ACCSCIENCE PUBLISHING

Cancer Plus

ORIGINAL RESEARCH ARTICLE

Unsaturated fatty acid metabolism and *FADS2* as candidate factors in triple-negative breast cancer

Ying Chen, Yu-Hui Ma, Zhi-hui Hu, Qi Liu, Bin Song*, and He Hao*

Cancer Center, Shanxi Bethune Hospital, Shanxi Academy of Medical Sciences, Third Hospital of Shanxi Medical University, Tongji Shanxi Hospital, Taiyuan, Shanxi, China

Abstract

Triple-negative breast cancer (TNBC) is aggressive and lacks effective targeted therapies. Using estrogen receptor (ER)-negative breast cancer as a proxy for TNBC, we explored circulating metabolite associations and downstream candidate genes. A two-sample Mendelian randomization (MR) analysis used large-scale metabolome genome-wide association study data (~1,400 metabolites). Candidate metabolites underwent Kyoto Encyclopedia of Genes and Genomes enrichment. Computationally predicted targets were intersected with common differentially expressed genes from three Gene Expression Omnibus datasets and Weighted Gene Co-expression Network Analysis modules. Least absolute shrinkage and selection operator-Cox and random survival forest prioritized genes in the Cancer Genome Atlas–TNBC. Single-cell RNA-seq (GSE176078) explored heterogeneity and pathways. MR identified 38 candidate metabolites nominally associated with ER-negative breast cancer risk, enriched in unsaturated fatty acid and arachidonic acid metabolism. Integration yielded 39 candidate genes, and survival-based machine learning prioritized *KRAS*, *MGST3*, *FADS2*, *UGT8*, *IDO1*, and *SRC*. Single-cell analysis showed higher *FADS2* in a presumed malignant epithelial subpopulation (log2FC = 2.53, false discovery rate < 0.05), associated with cholesterol biosynthesis, sterol metabolism, and cell-cycle programs. These exploratory findings nominate abnormal circulating unsaturated fatty acid metabolism and *FADS2* as candidate factors in ER-negative/TNBC biology. *FADS2* is a candidate biomarker and potential target requiring TNBC-specific genetic, patient-aware single-cell, prognostic, and experimental validation.

Keywords: Triple negative breast cancer; Estrogen receptor-negative breast cancer proxy; Mendelian randomization; Unsaturated fatty acid metabolism; *FADS2*; Single cell transcriptomics; Exploratory computational analysis

***Corresponding authors:**
He Hao
(haohe@sxbqeh.com.cn);
Bin Song
(songbin@sxbqeh.com.cn)

Citation: Chen Y, Ma Y_H, Hu Z, Liu Q, Song B, Hao H. Unsaturated fatty acid metabolism and FADS2 as candidate factors in triple-negative breast cancer. *Cancer Plus*.
doi: 10.36922/CP026340053

Received: August 23, 2026

Revised: September 27, 2026

Accepted: October 10, 2026

Published online: November 2, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Triplenegative breast cancer (TNBC) accounts for approximately 15%–20% of all breast cancers and is characterized by the absence of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2).[1,2] It is the most aggressive subtype, with limited targeted therapies and poor prognosis.[3] Metabolic reprogramming is now recognized as a hallmark of cancer, and altered circulating metabolite levels have been associated with the risk of various tumors.[4] In TNBC, both cancer-cell

metaprograms and the tumor microenvironment (TME) exhibit extensive reprogramming relative to normal breast tissues, contributing to intra-tumoral heterogeneity.[5] However, observational studies are prone to confounding and reverse causation, and the potential association between metabolites and ER-negative/TNBC biology remains to be rigorously evaluated.

Mendelian randomization (MR) uses genetic variants as instrumental variables (IVs), effectively reducing confounding bias, and has been widely applied to explore causal links between biomarkers and diseases.[6] Moreover, metabolites often exert their effects on tumors through downstream gene expression regulation[7]; therefore, integrating metabolite, transcriptomic, and single-cell data can help generate hypotheses about potential mechanisms.

The TME plays a critical role in cancer progression, and metabolic crosstalk among cancer-associated fibroblasts, immune cells, and tumor cells is increasingly recognized.[8] Recent advances in single-cell and spatial transcriptomics have revolutionized the understanding of tumor heterogeneity.[9] To our knowledge, few studies have integrated circulating metabolite data with single-cell-level candidate gene prioritization in TNBC.

This study aims to explore candidate associations between circulating metabolites and ER-negative/TNBC biology, identify candidate metabolic pathways, and prioritize candidate genes potentially regulated by these metabolites using multi-step bioinformatics analyses. Specifically, we focus on dissecting the expression heterogeneity and potential functions of key genes at single-cell resolution. Through this multidimensional analytical approach, our study generates hypotheses about potential links between circulating metabolites and ER-negative/TNBC biology and provides exploratory insights into the gene regulatory mechanisms at the single-cell level, potentially offering novel theoretical bases and targets for the diagnosis and treatment of TNBC.

2. Materials and methods

2.1. Mendelian randomization analysis of metabolites and estrogen receptor-negative breast cancer (proxy for triple-negative breast cancer)

Genetic association data for circulating metabolites were obtained from a large-scale plasma metabolome genome-wide association study (GWAS) covering approximately 1,400 metabolites. Independent genetic variants significantly associated with metabolite levels were selected as IVs using the following criteria: Genome-wide significance ($p < 5 \times 10^{-8}$), independence ($r^2 < 0.001$, linkage disequilibrium clumping window of 10,000 kb). The *F*-statistic was calculated, and variants with $F > 10$ were retained to avoid weak instrument bias. These IVs were used for exploratory two-sample MR analysis.

Outcome data for ER-negative breast cancer were derived from a GWAS of ER$^-$ breast cancer. Considering the limited availability of a dedicated large-scale GWAS specifically for TNBC (defined as ER$^-$/PR$^-$/HER2$^-$), we utilized the largest publicly available GWAS of ER$^-$ breast cancer as a proxy outcome. Given that TNBC constitutes a substantial proportion of clinically aggressive ER$^-$ breast cancers, this proxy approach has been used in previous two-sample MR studies; however, ER$^-$ breast cancer is not equivalent to TNBC because it includes HER2$^+$ disease**.** We explicitly acknowledge the potential outcome misclassification introduced by HER2$^+$ tumors. The direction of this bias is not predictable, and we interpret our findings with this caveat (Section 4). Allele harmonization was performed using the TwoSampleMR R package (v0.5.6). The inverse variance weighted (IVW) method was used as the primary estimate, with the weighted median method for sensitivity analysis and the MR–Egger intercept test to evaluate horizontal pleiotropy. The false discovery rate (FDR) was considered for multiple testing correction. However, because the downstream prioritization used nominal IVW $p < 0.05$, the MR findings should be regarded as exploratory. A metabolite was considered a candidate association when the IVW $p < 0.05$, the weighted median effect direction was consistent, and the MR–Egger pleiotropy $p > 0.05$. No definitive causal inference was made. Multivariable MR, colocalization, and MR–Pleiotropy Residual Sum and Outlier were not performed; these are important future directions.

2.2. Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis

Candidate metabolites were mapped where possible to Kyoto Encyclopedia of Genes and Genomes compound IDs. Enrichment analysis was performed using the clusterProfiler R package (v4.4.0) with all mappable metabolites as background. The significance threshold was set at adjusted $p < 0.05$. Candidate metabolites were classified and visualized using bar plots, circular plots, and forest plots. Enrichment results were considered exploratory.

2.3. Computational target prediction for lipid-related candidate metabolites

Based on the candidate metabolite associations from the MR analysis, candidate metabolites in the unsaturated fatty acid pathway (e.g., linoleic acid and arachidonic acid) were selected, and their Simplified Molecular Input Line Entry System or structural formulas were obtained.

Protein targets were computationally predicted using the SwissTargetPrediction and Similarity Ensemble Approach databases. Pathway-related genes were retrieved from GeneCards using the keywords "unsaturated fatty acid metabolism" and "arachidonic acid metabolism." The intersection of computationally predicted targets and pathway genes was taken as the candidate target set (Gene Set A). These predictions are hypothesis-generating and do not establish metabolite-driven gene regulation.

2.4. Triple-negative breast cancer differential expression analysis and weighted gene co-expression network analysis

Three expression datasets (GSE38959: 30 TNBC vs. 13 controls; GSE45827: 41 TNBC vs. 11 controls; GSE65194: 41 TNBC vs. 11 controls) were downloaded from the Gene Expression Omnibus (GEO) database. These datasets were generated on different platforms and may differ in control composition. Normalization and differential expression analysis were performed using the limma package (v3.52.0) with thresholds of adj.P.Val < 0.05 and |log2FC| > 0.585. This threshold was selected to capture moderate but potentially biologically relevant expression changes. Genes consistently differentially expressed across the three datasets were defined as Gene Set B; this intersection approach is exploratory and may be influenced by sample size and platform differences.

Weighted gene co-expression network analysis (WGCNA; WGCNA package v1.72) was performed on the GSE65194 dataset. The soft threshold was set to 4 (ensuring $R^2 > 0.9$ and high mean connectivity), and a weighted co-expression network was constructed to identify modules most positively correlated with the TNBC phenotype in this dataset (Gene Set C). The intersection of Gene Set B and Gene Set C yielded candidate TNBC-related genes (Gene Set D). The intersection of Gene Set A and Gene Set D gave the candidate gene set (Gene Set E). The large overlap between Gene Set B and Gene Set C may reflect module size and thresholding and requires independent validation.

2.5. Gene functional enrichment analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses were performed using clusterProfiler, with significance defined as $p < 0.05$ and $q < 0.05$. These enrichment results are exploratory and do not establish causal regulation. Visualization was performed using ggplot2 (v3.3.6).

2.6. Machine learning to prioritize candidate genes

RNA-sequencing expression data and corresponding clinical information for TNBC samples were downloaded from The Cancer Genome Atlas (TCGA) database. A total of 113 TNBC samples were included. The number of survival events and follow-up distribution were not sufficient for external validation. The expression matrix was matched with Gene Set E.

First, least absolute shrinkage and selection operator (LASSO) Cox regression (glmnet package v4.1) was used for candidate feature selection, with the optimal λ determined by ten-fold cross-validation. Subsequently, a random survival forest (RSF) model (randomForestSRC package v3.1) was used to evaluate variable importance. The intersection of non-zero coefficient genes from LASSO and the top 20 important genes from RSF was defined as the final candidate genes. Because both algorithms were applied to the same cohort without external validation, these genes should not be considered a validated prognostic signature.

2.7. Single-cell exploration of candidate gene expression heterogeneity

The TNBC single-cell dataset GSE176078 (eight tumor samples) was obtained from the GEO database. The number of cells per patient and patient-level clinical characteristics were not fully modeled. Quality control was performed using Seurat (v4.1.1): Cells with 200–7,000 detected genes and a mitochondrial gene proportion below 20% were retained. Principal component analysis dimensionality reduction (top 20 principal components), Louvain clustering (resolution 0.5), and Uniform Manifold Approximation and Projection (UMAP) visualization were performed. Cell types were annotated based on canonical marker genes: epithelial cells (*EPCAM*, *KRT8*), T cells (*CD3D*), and myeloid cells (*CD68*). Epithelial cells were annotated using canonical markers and are referred to as presumed malignant epithelial cells; malignant status was not confirmed by copy-number variation inference, and the remaining cells were defined as non-malignant cells. Expression differences of candidate genes between presumed malignant and non-malignant cells were compared using the Wilcoxon rank-sum test with FDR correction. These cell-level tests do not account for patient-level effects and may overstate significance.

2.8. Transcriptional-state comparison of *FADS2*-detected and *FADS2*-undetected epithelial cells

Within the epithelial cell population, cells were divided into *FADS2*-detected (expression count > 0) and *FADS2*-undetected (expression count = 0) groups. Differentially expressed genes between the two groups were identified using the FindMarkers function in Seurat (Wilcoxon rank-sum test, FDR < 0.05). This is an observational comparison and not a gene knockout. Zero counts may reflect dropout

or limited sequencing depth. This was followed by GO biological process enrichment analysis using clusterProfiler. These differences may also be influenced by library size, cell cycle, patient effects, and malignant subclone.

2.9. Statistical analysis

All analyses were exploratory. Continuous variables were compared using the Mann–Whitney *U* test, and categorical variables were compared using the χ^2 test. Survival analysis was performed using the Kaplan–Meier method with the log-rank test. All statistical analyses were performed with R software (v4.0.5, R Foundation for Statistical Computing, Austria), and two-sided $p < 0.05$ was considered statistically significant. No correction for patient-level pseudoreplication in single-cell comparisons was applied.

3. Results

3.1. Candidate metabolite associations with estrogen receptor-negative breast cancer risk identified with Mendelian randomization

Using the ER⁻ breast cancer GWAS as a proxy for TNBC risk, 42 metabolites were initially identified via IVW analysis as being nominally associated with ER⁻ breast cancer risk, used as a proxy for TNBC ($p < 0.05$). After requiring consistency of effect direction with the weighted median method and excluding metabolites with horizontal pleiotropy (MR–Egger $p > 0.05$), 38 candidate metabolites were retained (Figure 1A). These are exploratory, nominal associations and should not be interpreted as definitive causal effects. These metabolites included lipids, amino acid-related metabolites, and energy-related molecules. Polyunsaturated fatty acid metabolites (e.g., docosapentaenoate, docosatrienoate, and various arachidonoyl-containing phospholipids) accounted for 33% of the candidate metabolites. The effect sizes and confidence intervals of the top 20 metabolites are shown in a forest plot (Figure 1B), and scatter plots were further drawn for the top five metabolites (Figure 1C) to show the relationship between the genetic IV effect and the MR fitting results.

3.2. Unsaturated fatty acid and arachidonic acid metabolism highlighted in pathway enrichment of candidate metabolites

After classifying the 38 candidate metabolites obtained through final screening, it was found that they were mainly distributed in lipid metabolism categories such as polyunsaturated fatty acids/unsaturated fatty acids, arachidonic acid-related lipids, glycerophospholipids, sphingolipids, and acylcarnitines/fatty acid oxidation, as well as some tryptophan metabolism, amino acids/peptides, bile acids, carbohydrate metabolism, and microbial population/externally derived metabolism-related metabolites (Figure 2A, B). Overall, lipid metabolism-related molecules accounted for a relatively high proportion in the candidate metabolites, suggesting that abnormal lipid metabolism may be one feature of metabolic changes in this context. Further analysis from the perspective of functional modules revealed that these candidate metabolites mainly clustered in modules such as unsaturated fatty acid remodeling, arachidonic acid signal transduction, sphingolipid signaling pathways, and fatty acid β-oxidation. These modules are all closely related to membrane lipid remodeling, inflammatory signal regulation, cell proliferation, and energy supply, suggesting that lipid metabolism reprogramming may be relevant in this context. At the same time, different metabolites in terms of effect directions include both risk-related molecules and protective-related molecules, indicating that the metabolic network's regulation of tumor risk may have bidirectionality and complexity (Figure 2C). From representative metabolites, docosapentaenoate,

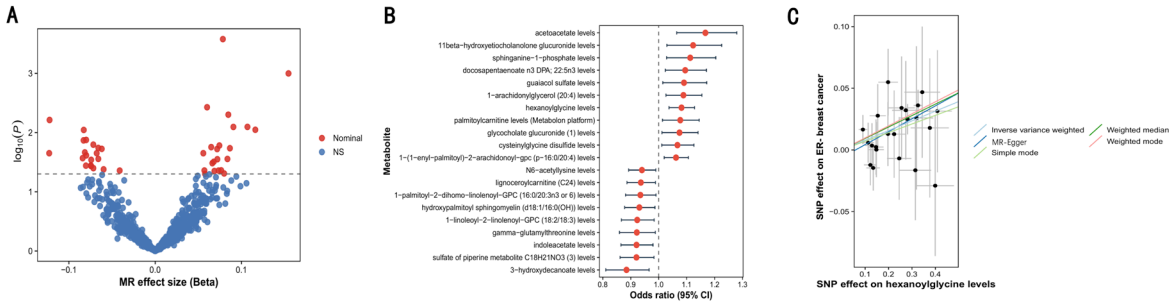


Figure 1. Results of MR analysis. (A) Volcano plot showing the effect size (odds ratio) and significance (–log10 *p*) for all metabolites. Red dots: Metabolites with inverse-variance-weighted $p < 0.05$ and consistent direction with the weighted median. These are nominal, exploratory associations. (B) Forest plot of the top 20 candidate metabolites. (C) Scatter plots of the top five metabolites.
Abbreviations: CI: Confidence interval; ER: Estrogen receptor; MR: Mendelian randomization; SNP: Single-nucleotide polymorphism.

docosatrienoate, and various arachidonoyl- or linoleoyl-related lipid molecules were all classified into the core category of lipid metabolism, further highlighting unsaturated fatty acid metabolism and arachidonic acid metabolism among the candidate metabolites. Overall, this analysis provided hypothesis-generating functional context for the candidate metabolites obtained through MR screening and helped guide subsequent candidate gene prioritization.

3.3. Computational prioritization of candidate targets associated with unsaturated fatty acids

Combining computational target prediction (SwissTargetPrediction and Similarity Ensemble Approach) with pathway gene retrieval (GeneCards) yielded 282 computationally predicted candidate targets (Gene Set A) highly relevant to unsaturated fatty acid metabolism, including key lipid metabolic enzymes, fatty acid transporters, and signaling molecules (Figure 3).

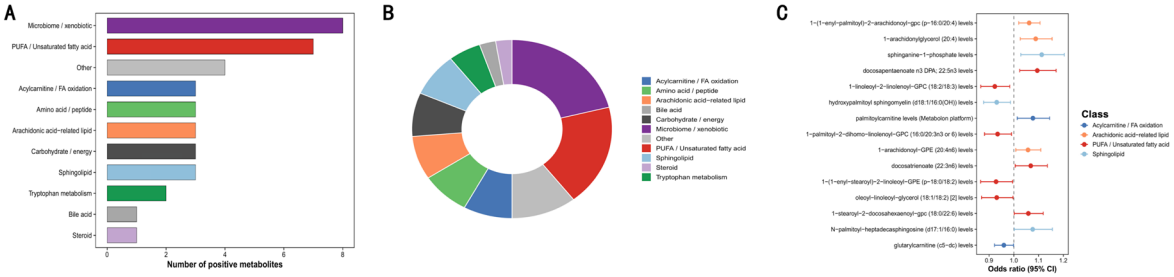


Figure 2. KEGG enrichment and classification of positive metabolites. (A) Bar plot of KEGG pathways enriched by candidate metabolites. (B) Circular plot of candidate metabolite categories. (C) Summary of core lipid metabolites.
Abbreviations: CI: Confidence interval; KEGG: Kyoto Encyclopedia of Genes and Genomes.

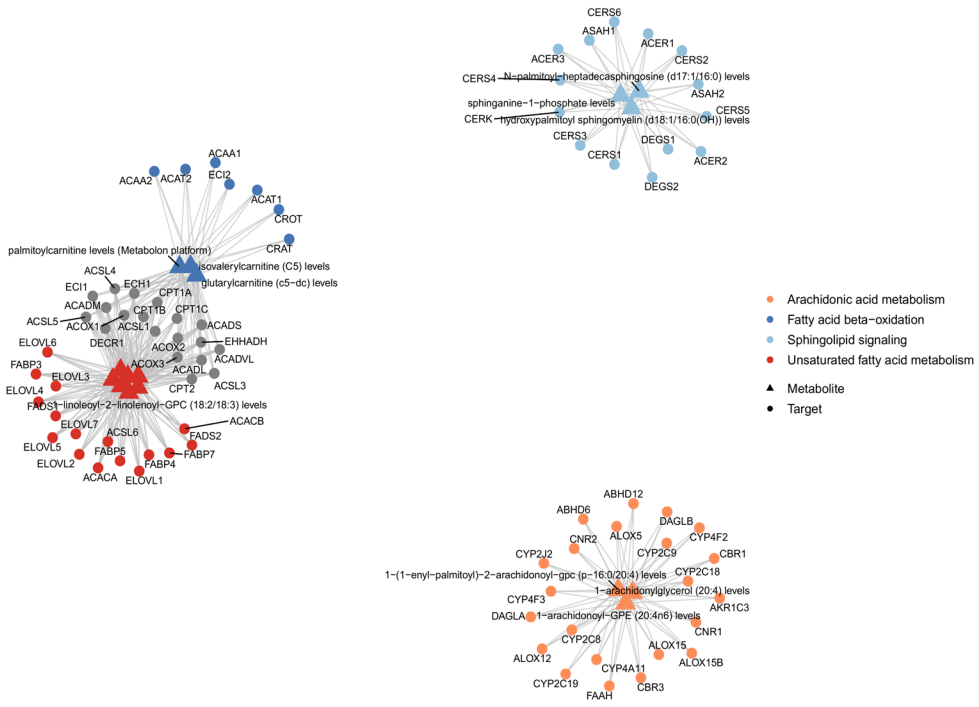


Figure 3. Metabolite-target network. Circles represent metabolites, diamonds represent predicted targets. Edge thickness indicates confidence score. Targets are computationally predicted and do not demonstrate regulation.

3.4. Identification of candidate triple-negative breast cancer-related modules using transcriptomic profiling and WGCNA

To explore the role of unsaturated fatty acid metabolism in TNBC, we conducted differential expression analysis using the limma package on three microarray datasets (GSE38959, GSE45827, and GSE65194). We then generated the volcano plots and heatmaps for visualization (Figure 4A). We further extracted the common differentially expressed genes from these three datasets to obtain 1,578 common differentially expressed genes, defined as Gene Set B (Figure 4B).

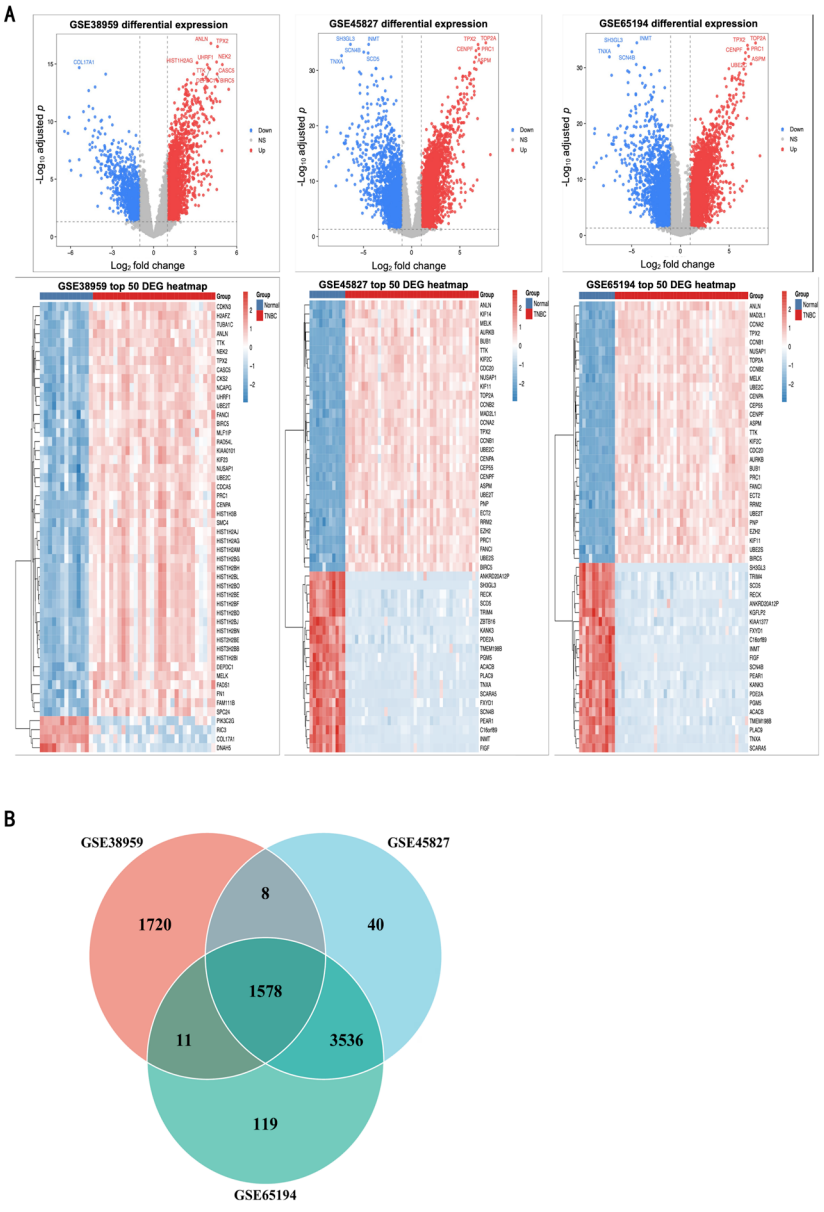


Figure 4. Differential expression analysis. (A) Volcano plot and heatmaps of three Gene Expression Omnibus datasets (GSE38959, GSE45827, and GSE65194). (B) Venn diagram of the intersection of differentially expressed genes.
Abbreviations: DEG: Differentially expressed gene; NS: Not significant; TNBC: Triple-negative breast cancer.

To further clarify the potential mechanisms and gene modules in TNBC, we performed batch normalization on the disease samples of dataset GSE65194. When R^2 > 0.9 and the average connectivity was high, we set the soft threshold to 4. After clustering with a high limit of 0.25 to merge strongly correlated modules, the initialized and merged modules were finally displayed under the clustering tree. WGCNA identified a module (MEblue) most positively correlated with TNBC (Figure 5A–C). The intersection of this module (Gene Set C) with Gene Set B gave candidate TNBC-related targets (Gene Set D, n = 1,204). The intersection of Gene Set A (n = 282) and Gene Set D (n = 1,204) yielded 39 candidate genes (Gene Set E). Functional enrichment of Gene Set E showed enrichment in mammary gland development, epithelial morphogenesis, PI3K–Akt, and p53 signaling pathways (Figure 5D). These results are exploratory and do not establish causal regulation.

3.5. Six candidate genes prioritized in machine learning

In the TCGA–TNBC cohort (113 samples), LASSO Cox regression selected 9 non-zero coefficient genes at the optimal λ.min = 0.0222 (Figure 6A, B). RSF variable importance ranking placed *FADS2*, *MGST3*, and *KRAS* among the top 20 (Figure 6C). The intersection of LASSO-selected genes (9 genes) and the top 20 RSF genes yielded six candidate genes: *KRAS*, *MGST3*, *FADS2*, *UGT8*, *IDO1*, and *SRC* (Figure 6D). These genes were selected in the same cohort without external validation and should not be considered a validated prognostic signature.

3.6. Heterogeneous *FADS2* expression in presumed malignant epithelial cells

After quality control, 99,086 high-quality single cells were obtained. UMAP clustering identified major cell types including epithelial cells (26,022 cells), T/natural killer cells, myeloid cells, and fibroblasts (Figure 7A, B). Epithelial cells were annotated using canonical markers and are referred to as presumed malignant epithelial cells; malignant status was not confirmed by copy-number variation inference. Comparison of candidate gene expression between presumed malignant and non-malignant cells showed that *FADS2* and *MGST3* expression was higher in malignant cells, with *FADS2* showing a log2FC of 2.53 (FDR < 0.05, Figure 7C). Feature plots revealed heterogeneous expression of *FADS2* in tumor cells.

3.7. Association between *FADS2* expression and sterol biosynthesis/cell-cycle programs

Within epithelial cells, *FADS2*-detected cells (5,775 cells) and *FADS2*-undetected cells (20,247 cells) were

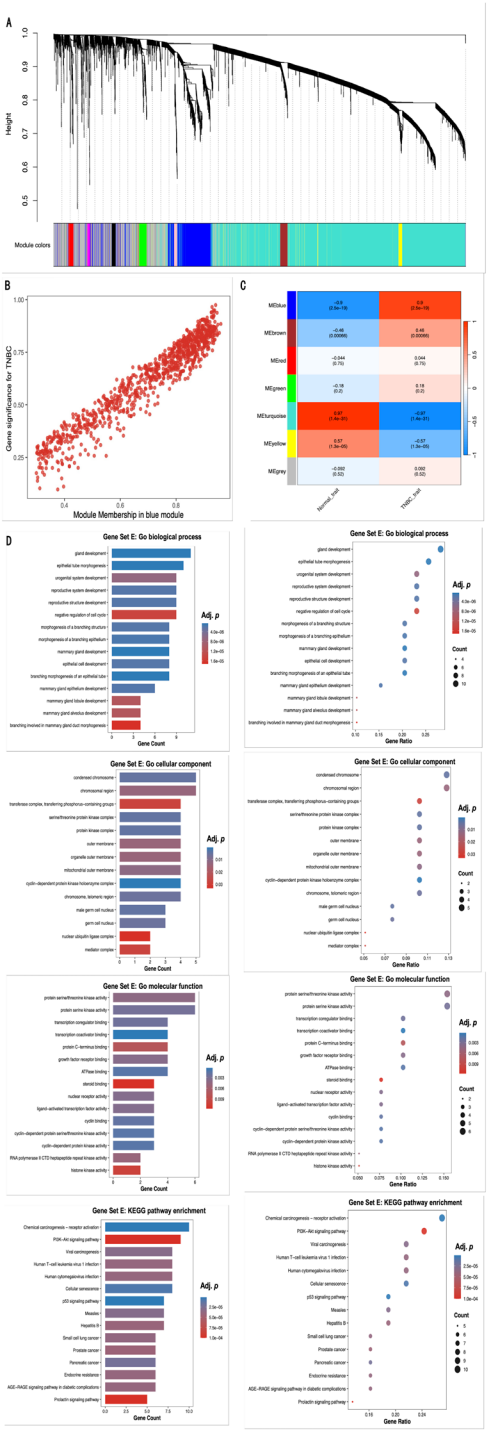


Figure 5. Weighted gene co-expression network analysis and integration with predicted targets. (A) Cluster dendrogram of genes with module colors. (B) Module membership in the blue module. (C) Module-trait correlations (red: positive correlation with triple-negative breast cancer [TNBC]). (D) Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment of Gene Set E.

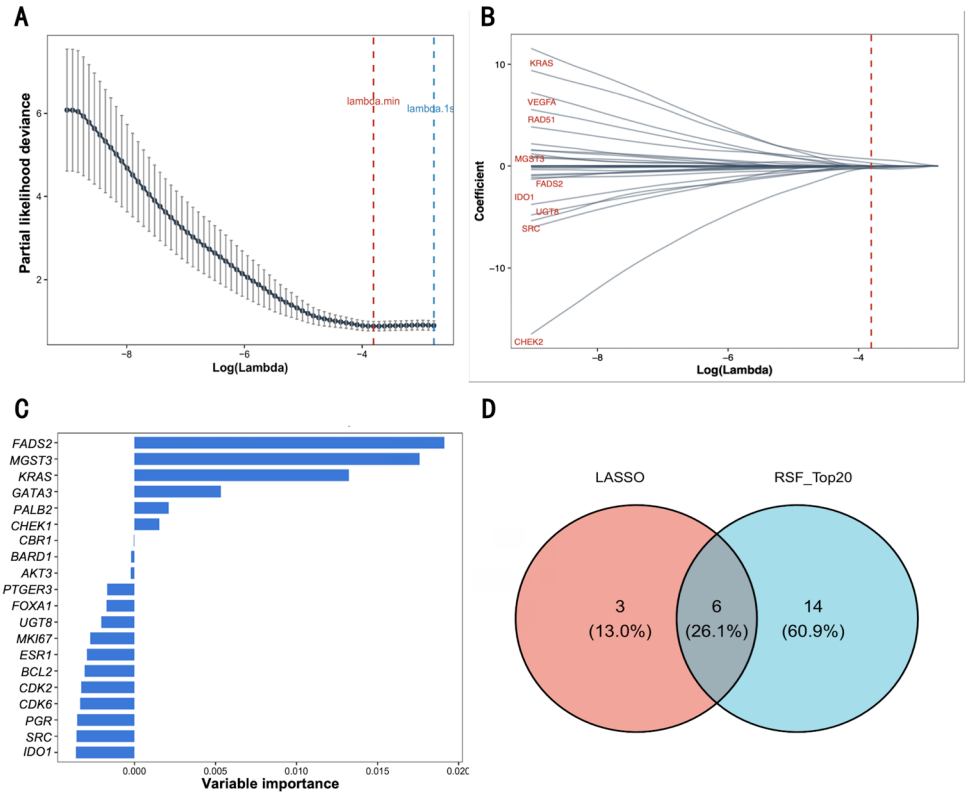


Figure 6. Machine learning for candidate gene prioritization. (A) Least absolute shrinkage and selection operator (LASSO) cross-validation curves (partial likelihood deviance). (B) LASSO coefficient paths as λ varies. (C) Random survival forest (RSF) variable importance plot. (D) Venn diagram of LASSO-selected and RSF-top 20 genes. No external validation was performed.

well separated on UMAP (Figure 8A). Differential expression analysis identified 1,203 upregulated and 845 downregulated genes in the *FADS2*-detected group (FDR < 0.05, Figure 8B). GO enrichment showed that genes associated with *FADS2* expression were enriched in cholesterol biosynthetic process, sterol metabolic process, steroid biosynthetic process, and acyl-coenzyme A metabolic process (Figure 8C). This is an observational comparison and not a gene knockout; zero counts may reflect dropout or limited sequencing depth. These differences may also be influenced by library size, cell cycle, patient effects, and malignant subclone. Additionally, some genes were enriched in chromosome segregation and nuclear division (cell cycle) processes.

4. Discussion

In this study, an integrative analysis of independent public datasets starting from MR and progressing through target prediction, differential expression, WGCNA, machine learning, and single-cell exploration was used to explore candidate associations between circulating metabolites and ER⁻/TNBC biology. Unsaturated fatty acid metabolism was identified as a candidate pathway. Six candidate genes were prioritized, among which *FADS2* expression was higher in presumed malignant epithelial cells at the single-cell level and was associated with lipid metabolic reprogramming.

The MR findings are consistent with emerging evidence linking lipid metabolism to breast cancer risk.[10] The enrichment of docosapentaenoate, docosatrienoate, and arachidonoyl-containing phospholipids suggests a potential role for long-chain polyunsaturated fatty acids and their remodeling pathways in ER⁻/TNBC biology.[11] This is biologically plausible because these fatty acids are precursors for bioactive lipids (e.g., eicosanoids) that modulate inflammation, cell proliferation, and angiogenesis.[12]

Although our machine learning pipeline prioritized six candidate genes, we prioritized *FADS2* (fatty acid desaturase 2) for further computational exploration. Unlike *KRAS* or *SRC*, which primarily transduce extracellular signals, or *IDO1*, which modulates immune checkpoints,

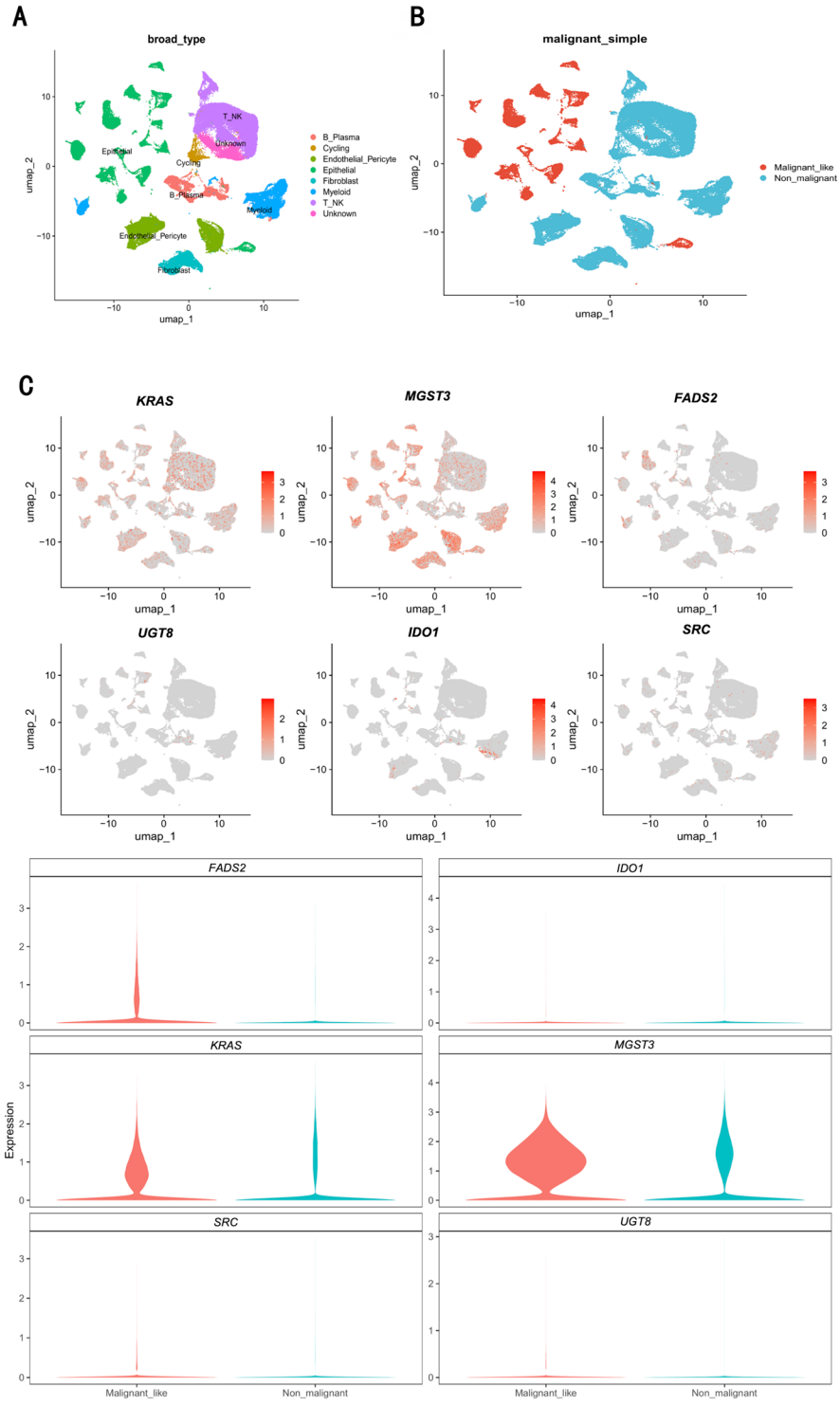


Figure 7. Single-cell landscape and candidate gene expression. (A) Uniform manifold approximation and projection plot of major cell types. (B) Distribution of presumed malignant (epithelial) vs. non-malignant cells. (C) FeaturePlot of six key genes and violin plots comparing expression of six candidate genes between presumed malignant and non-malignant cells.

FADS2 encodes a rate-limiting enzyme, delta-6 desaturase, which catalyzes the metabolism of unsaturated fatty acids.[13] Notably, these very fatty acids were identified as nominally associated with ER– breast cancer risk, used as a proxy for TNBC by our MR analysis. This raises the hypothesis that *FADS2* may link systemic lipid alterations to local malignant transformation. *FADS2* encodes a delta-6 desaturase that catalyzes the rate-limiting step in the conversion of linoleic acid and α-linolenic acid to longer chain polyunsaturated fatty acids.[13,14] Overexpression of *FADS2* has been reported in several cancers and is linked to poor prognosis.[15–18] However, this does not establish a prognostic role in the present cohort. The single-cell data showed that *FADS2* is not uniformly expressed in all presumed malignant cells but rather marks a subset of epithelial cells with high metabolic activity. This heterogeneity may reflect distinct metabolic states within the tumor, similar to the cancer-associated fibroblast heterogeneity described in gastric cancer.[19] The enrichment of cholesterol biosynthesis and sterol metabolism pathways in *FADS2*-expressing cells raises the possibility that *FADS2* may be associated with fatty acid desaturation and sterol synthesis, both of which are essential for membrane biogenesis and signaling in rapidly proliferating cancer cells.[17,20,21]

The other five candidate genes (*KRAS*, *MGST3*, *UGT8*, *IDO1*, *SRC*) also have plausible connections to TNBC. *KRAS* mutations are acquired or selected under endocrine therapy pressure in breast cancer, contributing to resistance.[22] In TNBC, mutant p53 protects tumor cells from ferroptosis by upregulating *Mgst3* via nuclear factor erythroid 2-related factor 2. *Mgst3* encodes a glutathione-dependent peroxidase that detoxifies lipid peroxides, thereby contributing to the survival of TNBC cells.[23] *UGT8* is a molecular marker of breast cancer malignancy and lung metastases,[24] and its inhibition suppresses basal-like breast cancer progression by attenuating the sulfatide-αVβ5 axis.[25] *IDO1* promoter methylation regulates *IDO1* expression in breast cancer, contributing to the development of the TME.[26] *SRC* is a non-receptor tyrosine kinase frequently activated in breast cancer.[27] It has been demonstrated that the oncogenic overexpression or activation of *SRC* plays critical roles in various aspects of breast cancer progression, including tumor initiation, growth, metastasis, drug resistance, and the regulation of stemness.[28] Their co-selection by both LASSO and RSF suggests potential relevance, but independent validation is required.

Our single-cell transcriptomic analysis further revealed pronounced expression heterogeneity of *FADS2* within the presumed malignant epithelial cell compartment, a finding that may provide hypothesis-generating insights into TNBC heterogeneity. TNBC is renowned for its extreme inter- and intra-tumor heterogeneity. According to established molecular frameworks, such as the Fudan Classification,[29] TNBC can be categorized into distinct subtypes, among which the Luminal Androgen Receptor (LAR) subtype is characterized by a prominent enrichment of lipid and steroid metabolism pathways.[30] The distinct *FADS2*-high epithelial cell cluster identified in our single-cell RNA sequencing analysis may resemble a lipid-driven phenotype similar to the LAR subtype; this requires formal androgen receptor (AR)/LAR-signature scoring, which is defined by active AR signaling and steroid metabolism addiction.[31] By utilizing *FADS2* expression to map metabolic subpopulations, our findings suggest a hypothesis that targeting fatty acid desaturation could be explored in future studies.

A compelling finding of our MR analysis is that systemic alterations in circulating unsaturated fatty acids and arachidonic acid metabolites may be associated with ER⁻/TNBC susceptibility. Systemic metabolic predispositions do not exist in isolation; instead, they maintain a profound crosstalk with the local TME.[32] Elevated circulating lipids may cross the endothelial barrier, serving as abundant raw

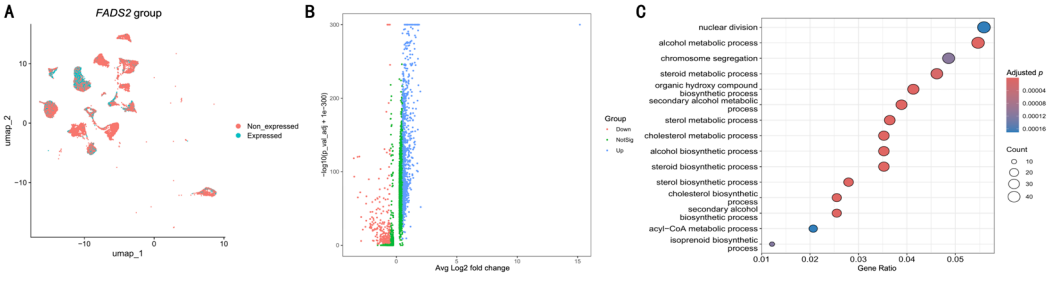


Figure 8. Transcriptional-state comparison of *FADS2*-detected and *FADS2*-undetected epithelial cells. (A) Uniform manifold approximation and projection of *FADS2*-detected vs. *FADS2*-undetected cells within the epithelial population. (B) Volcano plot of differential expression between *FADS2*-detected and *FADS2*-undetected cells. (C) Gene Ontology biological process enrichment of genes associated with *FADS2* expression. This is not a gene knockout.

materials that fuel the energy demands and membrane synthesis of lipid-addicted malignant cells. This potential systemic-to-local metabolic relationship may link patient-level metabolic traits and intra-tumoral programming,[33] offering a wider lens to understand how peripheral metabolic health shapes breast cancer initiation. These circulating lipids likely serve as the substrate pool for upregulated *FADS2* in presumed malignant cells, providing the raw material for the downstream cholesterol/sterol biosynthesis observed in our transcriptional-state comparison.

FADS2-high presumed malignant cells show enhanced cholesterol biosynthesis. Cholesterol is the essential precursor for all steroid hormones, including androgens,[34] which act as endogenous ligands for the AR. AR signaling is the core driver of the LAR subtype of TNBC.[31] Therefore, increased sterol synthesis may potentially contribute to AR activation through ligand availability. Beyond providing precursors, steroidogenic enzymes directly modulate AR function. For example, *AKR1C2* and *DHRS11* are upregulated in LAR tumors.[35,36] Both enzymes have been shown to regulate AR protein expression and nuclear localization. Cholesterol accumulation also remodels membrane lipid rafts, which are critical for AR membrane anchoring and efficient signal transduction.[37] This finding raises the intriguing possibility that *FADS2*-high presumed malignant cells, through enhanced cholesterol biosynthesis, may create a permissive metabolic environment for steroidogenesis, potentially relevant to the LAR subtype. This hypothesis, however, requires direct experimental validation. Nevertheless, we acknowledge that AR itself is a well-established transcriptional regulator of lipid metabolism. Thus, the hypothesized *FADS2*-sterol-AR relationship may involve bidirectional interactions, and our current computational framework cannot fully disentangle cause from consequence.

Intriguingly, the enrichment of cholesterol biosynthesis in *FADS2*-high presumed malignant cells extends beyond mere membrane biogenesis. In the context of TNBC, dysregulated sterol metabolism is tightly linked to the activation of the sterol regulatory element-binding protein signaling pathway and the synthesis of steroid hormones,[38,39] particularly in the LAR subtype. This metabolic rewiring not only supports rapid proliferation but may also confer resistance to ferroptosis by altering membrane lipid saturation status.[40] This hypothesis requires experimental validation, and it may be relevant to the poor prognosis associated with *FADS2* expression.

Compared to previous TNBC metabolomics studies, this work has several strengths. First, MR was used to explore genetically proxied associations, reducing confounding and reverse causation—an approach similar to recent studies that integrated MR with multi-omics. Second, multi-omics data (metabolomics, transcriptomics, single-cell) were integrated to build an integrative hypothesis-generating framework from metabolite to pathway to candidate gene. Third, candidate gene expression was explored at single-cell resolution, revealing intratumoral heterogeneity, analogous to the detailed metabolic molecular subtyping performed in nonalcoholic fatty liver disease and hepatocellular carcinoma.

Several limitations must be acknowledged. First and foremost, the MR estimates should be interpreted within the context of outcome definition. In the absence of a TNBC-specific GWAS, we relied on ER^- breast cancer GWAS as a proxy. We recognize that ER^- tumors are biologically heterogeneous and include $HER2^+$ subtypes, which may not fully recapitulate the metabolic etiology of TNBC. The direction of this misclassification bias is not predictable. Nevertheless, dedicated TNBC GWAS are urgently needed to validate our findings. Second, the direct regulatory relationship between unsaturated fatty acid metabolites and *FADS2* expression has not been experimentally validated. Multivariable MR, colocalization, and comprehensive pleiotropy analyses were not performed. Third, in the single-cell analysis, epithelial cells were designated as presumed malignant cells without additional evidence such as copy number variation inference. Cell-level single-cell tests may be affected by pseudoreplication; patient-aware pseudobulk or mixed-effects models are needed. Fourth, all conclusions are derived from computational analyses and lack in vitro or in vivo functional support. Fifth, the "virtual knockout" analysis is an observational comparison, not a gene perturbation. Zero counts may reflect dropout. Sixth, the LAR subtype hypothesis was not tested by AR/LAR-signature scoring. Seventh, no external validation of the candidate gene set was performed. More broadly, circulating biomarkers do not always show measurable associations with cancer risk; for example, blood zinc levels were not significantly associated with breast cancer incidence in a prospective cohort of *BRCA1* mutation carriers.[41]

Future studies should employ summary-data-based MR or expression quantitative trait locus colocalization to further explore the potential relationship between metabolites and *FADS2*. Functional experiments, such as *FADS2* knockdown or overexpression in TNBC cell lines (e.g., MDA-MB-231 and BT-549), combined with lipidomics and proliferation/migration assays, are needed to elucidate the molecular mechanism. Additionally, spatial transcriptomics could be used to map the spatial distribution of *FADS2*-expressing cells within the TME

and their interactions with cancer-associated fibroblasts and immune cells, as performed in the gastric cancer study.

5. Conclusion

This exploratory integrative analysis nominates abnormal unsaturated fatty acid metabolism and *FADS2* as candidate factors in ER⁻/TNBC biology. *FADS2* expression was higher in a subset of presumed malignant epithelial cells and was associated with transcriptional programs related to sterol metabolism and cell-cycle activity. These findings identify *FADS2* as a candidate biomarker and potential target requiring TNBC-specific genetic analyses, patient-aware single-cell validation, independent prognostic cohorts, and experimental perturbation.

Acknowledgments

None.

Funding

This work was supported by the Shanxi Bethune Hospital Research Initiation Grant 2025RC001, Research Project Supported by Shanxi Scholarship Council of China 2025-224, and 2025 Sanjin Talents Program for Young Top-notch Talents in Medical and Health Field (Grant No. SJYC2025528).

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Bin Song, He Hao
Formal analysis: Ying Chen, Yu-Hui Ma
Investigation: Zhi-hui Hu, Qi Liu
Methodology: Ying Chen, Yu-Hui Ma
Supervision: Bin Song and He Hao
Writing – original draft: Ying Chen
Writing – review & editing: Bin Song, He Hao

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The metabolomic GWAS summary statistics were obtained from the Canadian Longitudinal Study on Aging cohort GWAS. ER⁻ breast cancer GWAS data were sourced from the Integrative Epidemiology Unit OpenGWAS database (ebi-a-GCST005843) and the GWAS Catalog (GCST005076, GCST003842, GCST003845, GCST001930). Bulk transcriptomic (GSE38959, GSE45827, GSE65194) and single-cell (GSE176078) datasets are available from GEO (https://www.ncbi.nlm.nih.gov/geo/). TNBC RNA-sequencing and clinical data are from TCGA–breast invasive carcinoma via the Genomic Data Commons Data Portal (https://portal.gdc.cancer.gov/), with controlled access requiring Database of Genotypes and Phenotypes authorization.

References

1. Li Y, Zhang H, Merkher Y, *et al.* Recent advances in therapeutic strategies for triple-negative breast cancer. *J Hematol Oncol*. 2022;15(1):121.

   doi: 10.1186/s13045-022-01341-0

2. Asleh K, Riaz N, Nielsen TO. Heterogeneity of triple negative breast cancer: current advances in subtyping and treatment implications. *J Exp Clin Cancer Res*. 2022;41(1):265.

   doi: 10.1186/s13046-022-02476-1

3. Derakhshan F, Reis-Filho JS. Pathogenesis of triple-negative breast cancer. *Annu Rev Pathol*. 2022;17(1):181-204.

   doi: 10.1146/annurev-pathol-042420-093238

4. Mao Y, Xia Z, Xia W, Jiang P. Metabolic reprogramming, sensing, and cancer therapy. *Cell Rep*. 2024;43(12):115064.

   doi: 10.1016/j.celrep.2024.115064

5. Yan Y, Lin Y, Kumar T, *et al.* Ecotypes of triple-negative breast cancer in response to chemotherapy. *Nature*. 2026;654(8120):1088-1097.

   doi: 10.1038/s41586-026-10469-9

6. Ference BA, Holmes MV, Smith GD. Using Mendelian randomization to improve the design of randomized trials. *Cold Spring Harb Perspect Med*. 2021;11(7):a040980.

   doi: 10.1101/cshperspect.a040980

7. Li M, Yang Y, Xiong L, Jiang P, Wang J, Li C. Metabolism, metabolites, and macrophages in cancer. *J Hematol Oncol*. 2023;16(1):80.

   doi: 10.1186/s13045-023-01478-6

8. de Visser KE, Joyce JA. The evolving tumor microenvironment: from cancer initiation to metastatic outgrowth. *Cancer Cell*. 2023;41(3):374-403.

   doi: 10.1016/j.ccell.2023.02.016

9. Prakrithi P, Vo T, Xiong Z, *et al.* Unraveling lncRNA diversity at a single cell resolution and in a spatial context across different cancer types. *Nat Methods*. 2026;23(6):1236-1249.

   doi: 10.1038/s41592-026-03071-4

10. Lorito N, Subbiani A, Smiriglia A, *et al.* FADS1/2 control lipid metabolism and ferroptosis susceptibility in triple-

negative breast cancer. *EMBO Mol Med*. 2024;16(7):1533-1559.

doi: 10.1038/s44321-024-00090-6

11. Yu L, Liebenberg K, Shen Y, *et al.* Tumor-derived arachidonic acid reprograms neutrophils to promote immune suppression and therapy resistance in triple-negative breast cancer. *Immunity*. 2025;58(4):909-925.e7.

doi: 10.1016/j.immuni.2025.03.002

12. Currie E, Schulze A, Zechner R, Walther TC, Farese RV. Cellular fatty acid metabolism and cancer. *Cell Metab*. 2013;18(2):153-161.

doi: 10.1016/j.cmet.2013.05.017

13. Wang C, Hucik B, Sarr O, *et al.* Delta-6 desaturase (Fads2) deficiency alters triacylglycerol/fatty acid cycling in murine white adipose tissue. *J Lipid Res*. 2023;64(6):100376.

doi: 10.1016/j.jlr.2023.100376

14. MacLeod B, Wang C, Brown LH, *et al.* Fads2 knockout mice reveal that ALA prevention of hepatic steatosis is dependent on delta-6 desaturase activity. *J Lipid Res*. 2024;65(10):100642.

doi: 10.1016/j.jlr.2024.100642

15. Li YL, Tian H, Jiang J, Zhang Y, Qi XW. Multifaceted regulation and functions of fatty acid desaturase 2 in human cancers. *Am J Cancer Res*. 2020;10(12):4098-4111.

16. Xuan Y, Wang H, Yung MM, *et al.* SCD1/FADS2 fatty acid desaturases equipoise lipid metabolic activity and redox-driven ferroptosis in ascites-derived ovarian cancer cells. *Theranostics*. 2022;12(7):3534-3552.

doi: 10.7150/thno.70194

17. Kothapalli KSD, Park HG, Kothapalli NSL, Brenna JT. FADS2 function at the major cancer hotspot 11q13 locus alters fatty acid metabolism in cancer. *Prog Lipid Res*. 2023;92:101242.

doi: 10.1016/j.plipres.2023.101242

18. Zhao T, Gao P, Li Y, *et al.* Investigating the role of FADS family members in breast cancer based on bioinformatic analysis and experimental validation. *Front Immunol*. 2023;14:1074242.

doi: 10.3389/fimmu.2023.1074242

19. Zhang X, Ren B, Liu B, *et al.* Single-cell RNA sequencing and spatial transcriptomics reveal the heterogeneity and intercellular communication of cancer-associated fibroblasts in gastric cancer. *J Transl Med*. 2025;23(1):344-357.

doi: 10.1186/s12967-025-06376-8

20. Dadsena S, Cuevas Arenas R, Vieira G, Brodesser S, Melo MN, García-Sáez AJ. Lipid unsaturation promotes BAX and BAK pore activity during apoptosis. *Nat Commun*. 2024;15(1):4700.

doi: 10.1038/s41467-024-49067-6

21. Young RSE, Bowman AP, Williams ED, *et al.* Apocryphal FADS2 activity promotes fatty acid diversification in cancer. *Cell Rep*. 2021;34(6):108738.

doi: 10.1016/j.celrep.2021.108738

22. Razavi P, Chang MT, Xu G, *et al.* The genomic landscape of endocrine-resistant advanced breast cancers. *Cancer Cell*. 2018;34(3):427-438.e6.

doi: 10.1016/j.ccell.2018.08.008

23. Dibra D, Xiong S, Moyer SM, *et al.* Mutant p53 protects triple-negative breast adenocarcinomas from ferroptosis in vivo. *Sci Adv*. 2024;10(7):eadk1835.

doi: 10.1126/sciadv.adk1835

24. Dzięgiel P, Owczarek T, Plazuk E, *et al.* Ceramide galactosyltransferase (UGT8) is a molecular marker of breast cancer malignancy and lung metastases. *Br J Cancer*. 2010;103(4):524-531.

doi: 10.1038/sj.bjc.6605750

25. Cao Q, Chen X, Wu X, *et al.* Inhibition of UGT8 suppresses basal-like breast cancer progression by attenuating sulfatide-αVβ5 axis. *J Exp Med*. 2018;215(6):1679-1692.

doi: 10.1084/jem.20172048

26. Ding S, Yang R, Meng J, *et al.* Prognostic and immune correlation of IDO1 promoter methylation in breast cancer. *Sci Rep*. 2024;14(1):27836.

doi: 10.1038/s41598-024-79149-w

27. Finn RS. Targeting src in breast cancer. *Ann Oncol*. 2008;19(8):1379-1386.

doi: 10.1093/annonc/mdn291

28. Luo J, Zou H, Guo Y, *et al.* SRC kinase-mediated signaling pathways and targeted therapies in breast cancer. *Breast Cancer Res*. 2022;24(1):99.

doi: 10.1186/s13058-022-01596-y

29. Jiang YZ, Ma D, Suo C, *et al.* Genomic and Transcriptomic Landscape of Triple-Negative Breast Cancers: Subtypes and Treatment Strategies. *Cancer Cell*. 2019;35(3):428-440.e5.

doi: 10.1016/j.ccell.2019.02.001

30. Xiao Y, Ma D, Yang YS, *et al.* Comprehensive metabolomics expands precision medicine for triple-negative breast cancer. *Cell Res*. 2022;32(5):477-490.

doi: 10.1038/s41422-022-00614-0

31. Vtorushin S, Dulesova A, Krakhmal N. Luminal androgen receptor (LAR) subtype of triple-negative breast cancer: molecular, morphological, and clinical features. *J Zhejiang Univ Sci B*. 2022;23(8):617-624.

doi: 10.1631/jzus.B2200113

32. Elia I, Haigis MC. Metabolites and the tumour

microenvironment: from cellular mechanisms to systemic metabolism. *Nat Metab*. 2021;3(1):21-32.

doi: 10.1038/s42255-020-00317-z

33. Koundouros N, Poulogiannis G. Reprogramming of fatty acid metabolism in cancer. *Br J Cancer*. 2020;122(1):4-22.

doi: 10.1038/s41416-019-0650-z

34. Miller WL, Auchus RJ. The molecular biology, biochemistry, and physiology of human steroidogenesis and its disorders. *Endocr Rev*. 2011;32(1):81-151.

doi: 10.1210/er.2010-0013

35. Tanio M, Miyamoto Y, Saka T, *et al.* Kobochromone A, a polyphenol in Carex kobomugi, suppresses androgen signaling induced by 11-oxygenated androgens and enhances the efficacy of AKT inhibitors in triple-negative breast cancer cells. *Steroids*. 2025;223:109692.

doi: 10.1016/j.steroids.2025.109692

36. Li S, Lee W, Heo W, *et al.* AKR1C2 Promotes Metastasis and Regulates the Molecular Features of Luminal Androgen Receptor Subtype in Triple Negative Breast Cancer Cells. *J Breast Cancer*. 2023;26(1):60-76.

doi: 10.4048/jbc.2023.26.e1

37. Palmer CP, Mahen R, Schnell E, Djamgoz MBA, Aydar E. Sigma-1 receptors bind cholesterol and remodel lipid rafts in breast cancer cell lines. *Cancer Res*. 2007;67(23):11166-11175.

doi: 10.1158/0008-5472.CAN-07-1771

38. Lim SA, Wei J, Nguyen TLM, *et al.* Lipid signalling enforces functional specialization of Treg cells in tumours. *Nature*. 2021;591(7849):306-311.

doi: 10.1038/s41586-021-03235-6

39. Hu P, Zhou P, Sun T, Liu D, Yin J, Liu L. Therapeutic protein PAK restrains the progression of triple negative breast cancer through degrading SREBP-1 mRNA. *Breast Cancer Res*. 2023;25(1):151.

doi: 10.1186/s13058-023-01749-7

40. Liang D, Minikes AM, Jiang X. Ferroptosis at the intersection of lipid metabolism and cellular signaling. *Mol Cell*. 2022;82(12):2215-2227.

doi: 10.1016/j.molcel.2022.03.022

41. Matuszczak M, Kiljańczyk A, Marciniak W, *et al.* Zinc and its antioxidant properties: The potential use of blood zinc levels as a marker of cancer risk in BRCA1 mutation carriers. *Antioxidants*. 2024;13(5):609.

doi: 10.3390/antiox13050609