

Exploring the association between birthweight and breast cancer using summary statistics from a perspective of genetic correlation, mediation, and causality

Meng Zhang

Xuzhou Medical College: Xuzhou Medical University

Jiahao Qiao

Xuzhou Medical College: Xuzhou Medical University

Shuo Zhang

Xuzhou Medical College: Xuzhou Medical University

Ping Zeng (✉ zpstat@xzhmu.edu.cn)

Xuzhou Medical University <https://orcid.org/0000-0003-2710-3440>

Research Article

Keywords: Breast cancer, birthweight, fetal/maternal-specific effect, summary statistics, Mendelian randomization, genetic correlation, mediation, gene-based association analysis, age of menarche, age at menopause, pleiotropy analysis

Posted Date: January 18th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1145106/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Previous studies demonstrated a positive relationship between birthweight and breast cancer; however, inconsistent, sometimes even controversial, observations also emerged, and the nature of such relationship remains unknown.

Methods

Using summary statistics of birthweight (n=264,498 for own birthweight and 179,360 for offspring birthweight) and breast cancer (n=266,081), we assessed the fetal/maternal-specific genetic correlation between them via LDSC and prioritized fetal/maternal-specific pleiotropic genes through MAIUP. Relying on summary statistics we conducted Mendelian randomization (MR) analyses to evaluate the fetal/maternal-specific origin of causal relationship between birthweight, age of menarche, age at menopause and breast cancer.

Results

With summary statistics we identified a positive genetic correlation between fetal-specific birthweight and breast cancer ($r_g=0.123$ and $P=0.013$) as well as a negative but insignificant correlation between maternal-specific birthweight and breast cancer ($r_g=-0.068$, $P=0.206$); and detected 84 pleiotropic genes shared by fetal-specific birthweight and breast cancer, 49 shared by maternal-specific birthweight and breast cancer. We also revealed fetal-specific birthweight indirectly influenced breast cancer risk in adulthood via the path of age of menarche or age at menopause in terms of MR-based mediation analysis.

Conclusion

This study reveals that shared genetic foundation and causal mediation commonly drive the connection between the two traits, and that fetal/maternal-specific birthweight plays substantially distinct roles in such relationship. However, our work offers little supportive evidence for the fetal origins hypothesis of breast cancer originating in utero.

Background

Breast cancer remains the most frequent malignant tumor that occurs in the glandular epithelium of breast (1), and accounts for approximately 12% of the total 9.6 million deaths due to cancer (2). Since 1970s, the incidence rate of breast cancer worldwide has continued to increase; it is reported that one in

eight women suffered from this type of cancer in USA (3). Although mammary gland is not an important organ to maintain human life and breast cancer *in situ* is not fatal, breast cancer cells may lose characteristics of normal cells and is easy to fall off due to the loose connection among cells. Once falling off, cancer cells would spread throughout the body with blood or lymph, leading to cancer metastasis and thus endangering life (4). Over the past few decades the treatment of breast cancer has been advanced greatly, but the overall survival is still not optimistic (1, 5). Therefore, it is particularly important to understand the etiology, occurrence, and development of breast cancer for early prevention. Existing studies have identified a series of risk factors involved in breast cancer, including dietary habit, age at first birth, age at menarche, age at menopause, family history, excessive intake of exogenous hormones as well as genetic mutations such as *BRCA1*, *BRCA2*, and *PIK3CA* (1, 6–10).

However, these traditionally established risk factors during women's adult life appear not to adequately interpretate the occurrence pattern of breast cancer. To advance our understanding of disease causes, the relationship between breast cancer and early growth/development, perinatal intrauterine environments has been attracted much attention since 1990s (11–23). Among those, the association between birthweight and breast cancer has been attracted much research interest. Although a positive correlation between women's birthweight and breast cancer risk was discovered in some studies (11, 12, 20, 23–32), some others failed to replicate such connection or even detected inconsistent correlations in effect direction (13, 15, 21, 22, 33–41). These inconsistent findings may be partly due to potential confounding influences commonly arisen in observational studies, making it difficult to draw a definitive conclusion on the causal association between birthweight and breast cancer. In addition, it is not clear whether there exists a mediating association between the two traits (20, 42).

Furthermore, from a genetic perspective, it is also not known whether the observed co-existence of low/high birthweight and breast cancer is partly driven by causal association or shared genetic background between the two traits. All prior studies cannot distinguish the maternal-specific and fetal-specific effects of birthweight on breast cancer from each other. Compared to other factors, birthweight is a special exposure proxy genetically affected by both mother's and offspring's genotypes (43). Therefore, partitioning the overall effect of birthweight into maternal-specific and fetal-specific components holds the key for understanding the origin of the association between birthweight and breast cancer. Although the longitudinal cohort study can provide empirical evidence for causal inference, it requires large-scale populations and long-term follow-up before the onset of breast cancer (44); consequently, the implementation is not easy. In the traditional scenario, randomized controlled trial is the gold standard for inferring causality, but such study is also infeasible to investigate the causal association between birthweight and breast cancer (29). In addition, both the two types of studies cannot resolve the maternal-specific and fetal-specific impacts of birthweight on adult diseases including breast cancer.

The present work attempted to answer these critical questions via genetic analysis using summary-level data available from large-scale genome-wide association studies (GWASs). First, to assess the extent of genetic overlap shared between birthweight and breast cancer, we applied the cross-trait linkage disequilibrium score regression (LDSC) to quantify the genetic correlation between them (45). Second, we

employed a novel pleiotropy test method called MAIUP (Mixture Adjusted Intersect-Union Pleiotropy test) to determine pleiotropic genes (46–48). Third, to elaborate the causal association between birthweight and breast cancer, we resorted to apply Mendelian randomization (MR) methods (49–52). In the MR analysis, genetic variants, which are required to be associated with the exposure of focus, are used as instrumental variables, based on which the causal association between the exposure (e.g., birthweight) and the disease (e.g., breast cancer) can be inferred. Recently, one MR study was performed but found no evidence supporting the causal association between birthweight and breast cancer (53). However, that study did not explore the separate maternal-specific and fetal-specific effects of birthweight on breast cancer. The summary statistics of maternal/fetal-specific effects of SNPs (single nucleotide polymorphisms) on birthweight, released by a recent GWAS (43), offers us an unprecedented opportunity to untangle the maternal and fetal contributions of birthweight to breast cancer by using novel MR methods. Furthermore, as a byproduct of our MR analysis, we can evaluate the mediating relationship between birthweight and breast cancer, with age of menarche and age at menopause as two candidate mediators. The flow diagram of data process and statistical analysis for the present study is illustrated in Figure 1.

Methods

GWAS summary statistics

We obtained summary statistics (e.g., marginal effect size and standard error for SNPs) of birthweight from the largest GWAS to date published by the Early Growth Genetics consortium (43). This study analyzed 298,142 individuals' genotypes for their own birthweight and 210,267 individuals' genotypes for their offspring birthweight, with all being of European ancestry. To distinguish fetal and maternal genetic influence, the structural equation model was applied to a few of significant SNPs (54); however, it was computationally intensive and inflexible to jointly all available SNPs. A computationally quick weighted linear model was thus developed to estimate the effect size of SNPs on their own and offspring's birthweight across the entire genome (43). Then, the maternal-specific effect of SNP on birthweight adjusting for the genotypes of offspring and the fetal-specific effect of SNP on birthweight adjusting for the genotypes of mother were generated based on 264,498 individuals with their own birthweight and 179,360 individuals with offspring birthweight, which provides us an in-depth understanding of the biological regulation of birthweight and allows us to further investigate the origin of observed relationship between birthweight and breast cancer.

We yielded summary statistics of breast cancer from (53), where a total of 266,081 (142,798 cases and 123,283 controls) European ancestry female individuals and 12,937,276 SNPs were analyzed. In addition, we obtained summary statistics of age at menarche and age at menopause from the ReproGen consortium. Specifically, the GWAS of age at menarche included 329,345 European female individuals and 9,903,009 SNPs (55), while the GWAS of age at menopause included 69,360 European female individuals and 2,384,867 SNPs (56).

Genetic correlation estimation with LDSC

To assess the shared polygenic component between fetal/maternal-specific birthweight and breast cancer, we performed LDSC to estimate the overall genetic correlation (45). Conceptually, genetic correlation is defined as the correlation between SNP effect sizes of two traits to quantify the extent to which the same genetic variants have the same or similar impact on phenotypic variation (57, 58). Based on summary statistics, the cross-trait LDSC measures genome-wide genetic overlap using all available SNPs while taking LD in account. The LD score of every SNP was computed with genotypes of 7,120,251 common SNPs (with minor allele frequency >0.05 and the P value of the Hardy Weinberg equilibrium test $>1 \times 10^{-5}$) with a 10Mb window on 503 individuals of European descent in the 1000 Genomes Project (59). Then, the LDSC analysis proceeded by regressing the product of Z-statistics of the two traits on the LD score in a weighted manner via the python script offered by the developers with default parameter settings. The regression slope of LDSC provided an unbiased estimate for genetic correlation even when overlapping individuals existed between the two GWASs. Before the analysis, the stringent quality control (e.g., removing SNPs located within the MHC region) was carried out on summary statistics of birthweight and breast cancer following prior work (45).

Pleiotropic gene identification with MAIUP

Using summary statistics of birthweight and breast cancer, we attempted to further identify fetal/maternal-specific gene-level pleiotropic associations shared by the two traits. Here, pleiotropy is defined the phenomenon that a given gene is associated with both traits under investigation (60–63). Statistically, the presence of pleiotropy means that both the P values (say P_1 and P_2) of a particular gene should be equal or less than the preassigned significance level (say α); that is, $P_1 \leq \alpha$ and $P_2 \leq \alpha$ (H_{11}) need to be held simultaneously for pleiotropic association. In contrast, the absence of pleiotropy implies that at least one of the two P values would be larger than the significance level, which includes three sub-null scenarios; that is, (i) H_{00} : $P_1 > \alpha$ and $P_2 > \alpha$, (ii) H_{01} : $P_1 > \alpha$ and $P_2 < \alpha$, and (iii) H_{10} : $P_1 < \alpha$ and $P_2 > \alpha$. From a statistical perspective, it is easy to see that the pleiotropy detection can be viewed as a high-dimensional challenging issue of composite null hypothesis testing (46). To address this problem effectively, we employed a recently proposed pleiotropy test method called MAIUP to detect common genetic loci underlying birthweight and breast cancer (64). Methodologically, MAIUP is constructed under the principle of intersect-union test originally proposed within the framework of high-dimensional mediation analysis (47, 48, 64), which takes two sets of P values for each gene as input with a three-component mixture null distribution for its test statistics and generally behaves much better in power compared to other existing pleiotropy test methods. Technical details regarding MAIUP can be found in (64).

To generate P value for each gene in GWAS, we need to first integrate multiple SNP-level association signals into a single gene-level association signal. For this aim, we applied MAGMA (Multi-marker Analysis of GenoMic Annotation) which is a powerful SNP-set test method and can be efficiently conducted via user-friendly software (65). Due to population stratification, family structures, and cryptic relatedness (66–68), the empirical null distribution in MAGMA may be sometimes inflated. In order to correct such deviation, before the formal pleiotropy analysis we performed genomic control if the inflation was observed, which was measured by the genomic control inflation factor (>1.05) for chi-square

statistics (69, 70). Afterwards, P values for all genes were available for each trait. Depending on these P values, we conducted MAIUP to discover significant genes that were simultaneously associated with birthweight and breast cancer.

Mendelian randomization for causal association between birthweight, age at menarche, age at menopause, and breast cancer

We finally evaluated the causal association among the four traits using various MR methods. Following prior studies (43, 71, 72), we selected a set of independent birthweight-associated SNPs ($P < 6.60 \times 10^{-9}$ and $r^2 < 0.10$) as instruments. Specifically, the total of 104 instruments for fetal-specific birthweight included 63 fetal-effect specific SNPs, 26 SNPs exerting both fetal and maternal effects with the same effect direction and 15 SNPs exhibiting both fetal and maternal effects but with the opposite effect direction (Table S1); while the total of 72 instruments for maternal-specific birthweight included 31 maternal-effect specific SNPs, 26 SNPs exerting both fetal and maternal effects with the same effect direction, 15 SNPs exhibiting both fetal and maternal effects but with the opposite effect direction (Table S2). To avoid weak instrument bias, 71 SNPs with unclassified effect direction were excluded as fetal-specific or maternal-specific birthweight instruments. With these instruments of birthweight, we estimated the fetal/maternal-specific causal effect of birthweight on age at menarche, age at menopause, or breast cancer. In addition, to estimate the causal effect of age at menarche or age at menopause on breast cancer, we selected independent associated SNPs for age at menarche or age at menopause as candidate instruments by applying the clumping procedure in PLINK (73). We set the primary and secondary significance levels of indexed SNPs to 5×10^{-8} , r^2 to 0.001 and physical distance to 1Mb, with the 1000 Genomes Project as the reference panel.

We estimated the causal effect primarily using the IVW method (74, 75). To examine the causal relationship between birthweight and breast cancer while considering menarche age and menopausal age as potential confounding factors, we conducted the multivariable inverse-variance weighted method (76, 77). We also assessed the causal relationship between age at menarche (or/and age at menopause) and breast cancer assuming birthweight was a confounding factor. In order to avoid the impact of horizontal pleiotropy, we relied on a conservative strategy to exclude some candidate instruments that had a P value less than 0.05 after Bonferroni's correction (78–80).

Results

Estimated genetic correlation between birthweight and breast cancer

We observed there existed a substantial genetic correlation between fetal-specific birthweight and breast cancer ($r_g = 0.123$ and $P = 0.013$), in contrast to the negative but non-significant genetic correlation between maternal-specific birthweight and breast cancer ($r_g = -0.068$ and $P = 0.206$). This finding is slightly in contrast to results in prior work (43) where neither fetal-specific birthweight nor maternal-specific birthweight was genetically related to breast cancer ($r_g = 0.015$ and $P = 0.828$ for fetal birthweight; $r_g = -0.072$ and $P = 0.321$ for maternal birthweight); whereas both showed consistent direction for fetal-specific or

maternal-specific birthweight. The opposite direction in genetic correlation can be expected as the maternal-specific and fetal-specific SNP effects on birthweight are inversely correlated (43). In addition, we did not discover significant genetic correlation between birthweight and age at menarche (or age at menopause) (Table S3). These non-significant genetic correlations do not necessarily imply the absence of shared genetic component between birthweight and breast cancer (or the two ages) as r_g only measures the average genetic correlation of effect sizes for all SNPs across the whole genome, which does not capture detailed patterns for individual shared genetic loci. For example, the mixture of a significantly positive genetic correlation for a local region and a significantly negative genetic correlation for another local region would lead to a non-significant overall genetic correlation, as partly demonstrated by the chromosome-specific genetic correlation in Figure 2, where both negative and positive relationships were present across the chromosomes. Therefore, we cannot completely rule out the possibility that genes in some local genetic regions would be associated with both birthweight and breast cancer.

Pleiotropic genes for birthweight and breast cancer

Using MAIUP (64) we identified a large set of commonly associated genes that were shared by birthweight and breast cancer. Specifically, there were 84 pleiotropic genes (false discovery rate [FDR] <0.05) shared by fetal-specific birthweight and breast cancer (Table S4), while the number was 49 between maternal-specific birthweight and breast cancer (Table S5). Among the two sets of pleiotropic associations, there existed 16 common genes including *ANO8*, *BBS1*, *C15orf39*, *CDKAL1*, *DDA1*, *DPP3*, *FAM219B*, *GOLGA6C*, *GTPBP3*, *LOC100652768*, *MPI*, *PCSK7*, *PELI3*, *SCAMP2*, *TAGLN*, and *ZDHHC24*. Several genes were previously confirmed to have a connection with breast cancer. For example, it was revealed that, together with the molecular subtype, the expression signature of *BBS1* was significantly related to the bone metastasis status of breast cancer and encoded mainly membrane-bound molecules with molecular function of protein binding (81). A locus, rs9368197, located within *CDKAL1* (intron) was detected to be associated with increased breast cancer risk in European American women (82). As another example, *TAGLN* was identified as a target of DNA hypermethylation in breast cancer by using microarray expression profiling of AZA- or DMSO-treated breast cancer and non-tumorigenic breast cells (83).

To evaluate the similarity of genetic influence of these pleiotropic genes, for every shared gene we further calculated the Pearson's correlation coefficient of effect sizes between birthweight and breast cancer with local SNPs belonging to that gene. We found that most of these pleiotropic genes (~68.4%) – 47 out of 84 for fetal-specific birthweight and 44 out of 49 for maternal-specific birthweight – displayed a positive correlation in effect direction, meaning that they generally exerted consistent genetic impact on birthweight and breast cancer (Figure 3A-B). These genes with consistent effects are believed to contribute to the observed positive relationship between birthweight and breast cancer. The remaining pleiotropic genes demonstrated a negative correlation in SNP effect sizes between the two traits, indicating that these genes exhibit functionally different influence on birthweight and breast cancer. Note that, this phenomenon of antagonistic effects of shared genetic loci is also widely observed for other

genetically correlated traits such as psychiatric disorders (84, 85) and immune-mediated diseases (86–88).

We further compared the correlation coefficients for the 16 genes shared by fetal/maternal-specific birthweight with breast cancer. It is very interesting that the two sets of correlation coefficients were completely opposite, with a highly negative correlation between themselves (Figure 3C), indicating that these pleiotropic genes showed considerably distinct genetic impact on birthweight and breast cancer. For example, the genes *FAM219B* and *TAGLN* presented a negative correlation of SNP effect sizes between fetal-specific birthweight and breast cancer ($r=-0.122$ and 0.194 , respectively), but displayed a positive correlation between maternal-specific birthweight and breast cancer ($r=0.121$ and 0.168 , respectively).

Estimated causal effect with MR analyses

We conducted a set of MR analyses to assess the causal association between birthweight, breast cancer, age at menarche, and age at menopause. It is worth mentioning that we had chosen two different sets of instruments for birthweight: one set of SNPs with fetal-specific effect on birthweight and another set of SNPs with maternal-specific effect on birthweight (Tables S1-S2), which offered us an effective manner to untangle the origin of the observed relationship between birthweight and breast cancer. First, we carried out the univariate inverse-variance weighted (IVW) analysis to evaluate the impact birthweight on breast cancer, but failed to find evidence of causal relationship between birthweight and breast cancer ($P=0.806$ for the fetal-specific birthweight, $P=0.244$ for the maternal-specific birthweight). Based on an online simulation with the same sample size and proportion of disease cases used here (89), we had ~70% statistical power when the odds ratio was set to 0.90 (or 1.10) at the significance level of 0.05, largely ruling out the likelihood that the null causal association was due to low power. Then, we performed the univariate IVW analysis to assess the association between birthweight and age at menarche, and observed that fetal-specific birthweight was positively correlated to age at menarche ($\beta=0.089$, 95%CI: 0.019~0.0158), indicating higher birthweight can delay age at menarche in a fetal way; but we did not detect a substantial association between maternal-specific birthweight and age at menarche ($P=0.907$).

Next, we conducted the multivariate IVW analysis to assess the relationship between birthweight and age at menopause while controlling the influence of age at menarche. We discovered that there was no substantial causal association between birthweight and age at menopause ($P=0.927$ for the fetal-specific birthweight and $P=0.590$ for the maternal-specific birthweight) but found that age at menarche was a significant confounder with a positive effect on age at menopause ($\beta=0.111$ and $P=0.035$ for fetal-specific birthweight; $\beta=0.123$ and $P=0.062$ for maternal-specific birthweight). Note that, age at menarche can influence age at menopause but not vice versa. We also examined the relationship between age at menarche and breast cancer while adjusting for birthweight but did not observe obviously causal association between them ($P=0.525$ when controlling for fetal-specific birthweight, and $P=0.368$ when controlling for maternal-specific birthweight).

Finally, we evaluated the relationship between age at menopause and breast cancer while controlling for birthweight and age at menarche. We discovered that there only existed a significant positive correlation between age at menopause and breast cancer ($P=0.029$ for when adjusting for fetal-specific birthweight and age at menarche, $P=0.032$ for when adjusting for maternal-specific birthweight and age at menarche). Again, according to the naïve principle of mediation analysis, we can conclude that age at menarche and age at menopause mediated the impact of birthweight on adult breast cancer, among which the fetal role appeared much more evident. The associations identified by distinct MR analyses are demonstrated in Figure 4, with the detailed results further shown in Table S6.

Discussion

In the present study we have investigated the genetic correlation and causal association between birthweight and breast cancer. To our knowledge, the present work is among the first endeavor to study the relationship between the two traits by leveraging novel statistical methods with large-scale summary-level genetic data. As a result, we offered implicit answers for some key questions regarding such relationship between them. First, to understand whether the observed relationship is due to common genetic background, we employed LDSC (45) and identified a positively significant genetic correlation between fetal-specific birthweight and breast cancer as well as a negative but insignificant genetic correlation between maternal-specific birthweight and breast cancer. Moreover, using MAIUP (64) we showed that there were extensively common genetic loci underlying the two traits. Second, to examine whether the observed relationship represents a linear causality between birthweight and breast cancer, we carried out the MR analysis but did not identify a linear causal association, which is in agreement with the null finding obtained from another MR study (39). Third, to determine whether some growth traits and life processes may mediate the long-term impact of birthweight on breast cancer, we depended on the principle of mediation analysis (47, 90–93) and demonstrated that fetal-specific birthweight can indirectly influence breast cancer risk in adulthood via the path of age of menarche or age at menopause.

Unlike prior relevant studies (39), one of the remarkable strengths of our work is that we resolved the relative contributions of fetal and maternal genotypes on birthweight and employed fetal/maternal-specific effects of birthweight in our genetic overlap analysis as well as in our MR analysis, which provides us an unprecedented opportunity to untangle the origin of the relationship between birthweight and breast cancer (43, 71, 72). For example, we discovered that fetal-specific birthweight was genetically correlated to breast cancer in a positive direction, while maternal-specific birthweight showed a negative genetic correlation to breast cancer. In addition, as demonstrated, the pleiotropic genes shared between fetal-specific birthweight and breast cancer was not completely overlapped with those shared between maternal-specific birthweight and breast cancer, implying the diverse contribution of fetal-specific birthweight and maternal-specific birthweight to the observed relationship. Furthermore, we found that fetal-specific birthweight, rather than maternal-specific birthweight, was causally associated with age of menarche which further would affect age at menopause and breast cancer, indicating that, together the identified positive genetic correlation mentioned above, fetal-specific birthweight might exert a more pronounced influence on the development of breast cancer in later life compared to maternal-specific

birthweight. Together, these findings suggest that the growth environment in childhood might be very important for the development of breast cancer in adulthood. However, there is insufficient evidence of maternal-specific birthweight effect on offspring's breast cancer, implying that the maternal intrauterine environment does not seem to be the major determinant of the risk of breast cancer.

Some limitations of the current work should be mentioned. First, like other MR studies, we assumed a linear effect association between birthweight and breast cancer but cannot completely rule out the likelihood of nonlinear association between birthweight and breast cancer as suggested in prior studies (15, 29). Second, no data on the duration and severity of breast cancer can be available for us; therefore, we cannot assess the dose-response association between birthweight and breast cancer, which is an important aspect of causal inference. Third, due to unavailability of relevant data, we cannot further assess the impact of birthweight on distinct subtypes of breast cancer. Because of the same reason, we also cannot evaluate the association between birthweight and breast cancer stratified by the menopausal status (15), which may indicate various influence of birthweight on breast cancer (22, 23, 94). Fourth, because the traditional mediation test methods such as the Sobel test and the joint significance test (47), in our analysis we did employ any formal approaches but only applied the naïve principle that the presence of both the exposure-mediator effect and mediator-outcome effect indicates the existence of mediation effect. Therefore, powerful mediation test methods would be warranted for a single or only a few mediators under the summary-level framework, which is our ongoing work. Fifth, we here only two mediators (e.g., age of menarche and age at menopause) were considered; a more comprehensive evaluation of growth traits and life processes are projected to discover other causal paths from birthweight to breast cancer.

Overall, this study reveals that shared genetic foundation and causal mediation commonly drive the connection between the two traits, and that fetal/maternal-specific birthweight plays substantially distinct roles in such relationship. However, our work offers little supportive evidence for the fetal origins hypothesis of breast cancer originating in utero.

Abbreviations

LDSC

linkage disequilibrium score regression

GWAS

genome-wide association study

MAIUP

Mixture Adjusted Intersect-Union Pleiotropy test

MR

Mendelian randomization

MAGMA

Multi-marker Analysis of GenoMic Annotation

IVW

inverse variance weighted method
SNP
single nucleotide polymorphism

Declarations

Supplementary information

Supplementary File.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analyzed during this study are included in this article and its supplementary information files.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank the GWAS consortia of birthweight, breast cancer, age at menarche and age at menopause for making summary statistics publicly available for us and are grateful to all the investigators and participants contributed to those studies. The data analyses in the present study were carried out with the high-performance computing cluster that was supported by the special central finance project of local universities for Xuzhou Medical University.

Funding

The research of Ping Zeng was supported in part by the National Natural Science Foundation of China (82173630 and 81402765), the Youth Foundation of Humanity and Social Science funded by Ministry of Education of China (18YJC910002), the Natural Science Foundation of Jiangsu Province of China (BK20181472), the China Postdoctoral Science Foundation (2018M630607 and 2019T120465), the Six-Talent Peaks Project in Jiangsu Province of China (WSN-087), the Training Project for Youth Teams of Science and Technology Innovation at Xuzhou Medical University (TD202008).

Author's contributions

PZ conceived the idea for the study. PZ obtained the data. PZ, JQ and MZ cleared up the datasets; PZ, SZ, JQ and MZ performed the data analyses. PZ and MZ interpreted the results of the data analyses. PZ, JQ and MZ wrote the manuscript with the participation of other authors.

References

1. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. *The Lancet*. 2021;397(10286):1750–69.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018;68(6):394–424.
3. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J. Breast Cancer: Epidemiology and Etiology. *Cell Biochem Biophys*. 2015;72(2):333–8.
4. Peart O. Metastatic Breast Cancer. *Radiologic technology*. 2017;88(5):519M-39M.
5. Wörmann B. Breast cancer: basics, screening, diagnostics and treatment. *Medizinische Monatsschrift für Pharmazeuten*. 2017;40(2):55–64.
6. Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. *Int J Cancer*. 1990;46(5):796–800.
7. Lipworth L. Epidemiology of breast cancer. *European journal of cancer prevention: the official journal of the European Cancer Prevention Organisation (ECP)*. 1995;4(1).
8. Key TJ, Verkasalo PK, Banks E. Epidemiology of breast cancer. *The Lancet Oncology*. 2001;2(3):133–40.
9. Sasco AJ. Epidemiology of breast cancer: an environmental disease? *APMIS: acta pathologica, microbiologica, et immunologica Scandinavica*. 2001;109(5):321–32.
10. Parks RM, Derks MGM, Bastiaannet E, Cheung KL. Breast Cancer Epidemiology. In: Wyld L, Markopoulos C, Leidenius M, Senkus-Konefka E, editors. *Breast Cancer Management for Surgeons: A European Multidisciplinary Textbook*. Cham: Springer International Publishing; 2018. p. 19–29.
11. Barber LE, Bertrand KA, Rosenberg L, Battaglia TA, Palmer JR. Pre- and perinatal factors and incidence of breast cancer in the Black Women's Health Study. *Cancer causes & control: CCC*. 2019;30(1):87–95.
12. Michels KB, Trichopoulos D, Robins JM, Rosner BA, Manson JE, Hunter DJ, et al. Birthweight as a risk factor for breast cancer. *Lancet (London, England)*. 1996;348(9041):1542–6.
13. Sanderson M, Pérez A, Weriwoh ML, Alexander LR, Peltz G, Agboto V, et al. Perinatal factors and breast cancer risk among Hispanics. *J Epidemiol Glob Health*. 2013;3(2):89–94.
14. Sanderson M, Williams MA, Daling JR, Holt VL, Malone KE, Self SG, et al. Maternal factors and breast cancer risk among young women. *Paediatr Perinat Epidemiol*. 1998;12(4):397–407.

15. Sanderson M, Williams MA, Malone KE, Stanford JL, Emanuel I, White E, et al. Perinatal factors and risk of breast cancer. *Epidemiology (Cambridge, Mass)*. 1996;7(1):34–7.
16. Trichopoulos D. Passive smoking, birthweight, and oestrogens. *Lancet (London, England)*. 1986;2(8509):743.
17. Trichopoulos D. Hypothesis: does breast cancer originate in utero? *The Lancet*. 1990;335(8695):939–40.
18. Steer PJ. Maternal hemoglobin concentration and birth weight. *The American journal of clinical nutrition*. 2000;71(5):1285S-7S.
19. Kaijser M, Akre O, Cnattingius S, Ekblom A. Preterm birth, birth weight, and subsequent risk of female breast cancer. *Br J Cancer*. 2003;89(9):1664–6.
20. dos Santos Silva I, De Stavola BL, Hardy RJ, Kuh DJ, McCormack VA, Wadsworth MEJ. Is the association of birth weight with premenopausal breast cancer risk mediated through childhood growth? *Br J Cancer*. 2004;91(3):519–24.
21. Hodgson ME, Newman B, Millikan RC. Birthweight, parental age, birth order and breast cancer risk in African-American and white women: a population-based case-control study. *Breast cancer research: BCR*. 2004;6(6):R656-R67.
22. Luo J, Chen X, Manson JE, Shadyab AH, Wactawski-Wende J, Vitolins M, et al. Birth weight, weight over the adult life course and risk of breast cancer. *Int J Cancer*. 2020;147(1):65–75.
23. Zhou W, Chen X, Huang H, Liu S, Xie A, Lan L. Birth Weight and Incidence of Breast Cancer: Dose-Response Meta-analysis of Prospective Studies. *Clin Breast Cancer*. 2020;20(5):e555-e68.
24. Michels KB, Xue F, Terry KL, Willett WC. Longitudinal study of birthweight and the incidence of breast cancer in adulthood. *Carcinogenesis*. 2006;27(12):2464–8.
25. Xu X, Dailey AB, Peoples-Sheps M, Talbott EO, Li N, Roth J. Birth weight as a risk factor for breast cancer: a meta-analysis of 18 epidemiological studies. *J Womens Health (Larchmt)*. 2009;18(8):1169–78.
26. Silva IdS, De Stavola B, McCormack V. Birth size and breast cancer risk: re-analysis of individual participant data from 32 studies. *PLoS Med*. 2008;5(9):e193.
27. Wu AH, McKean-Cowdin R, Tseng C-C. Birth weight and other prenatal factors and risk of breast cancer in Asian-Americans. *Breast Cancer Res Treat*. 2011;130(3):917–25.
28. Vatten LJ, Maehle BO, Lund Nilssen TI, Tretli S, Hsieh Cc, Trichopoulos D, et al. Birth weight as a predictor of breast cancer: a case-control study in Norway. *Br J Cancer*. 2002;86(1):89–91.
29. Møller L, Olsen ML, Sørensen HT, Thulstrup AM, Olsen J, Olsen JH. Birth weight and risk of early-onset breast cancer (Denmark). *Cancer Causes Control*. 2003;14(1):61–4.
30. Ahlgren M, Sørensen T, Wohlfahrt J, Haflidadóttir A, Holst C, Melbye M. Birth weight and risk of breast cancer in a cohort of 106,504 women. *Int J Cancer*. 2003;107(6).
31. Ahlgren M, Melbye M, Wohlfahrt J, Sørensen TIA. Growth Patterns and the Risk of Breast Cancer in Women. *N Engl J Med*. 2004;351(16):1619–26.

32. Swerdlow AJ, Wright LB, Schoemaker MJ, Jones ME. Maternal breast cancer risk in relation to birthweight and gestation of her offspring. *Breast cancer research: BCR*. 2018;20(1):110.
33. Ekblom A, Adami HO, Trichopoulos D, Hsieh CC, Lan SJ. Evidence of prenatal influences on breast cancer risk. *The Lancet*. 1992;340(8826):1015–8.
34. Ekblom A, Hsieh CC, Lipworth L, Adami HQ, Trichopoulos D. Intrauterine environment and breast cancer risk in women: a population-based study. *J Natl Cancer Inst*. 1997;89(1):71–6.
35. Hilakivi-Clarke L, Forsén T, Eriksson JG, Luoto R, Tuomilehto J, Osmond C, et al. Tallness and overweight during childhood have opposing effects on breast cancer risk. *Br J Cancer*. 2001;85(11):1680–4.
36. Sanderson M, Shu XO, Jin F, Dai Q, Ruan Z, Gao YT, et al. Weight at birth and adolescence and premenopausal breast cancer risk in a low-risk population. *Br J Cancer*. 2002;86(1):84–8.
37. Hajiebrahimi M, Bahmanyar S, Oberg S, Iliadou AN, Cnattingius S. Breast cancer risk in opposite-sexed twins: influence of birth weight and co-twin birth weight. *J Natl Cancer Inst*. 2013;105(23):1833–6.
38. Andersen ZJ, Baker JL, Bihmann K, Vejborg I, Sørensen TIA, Lyng E. Birth weight, childhood body mass index, and height in relation to mammographic density and breast cancer: a register-based cohort study. *Breast cancer research: BCR*. 2014;16(1):R4-R.
39. Kar SP, Andrulis IL, Brenner H, Burgess S, Chang-Claude J, Conside D, et al. The association between weight at birth and breast cancer risk revisited using Mendelian randomisation. *Eur J Epidemiol*. 2019;34(6):591–600.
40. Spracklen CN, Wallace RB, Sealy-Jefferson S, Robinson JG, Freudenheim JL, Wellons MF, et al. Birth weight and subsequent risk of cancer. *Cancer Epidemiol*. 2014;38(5):538–43.
41. Le Marchand L, Kolonel LN, Myers BC, Mi MP. Birth characteristics of premenopausal women with breast cancer. *Br J Cancer*. 1988;57(4):437–9.
42. Bukowski R, Chlebowski RT, Thune I, Furberg AS, Hankins GD, Malone FD, et al. Birth weight, breast cancer and the potential mediating hormonal environment. *PLoS ONE*. 2012;7(7):e40199.
43. Warrington NM, Beaumont RN, Horikoshi M, Day FR, Helgeland Ø, Laurin C, et al. Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nat Genet*. 2019;51(5):804–14.
44. Hilakivi-Clarke L, Forsén T, Eriksson JG, Luoto R, Tuomilehto J, Osmond C, et al. Tallness and overweight during childhood have opposing effects on breast cancer risk. *British journal of cancer*. 2001;85(11):1680–4.
45. 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;47(11):1236–41.
46. Ray D, Chatterjee N. A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between Type 2 Diabetes and Prostate Cancer. *PLoS Genet*. 2020;16(12):e1009218.

47. Zeng P, Shao Z, Zhou X. Statistical methods for mediation analysis in the era of high-throughput genomics: current successes and future challenges. *Computational and Structural Biotechnology Journal*. 2021;19:3209–24.
48. Dai JY, Stanford JL, LeBlanc M. A Multiple-Testing Procedure for High-Dimensional Mediation Hypotheses. *J Am Stat Assoc*. 2020:1–16.
49. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res*. 2017;26(5):2333–55.
50. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol*. 2017;46(6):1985–98.
51. Bowden J, Del Greco M F, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I^2 statistic. *Int J Epidemiol*. 2016;45(6):1961–74.
52. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol*. 2017;46(6):1734–9.
53. Zhang H, Ahearn TU, Lecarpentier J, Barnes D, Beesley J, Qi G, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet*. 2020;52(6):572–81.
54. Warrington NM, Freathy RM, Neale MC, Evans DM. Using structural equation modelling to jointly estimate maternal and fetal effects on birthweight in the UK Biobank. *Int J Epidemiol*. 2018;47(4):1229–41.
55. Day FR, Thompson DJ, Helgason H, Chasman DI, Finucane H, Sulem P, et al. Genomic analyses identify hundreds of variants associated with age at menarche and support a role for puberty timing in cancer risk. *Nat Genet*. 2017;49(6):834–41.
56. Day FR, Ruth KS, Thompson DJ, Lunetta KL, Pervjakova N, Chasman DI, et al. Large-scale genomic analyses link reproductive aging to hypothalamic signaling, breast cancer susceptibility and BRCA1-mediated DNA repair. *Nat Genet*. 2015;47(11):1294–303.
57. van Rheenen W, Peyrot WJ, Schork AJ, Lee SH, Wray NR. Genetic correlations of polygenic disease traits: from theory to practice. *Nat Rev Genet*. 2019;20(10):567–81.
58. Lu H, Wang T, Zhang J, Zhang S, Huang S, Zeng P. Evaluating marginal genetic correlation of associated loci for complex diseases and traits between European and East Asian populations. *Hum Genet*. 2021;140(9):1285–97.
59. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature*. 2015;526(7571):68–74.
60. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW. Pleiotropy in complex traits: challenges and strategies. *Nat Rev Genet*. 2013;14(7):483–95.
61. Watanabe K, Stringer S, Frei O, Umićević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339–48.

62. Sivakumaran S, Agakov F, Theodoratou E, Prendergast JG, Zgaga L, Manolio T, et al. Abundant pleiotropy in human complex diseases and traits. *Am J Hum Genet.* 2011;89(5):607–18.
63. Zeng P, Hao X, Zhou X. Pleiotropic mapping and annotation selection in genome-wide association studies with penalized Gaussian mixture models. *Bioinformatics.* 2018;34(16):2797–807.
64. Wang T, Lu H, Zeng P. Identifying pleiotropic genes for complex phenotypes with summary statistics from a perspective of composite null hypothesis testing. *Brief Bioinform.* 2021:10.1093/bib/bbab389.
65. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS computational biology.* 2015;11(4):e1004219.
66. Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet.* 2003;361(9357):598–604.
67. Jiang Y, Epstein MP, Conneely KN. Assessing the impact of population stratification on association studies of rare variation. *Hum Hered.* 2013;76(1):28–35.
68. van den Berg S, Vandenplas J, van Eeuwijk FA, Lopes MS, Veerkamp RF. Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. *J Anim Breed Genet.* 2019;136(6):418–29.
69. Dadd T, Weale ME, Lewis CM. A critical evaluation of genomic control methods for genetic association studies. *Genet Epidemiol.* 2009;33(4):290–8.
70. Zeng P, Zhao Y, Qian C, Zhang L, Zhang R, Gou J, et al. Statistical analysis for genome-wide association study. *J Biomed Res.* 2015;29(4):285–97.
71. Yu X, Wei Y, Zeng P, Lei S. Birth weight is positively associated with adult osteoporosis risk: observational and Mendelian randomization studies. *J Bone Miner Res.* 2021;36(8):1469–80.
72. Yu X, Yuan Z, Lu H, Gao Y, Chen H, Shao Z, et al. Relationship between birth weight and chronic kidney disease: evidence from systematic review and two-sample Mendelian randomization analysis. *Hum Mol Genet.* 2020;29(13):2261–74.
73. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559–75.
74. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. *Int J Epidemiol.* 2017;46(6):1985–98.
75. Yavorska OO, Burgess S. MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int J Epidemiol.* 2017;46(6):1734–9.
76. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol.* 2015;181(4):251–60.
77. Yu X, Wang T, Chen Y, Shen Z, Gao Y, Xiao L, et al. Alcohol Drinking and Amyotrophic Lateral Sclerosis: An Instrumental Variable Causal Inference. *Ann Neurol.* 2020;88(1):195–8.
78. Yu X, Wang T, Chen Y, Shen Z, Gao Y, Xiao L, et al. Alcohol Drinking and Amyotrophic Lateral Sclerosis: An Instrumental Variable Causal Inference. *Ann Neurol.* 2020;88(1):195–8.

79. Zeng P, Wang T, Zheng J, Zhou X. Causal association of type 2 diabetes with amyotrophic lateral sclerosis: new evidence from Mendelian randomization using GWAS summary statistics. *BMC Med.* 2019;17(1):225.
80. Zeng P, Zhou X. Causal effects of blood lipids on amyotrophic lateral sclerosis: a Mendelian randomization study. *Hum Mol Genet.* 2019;28(4):688–97.
81. Savci-Heijink CD, Halfwerk H, Koster J, van de Vijver MJ. A novel gene expression signature for bone metastasis in breast carcinomas. *Breast Cancer Res Treat.* 2016;156(2):249–59.
82. Cheng TY, Shankar J, Zirpoli G, Roberts MR, Hong CC, Bandera EV, et al. Genetic variants in the mTOR pathway and interaction with body size and weight gain on breast cancer risk in African-American and European American women. *Cancer Causes Control.* 2016;27(8):965–76.
83. Sayar N, Karahan G, Konu O, Bozkurt B, Bozdogan O, Yulug IG. Transgelin gene is frequently downregulated by promoter DNA hypermethylation in breast cancer. *Clin Epigenetics.* 2015;7:104.
84. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, et al. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell.* 2019;179(7):1469-82.e11.
85. Rees E, Kirov G, Sanders A, Walters JTR, Chambert KD, Shi J, et al. Evidence that duplications of 22q11.2 protect against schizophrenia. *Mol Psychiatry.* 2014;19(1):37–40.
86. Baurecht H. Genome-wide comparative analysis of atopic dermatitis and psoriasis gives insight into opposing genetic mechanisms. *Am J Hum Genet.* 2015;96:104–20.
87. Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet.* 2008;17:R116-R21.
88. Schmitt J, Schwarz K, Baurecht H, Hotze M, Fölster-Holst R, Rodríguez E, et al. Atopic dermatitis is associated with an increased risk for rheumatoid arthritis and inflammatory bowel disease, and a decreased risk for type 1 diabetes. *J Allergy Clin Immunol.* 2016;137(1):130–6.
89. Brion M-JA, Shakhbazov K, Visscher PM. Calculating statistical power in Mendelian randomization studies. *Int J Epidemiol.* 2013;42(5):1497–501.
90. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. *Psychol Methods.* 2002;7(1):83–104.
91. MacKinnon DP, Fairchild AJ, Fritz MS. Mediation analysis. *Annu Rev Psychol.* 2007;58:593–614.
92. MacKinnon DP. *Introduction to statistical mediation analysis*: Routledge; 2008.
93. VanderWeele T. *Explanation in causal inference: methods for mediation and interaction*: Oxford University Press; 2015.
94. Tamimi RM, Spiegelman D, Smith-Warner SA, Wang M, Pazaris M, Willett WC, et al. Population Attributable Risk of Modifiable and Nonmodifiable Breast Cancer Risk Factors in Postmenopausal Breast Cancer. *Am J Epidemiol.* 2016;184(12):884–93.

Figures

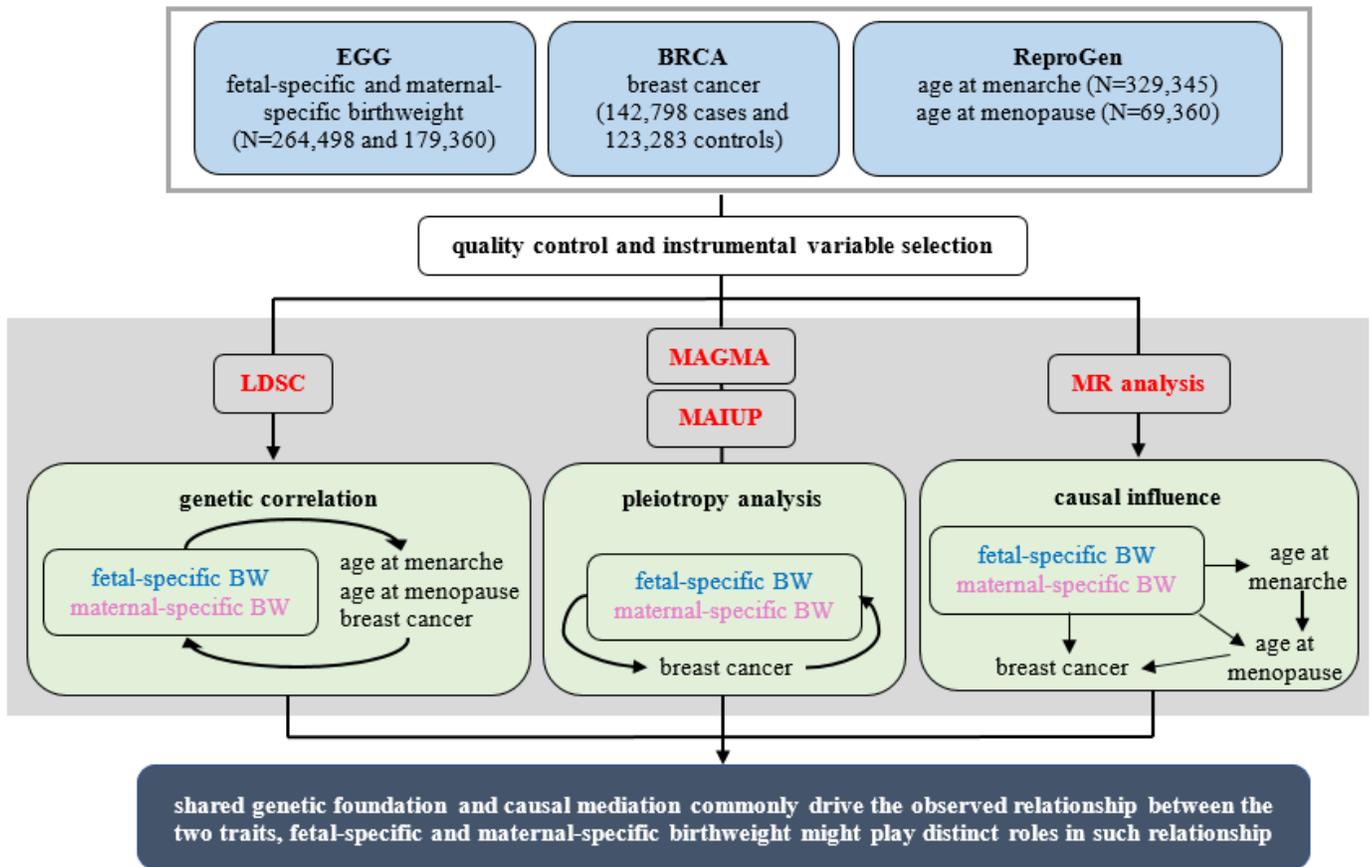


Figure 1

Flow diagram of data process and statistical analysis for the present study. quality control: exclude SNPs having no rs label and remove duplicated SNPs; LDSC: the cross-trait linkage disequilibrium score regression; MAGMA: Multi-marker Analysis of GenoMic Annotation; MAIUP: Mixture Adjusted Intersect-Union Pleiotropy test; MR: Mendelian randomization; BW: birthweight.

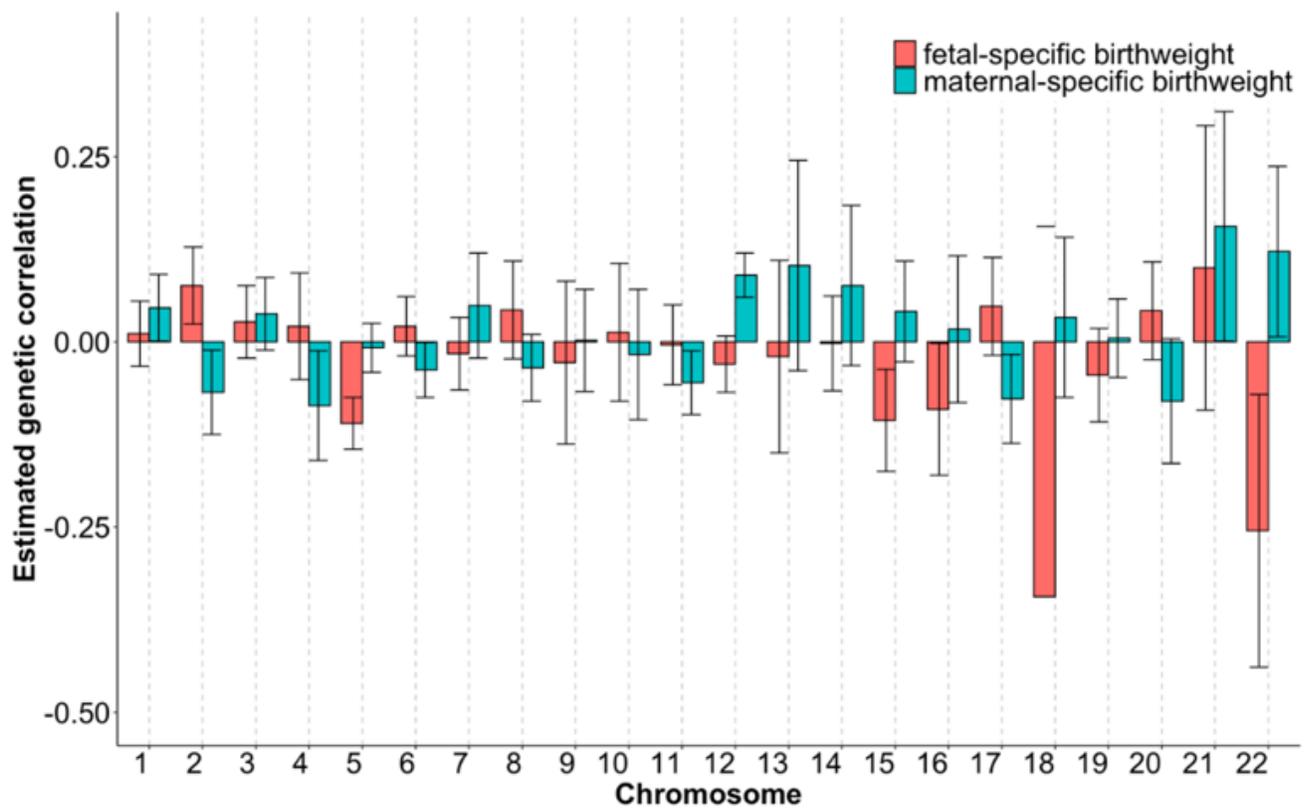


Figure 2

Estimated genetic correlation and 95% confidence intervals for each chromosome between fetal/maternal-specific birthweight and breast cancer.

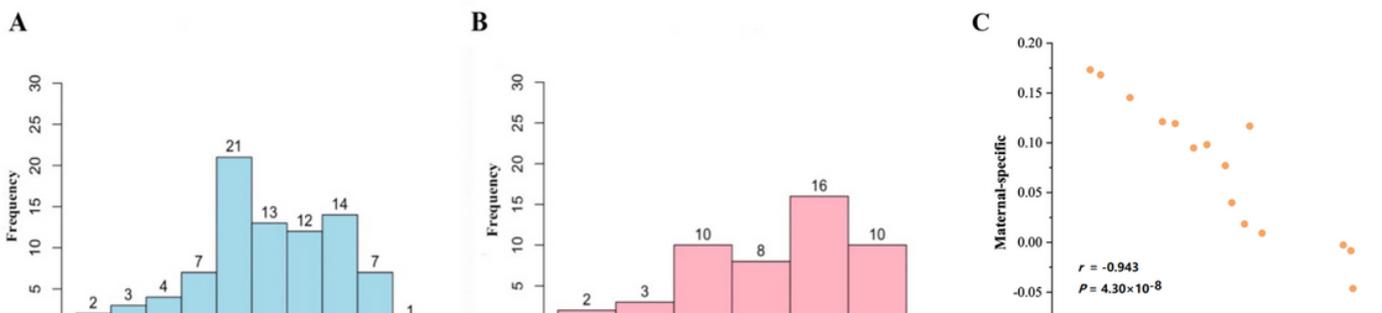


Figure 3

(A) Distribution of correlation coefficients of SNP effect sizes of pleiotropic genes between fetal-specific birthweight and breast cancer. (B) Distribution of correlation coefficients of SNP effect sizes of pleiotropic genes between maternal-specific birthweight and breast cancer. (C) Relationship between correlation coefficients of SNP effect sizes for the 16 genes shared by fetal-specific birthweight with breast cancer

and these of SNP effect sizes for the 16 genes shared by maternal-specific birthweight with breast cancer.

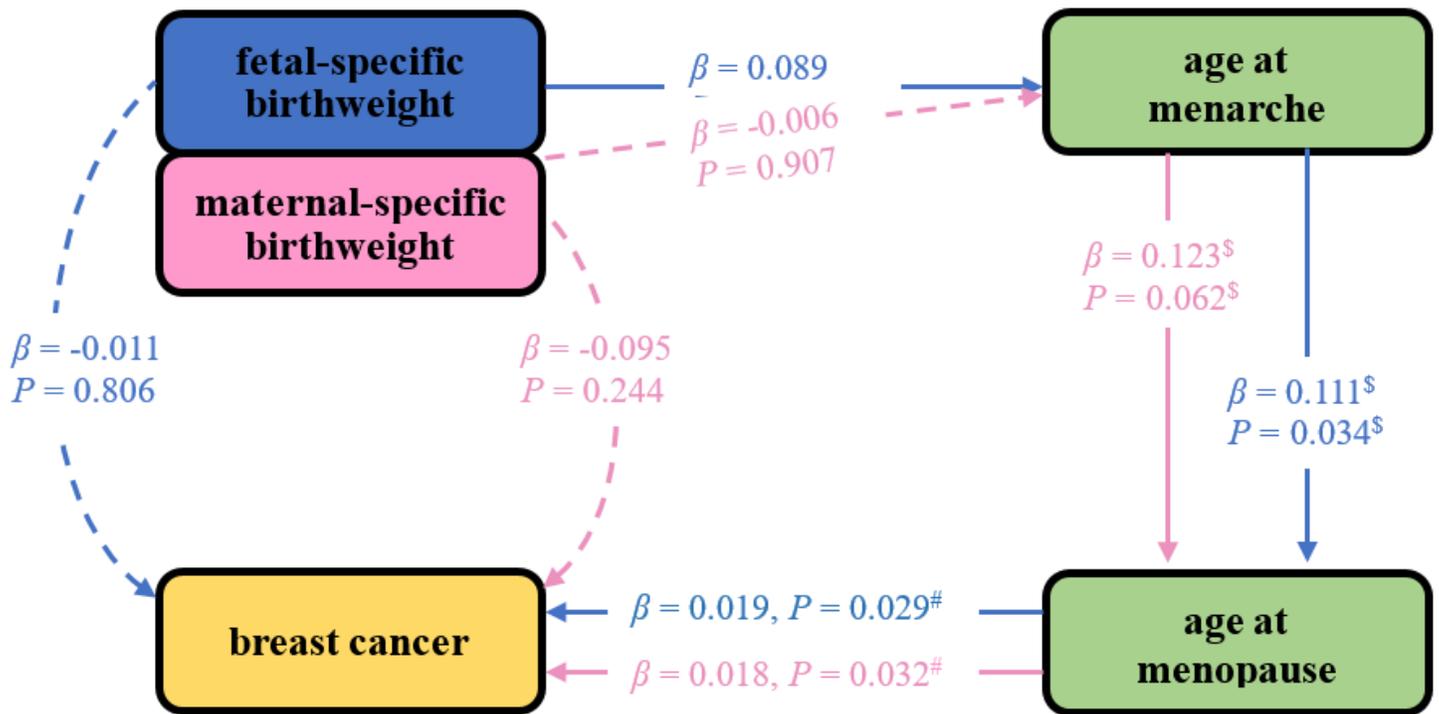


Figure 4

Estimated effect sizes and corresponding P values in the MR analysis. \$ indicates the estimated effect sizes were obtained while adjusting for fetal-specific or maternal-specific birthweight; # denotes the estimated effect sizes were obtained while controlling for fetal/maternal-specific birthweight and age at menarche. Note here that the color of arrow and number is consistent with that for fetal-specific or maternal-specific birthweight. Dot arrow stands for the absence of association, while solid arrow stands for the presence of association. To be concise and easy to understand, besides all significant associations, only a few of important null associations are shown.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementFile.docx](#)