Plasma Proteome-Genome Integration Reveals Novel Protein Biomarkers and Drug Targets Linked to Breast Cancer Survival

Keywords
Breast cancer, Plasma proteomes, Proteome-wide Mendelian randomization, Gene expression, Drug targets

Abstract

Introduction
Identifying new drug targets is essential for improving breast cancer survival. The proteome provides a rich source for potential therapeutic targets. This study aimed to identify protein markers and therapeutic targets for breast cancer by using proteome-wide Mendelian randomization (MR).

Material and methods
Protein quantitative trait loci (pQTL) data were obtained from four large-scaled proteomic studies, including 17,267 circulating protein markers. Genetic associations with breast cancer survival were derived from a large-scale GWAS meta-analysis (37,954 cases, 2,900 deaths). Proteome-wide MR was performed to assess the association between proteins and breast cancer survival, complemented by single-cell expression analysis to identify enriched cell types. Protein-protein interactions (PPI) and druggability assessments were also conducted to prioritize therapeutic targets.

Results
Gene prediction levels for 27 proteins were found to be associated with breast cancer survival. Among these, eight proteins (ADAM15, CD83, SH3BGRL3, SNCG, ANXA1, GRHPR, ALDH2, and MTHFD2) showed the strongest evidence of association, while four proteins (ARG2, RPL14, NFU1, and TXNL4B) demonstrated a strong but slightly weaker correlation. Notably, SH3BGRL3, GRHPR, ARG2, RPL14, NFU1, and TXNL4B were newly identified as circulating protein markers significantly associated with breast cancer prognosis. Druggability revealed that 13 of these proteins were already targeted by existing drugs, offering potential for breast cancer treatment.

Conclusions
We identified 27 genes associated with overall and subtype-specific breast cancer survival, providing potential prognostic biomarkers and therapeutic targets, and offering new avenues for improving breast cancer management.

Plasma Proteome-Genome Integration Reveals Novel Protein Biomarkers and Drug Targets Linked to Breast Cancer Survival

# These two authors contributed equally.

Abstract

Background

Identifying new drug targets is essential for improving breast cancer survival. The proteome provides a rich source for potential therapeutic targets. This study aimed to identify protein markers and therapeutic targets for breast cancer by using proteome-wide Mendelian randomization (MR).

Methods

Protein quantitative trait loci (pQTL) data were obtained from four large-scaled proteomic studies, including 17,267 circulating protein markers. Genetic associations with breast cancer survival were derived from a large-scale GWAS meta-analysis (37,954 cases, 2,900 deaths). Proteome-wide MR was performed to assess the association between proteins and breast cancer survival, complemented by single-cell expression analysis to identify enriched cell types. Protein-protein interactions (PPI) and druggability assessments were also conducted to prioritize therapeutic targets.

Results

Gene prediction levels for 27 proteins were found to be associated with breast cancer survival. Among these, eight proteins (ADAM15, CD83, SH3BGRL3, SNCG, ANXA1, GRHPR, ALDH2, and MTHFD2) showed the strongest evidence of association, while four proteins (ARG2, RPL14, NFU1, and TXNL4B) demonstrated a strong but slightly weaker correlation. Notably, SH3BGRL3, GRHPR, ARG2, RPL14, NFU1, and TXNL4B were newly identified as circulating protein markers significantly associated with breast cancer prognosis. Druggability revealed that 13 of these proteins were already targeted by existing drugs, offering potential for breast cancer treatment.

Conclusions

We identified 27 genes associated with overall and subtype-specific breast cancer survival, providing potential prognostic biomarkers and therapeutic targets, and offering new avenues for improving breast cancer management.

Keywords: Breast cancer, Plasma proteomes, Proteome-wide Mendelian randomization, Gene expression, Drug targets

Background

Breast cancer is the most common malignancy among women and remains the leading cause of cancer-related mortality worldwide, with an estimated 665,684 deaths in 2022 [1]. Clinically, breast cancer is categorized into estrogen receptor-positive (ER+) and estrogen receptor-negative (ER−) subtypes based on estrogen receptor expression [2]. Key pathological indicators play a critical role in diagnosis, prognostic assessment, and therapeutic decision-making, such as tumor subtype, histological grade, and ER/PR/HER2 status [3]. However, substantial inter-patient heterogeneity and variability in treatment response limit the prognostic accuracy of these conventional markers [4]. Thus, there is an urgent need for more precise prognostic biomarkers to guide individualized therapy [5].

Circulating proteins have emerged as promising biomarkers for disease diagnosis, prognosis, and therapeutic targeting [6, 7]. In breast cancer, proteomic studies have identified key candidates such as PRKDC, which shows elevated phosphorylation in the Basal-I subtype and may serve as a subtype-specific target [8]. However, studies specifically linking circulating proteins to breast cancer prognosis remain limited [9]. Mendelian randomization (MR) uses genetic variants as instrumental variables (IVs) to infer causal relationships between exposures (e.g., circulating proteins) and disease outcomes, minimizing confounding and reverse causation [10]. Integrating MR with large-scale plasma proteomics provides novel insights into the genetic basis of cancer progression [11].

In this study, we applied colocalization analysis, summary-based MR (SMR), heterogeneity in dependent instruments (HEIDI) tests, and two-sample MR (TSMR) to systematically identify circulating proteins associated with breast cancer survival. Additionally, single cell type expression analysis and draggability assessments were performed to explore their potential as targets for improving breast cancer

prognosis [12]. This study aimed to identify circulating proteins associated with breast cancer prognosis, providing insights for therapeutic development.

Materials and methods

The study design is presented in Figure 1. We sequentially applied Bayesian colocalization, SMR, HEIDI tests, and TSMR to validate the potential causal relationships between protein biomarkers and breast cancer survival, with additional validation using GTEx and eQTLgen data. To determine the tumor-specific expression patterns of the identified genes, we conducted single-cell RNA-seq analysis to explore cell type-specific enrichment in breast cancer tissues. Finally, we performed protein-protein interaction (PPI) and druggability analyses to evaluate their therapeutic potential.

Proteomic data source

Summary statistics of genetic associations with plasma proteins were extracted from four large proteomic studies: Alexander (2091 proteins) [13], deCODE (4907 proteins) [14], Fenland (3892 proteins) [15], and UKBPPP (1478 proteins) [16]. For external validation, eQTLgen (whole blood expression data from >31,000 individuals) and GTEx (multi-tissue expression data) were utilized.

Outcome data sources

The study included 37,954 breast cancer patients of European ancestry, with 2,900 deaths recorded. Subtype-specific analyses included 6,881 ER- patients (920 deaths) and 23,059 ER+ patients (1,333 deaths). Details on study populations, genotyping, and imputation methods are available in prior publications [17]. Ethics approvals and informed consent were obtained. Supplementary file 3 lists the sources and corresponding information of all aggregated statistical datasets used in this study.

Bayesian colocalization analysis

For each locus, the Bayesian method assessed the support for the following five exclusive hypotheses: 1) no association with either trait; 2) association with trait 1 only; 3) association with trait 2 only; 4) both traits are associated, but distinct causal variants were for two traits; and 5) both traits are associated, and the same shares causal variant for both traits. The analysis provides posterior probabilities for each hypothesis testing (H0, H1, H2, H3, and H4). We used the following prior probabilities: p1 = 10−4, p2 = 10−4 and p12 = 10−5. Colocalization was defined as PP4 > 0.5.

SMR analysis

MR analysis, treating plasma proteins as exposures and breast cancer survival as outcomes, used Bonferroni correction to adjust for multiple testing. Specifically, proteins were categorized into three groups: (1) no colocalization evidence, (2) moderate colocalization evidence (e.g., PP4 between 0.5 and 0.8), and (3) high colocalization evidence (e.g., PP4 ≥ 0.8). The Bonferroni correction set the significance threshold at $P < 0.05$.

The HEIDI test, applied when ≥3 SNPs were available, excluded associations with pleiotropy ($P_{HEIDI} < 0.05$). SNPs with high (r2 > 0.9) or weak (r2 < 0.05) linkage disequilibrium (LD)were excluded.

Selection of genetic instruments

In TSMR analysis, cis-pQTLs (± 1 MB of the gene) with $P < 5 \times 10^{-8}$ were used as IVs. Furthermore, SNPs with allele frequency differences greater than 0.2 between the pQTL data and the GWAS data were excluded. We permitted to exclude a maximum of 5% of SNPs based on allele frequency differences. IV strength was assessed using the F-statistic ($F = \beta^2/SE^2$), and SNPs with F < 10 were considered weak and excluded. Finally, the top-associated SNP with gene expression was selected as the genetic instrument [10, 12].

TSMR analysis

TSMR analysis was further conducted to verify the causal associations between proteins and breast cancer survival. The following criteria were used to select instruments and proteins: (i) SNPs associated with any protein were selected ($P < 5\times10^{-8}$); (ii) the SNPs and proteins within the Major Histocompatibility Complex (MHC) region (chr6: 25.5–34.0Mb) were excluded due to their complex LD structure; (iii) the LD clumping was then conducted to identify independent pQTLs for each protein ($r^2 < 0.01$); (iv) the R^2 and F-statistic ($R^2=2\times EAF\times(1-EAF)\times beta^2$; $F=R^2\times(N-2)/(1-R^2)$) were used to estimate the strength of genetic instruments, where R^2 was the proportion of the variability of the protein levels explained by each genetic instrument.

We performed sensitivity analyses using Cochran's Q, MR-Egger intercept, and Steiger filtering, with significance determined by corresponding p-values.

Single cell-type expression analysis

To explore cell type-specific expression, we analyzed single-cell RNA-seq data (GSE176078) from breast tumor tissues, focusing on genes with potential causal effects on breast cancer. Low-quality cells were filtered out, and the remaining data were log-normalized. To assess whether breast cancer survival‐associated genes are preferentially expressed in specific cell types within breast tumor tissue, differential expression analysis was performed using the Wilcoxon rank-sum test. The genes with an average Log_2 fold change (Log_2FC) more than 0.5 and a false discovery rate (FDR) adjusted P value less than 0.05 were identified as enrichment genes in a cell type.

Based on colocalisation analysis, MR analysis, and single-cell specificity analysis, we classified proteins into three distinct target groups. Those that passed all MR tests and exhibited cell type-specific

enrichment were assigned to Tier 1, those with PPH > 0.8 and cell type-specific enrichment were assigned to Tier 2, and proteins lacking single-cell expression or with moderate colocalisation evidence were assigned to Tier 3.

Immune Infiltration Analysis

To assess the relationship between prognostic protein-coding gene expression and immune cell infiltration in breast cancer, we employed the TIMER web tool (Tumor Immune Estimation Resource, http://cistrome.org/TIMER/) [18]. In this study, the "gene" module was used to evaluate the relationship between gene expression and immune cell infiltration.

PPI and druggability evaluation

PPI network were constructed using the STRING database (https://string-db.org/). To assess the druggability of identified proteins, we searched identified proteins in DrugBank, DGIdb, the ChEMBL and Dependency Map databases [19]. For proteins identified in drug databases, information on the drug name and the process of drug development was documented. To assess the potential druggability, we classified these proteins into four categories: 1) Approved; 2) in clinical trials; 3) Investigational; 4) Experimental.

Statistical analysis

MR analyses applied the Wald ratio (single SNP) and inverse-variance weighted (IVW) (≥2 SNPs) methods, complemented by MR-Egger and weighted median approaches to account for pleiotropy and ensure robust causal estimates [20, 21]. The results were presented as odds ratios per standard deviation increase in genetically determined plasma proteins. The above analyses were performed using "coloc", "TwoSampleMR", "Seurat", "SingleR" and other necessary packages in R (version 4.3.2) [22,

23].

Results

Colocalization analysis

We conducted a colocalization analysis to evaluate whether the observed associations between proteins and breast cancer survival or its subtypes were driven by shared genetic signals (Supplementary Table SI). Seven proteins showed strong colocalization evidence: ARG2, RPL14, and ACBD7 with overall survival; OPCML and DRAXIN with ER- survival; and NFU1 and TXNL4B with ER+ survival. Additionally, 41 proteins demonstrated moderate evidence of colocalization. The remaining proteins showed no evidence of colocalization with breast cancer survival.

SMR and HEIDI tests verified seven causal proteins

To validate the effect of proteins on breast cancer survival, we performed SMR and HEIDI analyses on 7522 proteins using data from four large cohorts. Among the 40 proteins that passed SMR analysis, only 4 of them (LDLRAP1, SKAP1, CSF2, SCLY) failed the HEIDI test ($P < 0.05$). Among the remaining proteins, 20 were identified as potentially associated with overall breast cancer survival (Supplementary Table SII). Subtype analysis revealed that 7 proteins might be associated with ER- breast cancer survival, while 9 proteins might be associated with ER+ breast cancer survival.

TSMR analysis

We identified 131 significant SNPs as IVs ($P < 5 \times 10^{-8}$), all with F-statistics >10 (Supplementary Table SIII). Eight proteins did not pass the TSMR analysis (P-adj > 0.05) and were excluded from further analysis. Using the Wald ratio or IVW methods with Bonferroni correction, we found that nine proteins (ADAM15, ARG2, CD83, CEP85, GORASP2, HAPLN1, LEFTY2, SH3BGRL3, and SNCG) were

associated with improved survival, and five to poorer outcomes (IL36A, PGM1, RPL14, SERPINB5, and UBE2F). In stratified analyses by breast cancer subtype, ALDH2, HAPLN1, MTHFD2, and TXNL4B were associated with improved ER+ survival, whereas MUC16 and NFU1 predicted worse outcomes. For ER- breast cancer, ANXA1 was linked to improved survival, while ALOX15B, CPA2, GRHPR, KLK14, and OPCML were associated with reduced survival (Supplementary Table SIV). These associations were generally consistent in the weighted median, and MR-Egger analyses. The results of the four main TSMR methods are shown in Supplementary Table SV. No heterogeneity and horizontal pleiotropy were found (Q_pval_Inverse.variance.weighted > 0.05, pval_Egger_intercept > 0.05) (Supplementary Table SVI).

In the external validation phase, we successfully replicated the causal association of GORASP2 and UBE2F with breast cancer survival, as well as MTHFD2 with ER+ breast cancer survival, using data from the eQTLgen (Figure 2). However, validation was not possible for nine proteins due to data unavailability, and several others did not replicate their causal associations in external datasets. These discrepancies may reflect dataset-specific differences such as sample size, or population characteristics. Colocatization, SMR and TSMR results are summarized for display in Supplementary Table SVII. Furthermore, we linked genetic effects to protein function and assessed the expression levels of predicted proteins across various tissues using the GTEx (Supplementary Figure S1).

Cell-type specificity expression in the breast cancer tissue

To investigate whether the 27 genes exhibited cell type-specific enrichment in breast cancer tissues, we conducted a single-cell expression analysis using single-cell RNA-seq data from the GEO database. Cells were clustered into 19 clusters and subsequently categorized into eight cell types: epithelial cells,

cycling cells, T cells, myeloid cells, B cells, plasmablasts, endothelial cells, and mesenchymal cells (Figure 3A). Figure 3 (B and C) shows the single-cell expression of these 27 genes in each cluster. Notably, IL6A was not included in the dataset, and neither KLK14 nor OPCML was expressed in any cell population. Eight genes demonstrated cell type-specific enrichment in breast cancer tissues, characterized by an average $Log_2FC > 0.5$ and $FDR < 0.05$ (Figure 3D).

Finally, guided by our colocalization analysis, MR analysis, and single-cell specificity analysis, we categorized the proteins into three distinct target groups, summarized in Supplementary Table SVIII. Tier 1 includes eight proteins that passed all tests (ADAM15, CD83, SH3BGRL3, SNCG, ANXA1, GRHPR, ALDH2, MTHFD2), and Tier 2 includes four proteins (ARG2, RPL14, NFU1, TXNL4B). Proteins lacking single-cell expression or supported by moderate colocalization evidence were assigned to Tier 3. Figure 4 shows the supporting evidence for colocalization between the 27 proteins and the results.

To explore the potential immunological role of proteins, we used TIMER to analyse the correlation between protein expression levels and tumor-infiltrating immune cell levels. Notably, in BRCA, tier 1 of proteins showed a significant association with immune infiltration (Supplementary Figure S2). These findings support the hypothesis that proteins may influence the tumor microenvironment by regulating the recruitment or activation of immune cells.

PPI and druggability evaluation on the potentials of therapeutic targets

PPI analyses revealed limited interactions between identified potentially pathogenic proteins, with only eight proteins interacting (Supplementary Figure S3). Several of these proteins are targeted by existing drugs approved for other indications, suggesting potential for repurposing in breast cancer. For instance, sulfasalazine (targeting CD83), kaempferol (ALOX15B), and eflornithine (ARG2) demonstrate

anti-inflammatory or anti-tumor properties. Additionally, cardiovascular agents like acetylsalicylic acid (ALOX15B) and nitroglycerin (ALDH2) may warrant further investigation. Lastly, hydrocortisone (ANXA1 target) could modulate the tumor microenvironment via metabolic and immune regulation. Further studies and clinical trials would be needed to confirm their applicability for breast cancer. A summary of investigational and approved drugs targeting the identified proteins is provided in Supplementary Table SIX.

Discussion

We systematically examined causal relationships between 17,267 circulating proteins and breast cancer survival using Bayesian colocalization, SMR, HEIDI tests, and TSMR, identifying 27 potential prognostic biomarkers. Six proteins (SH3BGRL3, GRHPR, ARG2, RPL14, NFU1, TXNL4B) were linked to breast cancer survival for the first time. Among these, eight proteins (ADAM15, CD83, SH3BGRL3, SNCG, ANXA1, GRHPR, ALDH2, and MTHFD2) showed the strongest overall survival associations, while four proteins (ARG2, RPL14, NFU1, and TXNL4B) demonstrated strong but slightly less robust associations. Subtype-stratified analyses revealed distinct patterns: ALDH2, HAPLN1, MTHFD2, and TXNL4B were associated with improved survival in ER+ breast cancer, whereas MUC16 and NFU1 were linked to worse prognosis. For ER- breast cancer, ANXA1 correlated with better survival, while ALOX15B, CPA2, GRHPR, KLK14, and OPCML were linked to poorer survival. To further explore clinical potential of these findings, we evaluated druggability and identified 13 proteins with approved or investigational therapeutic agents. These findings provide valuable insights into the molecular mechanisms underlying breast cancer prognosis and suggest potential therapeutic targets for improving patient outcomes.

ADAM15, particularly its isoform ADAM15-C, has been linked to improved survival after lymph node metastasis, likely through its effects on tumor growth and angiogenesis [24]. CD83, highly expressed in mature dendritic cells and axillary lymph nodes, enhances anti-tumor immunity and may serve as a novel prognostic marker for early metastasis [25, 26]. ANXA1, involved in both tumor growth and immune response, was associated with improved survival, possibly by reducing inflammation and activating M1 macrophages [27]. ALDH2, an alcohol metabolism enzyme, showed improved ER+ survival, likely through reduced oxidative stress and enhanced myeloid cell function in the tumor microenvironment [28, 29]. MTHFD2, as a key enzyme in folate metabolism, helps maintain cellular homeostasis, limit harmful senescence-associated effects, and thereby contribute to improved breast cancer prognosis [30]. Unfortunately, the relationship between SNCG expression and prognosis in our MR analysis appears inconsistent with previous studies and may be partly due to potential confounding or intermediate factors [31, 32]. Collectively, these findings not only highlight their potential as prognostic biomarkers but also provide insights into breast cancer progression and immune interactions, offering promising avenues for therapeutic intervention.

In addition to previously recognized prognostic proteins, we also identified several new biomarkers associated with breast cancer survival, including SH3BGRL3, GRHPR, ARG2, RPL14, NFU1, and TXNL4B, with SH3BGRL3 and GRHPR providing the most compelling evidence (Tier 1). SH3BGRL3 significantly correlates with epidermal growth factor receptor (EGFR) expression ($P < 0.0001$), suggesting involvement in EGFR-mediated oncogenic pathways, making it a promising therapeutic target [33].

GRHPR, a cytoplasmic glyoxylate-metabolizing enzyme, is negatively associated with survival in ER-negative breast cancer, indicating subtype-specific roles [34]. Its species-specific regulation, especially the lack of PPARα control in humans, calls for further study of its metabolic role in breast cancer [35]. Beyond these findings, ARG2, RPL14, NFU1, and TXNL4B also emerged as novel survival-associated proteins, warranting further research to determine their biological functions and potential therapeutic implications. These newly identified biomarkers expand our understanding of breast cancer prognosis, offering new avenues for both biomarker development and therapeutic intervention.

A key strength of this study is the subtype-stratified analysis, which revealed distinct biomarker-survival links across molecular subtypes and uncovered overlooked prognostic factors. Notably, we observed that not all circular proteins directly influence survival outcomes during tumor development and progression. This underscores the importance of incorporating prognostic data into MR studies and the need for further mechanistic and clinical validation to improve biomarker-based prognostication. Given that gene function can vary by cell type, we leveraged single-cell transcriptomic datasets to examine gene expression patterns at the cellular level. This analysis enhances our understanding of biomarker relevance in breast cancer and supports the development of more precise targeted therapies. Furthermore, TIMER-based immune infiltration analysis showed that these genes correlate with specific immune cells, suggesting they may shape the tumor microenvironment and have prognostic or therapeutic value.

However, several limitations of this study should also be considered. Firstly, caution is required when interpreting the posterior probability in colocalization (PH4). A low PH4 value may not indicate a lack of co-localization evidence if PH3 is also low due to insufficient power. Secondly, some causal

associations failed to replicate in eQTLgen, likely due to differences in population structure, limited power, or tissue origin. Notably, eQTLGen is based on whole blood, which may not capture regulatory effects relevant to breast tissue. Nonetheless, the successful replication of several key proteins, including GORASP2, UBE2F, and MTHFD2, supports the credibility of our main findings. Future studies utilizing larger, multi-tissue eQTL resources are warranted to improve replication accuracy. Thirdly, although our study provides preliminary evidence linking certain drug targets to breast cancer, these associations may be biased if the genetic instruments influence outcomes through pathways other than protein levels. Moreover, the biological mechanisms by which these proteins affect tumor progression remain unclear and require further validation through in vivo and in vitro studies. Unfortunately, the mechanisms by which these proteins affect tumor progression are not yet fully understood and require validation through in vivo and in vitro studies.

Conclusions

In this study, we identified 27 genes associated with overall and subtype-specific breast cancer survival across eight distinct cell types. These genes display varying effect sizes and unique associations with breast cancer, offering promising targets for both screening biomarkers and therapeutic drug development.

Abbreviations

ER	Estrogen receptor
EGFR	Epidermal growth factor receptor
FDR	False discovery rate
GWAS	Genome-wide association study
GEO	Gene Expression Omnibus

HEIDI	Heterogeneity in dependent instruments
IVs	Instrumental variables
IVW	Inverse Variance Weighted
LD	Linkage disequilibrium
Log_2FC	Log_2 fold change
MR	Mendelian randomization
MHC	Major histocompatibility complex
PPI	Protein-protein interaction
pQTL	Protein quantitative trait loci
SMR	summary database-based mendelian randomization
SNP	single nucleotide polymorphism
TSMR	two-sample mendelian randomization

Acknowledgements

We thank all participants, institutions, and their staff in the four proteomics studies, BCAC GWAS, the eQTLgen and the GTEx database for providing data.

Author contributions

Conceptualization: CZX and SLT; methodology, software, formal analysis, data curation, investigation and visualization: CZX and PQY; writing-original draft: CZX and PQY; writing-review and editing: SLT; supervision and funding acquisition: SLT.

Funding

Hunan Administration of Traditional Chinese Medicine (No. B2023062), and Natural Science Foundation of Hunan Province (No. 2024JJ8214).

Availability of data and materials

All packages used for data analysis in this study are open-source and were implemented in R software (version 4.3.2). The scRNA-seq data were sourced from the NCBI GEO database. All results are provided in the article and supplementary materials. further data are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The data used in this study have been ethically approved.

Consent for publication

Not applicable.

Competing interests

No competing interests.

Reference

[1] Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2024; 74: 229-63.

[2] Hong Z, Fang Z, Lei J, et al. The significance of Runx2 mediating alcohol-induced Brf1 expression and RNA Pol III gene transcription. Chem Biol Interact 2020; 323: 109057.

[3] Li H, Ju X, Zeng C, et al. Development and validation of a pathological model predicting the efficacy of neoadjuvant therapy for breast cancer based on RCB scoring. Arch Med Sci 2025; 21: 92-101.

[4] Rehman SU, Asel U, Abdullah M, et al. The development of predictive biomarkers and immunologic markers for breast cancer: current status and future perspectives. Braz J Biol 2025; 85: e292947.

[5] Nalejska E, Mączyńska E, Lewandowska MA. Prognostic and predictive biomarkers: tools in personalized oncology. Mol Diagn Ther 2014; 18: 273-84.

[6] Wang SE, Tan VY, Yarmolinsky J, et al. The effect of circulating proteins and their role in mediating adiposity's effect on endometrial cancer risk: Mendelian randomisation and colocalization analyses. Cancer Epidemiol Biomarkers Prev 2025.

[7] Fan KC, Chen SC, Yen IW, et al. Plasma angiopoietin-like protein 4 as a novel biomarker predicting 10-year mortality in a community-based population: a longitudinal cohort study. Arch Med Sci 2025; 21: 51-9.

[8] Krug K, Jaehnig EJ, Satpathy S, et al. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. Cell 2020; 183: 1436-56.e31.

[9] Morra A, Escala-Garcia M, Beesley J, et al. Association of germline genetic variants with breast cancer-specific survival in patient subgroups defined by clinic-pathological variables related to tumor biology and type of systemic treatment. Breast Cancer Res 2021; 23: 86.

[10] Chen J, Xu F, Ruan X, et al. Therapeutic targets for inflammatory bowel disease: proteome-wide Mendelian randomization and colocalization analyses. EBioMedicine 2023; 89: 104494.

[11] Zheng J, Haberland V, Baird D, et al. Phenome-wide Mendelian randomization mapping the influence of the plasma proteome on complex diseases. Nat Genet 2020; 52: 1122-31.

[12] Sun J, Zhao J, Jiang F, et al. Identification of novel protein biomarkers and drug targets for colorectal cancer by integrating human plasma proteome with genome. Genome Med 2023; 15: 75.

[13] Gudjonsson A, Gudmundsdottir V, Axelsson GT, et al. A genome-wide association study of serum proteins reveals shared loci with common diseases. Nat Commun 2022; 13: 480.

[14] Ferkingstad E, Sulem P, Atlason BA, et al. Large-scale integration of the plasma proteome with genetics and disease. Nat Genet 2021; 53: 1712-21.

[15] Pietzner M, Wheeler E, Carrasco-Zanini J, et al. Mapping the proteo-genomic convergence of human diseases. Science 2021; 374: eabj1541.

[16] Sun BB, Maranville JC, Peters JE, et al. Genomic atlas of the human plasma proteome. Nature 2018; 558: 73-9.

[17] Guo Q, Schmidt MK, Kraft P, et al. Identification of novel genetic markers of breast cancer survival. J Natl Cancer Inst 2015; 107.

[18] Li T, Fan J, Wang B, et al. TIMER: A Web Server for Comprehensive Analysis of Tumor-

Infiltrating Immune Cells. Cancer Res 2017; 77: e108-e10.

[19] Finan C, Gaulton A, Kruger FA, et al. The druggable genome and support for target identification and validation in drug development. Sci Transl Med 2017; 9.

[20] Shu L, Sun L, Yu C, Ren D, Zhang Y, Zheng P. Bidirectional two-sample Mendelian randomization analysis identifies protein C rather than protein S or antithrombin-III as associated with deep venous thrombosis. Arch Med Sci 2025; 21: 215-23.

[21] Zhang C, Qin F, Li X, Du X, Li T. Identification of novel proteins for lacunar stroke by integrating genome-wide association data and human brain proteomes. BMC Med 2022; 20: 211.

[22] Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014; 10: e1004383.

[23] Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat Immunol 2019; 20: 163-72.

[24] Zhong JL, Poghosyan Z, Pennington CJ, et al. Distinct functions of natural ADAM-15 cytoplasmic domain variants in human mammary carcinoma. Mol Cancer Res 2008; 6: 383-94.

[25] López C, Bosch R, Korzynska A, et al. CD68 and CD83 immune populations in non-metastatic axillary lymph nodes are of prognostic value for the survival and relapse of breast cancer patients. Breast Cancer 2022; 29: 618-35.

[26] Prechtel AT, Steinkasserer A. CD83: an update on functions and prospects of the maturation marker of dendritic cells. Arch Dermatol Res 2007; 299: 59-69.

[27] Araújo TG, Mota STS, Ferreira HSV, Ribeiro MA, Goulart LR, Vecchi L. Annexin A1 as a Regulator of Immune Response in Cancer. Cells 2021; 10.

[28] Hu J, Yang L, Peng X, et al. ALDH2 Hampers Immune Escape in Liver Hepatocellular Carcinoma through ROS/Nrf2-mediated Autophagy. Inflammation 2022; 45: 2309-24.

[29] Xu T, Guo J, Wei M, et al. Aldehyde dehydrogenase 2 protects against acute kidney injury by regulating autophagy via the Beclin-1 pathway. JCI Insight 2021; 6.

[30] Wang P, Fang Z, Pei W, et al. Senescence Reprogramming by MTHFD2 Deficiency Facilitates Tumor Progression. J Cancer 2024; 15: 6577-93.

[31] Bellou V, Belbasis L, Tzoulaki I, Evangelou E. Risk factors for type 2 diabetes mellitus: An

exposure-wide umbrella review of meta-analyses. PLoS One 2018; 13: e0194127.

[32] Zhao Y, Zhao L, Wang T, et al. The Herbal Combination Shu Gan Jie Yu Regulates the SNCG/ER-a/AKT-ERK Pathway in DMBA-Induced Breast Cancer and Breast Cancer Cell Lines Based on RNA-Seq and IPA Analysis. Integr Cancer Ther 2024; 23: 15347354241233258.

[33] Chiang CY, Pan CC, Chang HY, et al. SH3BGRL3 Protein as a Potential Prognostic Biomarker for Urothelial Carcinoma: A Novel Binding Partner of Epidermal Growth Factor Receptor. Clin Cancer Res 2015; 21: 5601-11.

[34] Zong C, Nie X, Zhang D, et al. Up regulation of glyoxylate reductase/hydroxypyruvate reductase (GRHPR) is associated with intestinal epithelial cells apoptosis in TNBS-induced experimental colitis. Pathol Res Pract 2016; 212: 365-71.

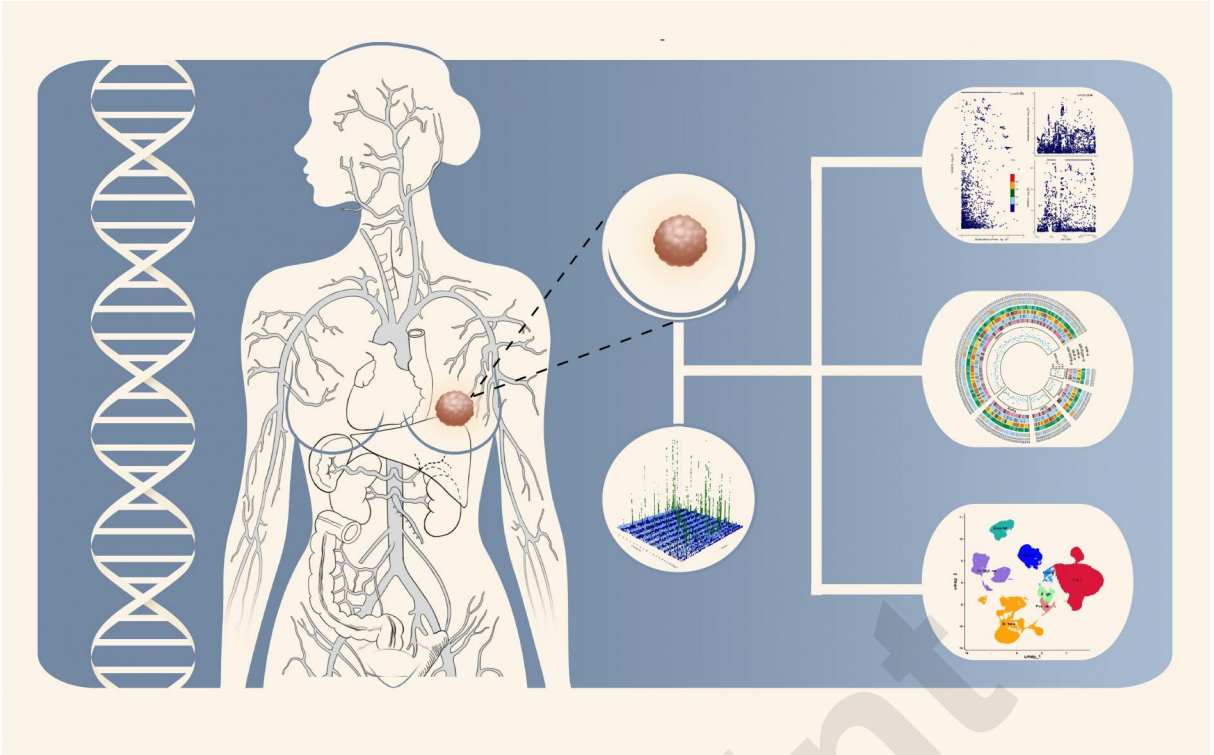
[35] Genolet R, Kersten S, Braissant O, et al. Promoter rearrangements cause species-specific hepatic regulation of the glyoxylate reductase/hydroxypyruvate reductase gene by the peroxisome proliferator-activated receptor alpha. J Biol Chem 2005; 280: 24143-52.

Figure 1 Study design. UKBPPP, UK Biobank Pharma Proteomics Project; BCAC, Breast Cancer Association Consortium; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instrument; MR, Mendelian Randomization.

Figure 2 The forest plot for SMR and TSMR analysis based on 27 proteins.

Figure 3 Single-cell type expression of 27 genes in breast cancer tissues. A – A total of 19 cell clusters and 8 cell types were identified. B, C – show the expression of protein coding genes in each cluster. D – Eight protein-coding genes had evidence of enrichment in a cell type at average $Log_2FC > 0.5$ and FDR < 0.05 level.

Figure 4 Support evidence for colocalization between proteins and outcomes. Circle size indicates the colocalization P value for H4 (colocalization analysis) and the colour of the circle indicate the classification of the evidence.



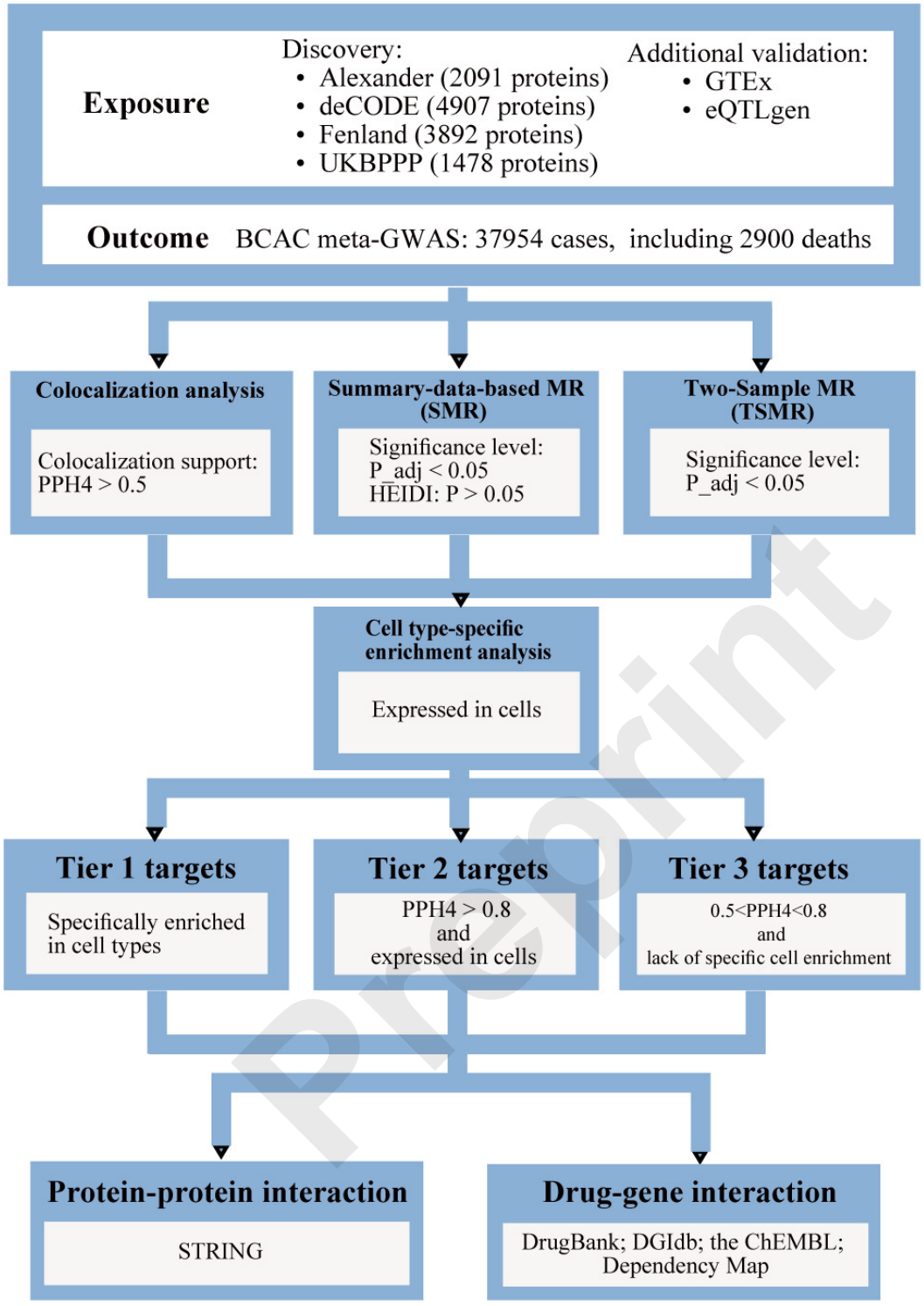


Fig. 1 Study design. UKBPPP, UK Biobank Pharma Proteomics Project; BCAC, Breast Cancer Association Consortium; GWAS, genome-wide association study; HEIDI, heterogeneity in dependent instrument; MR, Mendelian Randomization.

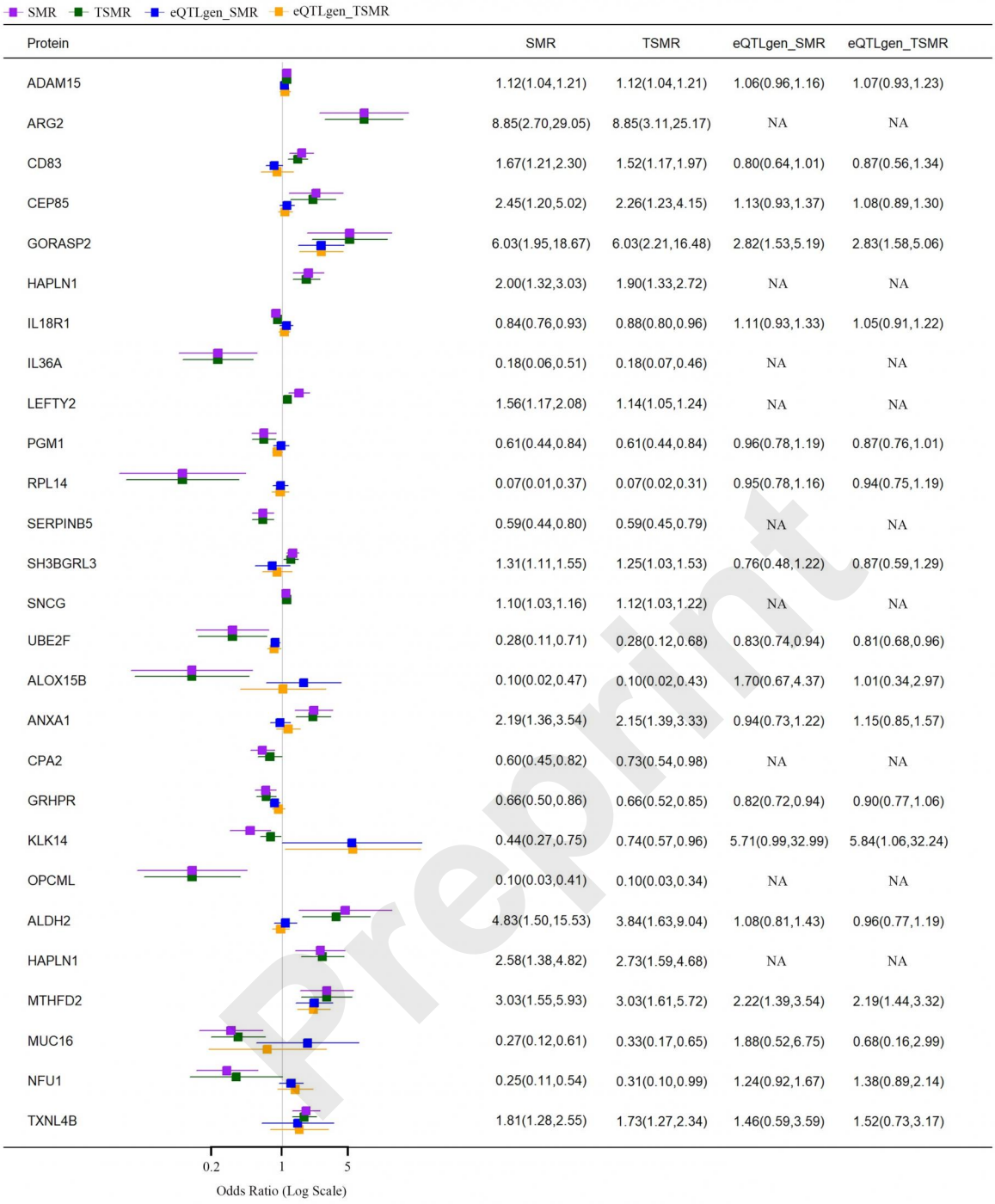


Figure 2. The forest plot for SMR and TSMR analysis based on 27 proteins.

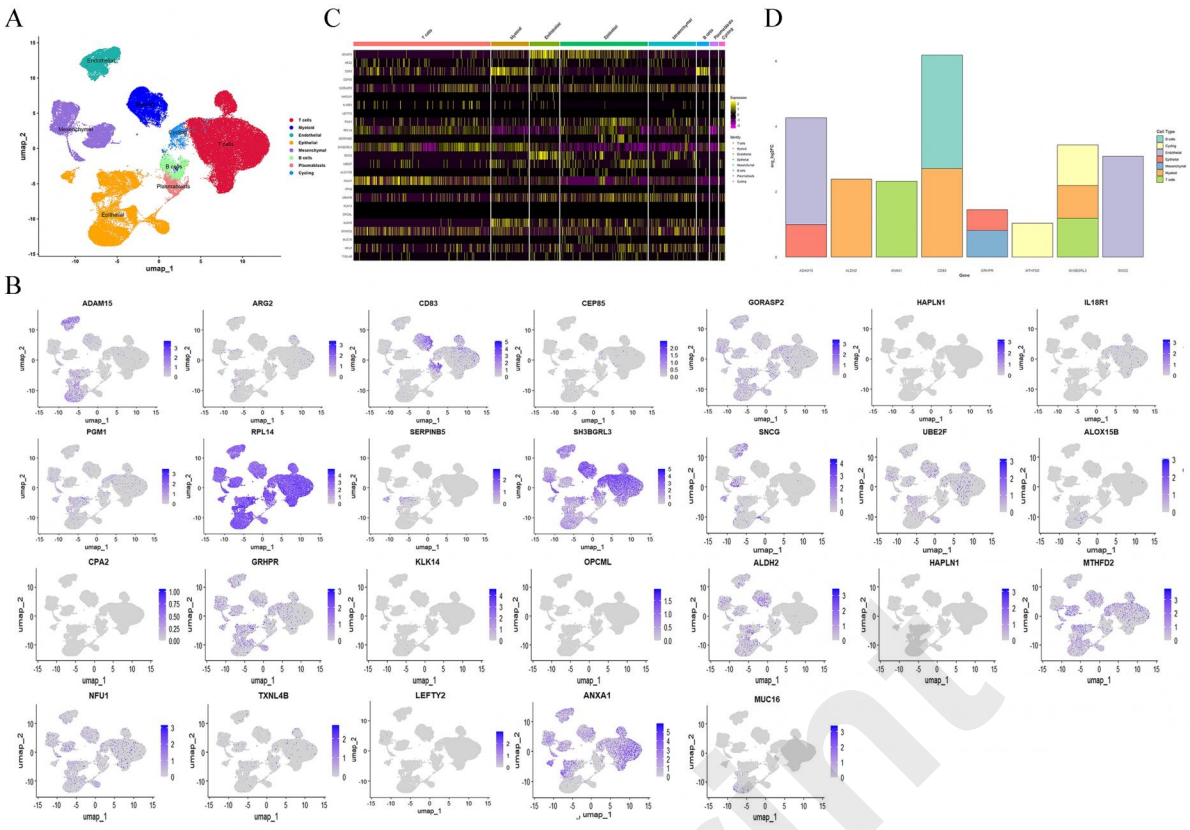


Figure 3. Single-cell type expression of 27 genes in breast cancer tissues. A – A total of 19 cell clusters and 8 cell types were identified. B, C – show the expression of protein coding genes in each cluster. D – Eight protein-coding genes had evidence of enrichment in a cell type at average Log2FC > 0.5 and FDR < 0.05 level.

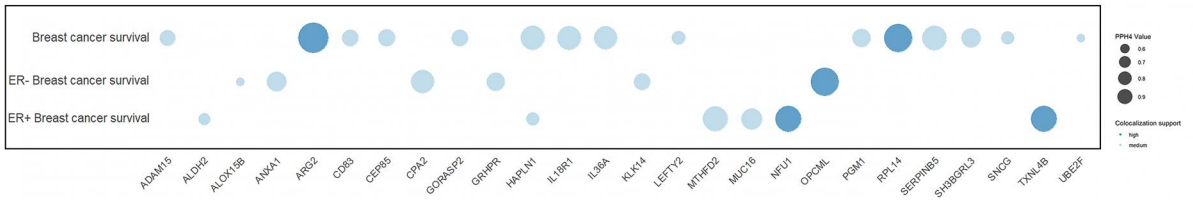


Figure 4 Support evidence for colocalization between proteins and outcomes. Circle size indicates the colocalization P value for H4 (colocalization analysis) and the colour of the circle indicate the classification of the evidence.