and T32DA016184 (Marceau), K01AA021113 (Palmer), MH092118 (Neiderhiser), T32MH19927 (Smith), DA023134 (Knopik).

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors. Informed consent was obtained from all individual participants included in the study.

References

Agrawal A, Knopik VS, Pergadia ML, Waldron M, Bucholz KK, Martin NG, Madden PA (2008) Correlates of cigarette smoking during pregnancy and its genetic and environmental overlap with nicotine dependence. Nicotine Tob Res 10(4):567–578. doi:10.1080/14622200801978672

Agrawal A, Scherrer JF, Grant JD, Sartor CE, Pergadia ML, Duncan AE, Xian H (2010) The effects of maternal smoking during pregnancy on offspring outcomes. Prev Med 50(1–2):13–18. doi:10.1016/j.ypmed.2009.12.009

Blumenfeld Y, Fan C, Sangkuhl K, Altman R, El-Sayed Y, Quake S (2009) 702: comprehensive maternal-fetal pharmacogenomics—a novel pharmacogenomic thumbprint. Am J Obstet Gynecol 201(6):S254. doi:10.1016/j.ajog.2009.10.719

Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, Davey Smith G (2013) Cohort profile: the 'children of the 90 s'—the index offspring of the avon longitudinal study of parents and children. Int J Epidemiol 42(1):111–127. doi:10.1093/ije/dys064

Cicchetti D and Rogosch FA (1996) Equifinality and multifinality in developmental psychopathology. Dev Psychopathol 8(04): 597–600. M593—510.1017/S0954579400007318. Retrieved from href = "http://dx.doi.org/10.1017/S0954579400007318

Derringer J, Krueger R, Dick D, Aliev F, Grucza R, Saccone S, Bierut L (2012) The aggregate effect of dopamine genes on dependence symptoms among cocaine users: cross-validation of a candidate system scoring approach. Behav Genet 42(4):626–635. doi:10.1007/s10519-012-9531-4

D'Onofrio BM, Turkheimer EN, Eaves LJ, Corey LA, Berg K, Solaas MH, Emery RE (2003) The role of the Children of Twins design in elucidating causal relations between parent characteristics and child outcomes. J Child Psychol Psychiatry 44(8):1130–1144

Dwyer JB, McQuown SC, Leslie FM (2009) The dynamic effects of nicotine on the developing brain. Pharmacol Ther 122(2):125–139. doi:10.1016/j.pharmthera.2009.02.003

Eaves L, Pourcain B, Smith G, York T, Evans D (2014) Resolving the effects of maternal and offspring genotype on dyadic outcomes in genome wide complex trait analysis ("M-GCTA"). Behav Genet 44(5):445–455. doi:10.1007/s10519-014-9666-6

Ernst M, Moolchan ET, Robinson ML (2001) Behavioral and neural consequences of prenatal exposure to nicotine. J Am Acad Child Adolesc Psychiatry 40(6):630–641

Fanaroff AA, Stoll BJ, Wright LL et al (2007) Trends in neonatal morbidity and mortality for very low birthweight infants. Am J Obstet Gynecol 196(2):147e.141–147e.148. doi:10.1016/j.ajog.2006.09.014

Freathy RM, Mook-Kanamori DO, Sovio U, et al (2010). Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. Nat Genet, 42(5):430–435. http://www.nature.com/ng/journal/v42/n5/suppinfo/ng.567_S1.html

Gaysina D, Fergusson DM, Leve LD, Horwood J, Reiss D, Shaw DS, Harold GT (2013) Maternal smoking during pregnancy and offspring conduct problems: evidence from 3 independent genetically sensitive research designs. JAMA Psychiatry 70(9):956–963. doi:10.1001/jamapsychiatry.2013.127

Hebel JR, Fox NL, Sexton M (1988) Dose-response of birth weight to various measures of maternal smoking during pregnancy. J Clin Epidemiol 41(5):483–489. doi:10.1016/0895-4356(88)90050-9

Herrmann M, King K, Weitzman M (2008) Prenatal tobacco smoke and postnatal secondhand smoke exposure and child neurodevelopment. Curr Opin Pediatr 20(2):184–190. doi:10.1097/MOP.0b013e3282f56165

Horikoshi M, Yaghootkar H, Mook-Kanamori DO, et al (2013) New loci associated with birth weight identify genetic links between intrauterine growth and adult height and metabolism. Nat Genet 45(1):76–82. http://www.nature.com/ng/journal/v45/n1/abs/ng.2477.html#supplementary-information

Hur J, Kim H, Ha EH, Park H, Ha M, Kim Y, Chang N (2013) Birth weight of Korean infants is affected by the interaction of maternal iron intake and GSTM1 polymorphism. J Nutr 143(1):67–73. doi:10.3945/jn.112.161638

Infante-Rivard C, Weinberg CR, Guiguet M (2006) Xenobiotic-metabolizing genes and small-for-gestational-age births: interaction with maternal smoking. Epidemiology 17(1):38–46

Juárez SP, Merlo J (2013) Revisiting the effect of maternal smoking during pregnancy on offspring birthweight: a quasi-experimental sibling analysis in Sweden. PLoS ONE 8(4):e61734. doi:10.1371/journal.pone.0061734

Juhasz G, Hullam G, Eszlari N, Gonda X, Antal P, Anderson IM, Bagdy G (2014) Brain galanin system genes interact with life stresses in depression-related phenotypes. Proc Natl Acad Sci 111(16):E1666–E1673. doi:10.1073/pnas.1403649111

Knopik VS (2009) Maternal smoking during pregnancy and child outcomes: real or spurious effect? Dev Neuropsychol 34(1):462593

Knopik VS, Heath AC, Marceau K, Palmer RH, McGeary JE, Todorov A, Schettini Evans A (2015a) Missouri mothers and their children: a family study of the effects of genetics and the prenatal environment. Twin Res Hum Genet 18(5):485–496

Knopik VS, Marceau K, Palmer RH, Smith TF, Heath AC (2015b) Maternal smoking during pregnancy and offspring birth weight: a genetically-informed approach comparing multiple raters. Behav Genet. doi:10.1007/s10519-015-9750-6

Kramer MS (1987) Determinants of low birth weight: methodological assessment and meta-analysis. Bull World Health Organ 65(5):663–737. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2491072/

Kramer MS, Goulet L, Lydon J, Seguin L, McNamara H, Dassa C, Koren G (2001) Socio-economic disparities in preterm birth: causal pathways and mechanisms. Paediatr Perinat Epidemiol 15(suppl 2):104–123

McIntire DD, Bloom SL, Casey BM, Leveno KJ (1999) Birth weight in relation to morbidity and mortality among newborn infants. N Engl J Med 340(16):1234–1238. doi:10.1056/NEJM199904223401603

McNeil TF (1995) Perinatal risk factors and schizophrenia: selective review and methodological concerns. Epidemiol Rev 17(1):107–112. http://epirev.oxfordjournals.org/content/17/1/107.short

Milberger S, Biederman J, Faraone SV, Chen L et al (1996) Is maternal smoking during pregnancy a risk factor for attention deficit hyperactivity disorder in children? Am J Psychiatry 153(9):1138–1142

Narusyte J, Neiderhiser JM, D'Onofrio B, Reiss D, Spotts EL, Ganiban J, Lichtenstein P (2008) Testing different types of genotype-environment correlation: an extended chidlren-of-twins model. Dev Psychol 44(6):1591–1603

Nukui T, Day RD, Sims CS, Ness RB, Romkes M (2004) Maternal/newborn GSTT1 null genotype contributes to risk of preterm, low birthweight infants. Pharmacogenetics 14(9):569–576

Pickett KE, Kasza K, Biesecker G, Wright RJ, Wakschlag LS (2009) Women who remember, women who do not: a methodological study of maternal recall of smoking in pregnancy. Nicotine Tob Res 11(10):1166–1174. doi:10.1093/ntr/ntp117

Plomin R (2014) Genotype-environment correlation in the era of DNA. Behav Genet 44(6):629–638. doi:10.1007/s10519-014-9673-7

Purcell SM, Wray NR, Stone JL, et al (2009) Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 460(7256):748–752. http://www.nature.com/nature/journal/v460/n7256/suppinfo/nature08185_S1.html

Reich W, Todd RD, Joyner CA, Neuman RJ, Heath AC (2003) Reliability and stability of mothers' reports about their pregnancies with twins. Twin Res Hum Genet 6(02):85–88. doi:10.1375/twin.6.2.85

Rice F, Harold GT, Bolvin J, Hay DF, Van den Bree M, Thapar A (2009) Disentangling prenatal and inherited influences in humans with an experimental design. PNAS 106(7):2464–2467

Salvatore JE, Aliev F, Edwards AC, Evans DM, Macleod J, Hickman M, Dick DM (2014) Polygenic scores predict alcohol problems in an independent sample and show moderation by the environment. Genes 5(2):330–346

Sharma E, Mustafa M, Pathak R, Guleria K, Ahmed RS, Vaid NB, Banerjee BD (2012) A case control study of gene environmental interaction in fetal growth restriction with special reference to organochlorine pesticides. Eur J Obstet Gynecol Reprod Biol 161(2):163–169. doi:10.1016/j.ejogrb.2012.01.008

Stone WL, Bailey B, Khraisha N (2014) The pathophysiology of smoking during pregnancy: a systems biology approach. Front Biosci (Elite Ed) 6:318–328

Thapar A, Fowler T, Rice F, et al (2003) Maternal smoking during pregnancy and attention deficit hyperactivity disorder symptoms in offspring. 1985–1989

Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F, Laitinen J (2010) Sequence variants at CHRNB3-CHRNA6 and CYP2A6 affect smoking behavior. Nat Genet 42(5):448–453

Ward C, Lewis S, Coleman T (2007) Prevalence of maternal smoking and environmental tobacco smoke exposure during pregnancy and impact on birth weight: retrospective study using Millennium Cohort. BMC Pub Health 7(1):81. http://www.biomedcentral.com/1471-2458/7/81

Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 88(1):76–82. doi:10.1016/j.ajhg.2010.11.011