



OPEN

Genetic correlation and Mendelian randomization analyses support causal relationships between dietary habits and age at menarche

Ruilong Guo^{1,3}, Ruoyang Feng^{2,3}, Jiong Yang¹, Yanfeng Xiao^{1✉} & Chunyan Yin^{1✉}

Dietary habits are essential in the mean age at menarche (AAM). However, the causal relationship between these factors remains unclear. Therefore, this study aimed to elucidate the genetic relationship between dietary habits and AAM. Genetic summary statistics for dietary habits were obtained from the UK Biobank. GWAS summary data for AAM was obtained from the ReproGen Consortium. Linkage disequilibrium score regression was used to test genetic correlations between dietary habits and AAM. The Mendelian randomization (MR) analyses used the inverse-variance weighted method. Genetic correlations with AAM were identified for 29 candidate dietary habits, such as milk type (skimmed, semi-skimmed, full cream; coefficient = 0.2704, $P_{ldsc} = 1.13 \times 10^{-14}$). MR evaluations revealed that 19 dietary habits were associated with AAM, including bread type (white vs. any other; OR 1.71, 95% CI 1.28–2.29, $P_{mr} = 3.20 \times 10^{-4}$), tablespoons of cooked vegetables (OR 0.437, 95% CI 0.29–0.67; $P_{mr} = 1.30 \times 10^{-4}$), and cups of coffee per day (OR 0.72, 95% CI 0.57–0.92, $P_{mr} = 8.31 \times 10^{-3}$). These results were observed to be stable under the sensitivity analysis. Our study provides potential insights into the genetic mechanisms underlying AAM and evidence that dietary habits are associated with AAM.

Keywords Dietary habits, Age at menarche, Linkage disequilibrium score regression, Mendelian randomization, GWAS

Menarche is a significant sign of pubertal onset, marking the beginning of a female's fertility and reproductive ability¹. The age at menarche (AAM) is a well-remembered and widely measured marker of female sexual development. AAM has been widely used in studies of female health², where it has been observed to correlate with a body-mass index (BMI)³, height⁴, fertility⁵, psychological health⁶, and cancer⁷. A combination of genetic and environmental factors determines AAM. Large-scale genomic analysis of AAM has identified hundreds of associated variants and determined the genetic mechanism underlying the role of AAM in breast and endometrial cancer risk⁸. However, the effects of environmental factors on AAM remain unclear.

Dietary habits are critical for human health and disease prevention⁹. Unhealthy dietary habits may lead to various diseases, including cardiovascular¹⁰, endocrine¹¹, and infertility¹². Dietary habits are an important non-genetic factor affecting AAM¹³, and different dietary habits have different effects¹⁴. Research on the relationship between dietary habits and AAM has focused on observational studies and studies of single food groups. However, a systematic exploration of the potential correlations between dietary habits and AAM remains lacking. It is important to note that most eating habits are correlated and heritable. Genome-wide association studies (GWAS) have been conducted to analyze individual macronutrients in five questionnaires on macronutrient intake¹⁵ and 24-h dietary recall¹⁶. A recent GWAS study of dietary habits based on the UK Biobank identified genetic associations between hundreds of dietary habits, providing the potential to study the causal relationship

¹Department of Pediatrics, Xi'an Jiaotong University Second Affiliated Hospital, Xi'an 710054, Shanxi, China. ²Department of Joint Surgery, Xi'an Jiaotong University Hong Hui Hospital, Xi'an 710054, Shanxi, China. ³These authors contributed equally: Ruilong Guo and Ruoyang Feng. ✉email: xiaoyanfenggroup@sina.com; yinchunyan0624@sina.com

between dietary habits and disease risk¹⁷. Dietary habits have been widely used to assess causal relationships with diseases or traits, such as migraines¹⁸, osteoporosis¹⁹, and cerebral cortex structure²⁰.

With the advent of genome-wide association studies and improvements in widely applicable tools, the use of GWAS data for correlation analysis between multiple traits is becoming more common. Linkage disequilibrium score regression (LDSC) is a widely used method to identify genetic correlations among complex traits and to distinguish inflated test statistics from confounding biases and polygenicity in GWAS data²¹. Using the aggregate data from GWAS, LDSC provides a simple and reliable method to screen thousands of traits simultaneously and determine their real genetic correlations²². However, LDSC can only analyze genetic correlations between traits. To understand the confounding factors in observational studies and determine causality, Mendelian randomization (MR) can be used with genetic variation as an instrumental variable (IV) to assess whether the observed associations between risk factors and outcomes are consistent with causal effects²³. MR has been used to identify reliable risk factors for various diseases²⁴. The combination of LDSC and MR analyses has been widely used to explore the associations between complex diseases and their risk factors²⁵.

In this study, we used LDSC to detect the genetic correlations between dietary habits and AAM. MR analysis assessed the causal relationship between the 143 dietary habits selected and AAM. Our results help to elucidate the potential genetic relationship between dietary habits and AAM.

Materials and methods

GWAS summary data

The genetic instruments for dietary habits were acquired from a public GWAS dataset¹⁷, which included 455,146 individuals from the UK Biobank. All individuals were 40–69 years old and lived in the UK between 2006 and 2010. Dietary habits were assessed using the UK Food Frequency Questionnaire (FFQ)²⁶. The information collected included the number of tablespoons of cooked vegetables eaten per day (field 1289), overall oily fish intake (field 1329), and data on foods that were never eaten (from the options dairy, eggs, sugar, and wheat) (field 6144). The GWAS data were processed as follows: Heritability measures were obtained using BOLT-Imm software (v.2.3.2)²⁷. Additional covariates in the BOLT-Imm analysis for both heritability and GWAS included the results from genotyping arrays and the first 10 genetic principal components (PCs) derived from a subset of unrelated European individuals using FlashPCA252, followed by projection of related individuals onto the PC space. Principal component analysis (PCA) generated 85 PC-DPs that captured the correlated structures between single food intake information and represented independent components of real-world eating habits¹⁷. A total of 814 independent loci (defined as > 500 kb apart) were identified, which exceeded the genome-wide significance ($p < 5.0 \times 10^{-8}$). A full description of the study design, sample characteristics, statistical analysis, and quality control can be obtained from the study results¹⁷.

GWAS summary data relating to AAM was obtained from the ReproGen Consortium to avoid sample overlap. This dataset²⁸ included 182,416 females of European descent from across 58 studies. Individuals who reported their age at menarche as < 9 or > 17 years were excluded from the analysis. Single nucleotide polymorphisms (SNPs) were excluded from the individual study datasets if they were poorly imputed or rare (MAF < 1%). We obtained 3915 SNPs associated with AAM ($p < 5 \times 10^{-8}$). A full description of the study design, sample characteristics, statistical analysis, and quality control can be obtained from the study results²⁸.

Genetic correlation analysis

The LD score regression (LDSC, v1.0.1, <https://github.com/bulik/ldsc>) software²¹ evaluated the genetic correlations between 143 dietary habits and AAM. LDSC is a useful approach for estimating the components of heritability and the genetic correlation and has been widely used to analyze complex diseases²¹. After strict Bonferroni correction, $p < 0.000350$ (0.05/143) was considered a significant association. The basic principle of the LDSC method is to use the test statistic of the expected value of the observed χ^2 SNP under the original hypothesis of no association. SNPs that mark more neighbors—and thus have higher LD scores—are more likely to mark one or more causal loci that influence the phenotype⁵. We restricted our analysis to Hapmap3 SNPs using pre-calculated European LD scores from the 1000 Genomes Project Phase 3 provided by LDSC.

Selection of instrumental variables (IVs)

The IVs used in MR analysis should meet three conditions: (1) they are correlated with exposure, (2) they are not associated with confounding factors, and (3) they are not related to outcome directly but are related through exposure²⁹. The IVs used in this study met the above conditions and are listed in Supplementary Table S2. All variables met the genome-wide significance threshold of $p < 5 \times 10^{-8}$. The parameters $kb = 10,000$ and $r^2 = 0.01$ were used to remove the linkage disequilibrium between each variable. F-statistics were computed to estimate whether a weak instrument bias was observed and to improve the power of the selected instrumental variables. The F-statistics for all IVs were above the threshold of 10^{30} .

Two-sample MR analysis

A two-sample MR analysis was used to evaluate the causal relationship between dietary habits and AAM. The SNPs used as IVs were within a distance of 10,000 kb and $r^2 > 0.001$. A two-sample MR package (version 0.5.6) was used to analyze MR³¹. Five models were used in the MR analysis: (1) the inverse-variance-weighted (IVW) model, (2) the weighted median estimator, (3) the MR-Egger regression method, (4) the simple mode, and (5) the weighted mode. The IVW model was used as the primary method to evaluate the causal effect of dietary habits on AAM. The significance level was taken as $p < 0.05$. Significantly associated dietary habit SNPs were further assessed using statistical analyses, including Cochran's Q test, a pleiotropy test, and a leave-one-out sensitivity test. However, if the pleiotropy test suggested the presence of pleiotropy ($p < 0.05$), MR Pleiotropy RESidual

Sum and Outlier (MR-PRESSO) was used to filter potential outliers and assist in correcting them³². Finally, the leave-one-out sensitivity analysis was performed to evaluate whether a single SNP provided significant results.

Institutional review board statement

Ethical approval was not applicable to our study as publicly available data were used for all analyses.

Informed consent statement

Informed consent statement was not applicable to our study as publicly available data were used for all analyses. Informed consent was obtained from all subjects involved in the original study.

Results

Genetic correlations between 143 dietary habits and AAM

The LDSC analysis identified 29 candidate dietary habits significantly associated with AAM ($P_{ldsc} < 3.50 \times 10^{-4}$, Fig. 1). These included milk types (skimmed, semi-skimmed, and full cream; coefficient = 0.2704; $P_{ldsc} = 1.13 \times 10^{-14}$) and PC1 (coefficient = -0.1699 ; $P_{ldsc} = 1.00 \times 10^{-10}$). PC1 is primarily defined by the type of bread consumed (whole grain/whole meal vs. white bread, two correlated FI-QTs contributing 15.4–15.8%). A further 41 dietary habits showed a suggestive association with AAM ($0.05 < P_{ldsc} < 3.50 \times 10^{-4}$), such as the frequency of adding salt to food (coefficient = 0.0862, $P_{ldsc} = 6.00 \times 10^{-4}$), PC24 (coefficient = 0.1093, $P_{ldsc} = 5.00 \times 10^{-4}$), and overall poultry intake (coefficient = -0.0945 , $P_{ldsc} = 3.80 \times 10^{-3}$). All genetic correlations between the 143 dietary habits and AAM are summarized in Supplementary Table S1.

Causal relationships between 143 dietary habits and AAM

The MR study identified causal relationships between 19 dietary habits and AAM. In addition, 15 dietary habits showed genetic correlations with AAM and therefore were deemed to have causal relationships with AAM. These included bread type (white vs. any other; OR 1.71, 95% CI 1.28–2.29; $P_{mr} = 3.20 \times 10^{-4}$), tablespoons of cooked vegetables per day (OR 0.437, 95% CI 0.29–0.67, $P_{mr} = 1.30 \times 10^{-4}$), milk type (skimmed, semi-skimmed, and full cream; OR 3.37, 95% CI 1.76–6.44, $P_{mr} = 2.50 \times 10^{-4}$), cups of coffee per day (OR 0.72, 95% CI 0.57–0.92, $P_{mr} = 8.31 \times 10^{-3}$), and PC3 (OR 1.14, 95% CI 1.03–1.27, $P_{mr} = 0.01$) (Fig. 2, Supplementary Tables S2–S3). PC3 is primarily defined by the spread type (butter vs. any other, one correlated FI-QT contributing over 10%)¹⁷.

Sensitivity analysis

All dietary habits identified as significantly associated with AAM were further analyzed using Cochran's Q test (Supplementary Table S4), the pleiotropy test (Supplementary Table S6), and the leave-one-out sensitivity analysis. The results of these tests were used to inform the MR method. In the absence of heterogeneity and pleiotropy, estimated IVW results were preferentially used, thus this method was used most frequently in this

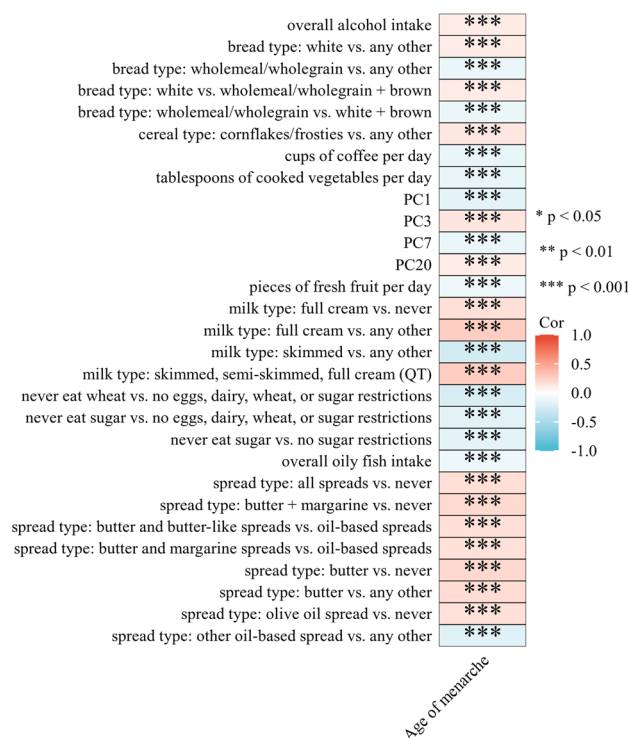


Figure 1. Heatmap of significant correlation ($P_{ldsc} < 3.50 \times 10^{-4}$) of LDSC analysis.

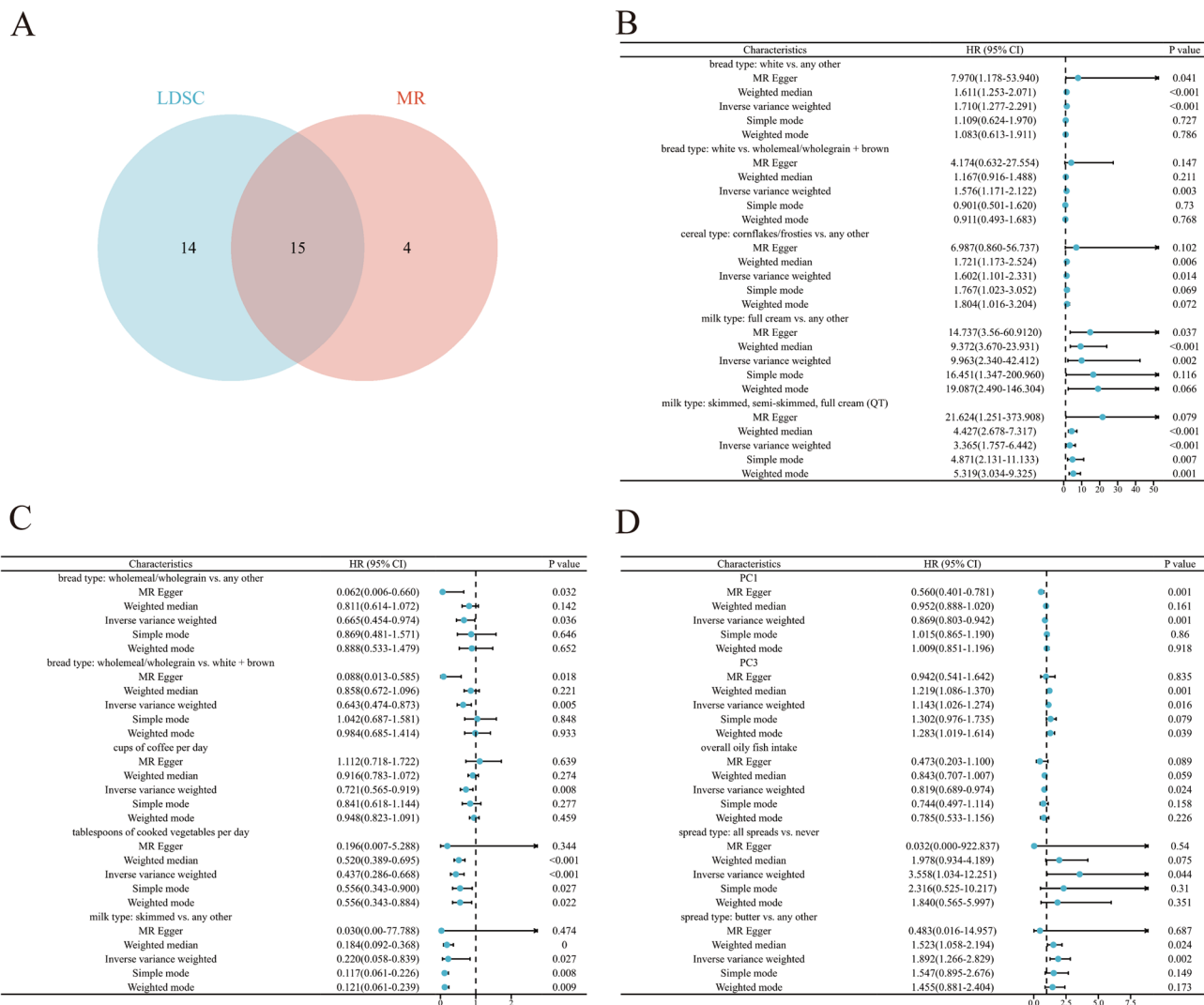


Figure 2. Results of MR analysis. (A) Venn plot of LDSC and MR. (B–D) Forest plots of MR results. MR Mendelian randomization, LDSC Linkage disequilibrium score regression, CI confidence interval.

study. When there was only heterogeneity, but no pleiotropy, a weighted median or random-effect IVW was used (Supplementary Table S5). Although a few results were heterogeneous, the direction of the effect obtained from these other methods was concordant with the IVW results. However, the MR-Egger method was used if the pleiotropy test suggested that the result was multi-efficacious. The leave-one-out sensitivity test results suggested the result should be considered reliable (Fig. 3, Supplementary Information 2).

Discussion

In this study, we performed LDSC and MR analyses to investigate the relationship between dietary habits and AAM. Using LDSC analysis, we identified 29 candidate dietary habits significantly associated with AAM. We further evaluated the causal relationship between dietary habits and AAM using MR analysis. After sensitivity analyses, we identified eight dietary habits that showed significant causal relationships with AAM: bread type (whole meal, whole grain, white, brown), cereal type (cornflakes, frosties), milk type (skimmed, semi-skimmed, full cream), spread type (butter, margarine, olive oil), overall oily fish intake, tablespoons of vegetables per day, and PC1 and PC3, which primarily represent bread type and spread type, respectively.

Diet plays a key role in human health, and a recent study showed that diet-associated risk is among the top five risks of attributable deaths worldwide³³. Dietary habits affect physical health in childhood³⁴ and adulthood, especially concerning chronic diseases³⁵. A pooled analysis of 2181 population-based studies³⁶ showed that lifelong health advantages and risks are affected by heterogeneous nutritional quality. Diet influences the age of puberty onset and further affects height gain during adolescence and late adolescence. Another prospective study of 3983 children showed that puberty started later in children with a high diet quality³⁷. A prospective cohort study of 215 girls showed that girls with a low intake of white meat (poultry and fish), fruits, and vegetables had an earlier AAM³⁸. In addition to dietary patterns, single dietary components can also affect AAM. A cohort study among Chinese children suggested that the higher the childhood soy intake, the later the puberty³⁹. According

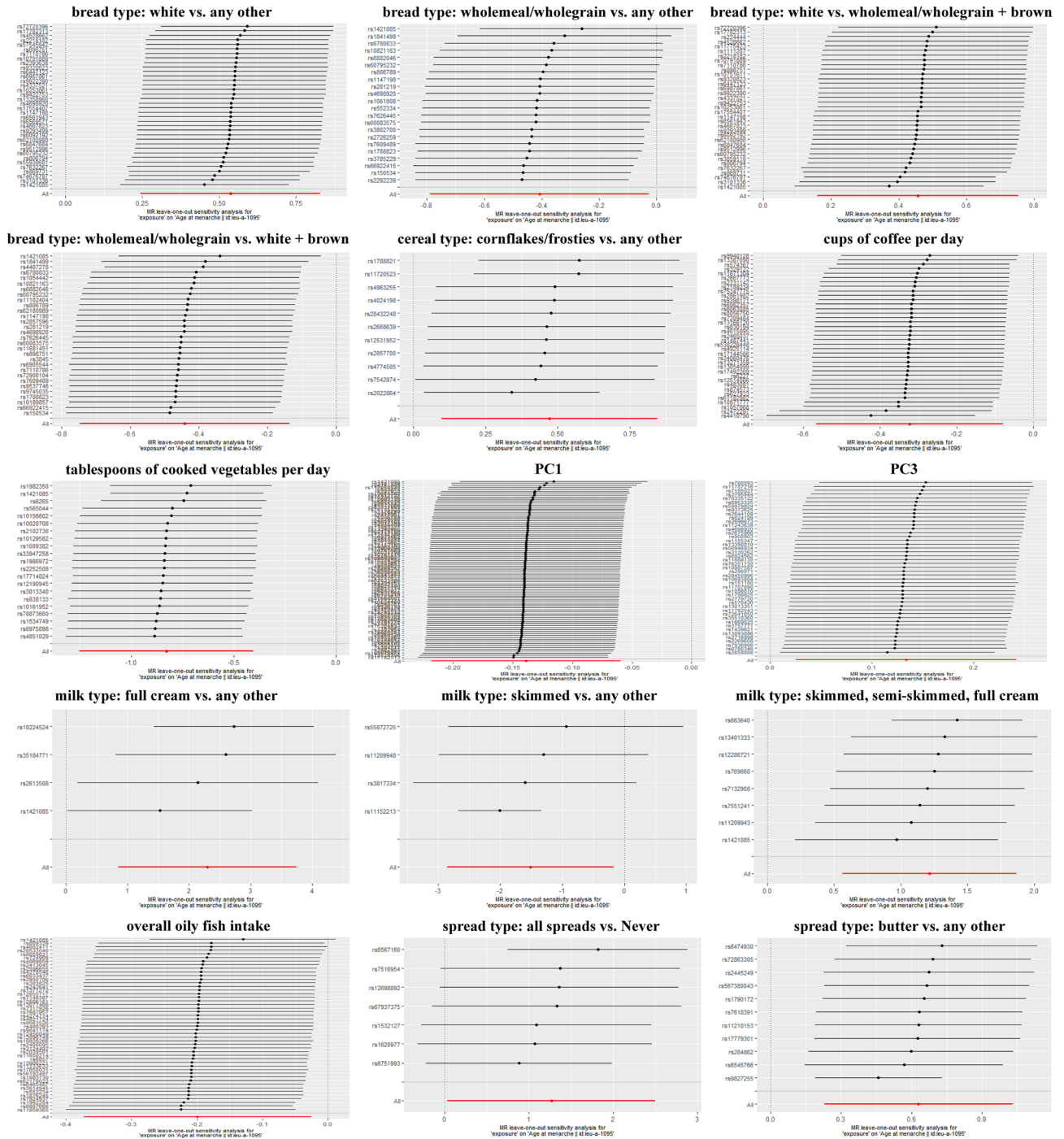


Figure 3. Leave-one-out sensitivity test.

to Cheng TS. et al., a higher dietary fat intake has been associated with earlier puberty⁴⁰. Although many cohort studies have observed the effects of different diets on AAM, the underlying mechanisms have not yet been elucidated. The effect of diet on AAM is influenced by many confounding factors, which are often difficult to control⁴¹. Therefore, the current research on the relationship between diet and AAM is mainly observational, and the results are often controversial. MR Analysis has a unique advantage in excluding confounding factors⁴². Although Cheng TS. et al. showed a causal effect of higher dihomo- γ -linolenic acid concentrations on earlier AAM using MR analysis⁴³. There have been no studies on the association between dietary habits in general and AAM. We used the most comprehensive dietary habit GWAS summary dataset to evaluate the relationship between dietary habits and AAM from a genetic perspective.

The effect of estrogenic endocrine disruptors (EEDs) on puberty has long been understood⁴⁴. Numerous observational studies have shown that girls chronically exposed to EEDs are more likely to experience early puberty^{45–47}, regardless of the incidence of obesity⁴⁸. The similar structure of EEDs to estrogen allows them to

bind and activate estrogen receptors, thus exerting effects similar to estrogen⁴⁹. Kisspeptin regulates puberty and fertility in humans⁵⁰ by stimulating GnRH neuronal activity and regulating ovulation by driving luteinizing hormone surges⁵¹. Exposure to EEDs affects the expression of kisspeptin and GnRH and influences the pulsatile release of GnRH, which subsequently alters gonadotropin levels⁵². Animal studies have shown that EEDs alter hypothalamic Kiss1 mRNA expression levels and kisspeptin fiber density while altering gonadotropin secretion and/or gonadotropin-releasing hormone neuronal activation⁵³. Coffee is one of the most widely consumed beverages worldwide⁵⁴ and contains various ingredients, including caffeine, carbohydrates, lipids, and proteins, and has a few estrogenic activities⁵⁵. In addition to caffeine, other components of coffee, such as aromatic acids, esters, and sterols, have estrogenic activity⁵⁶. Our results suggest that increased daily coffee intake leads to earlier AAM, which may be attributed to the estrogenic activity of coffee. Isoflavones as EEDs have the potential to modulate estrogen metabolism⁵⁷. Isoflavones are structurally similar to estrogens and can compete with endogenous estrogen for the estrogen receptor⁵⁸. Whole wheat bread contains high concentrations of isoflavones (average 450 mcg/100 g)⁵⁹ and is the primary source of isoflavones in the European diet⁶⁰. Thus, our results suggest that the intake of whole grain bread may result in earlier AAM compared to that associated with white bread. Similarly, increased daily vegetable intake was associated with earlier AAM. The consumption of phytoestrogen-containing vegetables, such as soy and soy products, may promote puberty onset by performing estrogen-like biological functions^{61,62}. The intake of vegetables, especially raw vegetables, may increase the risk of pesticide exposure^{63,64}, and the effects on puberty timing as a type of EED have been extensively studied^{65,66}.

Insulin-like growth factors (IGFs), including IGF-1 and IGF-2, are important peptides that regulate essential cellular activities⁶⁷. IGF-1 levels are associated with age and play a key role in growth⁶⁸. A few longitudinal studies have shown that higher IGF-1 levels are related to earlier puberty onset⁶⁹ and higher breast cancer risk⁷⁰. Oily fish were positively associated with circulating IGF-I concentrations⁷¹. Our results suggest that an increased intake of oily fish may lead to earlier AAM. In addition, several prospective cohort studies have shown that milk intake before puberty may accelerate AAM^{72,73}. Other studies have shown no effect of milk intake on AAM^{74,75}. Although milk contains IGF-1, studies have shown that humans do not absorb biologically significant levels of intact IGF-I from their food⁷⁶. Different types of milk may contain different levels of sugar, fats, proteins, and other trace elements. A study based on the Growth and Obesity Cohort showed that consumption of sugar-sweetened milk beverages led to earlier mammary gland development compared to that in low-fat dairy products, namely low-fat milk and yogurt⁷⁷. Another study showed that milk and butter consumption at the ages 3–5 was inversely related to breast development at age 10.8⁷⁸. Our results suggest that butter intake was negatively associated with AAM, while the intake of full-cream milk did not lead to earlier AAM compared to skimmed or semi-skimmed milk. The controversy over these results is mainly due to the complex composition of milk beverages⁷⁹. Therefore, stricter control of milk composition in the study cohort would be needed to produce more rigorous results. Cultural factors may play a confounding role; for example, girls with earlier AAM may avoid milk if they believe it is associated with acne, while girls with later AAM may consume more milk and grow further in height. These factors may lead to uncertainty in the causal associations determined by such studies.

Despite the lack of longitudinal studies, this work is suggestive of new directions for understanding the genetic mechanisms between diet habits and AAM, since we are the first to combine LDSC and MR analysis to investigate the causal association between dietary habits and AAM. However, our study has some limitations. Although eating habits may change in adulthood compared to childhood, eating habits formed during childhood have been shown to have a lasting effect on adult eating habits⁸⁰. We used data of eating habits in adulthood to reflect the exposure to certain products in childhood, due to the lack of large-scale heritability studies of childhood eating habits. First, dietary recall errors and individual differences in dose estimates are inevitable in an FFQ. Such reporting errors may lead to underestimating the true relationship between dietary habits and AAM. Second, people with AAM < 9 years or > 17 years were excluded. Therefore, the conclusions of this study are limited to people with AAM within the typical range and do not apply to people with abnormal pubertal initiation (precocious puberty or delayed pubertal development). Further, only European populations were included in this study, and as such, the conclusions should be interpreted with caution concerning other ethnic groups. Finally, the results of LDSC and MR only suggest possible genetic correlations and causal associations from a genetic perspective. Further experimentation is needed to confirm any underlying biological mechanisms indicated by these genetic associations. It is worth noting that the interpretation of genetic factors in dietary traits can be complex. Especially when AAM is also influenced by environmental factors. Therefore, more GWAS data on children and lifestyle questionnaires are expected to be generated. This will facilitate the study of genetic associations of lifestyle in children.

Conclusions

Based on GWAS summary data for dietary habits and AAM, we identified 29 candidate dietary habits that showed genetic associations with AAM. MR evaluations revealed that 19 dietary habits were causally associated with AAM, including bread type (whole meal, whole grain, white, brown), cereal type (cornflakes, frosties), milk type (skimmed, semi-skimmed, full cream), spread type (butter, margarine, olive oil), overall oily fish intake, and tablespoons of vegetables per day. These results provide a potential novel understanding of the genetic mechanisms underlying the relationship between diet habits and AAM.

Data availability

The datasets analyzed during the current study are available from the UK biobank (<http://geneatlas.roslin.ed.ac.uk/>) (fields: 20002), ReproGen Consortium (https://www.reprogen.org/data_download.html).

Received: 12 June 2023; Accepted: 5 April 2024

Published online: 10 April 2024

References

- Chen, L. *et al.* Age at menarche and risk of gestational diabetes mellitus: A prospective cohort study among 27,482 women. *Diabetes Care* **39**(3), 469–471 (2016).
- Hollis, B. *et al.* Genomic analysis of male puberty timing highlights shared genetic basis with hair colour and lifespan. *Nat. Commun.* **11**(1), 1536 (2020).
- Parikh, N. I. *et al.* Reproductive risk factors and coronary heart disease in the women's health initiative observational study. *Circulation* **133**(22), 2149–2158 (2016).
- Pickrell, J. K. *et al.* Detection and interpretation of shared genetic influences on 42 human traits. *Nat. Genet.* **48**(7), 709–717 (2016).
- Agarwala, A. *et al.* The use of sex-specific factors in the assessment of women's cardiovascular risk. *Circulation* **141**(7), 592–599 (2020).
- Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**(3), 343–352 (2019).
- Pashayan, N. *et al.* Personalized early detection and prevention of breast cancer: ENVISION consensus statement. *Nat. Rev. Clin. Oncol.* **17**(11), 687–705 (2020).
- Day, F. R. *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.* **49**(6), 834–841 (2017).
- Przybyłowicz, K. E. & Danielewicz, A. Eating habits and disease risk factors. *Nutrients* **14**(15), 3143 (2022).
- Mehta, L. S. *et al.* Cardiovascular disease and breast cancer: Where these entities intersect: A scientific statement from the American Heart Association. *Circulation* **137**(8), e30–e66 (2018).
- Nip, A. S. Y. *et al.* Disordered eating behaviors in youth and young adults with type 1 or type 2 diabetes receiving insulin therapy: The SEARCH for diabetes in youth study. *Diabetes Care* **42**(5), 859–866 (2019).
- Fontana, R. & Della Torre, S. The deep correlation between energy metabolism and reproduction: A view on the effects of nutrition for women fertility. *Nutrients* **8**(2), 87 (2016).
- Yermachenko, A. & Dvornyk, V. Nongenetic determinants of age at menarche: A systematic review. *Biomed. Res. Int.* **2014**, 371583 (2014).
- Szamreta, E. A. *et al.* Greater adherence to a Mediterranean-like diet is associated with later breast development and menarche in peripubertal girls. *Public Health Nutr.* **23**(6), 1020–1030 (2020).
- Meddens, S. F. W. *et al.* Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Mol. Psychiatry* **26**(6), 2056–2069 (2021).
- Merino, J. *et al.* Genome-wide meta-analysis of macronutrient intake of 91,114 European ancestry participants from the cohorts for heart and aging research in genomic epidemiology consortium. *Mol. Psychiatry* **24**(12), 1920–1932 (2019).
- Cole, J. B., Florez, J. C. & Hirschhorn, J. N. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nat. Commun.* **11**(1), 1467 (2020).
- Liu, X. *et al.* Association between dietary habits and the risk of migraine: A Mendelian randomization study. *Front. Nutr.* **10**, 1123657 (2023).
- Xu, J. *et al.* Assessing the association between important dietary habits and osteoporosis: A genetic correlation and two-sample Mendelian randomization study. *Nutrients* **14**(13), 2656 (2022).
- Li, C. *et al.* Unveiling the influence of daily dietary patterns on brain cortical structure: Insights from bidirectional Mendelian randomization. *Food Funct.* **14**(23), 10418–10429 (2023).
- Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**(3), 291–295 (2015).
- Lee, J. J. *et al.* The accuracy of LD Score regression as an estimator of confounding and genetic correlations in genome-wide association studies. *Genet. Epidemiol.* **42**(8), 783–795 (2018).
- Emdin, C. A., Khara, A. V. & Kathiresan, S. Mendelian randomization. *JAMA* **318**(19), 1925–1926 (2017).
- Huang, H. *et al.* Dissecting the association between psychiatric disorders and neurological proteins: A genetic correlation and two-sample bidirectional Mendelian randomization study. *Acta Neuropsychiatr.* **34**(6), 311–317 (2022).
- Yuan, G. *et al.* A large-scale genetic correlation scan between rheumatoid arthritis and human blood metabolites. *Ann. Hum. Genet.* **86**(3), 127–136 (2022).
- Block, G., Hartman, A. M. & Naughton, D. A reduced dietary questionnaire: Development and validation. *Epidemiology* **1**(1), 58–64 (1990).
- Loh, P. R. *et al.* Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**(3), 284–290 (2015).
- Perry, J. R. *et al.* Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**(7520), 92–97 (2014).
- Zou, X. L. *et al.* Childhood obesity and risk of stroke: A Mendelian randomisation analysis. *Front. Genet.* **12**, 727475 (2021).
- Pierce, B. L., Ahsan, H. & Vanderweele, T. J. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int. J. Epidemiol.* **40**(3), 740–752 (2011).
- Yavorska, O. O. & Burgess, S. MendelianRandomization: An R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* **46**(6), 1734–1739 (2017).
- Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**(2), 512–525 (2015).
- Global burden of 87 risk factors in 204 countries and territories, 1990–2019: A systematic analysis for the Global Burden of Disease Study 2019. *Lancet.* **396**(10258), 1223–1249 (2020).
- Stephenson, J. *et al.* Before the beginning: nutrition and lifestyle in the preconception period and its importance for future health. *Lancet* **391**(10132), 1830–1841 (2018).
- Health effects of dietary risks in 195 countries, 1990–2017: A systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* **393**(10184), 1958–1972 (2019).
- Height and body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: A pooled analysis of 2181 population-based studies with 65 million participants. *Lancet.* **396**(10261), 1511–1524 (2020).
- Duan, R. *et al.* The overall diet quality in childhood is prospectively associated with the timing of puberty. *Eur. J. Nutr.* **60**(5), 2423–2434 (2021).
- Martinez-Arroyo, A. *et al.* Lower adherence to a prudent dietary pattern is associated with earlier age at menarche in adolescents from the Growth and Obesity Chilean Cohort Study. *Front. Public Health* **10**, 995593 (2022).
- Xiong, J. *et al.* Prospective association of dietary soy and fibre intake with puberty timing: A cohort study among Chinese children. *BMC Med.* **20**(1), 145 (2022).
- Xu, Y. *et al.* Dietary fat and polyunsaturated fatty acid intakes during childhood are prospectively associated with puberty timing independent of dietary protein. *Nutrients* **14**(2), 275 (2022).

41. Qian, F. *et al.* Association between plant-based dietary patterns and risk of type 2 diabetes: A systematic review and meta-analysis. *JAMA Intern. Med.* **179**(10), 1335–1344 (2019).
42. Zhu, Z. *et al.* Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**(1), 224 (2018).
43. Cheng, T. S. *et al.* Prepubertal dietary and plasma phospholipid fatty acids related to puberty timing: Longitudinal cohort and mendelian randomization analyses. *Nutrients* **13**(6), 1868 (2021).
44. Massart, F. *et al.* How do environmental estrogen disruptors induce precocious puberty?. *Minerva Pediatr* **58**(3), 247–254 (2006).
45. Oskar, S. *et al.* Identifying environmental exposure profiles associated with timing of menarche: A two-step machine learning approach to examine multiple environmental exposures. *Environ. Res.* **195**, 110524 (2021).
46. Jensen, T. K. *et al.* Prenatal paraben exposure and anogenital distance and reproductive hormones during mini-puberty: A study from the Odense Child Cohort. *Sci. Total Environ.* **769**, 145119 (2021).
47. Gea, M. *et al.* Oestrogenic activity in girls with signs of precocious puberty as exposure biomarker to endocrine disrupting chemicals: A pilot study. *Int. J. Environ. Res. Public Health* **20**(1), 14 (2022).
48. Srilanchakon, K. *et al.* Higher phthalate concentrations are associated with precocious puberty in normal weight Thai girls. *J. Pediatr. Endocrinol. Metab.* **30**(12), 1293–1298 (2017).
49. Roy, J. R., Chakraborty, S. & Chakraborty, T. R. Estrogen-like endocrine disrupting chemicals affecting puberty in humans—A review. *Med. Sci. Monit.* **15**(6), 137–145 (2009).
50. Herber, C. B. *et al.* Estrogen signaling in arcuate Kiss1 neurons suppresses a sex-dependent female circuit promoting dense strong bones. *Nat. Commun.* **10**(1), 163 (2019).
51. Hellier, V. *et al.* Female sexual behavior in mice is controlled by kisspeptin neurons. *Nat. Commun.* **9**(1), 400 (2018).
52. Roepke, T. A. & Sadlier, N. C. REPRODUCTIVE TOXICOLOGY: Impact of endocrine disruptors on neurons expressing GnRH or kisspeptin and pituitary gonadotropins. *Reproduction* **162**(5), F131–f145 (2021).
53. Tena-Sempere, M. Kisspeptin/GPR54 system as potential target for endocrine disruption of reproductive development and function. *Int. J. Androl.* **33**(2), 360–368 (2010).
54. Park, S. Y. *et al.* Association of coffee consumption with total and cause-specific mortality among nonwhite populations. *Ann. Intern. Med.* **167**(4), 228–235 (2017).
55. Kitts, D. D. Studies on the estrogenic activity of a coffee extract. *J. Toxicol. Environ. Health* **20**(1–2), 37–49 (1987).
56. Kiyama, R. Estrogenic activity of coffee constituents. *Nutrients* **11**(6), 1401 (2019).
57. Dietz, B. M. *et al.* Botanicals and their bioactive phytochemicals for women's health. *Pharmacol. Rev.* **68**(4), 1026–1073 (2016).
58. Baglia, M. L. *et al.* The association of soy food consumption with the risk of subtype of breast cancers defined by hormone receptor and HER2 status. *Int. J. Cancer* **139**(4), 742–748 (2016).
59. Ward, H. A. & Kuhnle, G. G. Phytoestrogen consumption and association with breast, prostate and colorectal cancer in EPIC Norfolk. *Arch. Biochem. Biophys.* **501**(1), 170–175 (2010).
60. Adlercreutz, H. Epidemiology of phytoestrogens. *Baillieres Clin Endocrinol Metab* **12**(4), 605–623 (1998).
61. Patisaul, H. B. Effects of environmental endocrine disruptors and phytoestrogens on the kisspeptin system. *Adv. Exp. Med. Biol.* **784**, 455–479 (2013).
62. Adgent, M. A. *et al.* Early-life soy exposure and age at menarche. *Paediatr. Perinat. Epidemiol.* **26**(2), 163–175 (2012).
63. Jiang, M. *et al.* Detection of pesticide residues in vegetables sold in Changchun City, China. *J. Food Prot.* **84**(3), 481–489 (2021).
64. Tesi, G. O. *et al.* Human exposure to organochlorine pesticides in vegetables from major cities in south-south Nigeria. *Chemosphere* **303**(Pt 3), 135296 (2022).
65. Castiello, F. & Freire, C. Exposure to non-persistent pesticides and puberty timing: A systematic review of the epidemiological evidence. *Eur. J. Endocrinol.* **184**(6), 733–749 (2021).
66. Sakali, A. K. *et al.* Effects on puberty of nutrition-mediated endocrine disruptors employed in agriculture. *Nutrients* **13**(11), 4184 (2021).
67. Jones, J. I. & Clemmons, D. R. Insulin-like growth factors and their binding proteins: biological actions. *Endocr. Rev.* **16**(1), 3–34 (1995).
68. Lee, S. W. *et al.* Plasma levels of insulin-like growth factor-1 and insulin-like growth factor binding protein-3 in women with cervical neoplasia. *J. Gynecol. Oncol.* **21**(3), 174–180 (2010).
69. Baier, I. *et al.* Higher prepubertal IGF-1 concentrations associate to earlier pubertal tempo in both sexes. *Horm. Res. Paediatr.* **96**, 404–411 (2022).
70. Biro, F. M. *et al.* Pubertal growth, IGF-1, and windows of susceptibility: Puberty and future breast cancer risk. *J. Adolesc. Health* **68**(3), 517–522 (2021).
71. Watling, C. Z. *et al.* Associations between food group intakes and circulating insulin-like growth factor-I in the UK Biobank: A cross-sectional analysis. *Eur. J. Nutr.* **62**(1), 115–124 (2023).
72. Ramezani Tehrani, F. *et al.* Intake of dairy products, calcium, magnesium, and phosphorus in childhood and age at menarche in the Tehran Lipid and Glucose Study. *PLoS One* **8**(2), e57696 (2013).
73. Wiley, A. S. Milk intake and total dairy consumption: Associations with early menarche in NHANES 1999–2004. *PLoS One* **6**(2), e14685 (2011).
74. Du, X. *et al.* School-milk intervention trial enhances growth and bone mineral accretion in Chinese girls aged 10–12 years in Beijing. *Br. J. Nutr.* **92**(1), 159–168 (2004).
75. Carwile, J. L. *et al.* Milk consumption after age 9 years does not predict age at menarche. *J. Nutr.* **145**(8), 1900–1908 (2015).
76. Juskevich, J. C. & Guyer, C. G. Bovine growth hormone: Human food safety evaluation. *Science* **249**(4971), 875–884 (1990).
77. Gaskins, A. J. *et al.* Dairy intake in relation to breast and pubertal development in Chilean girls. *Am. J. Clin. Nutr.* **105**(5), 1166–1175 (2017).
78. Schraw, J. M. *et al.* In utero preeclampsia exposure, milk intake and pubertal development. *Reprod. Toxicol.* **54**, 19–25 (2015).
79. Oliveira, M. C. *et al.* Bovine milk extracellular vesicles are osteoprotective by increasing osteocyte numbers and targeting RANKL/OPG system in experimental models of bone loss. *Front. Bioeng. Biotechnol.* **8**, 891 (2020).
80. Anders, S. & Schroeter, C. Diabetes, diet-health behavior, and obesity. *Front. Endocrinol. (Lausanne)* **6**, 33 (2015).

Acknowledgements

We are indebted to all the individuals who participated in and helped with our research.

Author contributions

Author R.G. and R.F. collected and processed the data and wrote the article. J.Y. provided language help and writing assistance. C.Y. and Y.X. designed the study.

Funding

This work was supported by the National Natural Science Foundation of China (General Program: 82373594).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-024-58999-4>.

Correspondence and requests for materials should be addressed to Y.X. or C.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024