Journal of Health, Population and Nutrition

RESEARCH

Open Access

Exploring the genetic intersection of dried fruit intake and breast cancer risk: a multi-trait genomic analysis with epidemiological context

Keyan Zou[1,2], Yi Zheng[3,4], Xiaohong Ren[1] and Wei Cui[1*]

Abstract

Background Epidemiological studies have suggested that dried fruit intake may be associated with a lower risk of breast cancer (BC), yet the genetic mechanisms underlying this association remain unclear. This study aimed to explore the potential genetic relationship between dried fruit intake and BC susceptibility.

Methods We conducted a comprehensive genetic analysis using genome-wide association study (GWAS) data for dried fruit intake and BC, focusing on individuals of European ancestry. We assessed genome-wide and region-specific genetic correlations using several complementary methods, including linkage disequilibrium score regression and regional genetic mapping. To identify shared genetic regions, we applied statistical approaches that integrate information across traits and improve the detection of common genetic signals.

Results Modest but statistically significant negative genetic correlations were observed between dried fruit intake and BC, including its subtypes. Regional analysis revealed shared signals across multiple chromosomal regions. Five candidate loci were consistently identified as shared between the traits, including BCL11A, MAD1L1, MLLT10, JMJD1C, and RP11-795H16.2.

Conclusion This study provides evidence for a genome-wide genetic link between dried fruit intake and BC risk, identifying several loci that may be shared between the traits. These findings may help improve our understanding of BC development and offer preliminary leads for future dietary prevention and personalized interventions, pending further experimental validation.

Keywords Dried fruit intake, Breast cancer, Comorbidity gene, Genetic overlap, Genetic structure

*Correspondence:
Wei Cui
zwbtymz@163.com
[1]Preventive Medicine Department, Jiangmen Wuyi Hospital of TCM Affiliated Jiangmen TCM Hospital of Ji'nan University, Jiangmen, China
[2]The Second Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China
[3]Joint and Pediatric Orthopedics, Jiangmen Wuyi Hospital of TCM Affiliated Jiangmen TCM Hospital of Ji'nan University, Jiangmen, China
[4]The First Clinical Medical College, Guangzhou University of Chinese Medicine, Guangzhou, China

Introduction

According to the most recent epidemiological statistics, breast cancer (BC) remains the leading malignancy in terms of incidence among women worldwide. Projections from the American Cancer Society estimate that in 2024, the number of newly diagnosed BC cases in the United States will reach 313,510, with the vast majority occurring in female patients (310,720 cases) and a smaller proportion in male patients (2,790 cases). Concurrently, mortality figures are expected to total 42,780, comprising 42,250 female and 530 male fatalities [1]. In terms of clinical presentation, the disease manifests with diverse characteristics, the most notable of which include localized breast masses, skin dimpling or retraction, a distinctive orange-peel texture, and regional lymph node enlargement, particularly in the axillary lymph nodes [2]. Investigations into pathogenic determinants have indicated that, alongside immutable intrinsic factors (such as gender, age, and genetic predisposition), a variety of external influences also contribute markedly to disease progression, predominantly encompassing dietary habits, lifestyle choices, and environmental exposures [3–5]. As a typical component of the Mediterranean diet, dried fruit intake not only reflects overall lifestyle and dietary patterns but has also attracted considerable attention due to its rich content of dietary fiber, polyphenols, and antioxidants—compounds that may have potential anti-cancer effects [6–8]. However, findings across observational studies remain inconsistent. For instance, a large prospective cohort study involving over 85,000 women found no significant association between dried fruit intake and BC risk, even with long-term follow-up (HR = 0.98, 95% CI: 0.93–1.04) [9]. Despite some inconsistencies in observational findings, our decision to focus on dried fruit is well-justified by two recent Mendelian randomization (MR) studies that systematically examined a range of dietary factors in relation to BC risk. Importantly, dried fruit emerged as the only dietary component showing a consistent and statistically significant inverse association with BC across both studies [10, 11]. This robust evidence provides a strong rationale for prioritizing dried fruit over other dietary exposures in our analysis. Furthermore, dried fruits are uniquely rich in bioactive compounds such as polyphenols, which have been implicated in antioxidant and anti-inflammatory pathways potentially relevant to BC prevention [12, 13]. These factors collectively highlight the importance of investigating the genetic mechanisms linking dried fruit intake and breast cancer susceptibility. To better explore the genetic architecture between dried fruit intake and breast cancer risk, we plan to conduct a comprehensive genome-wide study. Unlike previous observational studies, which are often influenced by confounding factors, the GWAS approach minimizes these biases, providing a more robust analysis and results.

With the advancement of genome-wide association studies (GWAS), a progressively clearer picture has emerged regarding the genetic variation landscape linked to BC susceptibility and dietary behaviors, including dried fruit intake, thereby laying a robust foundation for examining their genetic correlations. In recent years, the advent of innovative genetic statistical methodologies has enabled the assessment of genetic overlap across multiple phenotypes. These approaches have been effectively utilized to uncover genetic associations among diverse complex traits, such as the shared genetic loci observed between alcohol and cheese consumption and inflammatory bowel disease [14], the genetic correlations identified between common epilepsy and cognitive ability [15], and the common genetic underpinnings found between inflammatory bowel disease and systemic lupus erythematosus [16]. In this investigation, these well-established genetic statistical techniques were systematically applied to comprehensively explore the nature of the genetic overlap between BC susceptibility and dried fruit intake behavior.

A multi-dimensional genetic statistical approach was implemented to investigate the patterns of genetic associations between the two target phenotypes, based on genome-wide association study (GWAS) data derived exclusively from individuals of European ancestry. At the genome-wide scale, linkage disequilibrium score regression (LDSC) [17] and the high-dimensional likelihood (HDL) model [18] were employed to assess the overall strength of genetic correlation, whereas local genetic variation analysis (LAVA) [19] was applied to identify regional genetic association characteristics. To pinpoint specific shared genetic variant loci, the conditional/conjunctional false discovery rate (cond/conjFDR) method [20] was integrated alongside the multi-trait analysis of GWAS (MTAG) [21]. The conjFDR approach not only evaluates the extent of polygenic overlap but also serves as an effective means of screening shared genetic risk loci [20]. Recognized as a widely utilized analytical tool, MTAG offers distinct advantages in detecting comorbidity risk loci [21]. This comprehensive analytical framework is expected to contribute to a clearer understanding of the genetic association mechanisms linking BC susceptibility to dried fruit intake behavior.

Methods and materials

GWAS data

All genomic data used in this study were sourced from the IEU GWAS database (https://gwas.mrcieu.ac.uk/). Appropriate datasets were selected for analysis based on criteria such as publication date, sample size, and number of SNPs. Specifically, BC data were derived from

the Breast Cancer Association Consortium (BCAC), based on a 2017 meta-analysis that integrated data from OncoArray, iCOGS, and previous GWAS. The corresponding GWAS identifiers are as follows: overall BC (N_case = 122,977; N_control = 105,974; ID: "ieu-a-1126"), estrogen receptor-positive BC (ER^+ BC) (N_case = 69,501; N_control = 105,974; ID: "ieu-a-1127"), and estrogen receptor-negative BC (ER^- BC) (N_case = 21,468; N_control = 105,974; ID: "ieu-a-1128") [22].

Data on dried fruit intake (421,764 cases and 78,236 controls) were obtained from the GWAS dataset "ukbb-16576," released by the MRC-IEU in 2018. This study recruited participants aged 40 to 69 between 2006 and 2010 [23].

We performed rigorous quality control on all GWAS datasets. SNPs with minor allele frequency (MAF) > 0.01 in Europeans (based on the 1000 Genomes Project Phase 3 reference panel) and biallelic variants were retained, while SNPs in high-LD regions were excluded. Data were aligned to the hg19 reference genome, and variants without rsID or with duplicate rsIDs were removed. Analyses were restricted to individuals of European ancestry to reduce population stratification.

To control false positives, we applied Benjamini–Hochberg FDR correction in both LAVA and MTAG analyses. For condFDR/conjFDR, a significance threshold of conjFDR < 0.05 was used, consistent with prior studies [14, 24]. No FDR correction was applied to LDSC and HDL analyses, as these estimate genome-wide correlations rather than SNP-level associations. A schematic overview of the study design is provided in Fig. 1.

Global genetic correlation analyses

LDSC was used to estimate genome-wide genetic correlations between traits based on GWAS summary statistics [17]. LDSC regresses SNP effect sizes (Z-scores) on their LD scores, with the slope reflecting genetic covariance and the intercept accounting for potential biases such as sample overlap and population stratification. Genetic correlation is then obtained by scaling the covariance with trait heritabilities. A key advantage of LDSC is its ability to estimate genetic correlations without individual-level data, making it suitable for large-scale trait screening.

HDL analysis was also employed [18]. HDL uses a maximum likelihood framework and full LD matrices to

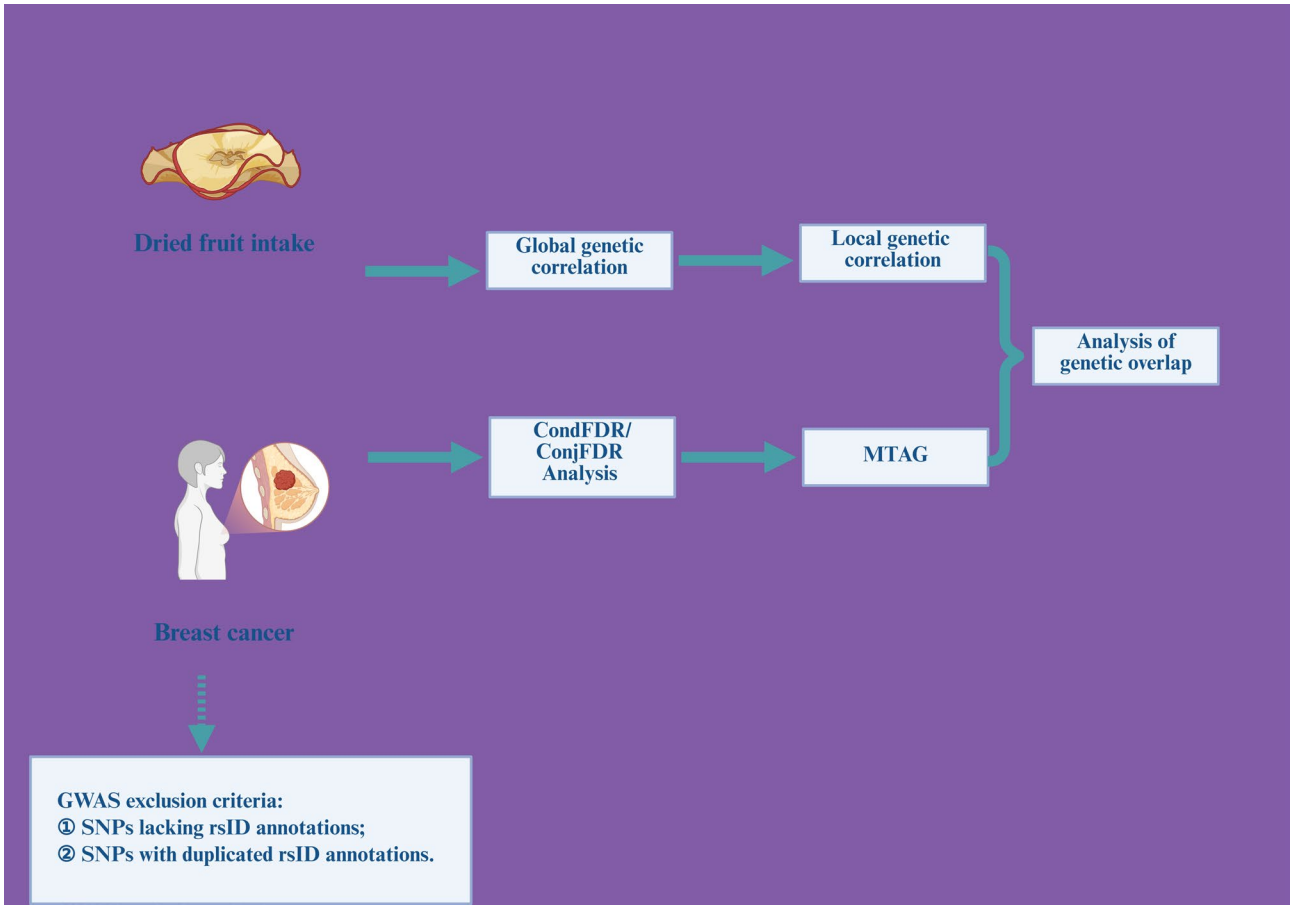

Fig. 1 Flowchart of the study. This figure is created by biorender

estimate genetic correlations, offering improved accuracy over LDSC. After quality control of GWAS summary statistics, an LD matrix is constructed, followed by eigen-decomposition to extract key components. These are used to estimate genetic covariance, which is then standardized to derive the genetic correlation coefficient. Unlike LDSC, HDL leverages the full LD structure rather than only diagonal elements, reducing bias.

To enhance robustness, both LDSC and HDL were applied in parallel. LDSC provides computational efficiency, while HDL offers higher precision. Their combined use supports consistency and reliability of the genetic correlation estimates.

Local genetic correlation analyses

The LAVA method identifies local genetic correlations between two traits through a series of straightforward steps [19]. First, genotype-phenotype correlation data are extracted from GWAS summary statistics, and these are used to calculate SNP effects. By applying linear regression models, LAVA maps the relationships between phenotypes and genotypes, allowing for the estimation of local genetic variance and gene effects for each trait. Based on these estimates, a matrix representing genetic effects in specific regions of the genome is created. Local genetic correlations are then calculated by determining the local genetic covariance between the traits. Unlike traditional global genetic correlation methods, LAVA can detect genetic overlap within smaller genomic regions, even if no significant global correlation is found. Additionally, LAVA can analyze conditional genetic relationships, enabling investigations into the genetic associations among multiple traits using partial correlation or multiple regression. This approach provides a more detailed view of complex genetic interactions, improving the detection of local genetic structures between traits, especially in cases where genetic influences are shared within particular genomic regions or gene loci. As a result, LAVA offers new insights into the genetic mechanisms of complex traits.

CondFDR/ConjFDR analysis

Conditional and conjunctional FDR (condFDR and conjFDR) are statistical approaches designed to identify shared genetic loci across traits using GWAS summary statistics [20]. condFDR improves discovery power by leveraging associations with a secondary phenotype to re-rank SNP associations for a primary trait, based on conditional cumulative distribution functions. conjFDR extends this by identifying SNPs jointly associated with both traits, using the maximum FDR value across traits to control for false discoveries.

These methods offer advantages over traditional approaches such as colocalization, including not requiring LD matrices, allowing for genome-wide application, and detecting both concordant and discordant effects. However, they are limited by sensitivity to long-range LD, inability to pinpoint causal variants, and dependence on GWAS sample size and quality.

Cross-trait meta-analysis

MTAG is a statistical method for jointly analyzing summary statistics from genetically correlated traits [21]. It assumes a shared variance-covariance structure of SNP effect sizes across traits, allowing improved power and trait-specific effect estimates. The method involves estimating GWAS error covariance via LD score regression, deriving the effect-size covariance matrix using the method of moments, and calculating adjusted SNP effects for each trait.

Compared to colocalization, MTAG enhances association detection by leveraging cross-trait genetic correlations and accounting for sample overlap, making it well suited for genome-wide discovery. However, MTAG does not confirm shared causal variants and assumes a common effect-size structure, which—if violated—may inflate false positives, particularly when GWAS traits differ in power. MTAG performance is also sensitive to GWAS data quality.

For the genetic variants identified through condFDR/conjFDR methods and MTAG analysis, systematic gene annotation was conducted utilizing the SNP2GENE functional module of the FUMA platform [25].

Results

Global genetic correlation

Through systematic genetic correlation analysis, significant genetic relationships were identified between BC and its subtypes in relation to dried fruit intake. The analysis demonstrated a negative genetic association between dried fruit intake and BC risk (Rg = -0.0905, $P = 0.0002$). Although statistically significant, this effect size is modest, suggesting a weak but potentially meaningful inverse genetic relationship. Additional subgroup analysis further validated this negative correlation in both ER^+ BC (Rg = -0.094, $P = 0.0036$) and ER^- BC (Rg = -0.1814, $P = 9.1079e\text{-}08$), with a slightly stronger genetic association observed in ER – BC (Table 1).

To enhance the robustness of the findings, independent validation was conducted using the HDL method. The cross-validation results showed high consistency with the primary analysis, reinforcing the presence of stable genetic associations between dried fruit intake and BC and its subtypes (Table 1). Taken together, these findings suggest a modest but consistent inverse genetic correlation, offering novel insights into the potential shared genetic architecture linking dried fruit intake behavior and BC susceptibility.

Table 1 Genetic correlation of BC(including ER^+BC and ER^-BC) and Dried fruit intake

Trait1	Trait2	LDSC-Rg	LDSC-*P*	HDL-Rg	HDL-*P*
BC	Dried fruit intake	-0.0905	0.0002	-0.0851	5.72e-10
ER^+BC	Dried fruit intake	-0.094	0.0036	-0.0617	5.97e-06
ER^-BC	Dried fruit intake	-0.1814	9.1079e-08	-0.1597	1.09e-13

BC, Breast cancer; ER^+, estrogen receptor-positive; ER^-, estrogen receptor-negative; Rg, genetic correlation. LDSC, Linkage Disequilibrium Score Regression; HDL, High-Dimensional Likelihood; Rg: Correlation between two traits, rg ranges from – 1 to 1, and the closer the value is to 1 or -1, the stronger the correlation is (plus or minus represents positive and negative correlation); P: The P-value of genetic correlation

The genetic correlations observed in this study were modest (e.g., $Rg \approx -0.09$), potentially reflecting polygenicity, residual confounding, or linkage disequilibrium (LD) artifacts rather than strong causality. Such small effect sizes are common in studies involving complex traits and dietary exposures. For instance, prior research has reported similar genetic correlations between cheese intake and inflammatory bowel disease ($Rg = -0.0929$) [14], and between alcohol consumption and cardiovascular disease ($Rg = 0.08$) [26]. Although statistically significant, these modest correlations may have limited biological or clinical implications. Biologically, they may indicate shared pathways influenced by multiple small-effect variants. Clinically, while they are unlikely to inform individual-level prediction, they may still offer insights at the population level. Therefore, these findings should be considered exploratory and hypothesis-generating, requiring further validation through functional studies, multi-omics approaches, and replication in diverse populations.

Local genetic correlation

Through local genetic correlation analysis, intricate association patterns between BC and dried fruit intake were identified at the chromosomal level. A total of twelve chromosomal regions exhibited significant correlations between BC and dried fruit intake, with five regions displaying positive correlations (located on chromosomes 1, 2, 3, and 18) and seven regions demonstrating negative correlations (located on chromosomes 4, 7, 11, 16, and 17) (Fig. 2A, Supplementary Table 1).

In the local genetic correlation analysis involving ER^+ BC, ER^- BC, and dried fruit intake, six significant correlation regions were identified for each subtype. Among these regions, an equal number of positive correlations (located on chromosomes 3, 16, and 21) and negative correlations (located on chromosomes 7, 10, and 16) were observed (Fig. 2B-C, Supplementary Tables 2–3).

ConjFDR analysis identifies shared genomic loci between two traits

Quantile-quantile (Q-Q) plot analysis identified significant genetic associations between BC and dried fruit intake (Fig. 3). As the significance level of genetic associations for one trait increased, the association values for the other trait exhibited a progressive leftward deviation. This observed deviation pattern not only signified strong genetic connections between the two traits but also implied the potential presence of shared disease-related genetic variant loci, offering critical evidence for further elucidation of the common genetic mechanisms underlying BC and dried fruit intake.

To investigate the potential genetic associations between dried fruit intake and BC, along with its subtypes, a comprehensive genetic analysis was performed utilizing the conjFDR method. At a statistical significance threshold of conjFDR < 0.05, a total of 32 shared genetic variants were identified, demonstrating simultaneous associations with both dried fruit intake and BC risk (Fig. 4A, Supplementary Table 4).

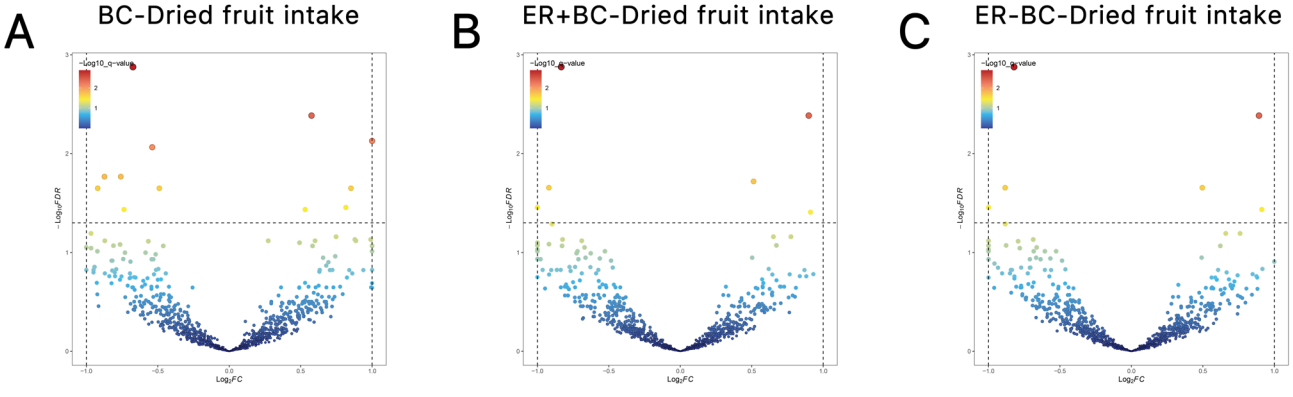

Fig. 2 LAVA analysis of BC and dried fruit intake. The x-axis stands for the local genetic correlation and the y-axis represents -log10 transformed LAVA values with a dotted horizontal line reflecting significance. Each dot represents a SNP and the border indicates a lead SNP.The dashed line indicates the expected line with a correction P of 0.05. **(A)** Local genetic correlation between BC and dried fruit intake. **(B)** Local genetic correlation between ER^+ BC and dried fruit intake. **(C)** Local genetic correlation between ER^-BC and dried fruit intake. LAVA, local variant association; BC, Breast cancer; ER^+, estrogen receptor-positive; ER^-, estrogen receptor-negative

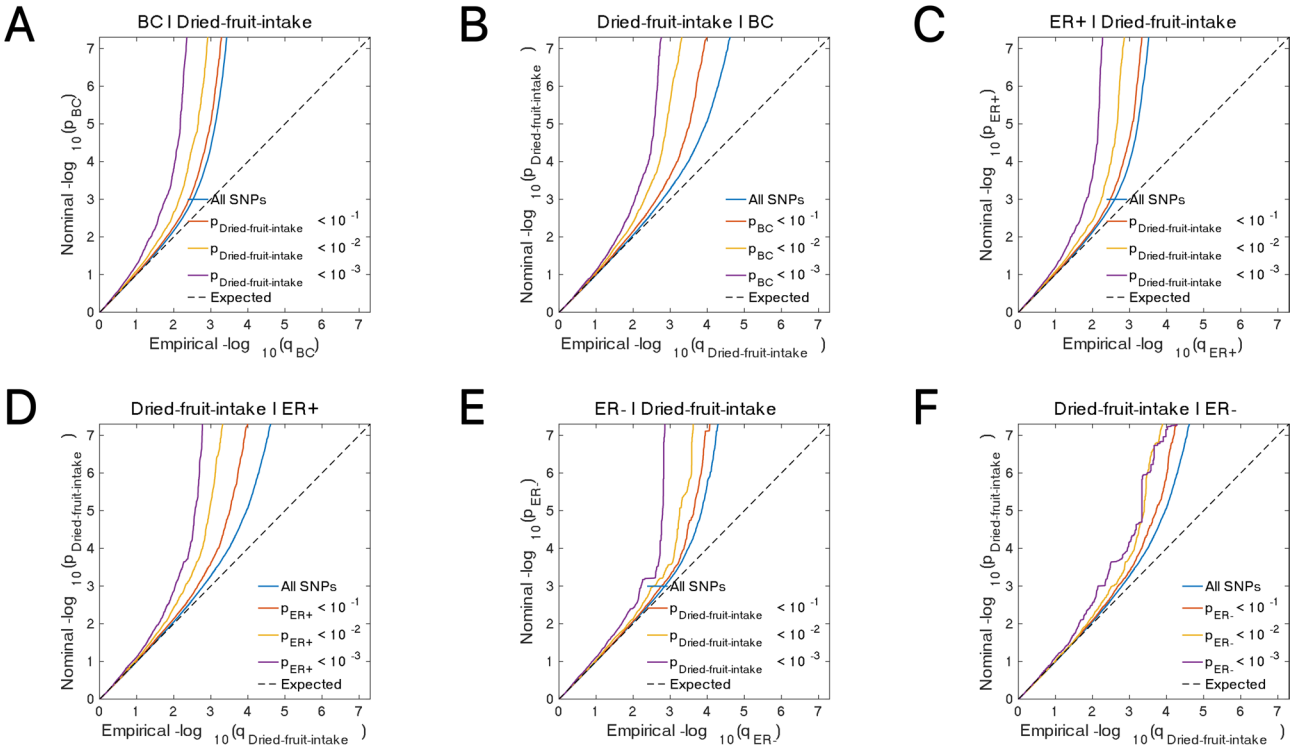

Fig. 3 A conditional Q–Q plot illustrating the observed − log10(p) values of single SNP associations for dried fruit intake and BC, plotted against their expected distribution under the null hypothesis. SNPs are stratified based on their significance levels in the conditional phenotype ($p \leq 0.1$, $p \leq 0.01$, and $p \leq 0.001$). The blue line represents all SNPs, while the dashed line indicates the null expectation. **(A)** BC-Dried fruit intake. **(B)** Dried fruit intake-BC. **(C)** ER^+ BC-Dried fruit intake. **(D)** Dried fruit intake-ER^+ BC. **(E)** ER^-BC-Dried fruit intake. **(F)** Dried fruit intake-ER^-BC. BC, Breast cancer; ER^+, estrogen receptor-positive; ER^-, estrogen receptor-negative

Through an extensive examination of BC subtypes, a marked disparity in shared genetic architecture was observed between ER^+ and ER^- BC. Specifically, 24 overlapping genetic regions were identified between dried fruit intake and ER^+ BC, whereas only 3 shared genetic variants were detected in ER^- BC (Fig. 4B–C, Supplementary Tables 5–6). This stark contrast (24 vs. 3 loci) strongly suggests the presence of subtype-specific genetic mechanisms underlying the relationship between dried fruit intake and BC risk. These systematic GWAS findings provide essential genetic evidence to support the biological distinction between ER^+ and ER^- BC in the context of dietary behavior.

MTAG

The FUMA system was utilized to perform systematic functional annotation of the MTAG output results. This analysis identified 36 shared genetic variant loci associated with both BC and dried fruit intake (Fig. 5A, Supplementary Table 7). Notably, through cross-validation integrating the conjFDR method and MTAG analysis, five key genes— B-cell CLL/lymphoma 11 A (BCL11A), Mitotic Arrest Deficient-Like 1 (MAD1L1), Myeloid/Lymphoid or Mixed-Lineage Leukemia; Translocated To, 10 (MLLT10), Jumonji Domain Containing 1 C (JMJD1C), and RP11-795H16.2—were successfully detected (Fig. 5B). Among them, BCL11A has been implicated in breast cancer progression through its role in stem cell regulation [27]; JMJD1C are transcriptional regulators with reported relevance in both tumorigenesis and endocrine function [28]; MAD1L1、MLLT10 and RP11-795H16.2 is increasingly recognized for its regulatory role in cancer biology. These findings provide important functional context and suggest that the shared loci may influence both dietary behavior and breast cancer risk through interconnected biological pathways.

In the subtype stratification analysis, MTAG identified 36 shared genetic susceptibility loci associated with both ER^+ BC and dried fruit intake behavior (Fig. 5C, Supplementary Table 8). Further cross-validation utilizing the conjFDR method confirmed significant associations for three genes—MAD1L1, MLLT10, and JMJD1C—across both statistical approaches (Fig. 5D). Conversely, MTAG analysis of ER^- BC detected 35 shared genetic variant loci linked to dried fruit intake (Fig. 5E, Supplementary Table 9). However, none of these loci were independently validated through the conjFDR method (Fig. 5F).

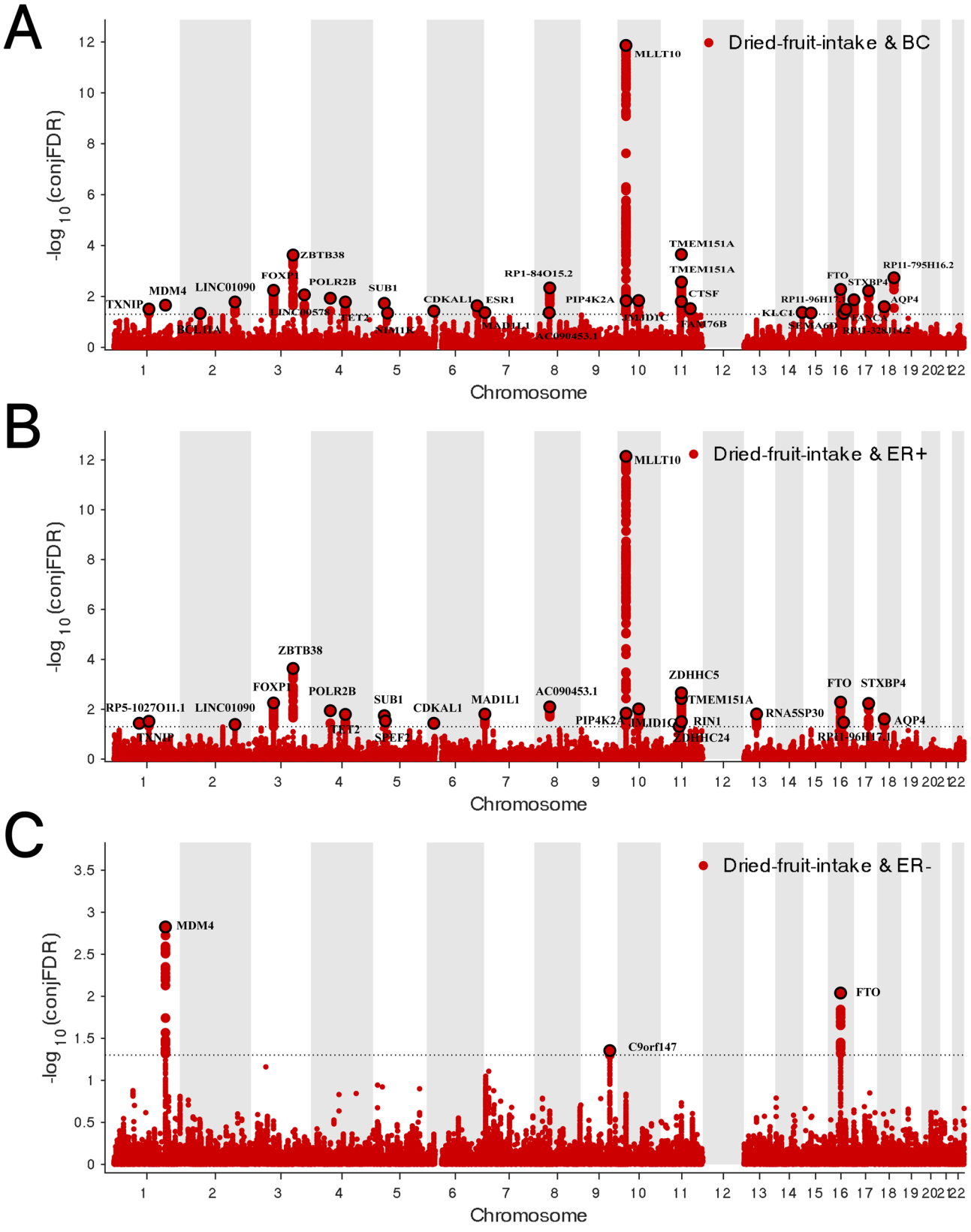

Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 The x-axis stands for the chromosomalnumber and position and the y-axis represents -log10 transformed conjFdr values with a dotted horizontal line reflecting significance. Each dot represents a SNP and the border indicates a lead SNP. **(A)** ConjFDR Manhattan plot of BC and dried fruit intake. **(B)** ConjFDR Manhattan plot of ER^+ BC and dried fruit intake. **(C)** ConjFDR Manhattan plot of ER^-BC and dried fruit intake.The shared risk loci between dried fruit intake and BC, ER^+ BC and ER^-BC were marked. The statistically significant causality is defined to be conjFDR < 0.05. ConjFDR: conjunctional FDR; BC, Breast cancer; ER^+, estrogen receptor-positive; ER^-, estrogen receptor-negative

Discussion

The genetic relationship between dried fruit intake and BC, alongside its subtypes, was examined via GWAS analysis. A genome-wide investigation identified significant inverse genetic correlations among these traits. Moreover, an in-depth local genetic correlation analysis uncovered notable association signals linking dried fruit intake with BC across several chromosomal regions. The genetic overlap between these two phenotypes was further corroborated through Q-Q plot analysis. Subsequently, by employing a dual validation strategy incorporating conjFDR and MTAG, five potential shared genetic loci were effectively pinpointed, specifically BCL11A, MAD1L1, MLLT10, JMJD1C, and RP11-795H16.2. These discoveries provide preliminary evidence for understanding the genetic mechanisms that may underlie the association between dried fruit intake and BC susceptibility.

Current evidence regarding the potential protective role of dried fruit intake against BC susceptibility remains limited and inconclusive. A cohort study involving 35,372 female participants suggested that individuals with regular dried fruit intake may exhibit up to a 40% reduction in BC risk [29]. Similarly, large-scale prospective cohort investigations have indicated that women consuming at least 28 g of dried fruits per week were associated with a statistically significant lower risk of BC (hazard ratio [HR] = 0.83, 95% confidence interval [CI]: 0.74–0.93) [30]. This apparent protective association appeared more pronounced among postmenopausal women, which may be attributed to the presence of polyunsaturated fatty acids, phytosterols, and antioxidants in dried fruits [31]. Systematic meta-analyses have also reported a significant association between long-term (> 10 years) habitual dried fruit intake and reduced BC risk (relative risk [RR] = 0.85, 95% CI: 0.76–0.95) [32]. Furthermore, some clinical intervention trials have suggested that supplementation with mixed nuts, alongside standard treatment, may improve inflammatory markers and oxidative stress indices in BC patients [33]. Previous studies have reported a potential dose-response relationship, suggesting that the protective effect of dried fruit intake may depend on the duration and quantity of consumption [34]. In addition, recent Mendelian randomization analyses have provided preliminary evidence for an inverse causal association between dried fruit intake and BC risk [10, 11], which supports our current findings. However, it is important to acknowledge that not all studies have observed a significant association. For instance, a large prospective cohort study involving over 85,000 women found no statistically significant relationship between dried fruit consumption and breast cancer incidence over a 14.8-year follow-up (HR = 0.98, 95% CI: 0.93–1.04). These null findings may reflect the complexity of dietary exposures and could be influenced by residual confounding, measurement errors, or differences in population characteristics. While our study identified modest but significant genetic correlations, these results should be interpreted with caution. Genetic approaches, unlike traditional observational studies, are less susceptible to confounding and reverse causation, and can detect weak yet consistent signals of genetic overlap [9]. By employing multiple complementary statistical methods, our findings provide preliminary evidence suggesting a shared genetic architecture between dried fruit intake and BC susceptibility. Nonetheless, these results remain exploratory and require further validation through functional studies and replication in multiethnic cohorts.

In investigations of shared susceptibility loci, BCL11A has emerged as a potentially important transcription factor due to its complex regulatory functions, which may be implicated in the onset and progression of BC. Evidence from both in vitro and in vivo studies has suggested that suppression of BCL11A expression could reduce the tumorigenic potential of BC cell lines and possibly promote tumor regression [27]. Conversely, upregulation of BCL11A has been associated with enhanced tumor formation, and may correlate positively with histological grade and clinical stage [35, 36], with higher expression levels often considered indicative of poor prognosis. Notably, BCL11A is hypothesized to play a role in regulating breast cancer stem cell (BCSC) differentiation [27]. A study by Moody et al. proposed that BCL11A may influence the transcription of genes critical for BCSC development through its interactions with histone-modifying complexes such as Polycomb Repressive Complex 2 (PRC2), the Nucleosome Remodeling and Deacetylase Complex (NuRD), and the Switch-Independent 3 A Histone Deacetylase Complex (SIN3A) [37]. In addition, Chen et al. reported that BCL11A expression in BC may be regulated by circular non-coding RNA derived from epithelial-stromal interaction 1 (circEPSTI1), a pathway potentially linked to cell proliferation and apoptosis [36]. Recent studies have also indicated that BCL11A could be involved in metabolic homeostasis through interactions with bioactive molecules—several of which are abundantly present in specific types of dried fruits

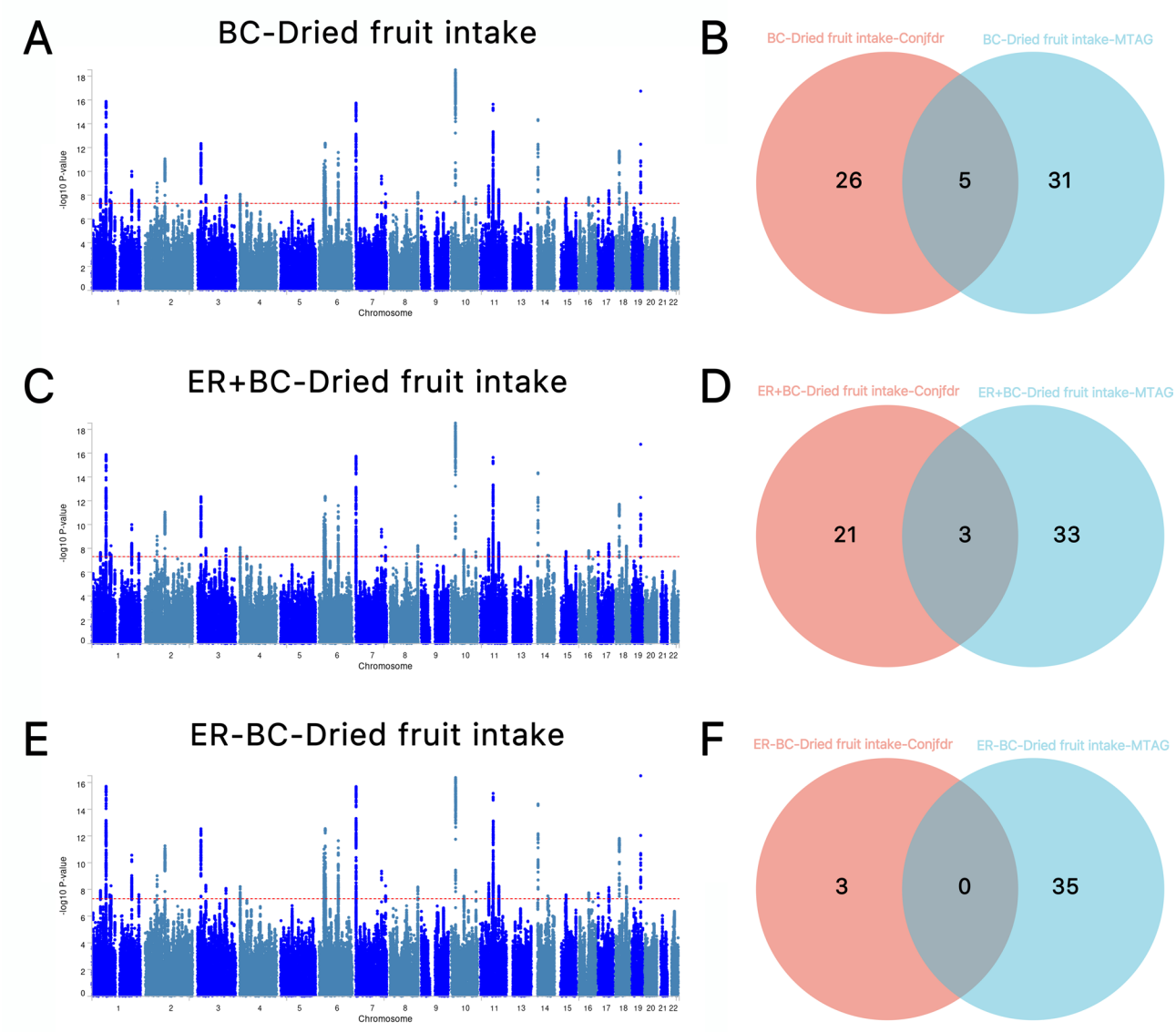

Fig. 5 The x-axis stands for the chromosomalnumber and position and the y-axis represents -log10 transformed MTAG values with a dotted horizontal line reflecting significance. Each dot represents a SNP and the border indicates a lead SNP. **(A)** Manhattan map of genetic risk loci for BC and dried fruit intake by MTAG. **(B)** Intersection gene map of BC and dried fruit intake after conjfdr and MTAG analysis. **(C)** Manhattan map of genetic risk loci for ER^+ BC and dried fruit intak by MTAG. **(D)** Intersection gene map of ER^+ BC and dried fruit intak after conjfdr and MTAG analysis. **(E)** Manhattan map of genetic risk loci for ER^-BC and dried fruit intake by MTAG. **(F)** Intersection gene map of ER^-BC and dried fruit intake after conjfdr and MTAG analysis. MTAG, multi-trait analysis of GWAS; BC, Breast cancer; ER^+, estrogen receptor-positive; ER^-, estrogen receptor-negative

[38]. Polyphenolic compounds in dried fruits, such as quercetin and ellagic acid, may influence breast cancer development by modulating the interaction between BCL11A and epigenetic enzymes like Histone Deacetylases (HDACs) and DNA Methyltransferases (DNMTs). These compounds can alter chromatin accessibility and regulate gene expression, thereby affecting the severity of breast cancer lesions [39]. BCL11A has been shown to regulate breast cancer stem cell (BCSC) differentiation. Moody et al. (2015) reported that BCL11A maintains stem-like properties in triple-negative breast cancer by interacting with chromatin remodeling complexes (e.g., PRC2, NuRD, SIN3A), influencing genes related to self-renewal and cell fate. Knockdown of BCL11A reduced tumorsphere formation and stem cell marker expression, highlighting its role in BCSC maintenance [35]. JMJD1C, an epigenetic regulator belonging to the Jumonji C class demethylases, is primarily implicated in the modulation of histone methylation states. Recent investigations have demonstrated that JMJD1C exerts a substantial influence on BC progression, being strongly associated with BC invasiveness, metastatic capability, and resistance to therapeutic agents [40]. Through its demethylation activity, JMJD1C governs gene transcriptional regulation,

particularly affecting the expression of genes involved in the cell cycle, proliferation, and apoptosis [28]. Additionally, JMJD1C has been reported to facilitate BC immune evasion mechanisms and augment tumor cell immune tolerance by orchestrating epigenetic modifications, such as histone H3K9 methylation [41]. Nutritional components in dried fruits, such as vitamin E and unsaturated fatty acids, may influence breast cancer risk by regulating the transcriptional and epigenetic networks mediated by JMJD1C, thereby affecting cell proliferation and apoptosis associated with BC [42, 43]. The identified genetic loci, their functional annotations, and literature links are provided in Supplementary Table 10. The research on the roles of MAD1L1, MLLT10, and RP11-795H16.2 in both BC and dried fruit intake, as well as how dried fruit components influence BC development through BCL11A and JMJD1C, is still limited. Direct experimental studies are needed in the future to validate these findings.

Although this study integrated multiple high-throughput computational approaches and yielded valuable scientific insights, several methodological constraints warrant consideration. First, existing statistical methodologies are unable to fully mitigate the potential influence of LD effects on the results. Second, the partial overlap of samples within large-scale cohort studies remains an unavoidable challenge, which could introduce biases affecting the precision of statistical inference. Moreover, the modest genetic correlations observed in this study (e.g., Rg ≈ −0.09) may reflect the polygenic architecture of the traits or residual confounding, and—given the reliance on summary statistics and methods like conjFDR/MTAG—could also be influenced by LD artifacts, rather than indicating strong causal relationships. Thus, these findings should be interpreted with caution. The GWAS datasets used in this study were not stratified by sex, limiting our ability to explore sex-specific differences in genetic susceptibility to BC. Given the sexual dimorphism in BC, the lack of sex stratification may have overlooked important sex-specific genetic associations. Future studies should incorporate sex-stratified analyses to address this limitation. A potential limitation is residual confounding from lifestyle factors. Since GWAS summary data lack full adjustment for behaviors like physical activity or alcohol intake, some observed genetic correlations may reflect unmeasured health-related traits. Due to the limited availability of GWAS data from non-European populations, this study was restricted to individuals of European ancestry. As genetic structures—such as allele frequencies and LD patterns—may vary across populations, the generalizability of our findings is limited. Future multiethnic studies are needed to validate and extend these results. We also acknowledge that genetic structures may differ across populations, which could impact the global applicability of these results. We look forward to the availability of data from other populations in the future to assess the broader applicability of our findings. Finally, the study lacks direct functional validation of the identified shared genes, which may limit the interpretation of their biological relevance. Future experimental studies, such as CRISPR-based functional assays and nutrigenomics analysis in relevant tissues, are needed to validate the roles of these genes and elucidate their potential mechanisms.

Conclusion

In conclusion, this exploratory analysis reveals a potential genetic overlap between dried fruit intake and BC, offering initial insights into possible molecular links between dietary behavior and disease susceptibility. Through multi-tiered genomic investigations, several putative shared genetic loci, including BCL11A, MAD1L1, MLLT10, JMJD1C, and RP11-795H16.2, have been identified. While these findings are statistically significant, further functional validation is required to confirm their biological relevance. If confirmed in future studies, these loci could provide valuable clues for developing more targeted hypotheses regarding diet–cancer interactions, with potential applications in personalized nutrition research.

Abbreviations

BC	Breast cancer
GWAS	Genome-wide association studies
Cond/conjFDR	Conditional/conjunctional FDR
MTAG	Multi-trait analysis of GWAS
MR	Mendelian randomization
LDSC	Linkage disequilibrium score regression
HDL	High-dimensional likelihood
LAVA	Local genetic variation analysis
BCAC	Breast Cancer Association Consortium
ER + BC	Estrogen receptor-positive BC
ER^- BC	Estrogen receptor-negative BC
MAF	Minor allele frequency
1KGP	1000 Genomes Project
FDR	False discovery rate
SNP	Single nucleotide polymorphism
LD	Linkage disequilibrium
Q-Q	Quantile-quantile
BCL11A	B-cell CLL/lymphoma 11 A
MAD1L1	Mitotic Arrest Deficient-Like 1
MLLT10	Myeloid/Lymphoid or Mixed-Lineage Leukemia; Translocated To, 10
JMJD1C	Jumonji Domain Containing 1 C
HR	Hazard ratio
CI	Confidence interval
RR	Relative risk
PRC2	Polycomb Repressive Complex 2
NuRD	Nucleosome Remodeling and Deacetylase Complex
SIN3A	Switch-Independent 3 A Histone Deacetylase Complex
CircEPSTI1	Circular non-coding RNA derived from epithelial-stromal interaction 1
HDACs	Histone Deacetylases
DNMTs	DNA Methyltransferases
BCSC	Breast cancer stem cell

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s41043-025-01059-y.

Supplementary Material 1

Acknowledgements
The authors thank Bullet Edits Limited for the linguistic editing and proofreading of the manuscript, and also thanks Biorender (biorender.com) for providing the graphical tools used in this study.

Author contributions
Keyan Zou: Conceptualization, methodology, formal analysis, data curation, writing-original draft preparation; Yi Zheng: writing-original draft preparation, visualization; Xiaohong Ren: Data curation, writing-original draft preparation, visualization; Wei Cui: writing-review and editing, Data curation, visualization. All authors contributed to the article and approved the final version of the manuscript.

Funding
This study did not receive any funding in any form.

Data availability
All the GWAS data and statistical software used in this study were publicly available (which can be accessed through the following URLs), and all the generated results in this study were provided in the main text and supplemental data. IEU database: https://gwas.mrcieu.ac.uk/. LDSC: https://github.com/bulik/ldsc. HDL: https://github.com/zhenin/HDL. LAVA: https://github.com/josefin-werme/LAVA. ConjFDR: https://github.com/precimed/pleiofdr. MTAG: https://github.com/jonjala/mtag. FUMA: https://fuma.ctglab.nl.

Declarations

Competing interests
The authors declare no competing interests.

Ethics approval and consent to participate
Not applicable. We used publicly available data that were obtained with ethical approval from their respective institutional review boards and informed consent from all participants. No administrative permissions were required for accessing the data.

Clinical trial number
Not applicable.

Received: 8 May 2025 / Accepted: 17 August 2025
Published online: 31 August 2025

References

1. Siegel RL, Giaquinto AN, Jemal A, Cancer statistics. 2024. CA Cancer J Clin (2024) 74:12–49. https://doi.org/10.3322/caac.21820
2. Loibl S, Poortmans P, Morrow M, Denkert C, Curigliano G. Breast cancer. Lancet. 2021;397:1750–69. https://doi.org/10.1016/S0140-6736(20)32381-3.
3. Toledo E, Salas-Salvadó J, Donat-Vargas C, Buil-Cosiales P, Estruch R, Ros E, Corella D, Fitó M, Hu FB, Arós F, et al. Mediterranean diet and invasive breast cancer risk among women at high cardiovascular risk in the PREDIMED trial: A randomized clinical trial. JAMA Intern Med. 2015;175:1752–60. https://doi.org/10.1001/jamainternmed.2015.4838.
4. Hiatt RA, Brody JG. Environmental determinants of breast cancer. Annu Rev Public Health. 2018;39:113–33. https://doi.org/10.1146/annurev-publhealth-040617-014101.
5. Campbell NJ, Barton C, Cutress RI, Copson ER. Impact of obesity, lifestyle factors and health interventions on breast cancer survivors. Proc Nutr Soc. 2023;82:47–57. https://doi.org/10.1017/S0029665122002816.
6. Bolling BW, Aune D, Noh H, Petersen KS, Freisling H. Dried fruits, nuts, and cancer risk and survival: A review of the evidence and future research directions. Nutrients. 2023;15:1443. https://doi.org/10.3390/nu15061443.
7. Lalioti A, Verzeletti L, Tiberio P, Gerosa R, Gaudio M, Saltalamacchia G, Pastore M, Zambelli A, Santoro A, De Sanctis R. Common misconceptions about diet and breast cancer: an unclear issue to dispel. Cancers. 2024;16:306. https://doi.org/10.3390/cancers16020306.
8. Ruban M, Pozhidaeva E, Bolotina L, Kaprin A. The role of diet and nutrition in cancer development and management: from molecular mechanisms to personalized interventions. Foods. 2025;14:1788. https://doi.org/10.3390/foods14101788.
9. Prentice RL, Caan B, Chlebowski RT, Patterson R, Kuller LH, Ockene JK, Margolis KL, Limacher MC, Manson JE, Parker LM, et al. Low-fat dietary pattern and risk of invasive breast cancer: the women's health initiative randomized controlled dietary modification trial. JAMA. 2006;295:629–42. https://doi.org/10.1001/jama.295.6.629.
10. Wang X, Chen L, Cao R, Ma R, Liu Y, Zhao Q, Du Y. Associations of 10 dietary habits with breast cancer: A Mendelian randomization study. Front Nutr. 2023;10. https://doi.org/10.3389/fnut.2023.1215220.
11. Yu C, Xu J, Xu S, Peng H, Tang L, Sun Z, Chen W. Causal relationship between dietary factors and breast cancer risk: A Mendelian randomization study. Heliyon. 2023;9. https://doi.org/10.1016/j.heliyon.2023.e20980.
12. Abdal Dayem A, Choi HY, Yang G-M, Kim K, Saha SK, Cho S-G. The anti-cancer effect of polyphenols against breast cancer and cancer stem cells: molecular mechanisms. Nutrients. 2016;8:581. https://doi.org/10.3390/nu8090581.
13. Zhou Y, Zheng J, Li Y, Xu D-P, Li S, Chen Y-M, Li H-B. Natural polyphenols for prevention and treatment of cancer. Nutrients. 2016;8:515. https://doi.org/10.3390/nu8080515.
14. Huang Z, Yuan W. Exploring genetic structures and shared sites between alcohol, cheese intake, and inflammatory bowel disease. Front Nutr. 2025;12:1468457. https://doi.org/10.3389/fnut.2025.1468457.
15. Karadag N, Hagen E, Shadrin AA, van der Meer D, O'Connell KS, Rahman Z, Kutrolli G, Parker N, Bahrami S, Fominykh V, et al. Unraveling the shared genetics of common epilepsies and general cognitive ability. Seizure. 2024;122:105–12. https://doi.org/10.1016/j.seizure.2024.09.016.
16. Yuan W, Luo Q, Wu N. Investigating the shared genetic basis of inflammatory bowel disease and systemic lupus erythematosus using genetic overlap analysis. BMC Genomics. 2024;25:868. https://doi.org/10.1186/s12864-024-10787-0.
17. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh P-R, ReproGen Consortium PG, Consortium, Genetic Consortium for Anorexia Nervosa of the Wellcome Trust Case Control Consortium 3, Duncan L, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47:1236–41. https://doi.org/10.1038/ng.3406.
18. Ning Z, Pawitan Y, Shen X. High-definition likelihood inference of genetic correlations across human complex traits. Nat Genet. 2020;52:859–64. https://doi.org/10.1038/s41588-020-0653-y.
19. Werme J, van der Sluis S, Posthuma D, de Leeuw CA. An integrated framework for local genetic correlation analysis. Nat Genet. 2022;54:274–82. https://doi.org/10.1038/s41588-022-01017-y.
20. Smeland OB, Frei O, Shadrin A, O'Connell K, Fan C-C, Bahrami S, Holland D, Djurovic S, Thompson WK, Dale AM, et al. Discovery of shared genomic loci using the conditional false discovery rate approach. Hum Genet. 2020;139:85–94. https://doi.org/10.1007/s00439-019-02060-2.
21. Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Nguyen-Viet TA, Wedow R, Zacher M, Furlotte NA, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. Nat Genet. 2018;50:229–37. https://doi.org/10.1038/s41588-017-0009-4.
22. Michailidou K, Lindström S, Dennis J, Beesley J, Hui S, Kar S, Lemaçon A, Soucy P, Glubb D, Rostamianfar A, et al. Association analysis identifies 65 new breast cancer risk loci. Nature. 2017;551:92–4. https://doi.org/10.1038/nature24284.
23. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK biobank resource with deep phenotyping and genomic data. Nature. 2018;562:203–9. https://doi.org/10.1038/s41586-018-0579-z.
24. Lu W, He X, Peng H, Lei P, Liu J, Ding Y, Yan B, Ma X, Yang J. Exploring the shared genetic architecture between testosterone traits and major depressive disorder. BMC Psychiatry. 2025;25:651. https://doi.org/10.1186/s12888-025-07096-5.
25. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun. 2017;8:1826. https://doi.org/10.1038/s41467-017-01261-5.
26. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, Datta G, Davila-Velderrain J, McGuire D, Tian C, et al. Association studies of up to 1.2 million individuals

yield new insights into the genetic etiology of tobacco and alcohol use. Nat Genet. 2019;51:237–44. https://doi.org/10.1038/s41588-018-0307-5.
27. Yeo SK, Guan J-L. Breast cancer: multiple subtypes within a tumor? Trends Cancer. 2017;3:753–60. https://doi.org/10.1016/j.trecan.2017.09.001.
28. Manni W, Jianxin X, Weiqi H, Siyuan C, Huashan S. JMJD family proteins in cancer and inflammation. Signal Transduct Target Ther. 2022;7:1–22. https://doi.org/10.1038/s41392-022-01145-1.
29. Dunneram Y, Greenwood DC, Cade JE. Diet and risk of breast, endometrial and ovarian cancer: UK women's cohort study. Br J Nutr. 2019;122:564–74. https://doi.org/10.1017/S0007114518003665.
30. Wu L, Wang Z, Zhu J, Murad AL, Prokop LJ, Murad MH. Nut consumption and risk of cancer and type 2 diabetes: a systematic review and meta-analysis. Nutr Rev. 2015;73:409–25. https://doi.org/10.1093/nutrit/nuv006.
31. Falasca M, Casari I, Maffucci T. Cancer chemoprevention with nuts. JNCI: J Natl Cancer Inst. 2014;106:dju238. https://doi.org/10.1093/jnci/dju238.
32. van den Brandt PA, Nieuwenhuis L. Tree nut, peanut, and peanut butter intake and risk of postmenopausal breast cancer: the Netherlands cohort study. Cancer Causes Control. 2017;29:63. https://doi.org/10.1007/s10552-017-0979-7.
33. Fadelu T, Zhang S, Niedzwiecki D, Ye X, Saltz LB, Mayer RJ, Mowat RB, Whittom R, Hantel A, Benson AB, et al. Nut consumption and survival in patients with stage III colon cancer: results from CALGB 89803 (alliance). J Clin Oncol: Off J Am Soc Clin Oncol. 2018;36:1112–20. https://doi.org/10.1200/JCO.2017.75.5413.
34. Aune D, Keum N, Giovannucci E, Fadnes LT, Boffetta P, Greenwood DC, Tonstad S, Vatten LJ, Riboli E, Norat T. Nut consumption and risk of cardiovascular disease, total cancer, all-cause and cause-specific mortality: a systematic review and dose-response meta-analysis of prospective studies. BMC Med. 2016;14:207. https://doi.org/10.1186/s12916-016-0730-3.
35. Khaled WT, Choon Lee S, Stingl J, Chen X, Raza Ali H, Rueda OM, Hadi F, Wang J, Yu Y, Chin S-F, et al. BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. Nat Commun. 2015;6:5987. https://doi.org/10.1038/ncomms6987.
36. Chen B, Wei W, Huang X, Xie X, Kong Y, Dai D, Yang L, Wang J, Tang H, Xie X. circEPSTI1 as a prognostic marker and mediator of triple-negative breast cancer progression. Theranostics. 2018;8:4003–15. https://doi.org/10.7150/thno.24106.
37. Moody RR, Lo M-C, Meagher JL, Lin C-C, Stevers NO, Tinsley SL, Jung I, Matvekas A, Stuckey JA, Sun D. Probing the interaction between the histone methyltransferase/deacetylase subunit RBBP4/7 and the transcription factor BCL11A in epigenetic complexes. J Biol Chem. 2018;293:2125–36. https://doi.org/10.1074/jbc.M117.811463.
38. Zhang H, Zeng J, Zhang F, Liu J, Liang L. Role of B-cell lymphoma/leukemia 11A in normal and malignant hematopoiesis. Biology. 2025;14:26. https://doi.org/10.3390/biology14010026.
39. Nandakumar V, Vaid M, Katiyar SK. (-)-epigallocatechin-3-gallate reactivates silenced tumor suppressor genes, Cip1/p21 and p16INK4a, by reducing DNA methylation and increasing histones acetylation in human skin cancer cells. Carcinogenesis. 2011;32:537–44. https://doi.org/10.1093/carcin/bgq285.
40. Xu R, Lan H, Zhang L, Yang S, Mao Y, Che H. circ_JMJD1C expedites breast cancer progression by regulating miR-182-5p/JMJD1C/SOX4 axis. Cell Mol Biol (Noisy-Le-Gd Fr) (2024) 70:204–11. https://doi.org/10.14715/cmb/2024.70.3.31
41. Sui Y, Gu R, Janknecht R. Crucial functions of the JMJD1/KDM3 epigenetic regulators in cancer. Mol Cancer Res: MCR. 2021;19:3–13. https://doi.org/10.1158/1541-7786.MCR-20-0404.
42. Mocchegiani E, Costarelli L, Giacconi R, Malavolta M, Basso A, Piacenza F, Ostan R, Cevenini E, Gonos ES, Franceschi C, et al. Vitamin E-gene interactions in aging and inflammatory age-related diseases: implications for treatment. A systematic review. Ageing Res Rev. 2014;14:81–101. https://doi.org/10.1016/j.arr.2014.01.001.
43. Sabaté J, Oda K, Ros E. Nut consumption and blood lipid levels: A pooled analysis of 25 intervention trials. Arch Intern Med. 2010;170:821–7. https://doi.org/10.1001/archinternmed.2010.79.

Publisher's note