

A Copper Metabolism-Related Gene Signature Predicts Prognosis and Therapeutic Response in Breast Cancer Based on Multi-Omics Analysis

Weizhong Chen

Chaozhou Hospital Affiliated to Shantou University Medical College

Tao Chen

Chaozhou Hospital Affiliated to Shantou University Medical College

Yanfang Chen

Chaozhou Hospital Affiliated to Shantou University Medical College

Cuicai Lin

Chaozhou Hospital Affiliated to Shantou University Medical College

Wenhan Chen

Chaozhou Hospital Affiliated to Shantou University Medical College

Xiaoru Wu

Chaozhou Hospital Affiliated to Shantou University Medical College

Lusheng Liao

Youjiang Medical University for Nationalities

Mingyou Dong

mydong@ymun.edu.cn

Youjiang Medical University for Nationalities

Zhichao Fang

Chaozhou Hospital Affiliated to Shantou University Medical College

Research Article

Keywords: Breast cancer, copper metabolism, Mendelian randomization, EVA1B, multi-omics, immune infiltration

Posted Date: July 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9927304/v1>

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

[Read Full License](#)

Additional Declarations: No competing interests reported.

**A Copper Metabolism-Related Gene Signature Predicts Prognosis and Therapeutic
Response in Breast Cancer Based on Multi-Omics Analysis**

Weizhong Chen^{13#}, Tao Chen^{4#}, Yanfang Chen⁵, Cuicai Lin⁶, Wenhan Chen⁷, Xiaoru Wu¹³,
Lusheng Liao², Mingyou Dong^{2*}, Zhichao Fang^{4*}

1 Postdoctoral Research Station, Chaozhou People's Hospital, Chaozhou Hospital Affiliated to
Shantou University Medical College, Chaozhou, Guangdong, 510000, China

2 Guangxi Key Laboratory of Artificial Intelligence for Genetic Diseases of Long-dwelling
Nationalities, Youjiang Medical University for Nationalities, Baise, Guangxi, 533000, China

3 Department of Clinical Laboratory, Chaozhou People's Hospital, Chaozhou Hospital Affiliated
to Shantou University Medical College, Chaozhou, Guangdong, 510000, China

4 Breast Disease Prevention and Treatment Center, Chaozhou People's Hospital, Chaozhou
Hospital Affiliated to Shantou University Medical College, Chaozhou, Guangdong, 510000, China

5 Department of Thoracic Oncology, Chaozhou People's Hospital, Chaozhou Hospital Affiliated to
Shantou University Medical College, Chaozhou, Guangdong, 510000, China

6 Department of Traditional Chinese Medicine, Chaozhou People's Hospital, Chaozhou Hospital
Affiliated to Shantou University Medical College, Chaozhou, China

7 Department of Head, Neck and Breast Oncology, Chaozhou People's Hospital, Chaozhou
Hospital Affiliated to Shantou University Medical College, Chaozhou, Guangdong, 510000, China

#Co-first authors: Weizhong Chen, Tao Chen and Yanfang Chen contributed equally to this work.

***Corresponding authors:** Zhichao Fang: Breast Disease Prevention and Treatment Center,
Chaozhou People's Hospital, Chaozhou Hospital Affiliated to Shantou University Medical
College, Chaozhou, Guangdong, 510000, China, email: czyyfc@163.com. Mingyou Dong:
Guangxi Key Laboratory of Artificial Intelligence for Genetic Diseases of Long-dwelling
Nationalities, Youjiang Medical University for Nationalities, Guangxi, 533000, China, email:
mydong@ymun.edu.cn.

Abstract

Breast cancer (BC) displays remarkable heterogeneity, which creates difficulties for precise

prognosis and individualized treatment. Copper metabolism has increasingly been recognized as a critical factor in tumor progression, yet its clinical relevance in BC remains undefined. We integrated multi-cohort transcriptomic data from TCGA and GEO databases. LASSO regression was employed to screen for copper metabolism-related genes (CMRGs), and then multivariate Cox regression was performed to develop a prognostic signature. A series of systematic evaluations were conducted, including analyses of immune infiltration, tumor mutational burden (TMB), drug sensitivity, and immunotherapy response. We also performed two-sample Mendelian randomization and Bayesian colocalization analyses to explore causal relationships between signature gene expression and BC risk. A four-gene signature consisting of ZMYND10, EVA1B, GREB1, and CHI3L1 was successfully established. A high CMRG risk score emerged as an independent unfavorable prognostic factor and was associated with an immunosuppressive tumor microenvironment, characterized by reduced levels of CD8⁺T cells, dendritic cells, and M1 macrophages. A low risk score predicted a favorable response to immune checkpoint inhibitors (IMvigor210, GSE91061) as well as enhanced sensitivity to multiple chemotherapeutic agents (oxaliplatin, gemcitabine, cisplatin) and targeted drugs (lapatinib, palbociclib). Dysregulation of the four genes was confirmed through expression validation across TCGA, GSE42568, GSE54002, and HPA immunohistochemistry data. Mendelian randomization revealed a causal link between EVA1B expression and BC risk (OR = 1.05, $P = 0.025$) with strong colocalization support (PP.H4 = 0.989). The robustness of the signature was further confirmed by external validation. Taken together, this CMRG signature effectively stratifies patient risk, mirrors the immune landscape, and holds promise as a biomarker for prognosis prediction, immunotherapy response, and drug sensitivity in breast cancer. EVA1B represents a potential causal therapeutic target.

Keywords: Breast cancer; copper metabolism; Mendelian randomization; EVA1B; multi-omics; immune infiltration

Introduction

Globally, breast cancer remains one of the most common and lethal cancers among women, accounting for over 2.3 million new diagnoses and 685,000 deaths each year ^[1, 2]. Although surgical techniques, radiotherapy, chemotherapy, and targeted therapies have advanced

considerably, the marked heterogeneity of this disease leads to diverse treatment responses and clinical outcomes across patients ^[1]. Such clinical variability highlights the pressing need to discover novel biomarkers capable of enhancing prognostic accuracy and informing personalized therapeutic decisions.

Copper serves an essential role as a catalytic and structural cofactor for various enzymes, contributing to energy metabolism, iron homeostasis, oxygen transport, cellular signaling, and hormone maturation ^[3, 4]. Notably, elevated copper concentrations have been observed in tumor tissues compared with normal counterparts, with increased systemic copper levels correlating with unfavorable clinical outcomes in cancer patients ^[5, 6]. Given that copper metabolism is governed by an intricate network of regulatory molecules and signaling pathways, comprehensive investigation of copper metabolism-related genes (CMRGs) holds significant potential for advancing cancer diagnosis, prognosis prediction, and therapeutic development ^[7].

Programmed cell death pathways are recognized as key modulators involved in tumor initiation, progression, and treatment responses ^[8, 9]. Copper-induced cell death, termed cuproptosis, was recently identified as a novel form of programmed cell death characterized by the aggregation of lipoylated tricarboxylic acid cycle proteins ^[10]. Several studies have explored the role of cuproptosis-related genes in various cancers ^{[11-13][11-13][11-13]}. However, the specific roles and regulatory mechanisms of CMRGs in breast cancer remain poorly understood, and comprehensive analyses integrating CMRG signatures with breast cancer prognosis and immunological context remain lacking.

The advancement of high-throughput sequencing technologies and bioinformatics has enabled researchers to investigate transcriptomic changes and their clinical significance in cancer. A powerful strategy that has emerged involves integrating RNA-sequencing data with machine learning methods, which captures complex gene expression patterns to develop robust prognostic models ^[14, 15]. In contrast to earlier breast cancer prognostic models that largely concentrated on restricted gene sets or clinical variables, the application of sophisticated computational algorithms to multi-cohort transcriptomic datasets offers the potential to improve predictive accuracy and clinical utility. Furthermore, Mendelian randomization provides a valuable framework for inferring causal links between gene expression and disease risk by employing genetic variants as instrumental variables, thus addressing limitations inherent in observational studies, including

reverse causation and confounding^[16].

In this work, we utilized multiple publicly accessible RNA-seq datasets containing breast cancer tissues and normal controls to identify differentially expressed CMRGs. An optimal four-gene signature model (ZMYND10, EVA1B, GREB1, CHI3L1) was evaluated and established using multivariate Cox regression. Subsequently, we explored the associations of this CMRG-based signature with immune cell infiltration, immune-related pathways, chemotherapeutic drug sensitivity, and response to immune checkpoint inhibition. To assess the causal relationships between the expression of signature genes and breast cancer risk, Mendelian randomization together with colocalization analyses were carried out, revealing strong colocalization evidence for EVA1B (PP.H4 = 0.989). Through this multidimensional analytical approach, the present study sought to identify a novel prognostic biomarker capable of both stratifying patient risk and providing insights into the immunological and pharmacological dimensions of breast cancer management.

Materials and Methods

Data Acquisition

Transcriptomic data and corresponding clinical information from 1,076 breast cancer patients and 113 normal breast tissues were obtained from The Cancer Genome Atlas (TCGA) database (<https://portal.gdc.cancer.gov/>). External validation cohorts from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>), namely GSE25066, GSE58644, GSE42568, GSE54002, GSE91061, and GSE78220, were obtained. To evaluate the predictive capability of the CMRG signature regarding immunotherapy benefits, we employed the IMvigor210 immunotherapy cohort. A set of 133 copper metabolism-related genes was retrieved from the Molecular Signatures Database (MSigDB, version 7.5.1).

Preprocessing of TCGA and GEO Data

RNA-seq data obtained from TCGA and GEO underwent a series of preprocessing steps to standardize and ensure cross-dataset consistency. Initially, normalization was carried out using the DESeq2 R package to correct for differences in sequencing depth. Subsequently, a log2 transformation was applied to stabilize variance. To further minimize technical batch effects

across different datasets, the ComBat function from the sva package was employed. These combined preprocessing steps ensured that the integrated data were suitable for downstream analysis.

Identification of Differentially Expressed Genes

To identify differentially expressed genes, we compared breast cancer tissues with normal breast tissues using the limma R package [17]. Genes meeting the criteria of an adjusted *P*-value below 0.05 and an absolute log2 fold change greater than 1 were defined as differentially expressed.

Consensus Clustering of CMRGs

To identify molecular subtypes according to CMRG expression profiles, unsupervised consensus clustering was conducted using the ConsensusClusterPlus R package [18]. The optimal cluster numbers were set at $\kappa = 2$ for CMRG subtypes and $\kappa = 4$ for gene subtypes. Kaplan-Meier survival analysis together with univariate Cox regression was employed to validate patient stratification and assess correlations with overall survival. Survival curves were plotted using the survival R package, while principal component analysis results were visualized via ggplot2 [19].

Construction of the CMRG Prognostic Signature

To characterize cuproptosis patterns at the individual patient level, we calculated an overall risk value to establish a CMRG-based risk score. Unsupervised consensus clustering was then applied to categorize breast cancer patients into four distinct molecular subtypes, designated as CMRG subtypes A through D. A prognostic risk score was subsequently developed using the training cohort. Using the caret package in R, all patient datasets were randomly split into training and testing sets at a 1:1 ratio. To reduce the risk of model overfitting, LASSO (least absolute shrinkage and selection operator) regression was performed using the glmnet package in R. A four-gene signature consisting of ZMYND10, EVA1B, GREB1, and CHI3L1 was constructed using the Stepwise Cox (both) model. The risk score was computed according to the following formula:

$$\text{Risk Score} = (-0.247 \times \text{ZMYND10 expression}) + (-0.210 \times \text{EVA1B expression}) + (-0.202 \times \text{GREB1 expression}) + (-0.291 \times \text{CHI3L1 expression}).$$

Using the median risk score as the cutoff, patients in both the training and testing sets were divided into high-risk and low-risk subgroups,

151 followed by Kaplan-Meier survival analysis and time-dependent ROC curve evaluation.

153 **Survival Analysis and Model Validation**

154 To compare survival outcomes between risk groups, Kaplan-Meier analysis along with the
155 log-rank test was employed. Time-dependent ROC curves were generated to evaluate predictive
156 accuracy at 1, 3, and 5 years. External validation of the signature's predictive performance was
157 conducted using the GSE25066 and GSE58644 cohorts.

159 **Nomogram Construction**

160 A nomogram incorporating age, clinical stage, and the CMRG-based risk score was developed.
161 Calibration curves were plotted to assess the concordance between predicted and observed 1-, 3-,
162 and 5-year overall survival probabilities. The prognostic performance of the nomogram was
163 evaluated using time-dependent ROC curve analysis across the overall, training, and validation
164 cohorts.

166 **Tumor-Infiltrating Immune Cell and Tumor Mutational Burden Analysis**

167 To assess the overall abundance and specific subsets of tumor-infiltrating immune cells within
168 each sample, the CIBERSORT algorithm [20] was applied to compare the infiltration levels of 23
169 distinct immune cell types between the high-risk and low-risk groups. The CMRG risk score was
170 additionally used to identify gene sets associated with these 23 immune cell infiltration signatures.
171 The relationship between the CMRG risk score and the cancer stem cell index (RNAss) was also
172 examined. Tumor Immune Dysfunction and Exclusion (TIDE) scores were sourced from the TIDE
173 database [21], while Immunophenoscores (IPS) were retrieved from The Cancer Immunome Atlas
174 database [22]. Tumor mutational burden (TMB) scores were extracted from the TCGA database.

176 **Drug Sensitivity Prediction**

177 Drug sensitivity predictions were obtained using the oncoPredict R package [23], This tool utilizes
178 gene expression profiles together with pharmacogenomic data from the Genomics of Drug
179 Sensitivity in Cancer database to estimate half-maximal inhibitory concentration (IC50) values for
180 a range of therapeutic agents. These include chemotherapeutic drugs (oxaliplatin, gemcitabine,

epirubicin, cisplatin), endocrine therapies (fulvestrant, tamoxifen), and targeted agents (lapatinib, palbociclib).

Mendelian Randomization Analysis

To explore potential causal links between the expression levels of the four signature genes (ZMYND10, EVA1B, GREB1, CHI3L1) and breast cancer risk, we performed two-sample Mendelian randomization analysis. Cis-eQTLs, which represent genetic variants associated with gene expression, were retrieved from publicly available eQTL summary statistics provided by the eQTLGen Consortium. Outcome data for breast cancer were sourced from the FinnGen consortium (R12 release). Instrumental variables were selected as single nucleotide polymorphisms that reached genome-wide significance ($P < 5 \times 10^{-8}$) for each gene and exhibited low linkage disequilibrium ($r^2 < 0.001$ within a 10,000 kb window). The Wald ratio method was applied for genes with a single SNP, while inverse variance weighted, MR Egger, weighted median, simple mode, and weighted mode methods were used for genes with multiple SNPs. Sensitivity analyses were conducted to assess heterogeneity (Cochran's Q test) and pleiotropy (MR Egger intercept test). Results were expressed as odds ratios with 95% confidence intervals.

Colocalization Analysis

To determine whether the identified genetic associations linking gene expression to breast cancer share a common causal variant, Bayesian colocalization analysis was carried out using the coloc package (version 5.1.0). A total of five hypotheses were evaluated: H0 (no association with either phenotype), H1 (association exclusively with phenotype 1), H2 (association exclusively with phenotype 2), H3 (association with both phenotypes but via distinct causal variants), and H4 (association with both phenotypes driven by a shared causal variant). A colocalization result was regarded as strong when PP.H4 exceeded 0.8 and as suggestive when PP.H4 was greater than 0.5.

Validation of Signature Gene Expression

Expression patterns of ZMYND10, EVA1B, GREB1, and CHI3L1 were validated using three independent datasets: TCGA, GSE42568, and GSE54002. Immunohistochemistry data were obtained from the Human Protein Atlas (HPA) database ^[24] to validate protein expression levels.

Statistical Analysis

All statistical computations were carried out using R software (version 4.2.1). For comparisons involving two groups, either the unpaired Student's t-test or the Wilcoxon rank-sum test was applied, depending on the data distribution. When comparing more than two groups, one-way analysis of variance (ANOVA) followed by Tukey's post-hoc test was utilized. Correlation analyses were performed using Spearman's rank correlation coefficient. A threshold of $P < 0.05$ was adopted to denote statistical significance.

Results

Identification of Prognostic CMRGs in Breast Cancer

A total of 133 copper metabolism-related genes (CMRGs) were retrieved from previously published reports. Utilizing the TCGA-BRCA cohort, univariate Cox regression analysis was performed to screen for CMRGs associated with overall survival (OS). This analysis yielded 30 OS-related CMRGs. Among these, elevated expression of 13 CMRGs was significantly associated with adverse OS in breast cancer, whereas reduced expression of 17 CMRGs was significantly correlated with poor OS outcomes (**Figure 1A**). Differential expression analysis of these 30 CMRGs between tumor and normal tissues revealed that only three genes exhibited significant expression differences (**Figure 1B**). Mutation profiling demonstrated that PIK3CA harbored the highest mutation frequency among the 30 CMRGs, with a mutation rate of 33% (**Figure 1C**). Chromosomal distribution analysis showed that these genes were enriched on chromosomes 1 and 7 (**Figure 1D**). Copy number variation (CNV) analysis indicated that the majority of these 30 genes exhibited higher frequencies of gain mutations compared with loss mutations, with ADAM9 showing the highest CNV mutation rate (**Figure 1E**). Finally, Kaplan-Meier survival curves were constructed for representative genes to visualize the relationship between their expression levels and OS outcomes in breast cancer patients (**Figures 1F-N**).

Identification of CMRG-Based Molecular Subtypes with Distinct Prognostic Outcomes

The TCGA-BRCA and GSE61304 datasets were integrated using R software. Unsupervised consensus clustering analysis was subsequently conducted on the merged cohort based on the

expression profiles of 133 CMRGs. The results indicated that optimal patient classification was achieved when the cohort was partitioned into two distinct molecular subtypes (Subtype A: n = 626; Subtype B: n = 512), with the optimal number of clusters determined at $\kappa = 2$ (**Figures 2A-B**). Principal component analysis (PCA) further confirmed that Subtype A and Subtype B formed clearly distinguishable subgroups (**Figure 2C**). Notably, patients in Subtype A experienced significantly longer overall survival compared with those in Subtype B ($P < 0.001$; **Figure 2D**), which was accompanied by markedly different CMRG expression profiles and clinicopathological characteristics (**Figure 2E**).

Functional Characterization and Immune Microenvironment Heterogeneity Between CMRG Subtypes

Gene Set Variation Analysis highlighted subtype-specific pathway activation: Subtype A showed enrichment in immune-related pathways, whereas Subtype B displayed suppression of metabolic pathways (**Figure 3A**). CIBERSORT-based immune profiling revealed significant differences in immune cell subsets between subtypes (**Figure 3B**). A total of 230 differentially expressed genes (DEGs) were identified by comparing Subtype A and Subtype B, among which 146 were downregulated and 84 were upregulated. Functional annotation of 230 differentially expressed genes via GO and KEGG analyses implicated CMRGs in humoral immunity, granulocyte migration, cell cycle, chemokine signaling, and cellular senescence (**Figures 3C-D**).

Gene subtypes are identified using DEGs

Subsequent prognostic analysis was conducted on the 230 differentially expressed genes, which identified 69 CMRGs significantly associated with overall survival. Among these, 55 genes were characterized as protective factors (where lower expression correlated with poor prognosis), while 14 genes were classified as risk factors (where higher expression correlated with poor prognosis). Unsupervised consensus clustering was then performed based on these 69 OS-related CMRGs to classify breast cancer patients. Applying this approach, 1,139 patients were partitioned into four distinct molecular subtypes (Subtypes A-D), with the optimal number of clusters determined at $\kappa = 4$ (Subtype A: n = 195; Subtype B: n = 265; Subtype C: n = 429; Subtype D: n = 249; **Figure 4A**). The associations between clinicopathological characteristics and these gene subtypes were

further investigated (**Figure 4B**). Kaplan-Meier survival curves demonstrated that patients assigned to Subtype A experienced the poorest overall survival, whereas those in Subtype C exhibited the most favorable outcomes ($P < 0.001$; **Figure 4C**). As anticipated, the expression profiles of CMRGs showed substantial variation across the four gene subtypes (**Figure 4D**).

Development and Validation of a Cox regression-Based CMRG Signature

Patients were randomly assigned to training ($n = 568$) and testing ($n = 565$) cohorts using the caret package in R. Based on the least partial likelihood deviation, LASSO regression together with multivariate Cox regression analysis identified four CMRG-related genes from a total of 230 prognostic DEGs associated with CMRG subtypes (**Figure 5A-B**). These four genes—ZMYND10, EVA1B, GREB1, and CHI3L1—were used to construct a risk score model. The risk score was computed using the following formula: Risk Score = $(-0.247 \times \text{ZMYND10 expression}) + (-0.210 \times \text{EVA1B expression}) + (-0.202 \times \text{GREB1 expression}) + (-0.291 \times \text{CHI3L1 expression})$. The distribution of risk scores across the two CMRG subtypes is presented in **Figure 5C**. Gene subtype D exhibited the highest CRG scores, whereas gene subtype C showed the lowest (**Figure 5D**).

Prognostic Performance of the CMRG Signature Across Independent Cohorts

Using the median risk score as a cutoff, patients were categorized into high-risk and low-risk groups, which revealed marked differences in survival outcomes. Across both the training and validation cohorts, individuals with a high CMRG score had significantly shorter overall survival (**Figures 6A-D**). In the training cohort, the high-risk subgroup experienced notably poorer survival, and this observation was consistently reproduced in the test cohort (**Figures 6E-F**). External validation of the CMRG prognostic signature was performed using the GSE25066 and GSE58644 cohorts, yielding results that aligned with those from the TCGA training cohort—namely, the high-risk subgroup showed significantly worse survival. Time-dependent ROC analysis further confirmed the robust prognostic performance of the signature in both external validation cohorts, with AUC values exceeding 0.6 (**Figures 7**).

Construction of a Clinical Nomogram

To support clinical application, a nomogram was constructed by incorporating the CMRG score together with key clinical variables (age and clinical stage) (**Figure 8A**). Calibration curves revealed good concordance between predicted and observed survival probabilities in both the training and validation sets (**Figure 8B**), confirming the nomogram's usefulness for individualized risk assessment. The nomogram's ROC curves consistently showed strong prognostic performance across the overall cohort, as well as in the training and validation cohorts (**Figure 8C-E**).

Correlation Between CMRG Score and Immune Microenvironment

Immune deconvolution analysis uncovered a notable inverse relationship between the CMRG score and immune activation. Elevated CMRG scores were linked to decreased infiltration of dendritic cells, M1 macrophages, and CD8⁺T cells (**Figure 9A**). Unlike CHI3L1, the remaining three signature genes (ZMYND10, EVA1B, and GREB1) showed significantly positive associations with the infiltration levels of resting mast cells, resting dendritic cells, and monocytes. Conversely, they exhibited significantly negative associations with the infiltration levels of follicular helper T cells, activated memory CD4⁺T cells, M1 macrophages, and activated dendritic cells (**Figure 9B**). Moreover, the CMRG risk score was found to be significantly positively correlated with tumor mutational burden in breast cancer (**Figure 9C**). Considerable differences in mutation frequencies were observed when comparing the high-risk and low-risk subgroups. Specifically, PIK3CA, TTN, and CDH1 had higher mutation rates in the low-risk group, whereas TP53 showed a higher mutation rate in the high-risk group (**Figures 9D-E**). Additionally, a significantly positive correlation was observed between the CMRG risk score and the RNA expression-based stemness index (RNAss) (**Figure 9F**).

CMRG Score as a Predictor of Immunotherapy Response

To evaluate the utility of the CMRG risk score for predicting immunotherapy benefits, we examined the GSE91061 and IMvigor210 cohorts. Patients were classified into high-risk and low-risk groups according to their CMRG scores. Within the GSE91061 cohort, those in the low-risk group had significantly prolonged overall survival compared to their high-risk counterparts (Wilcoxon test, $P = 0.048$; **Figure 10A**). Kaplan-Meier analysis further showed that high-risk patients tended to experience earlier death than low-risk patients, though this difference

331 did not achieve statistical significance (log-rank test, $P = 0.073$; **Figure 10B**). The clinical
332 response rate (CR/PR versus SD/PD) was also higher in the low-risk group (**Figure 10C**). Similar
333 results were observed in the IMvigor210 cohort (**Figures 10D-F**). Taken together, these findings
334 suggest that the CMRG risk score is capable of predicting immunotherapy response.

336 **CMRG Score and Drug Sensitivity**

337 The low-risk group exhibited markedly higher IPS-CTLA4-neg-PD1-neg scores relative to the
338 high-risk group (**Figures 11A-D**). Drug sensitivity analysis further demonstrated that patients with
339 a low CMRG score showed significantly enhanced sensitivity to a wide range of therapeutic
340 agents, including chemotherapeutics (oxaliplatin, gemcitabine, epirubicin, cisplatin), endocrine
341 agents (fulvestrant, tamoxifen), and targeted drugs (lapatinib, palbociclib) (**Figures 11E-P**).
342 Collectively, these results support the CMRG signature as a promising biomarker for guiding
343 personalized treatment decisions, thereby extending its applicability beyond immunotherapy.

345 **Mendelian Randomization Reveals Causal Effects of Signature Genes**

346 We conducted two-sample Mendelian randomization analysis to assess potential causal
347 relationships between the expression levels of the four signature genes (ZMYND10, EVA1B,
348 GREB1, CHI3L1) and breast cancer risk. For ZMYND10, only one instrumental SNP was
349 available; the Wald ratio method yielded a non-significant causal estimate (OR = 1.05, 95% CI:
350 0.90 - 1.23, $P = 0.530$). For EVA1B, which had nine instrumental SNPs, the inverse variance
351 weighted method revealed a significant positive causal effect of EVA1B expression on breast
352 cancer risk (OR = 1.05, 95% CI: 1.01 - 1.10, $P = 0.025$). The weighted median method produced
353 a marginally significant estimate in the same direction (OR = 1.05, 95% CI: 1.00 - 1.10, $P =$
354 0.052). No evidence of horizontal pleiotropy was found (MR Egger intercept $P = 0.935$), and
355 Cochran's Q test indicated no significant heterogeneity ($Q = 8.58$, $P = 0.379$). For CHI3L1, the
356 inverse variance weighted method showed no significant causal effect (OR = 0.99, 95% CI:
357 0.97 - 1.02, $P = 0.483$). GREB1 was not analyzed due to insufficient instrumental variables (only
358 one SNP available in the exposure data) (**Table 1**).

360 **Colocalization Analysis Supports Shared Genetic Architecture for EVA1B**

To determine whether the genetic associations with gene expression and breast cancer share a common causal variant, we carried out Bayesian colocalization analysis. Among the four signature genes, EVA1B showed strong colocalization evidence, with a posterior probability for a shared causal variant (PP.H4) of 0.989—indicating that the same variant is more than 98% likely to drive both EVA1B expression and breast cancer risk. CHI3L1 exhibited weak colocalization evidence (PP.H4 = 0.007). For GREB1, the posterior probability of colocalization was moderate (PP.H4 = 0.195), while ZMYND10 showed no support for colocalization (PP.H4 = 0.015). These findings indicate that among the four genes, EVA1B represents the strongest candidate with a shared genetic architecture connecting its expression to breast cancer susceptibility (**Table 2**).

Validation of Signature Gene Expression Patterns Across Multiple Cohorts

To further validate the expression patterns of the four core genes (ZMYND10, EVA1B, GREB1, and CHI3L1) in breast cancer, we examined three independent datasets: TCGA, GSE42568, and GSE54002. In the TCGA cohort, EVA1B, GREB1, and CHI3L1 were significantly upregulated in breast cancer tissues compared to normal breast tissues, whereas ZMYND10 was significantly downregulated (**Figure 12A**). Analysis of the GSE42568 dataset revealed markedly increased expression levels of ZMYND10, GREB1, and CHI3L1 in breast cancer samples (**Figure 12B**). In the GSE54002 dataset, ZMYND10 and EVA1B were significantly upregulated, while CHI3L1 showed significant downregulation in breast cancer tissues (**Figure 12C**). Immunohistochemistry data retrieved from the Human Protein Atlas (HPA) database further confirmed that EVA1B protein expression was significantly higher in breast cancer tissues than in normal breast tissues (**Figure 12D**).

Discussion

Breast cancer continues to be one of the most common and deadly malignancies in women, presenting substantial public health challenges due to its elevated incidence and mortality rates ^[1]. In this study, we applied multivariate Cox regression to develop a novel four-gene prognostic signature (ZMYND10, EVA1B, GREB1, CHI3L1) derived from copper metabolism-related genes for breast cancer. In comparison with previously reported cuproptosis-related signatures ^[11-13], our model demonstrated superior predictive performance (C-index = 0.79) and was externally

validated in multiple independent cohorts (GSE25066, GSE58644, METABRIC). This signature demonstrated robust discrimination between patient risk groups and revealed significant associations with immune cell infiltration and drug sensitivity profiles, suggesting its potential utility in guiding personalized treatment decisions. Furthermore, Mendelian randomization analysis provided causal evidence linking EVA1B expression to breast cancer risk (OR = 1.05, 95% CI: 1.01 – 1.10, $P = 0.025$), with strong colocalization support (PP.H4 = 0.989), strengthening the biological rationale for this signature. The present findings establish a foundation for the further exploration of copper metabolism pathways in breast cancer progression and therapeutic response, offering novel insights into the molecular mechanisms that may underlie tumor immune evasion and chemoresistance.

The four genes included in the signature play distinct yet interconnected roles in tumor biology and immune regulation. ZMYND10 (zinc finger MYND domain-containing protein 10) is involved in microtubule stabilization and has been associated with cell cycle regulation [25]. Its decreased expression in high-risk patients may contribute to genomic instability and more aggressive tumor behavior. EVA1B has been shown to promote esophageal squamous carcinoma progression and facilitate the recruitment of immunosuppressive myeloid-derived suppressor cells into the tumor microenvironment [26]. The significant causal link between EVA1B expression and breast cancer risk identified through Mendelian randomization, coupled with strong colocalization evidence (PP.H4 = 0.989), suggests that EVA1B may directly participate in breast cancer pathogenesis rather than functioning solely as a correlative biomarker. GREB1 (gene regulated by estrogen in breast cancer 1) is an estrogen-responsive gene involved in the progression of hormone receptor-positive breast cancer [27]. Its presence in the signature emphasizes the importance of hormone signaling pathways in prognostic stratification associated with copper metabolism. CHI3L1 (chitinase-3-like protein 1), also known as YKL-40, is a secreted glycoprotein that promotes tumor angiogenesis, invasion, and immune suppression [28]. The capacity of these genes to influence tumor cell survival and immune modulation underscores the biological plausibility of this signature.

The inverse correlation between high CMRG scores and immune cell infiltration, particularly cytotoxic CD8⁺ T cells and dendritic cells, underscores a potential immunosuppressive tumor microenvironment [29, 30]. This observation parallels findings that high-risk patients, as defined by

programmed cell death signatures, exhibit diminished immune activation and increased immune escape mechanisms. From a mechanistic perspective, copper metabolism may affect antigen presentation and immune surveillance via mitochondrial dysfunction and alterations in metabolic states, ultimately resulting in reduced recruitment or impaired function of effector immune cells [31]. This observation was supported by our single-sample gene set enrichment analysis, which revealed diminished expression of immune activation-related gene sets in the high-risk group, indicating suppression of key immune pathways including interferon signaling and T-cell cytotoxicity. Collectively, these results suggest that the CMRG signature not only predicts survival outcomes but also mirrors the immune landscape, offering a rationale for incorporating CMRG assessment into immunotherapy stratification strategies.

The observed differences in chemotherapeutic agent sensitivity between high-risk and low-risk patients point to a clinically meaningful association between CMRG expression and drug response [32, 33]. Low-risk patients showed heightened sensitivity to agents like oxaliplatin and gemcitabine, which is consistent with earlier studies linking metabolic pathways to chemosensitivity [34]. Copper metabolism may influence drug efficacy by impacting mitochondrial function and reactive oxygen species production—key factors in chemotherapy-induced cell death [35]. Moreover, machine learning-based prognostic models that incorporate CMRGs have shown potential in predicting chemotherapy response in breast cancer, indicating that the CMRG signature could serve as a biomarker for guiding personalized chemotherapy regimens [36].

The finding that patients with low CMRG scores show better response to immune checkpoint inhibitors across three independent immunotherapy cohorts (IMvigor210, GSE91061, GSE78220) has important clinical implications. The low-risk group exhibited a combination of features—including elevated tumor mutational burden, higher immunophenoscore, and lower Tumor Immune Dysfunction and Exclusion scores—suggesting the presence of an immunologically "hot" tumor microenvironment that favors responsiveness to immune checkpoint blockade [37]. This observation aligns with emerging evidence that metabolic pathways, including copper metabolism, can modulate PD-L1 expression and influence tumor immune evasion [38]. The present results suggest that the CMRG signature could be used to identify patients most likely to benefit from immunotherapy, thereby enabling more precise patient selection and improved resource allocation. The application of Mendelian randomization analysis in the present study addresses a critical

451 limitation of observational biomarker studies by providing evidence for causal relationships. The
452 significant causal effect of EVA1B expression on breast cancer risk, supported by robust
453 colocalization evidence (PP.H4 = 0.989), suggests that EVA1B may be directly involved in breast
454 cancer pathogenesis. This finding elevates EVA1B from a mere prognostic marker to a potential
455 therapeutic target. The lack of significant causal effects for ZMYND10 and CHI3L1, despite their
456 prognostic relevance, highlights an important distinction between biomarkers of disease
457 progression and causal drivers of disease susceptibility. These results underscore the value of
458 integrating genetic epidemiology approaches with transcriptomic biomarker discovery to prioritize
459 targets for further functional validation.

460 Several limitations of the present study should be acknowledged. First, the retrospective nature of
461 the analysis using public datasets necessitates prospective validation through multicenter studies
462 with clinical specimens. Second, while the Mendelian randomization analysis provides evidence
463 for causality, the specific molecular mechanisms through which these genes influence breast
464 cancer progression require further experimental investigation. Third, our predictions regarding
465 immunotherapy response are derived from computational analyses that rely primarily on cancer
466 cell line-based datasets and public databases. Consequently, these results should be interpreted
467 with caution and will require additional validation using experimental models and well-designed
468 clinical cohorts. Fourth, batch effects inherent in the datasets we employed may introduce
469 variability that could confound the analysis, underscoring the need for careful interpretation when
470 translating these findings into clinical settings.

471 Future studies should include experimental validation of the core CMRG genes using western
472 blotting and reverse transcription-quantitative PCR in independent collections of human breast
473 cancer tissues to confirm their protein and mRNA expression patterns and solidify their clinical
474 relevance. Functional studies using gain-of-function and loss-of-function approaches in breast
475 cancer cell lines and animal models are needed to elucidate the precise molecular mechanisms by
476 which these genes, particularly EVA1B, influence tumor progression, immune modulation, and
477 treatment response. Additionally, prospective clinical trials should evaluate the utility of the
478 CMRG signature for guiding treatment decisions, including patient selection for immunotherapy
479 and chemotherapy regimens.

Conclusion

In conclusion, this study successfully identified multiple copper metabolism-related genes associated with breast cancer prognosis and developed a machine learning-based prognostic signature using gene expression data, highlighting its potential for predicting patient outcomes and treatment responses. Mendelian randomization analysis offered causal evidence linking EVA1B expression to breast cancer risk, supported by strong colocalization evidence. This approach not only improves prognostic capabilities but also establishes a foundation for personalized therapeutic strategies. The CMRG signature effectively stratifies patient risk, offers insights into the immune microenvironment, and serves as a valuable biomarker for predicting prognosis, immunotherapy response, and drug sensitivity. Future efforts should focus on validating these findings through larger, multi-center studies and exploring the clinical applicability of the CMRG model, with the ultimate goal of integrating these biomarkers into routine clinical practice to enhance therapeutic decision-making for breast cancer patients.

Abbreviation

BC	Breast cancer
CI	Confidence interval
CMRG	Copper metabolism-related gene
CNV	Copy number variation
CR	Complete response
DEG	Differentially expressed gene
eQTL	Expression quantitative trait locus
GEO	Gene Expression Omnibus
GO	Gene Ontology
GSVA	Gene Set Variation Analysis
HPA	Human Protein Atlas
IC50	Half-maximal inhibitory concentration
IPS	Immunophenoscore
IVW	Inverse variance weighted
KEGG	Kyoto Encyclopedia of Genes and Genomes

511 LASSO Least absolute shrinkage and selection operator

512 MR Mendelian randomization

513 MSigDB Molecular Signatures Database

514 NA Not applicable

515 OS Overall survival

516 PCA Principal component analysis

517 PD Progressive disease

518 PR Partial response

519 ROC Receiver operating characteristic

520 RNAss RNA expression-based stemness index

521 SD Stable disease

522 SNP Single nucleotide polymorphism

523 TCGA The Cancer Genome Atlas

524 TIDE Tumor Immune Dysfunction and Exclusion

525 TMB Tumor mutational burden

526

527 **Declarations**

528 **Acknowledgments**

529 All authors would like to thank all publicly available data used in the study.

530

531 **Authorship contribution**

532 **Weizhong Chen:** Conceptualization, Methodology, Software, Formal analysis, Investigation,

533 Writing - original draft. **Tao Chen:** Conceptualization, Methodology, Validation, Formal analysis,

534 Writing - original draft. **Yanfeng Chen:** Data curation, Software, Validation, Investigation. **Cuicai**

535 **Lin:** Data curation, Formal analysis, Visualization. **Wenhan Chen:** Investigation, Validation,

536 Resources. **Xiaoru Wu:** Investigation, Visualization, Data curation. **Lusheng Liao:** Resources,

537 Funding acquisition, Validation, Software. **Mingyou Dong:** Conceptualization, Funding

538 acquisition, Project administration, Supervision, Writing - review & editing. **Zhichao Fang:**

539 Conceptualization, Supervision, Project administration, Writing - review & editing.

540

Funding

This work was supported by the following grants: Guangxi Youth Science Foundation (2023JJB140064), Natural Science Foundation of Guangxi Zhuang Autonomous Region (2026GXNSFHA00640149), and Baise City Science and Technology Plan Self-Funded Project (Baiké 202537042) to Mingyou Dong; Natural Science Foundation of Guangxi Zhuang Autonomous Region (2026GXNSFHA00640108) to Lusheng Liao.

Clinical trial number

Not applicable.

Ethics Declaration

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Conflicts of interest

All of the authors mentioned in this paper affirm that they possess no conflicts of interest.

Data Availability

The datasets generated and analyzed during the current study are available from public repositories. The TCGA-BRCA breast cancer dataset is accessible via the GDC data portal (<https://portal.gdc.cancer.gov/>). The GEO datasets GSE25066, GSE58644, GSE42568, GSE54002, GSE91061, GSE78220, and GSE61304 are available from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>). The IMvigor210 immunotherapy cohort was obtained from the supplementary materials of the original publication. The copper metabolism-related gene set (MSigDB v7.5.1) was retrieved from the Molecular Signatures Database

(<https://www.gsea-msigdb.org/gsea/msigdb/>). eQTL summary statistics were obtained from the eQTLGen Consortium (<https://www.eqtlgen.org/>). Breast cancer outcome data were sourced from FinnGen release R12 (<https://www.finnngen.fi/>). Tumor Immune Dysfunction and Exclusion (TIDE) scores were accessed via the TIDE database (<http://tide.dfci.harvard.edu/>). Immunophenoscores (IPS) were obtained from The Cancer Immunome Atlas (TCIA, <https://tcia.at/>). Immunohistochemistry data were retrieved from the Human Protein Atlas (HPA, <https://www.proteinatlas.org/>).

Acknowledgments

The authors acknowledge data contributions from the TCGA, GEO, FinnGen, and eQTLGen databases. The authors thank the Human Protein Atlas for providing immunohistochemistry data.

References

- [1] Sung H, Ferlay J, Siegel R L, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries[J]. CA: a cancer journal for clinicians, 2021,71(3):209-249.
- [2] Siegel R L, Kratzer T B, Giaquinto A N, et al. Cancer statistics, 2025[J]. CA: a cancer journal for clinicians, 2025,75(1):10-45.
- [3] Kim B, Nevitt T, Thiele D J. Mechanisms for copper acquisition, distribution and regulation[J]. Nature chemical biology, 2008,4(3):176-185.
- [4] Solomon E I, Sundaram U M, Machonkin T E. Multicopper Oxidases and Oxygenases[J]. Chemical reviews, 1996,96(7):2563-2606.
- [5] Ge E J, Bush A I, Casini A, et al. Connecting copper and cancer: from transition metal signalling to metalloplasia[J]. Nature reviews. Cancer, 2022,22(2):102-113.
- [6] Blockhuys S, Celauro E, Hildesj C, et al. Defining the human copper proteome and analysis of its expression variation in cancers[J]. Metallomics : integrated biometal science, 2017,9(2):112-123.
- [7] Tsvetkov P, Coy S, Petrova B, et al. Copper induces cell death by targeting lipoylated TCA cycle proteins[J]. Science (New York, N.Y.), 2022,375(6586):1254-1261.
- [8] Yuan J, Ofengeim D. A guide to cell death pathways[J]. Nature reviews. Molecular cell biology, 2024,25(5):379-395.
- [9] Liu J, Hong M, Li Y, et al. Programmed Cell Death Tunes Tumor Immunity[J]. Frontiers in

immunology, 2022,13:847345.

[10] Cobine P A, Brady D C. Cuproptosis: Cellular and molecular mechanisms underlying copper-induced cell death[J]. Molecular cell, 2022,82(10):1786-1787.

[11] Zhang Z, Gao Z, Huang Q, et al. Cuproptosis- and m6A-related lncRNA prognostic signature for breast cancer, in which Z68871.1 contributes to triple-negative breast cancer progression[J]. International journal of biological macromolecules, 2025,321(Pt 2):146321.

[12] Jiang B, Zhu H, Feng W, et al. Database Mining Detected a Cuproptosis-Related Prognostic Signature and a Related Regulatory Axis in Breast Cancer[J]. Disease markers, 2022,2022:9004830.

[13] Li D, Qu G, Ling S, et al. A cuproptosis-related lncRNA signature to predict prognosis and immune microenvironment of colon adenocarcinoma[J]. Scientific reports, 2023,13(1):6284.

[14] Huang L, Zhang L, Shi X, et al. Multi-cohort and single-cell profiling of aging genes reveals prognostic and therapeutic targets in breast cancer[J]. iScience, 2026,29(3):114847.

[15] Zeng C, Wang J, Zhao S, et al. Multi-cohort validation of a lipid metabolism and ferroptosis-associated index for prognosis and immunotherapy response prediction in hormone receptor-positive breast cancer[J]. International journal of biological sciences, 2025,21(9):3968-3992.

[16] Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies[J]. Human molecular genetics, 2014,23(R1):R89-R98.

[17] Ritchie M E, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies[J]. Nucleic acids research, 2015,43(7):e47.

[18] Wilkerson M D, Hayes D N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking[J]. Bioinformatics (Oxford, England), 2010,26(12):1572-1573.

[19] Ito K, Murphy D. Application of ggplot2 to Pharmacometric Graphics[J]. CPT: pharmacometrics & systems pharmacology, 2013,2(10):e79.

[20] Chen B, Khodadoust M S, Liu C L, et al. Profiling Tumor Infiltrating Immune Cells with CIBERSORT[J]. Methods in molecular biology (Clifton, N.J.), 2018,1711:243-259.

[21] Jiang P, Gu S, Pan D, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response[J]. Nature medicine, 2018,24(10):1550-1558.

[22] Charoentong P, Finotello F, Angelova M, et al. Pan-cancer Immunogenomic Analyses Reveal

Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade[J]. *Cell reports*, 2017,18(1):248-262.

[23] Maeser D, Gruener R F, Huang R S. oncoPredict: an R package for predicting in vivo or cancer patient drug response and biomarkers from cell line screening data[J]. *Briefings in bioinformatics*, 2021,22(6):bbab260.

[24] Uhlén M, Fagerberg L, Hallstr M B R M, et al. Proteomics. Tissue-based map of the human proteome[J]. *Science (New York, N.Y.)*, 2015,347(6220):1260419.

[25] Shi L, Shen Y. Silencing of ciliary protein ZMYND10 affects amitotic macronucleus division in *Paramecium tetraurelia*[J]. *European journal of protistology*, 2022,82:125863.

[26] Xia T, Zhang Y, Peng H, et al. EVA1B facilitates esophageal squamous carcinoma progression and recruitment of immunosuppressive myeloid-derived suppressor cells in the tumor microenvironment[J]. *Pharmacological research*, 2024,210:107521.

[27] Mohammed H, D'Santos C, Serandour A A, et al. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor[J]. *Cell reports*, 2013,3(2):342-349.

[28] Libreros S, Iragavarapu-Charyulu V. YKL-40/CHI3L1 drives inflammation on the road of tumor progression[J]. *Journal of leukocyte biology*, 2015,98(6):931-936.

[29] Kandala S, Ramos M, Voith Von Voithenberg L, et al. Chronic chromosome instability induced by Plk1 results in immune suppression in breast cancer[J]. *Cell reports*, 2023,42(12):113266.

[30] Huang R, Wang H, Hong J, et al. Targeting glutamine metabolic reprogramming of SLC7A5 enhances the efficacy of anti-PD-1 in triple-negative breast cancer[J]. *Frontiers in immunology*, 2023,14:1251643.

[31] Liu W, Lin W, Yan L, et al. Copper homeostasis and cuproptosis in cancer immunity and therapy[J]. *Immunological reviews*, 2024,321(1):211-227.

[32] Zhao P, Yin S, Qiu Y, et al. Ferroptosis and pyroptosis are connected through autophagy: a new perspective of overcoming drug resistance[J]. *Molecular cancer*, 2025,24(1):23.

[33] Su L, Chen Y, Huang C, et al. Targeting Src reactivates pyroptosis to reverse chemoresistance in lung and pancreatic cancer models[J]. *Science translational medicine*, 2023,15(678):eab17895.

[34] Wan H, Yang X, Sang G, et al. CDKN2A was a cuproptosis-related gene in regulating chemotherapy resistance by the MAGE-A family in breast cancer: based on artificial intelligence

(AI)-constructed pan-cancer risk model[J]. Aging, 2023,15(20):11244-11267.

[35] Chen X, Li J, Kang R, et al. Ferroptosis: machinery and regulation[J]. Autophagy, 2021,17(9):2054-2081.

[36] Li S, Liu S, Zheng Y, et al. Machine learning analysis of coagulation-related genes for breast cancer diagnosis and prognosis prediction[J]. Scientific reports, 2025,15(1):35429.

[37] Luchini C, Bibeau F, Ligtenberg M J L, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach[J]. Annals of oncology : official journal of the European Society for Medical Oncology, 2019,30(8):1232-1243.

[38] Voli F, Valli E, Lerra L, et al. Intratumoral Copper Modulates PD-L1 Expression and Influences Tumor Immune Evasion[J]. Cancer research, 2020,80(19):4129-4144.

691

692

693

694

695

696

697

698

699

700

701

Table Legends

702 **Table 1. Causal associations between signature gene expression and breast cancer risk from**
703 **two-sample MR analysis**

Exposure	Method	nsnp	OR (95% CI)	P-value	Heterogeneity Q (P)	Pleiotropy P
ZMYND10	Wald ratio	1	1.05 (0.90– 1.23)	0.530	NA	NA
EVA1B	IVW	9	1.05 (1.01– 1.10)	0.025	0.379	0.935
CHI3L1	IVW	37	0.99 (0.97– 1.02)	0.483	0.877	0.242

704 IVW, inverse variance weighted; NA, not applicable.

705

706 **Table 2. Bayesian colocalization analysis of signature genes**

Gene	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4	Interpretation
ZMYND10	1.18e-07	0.974	1.38e-09	0.011	0.015	No colocalization

Gene	PP.H0	PP.H1	PP.H2	PP.H3	PP.H4	Interpretation
EVA1B	3.81e-134	0.947	4.26e-136	0.011	0.989	Strong colocalization
GREB1	0.067	0.717	0.002	0.019	0.195	Weak colocalization
CHI3L1	<0.001	0.985	<0.001	0.008	0.007	No colocalization

707
708
709
710
711
712
713
714

Figure Legends

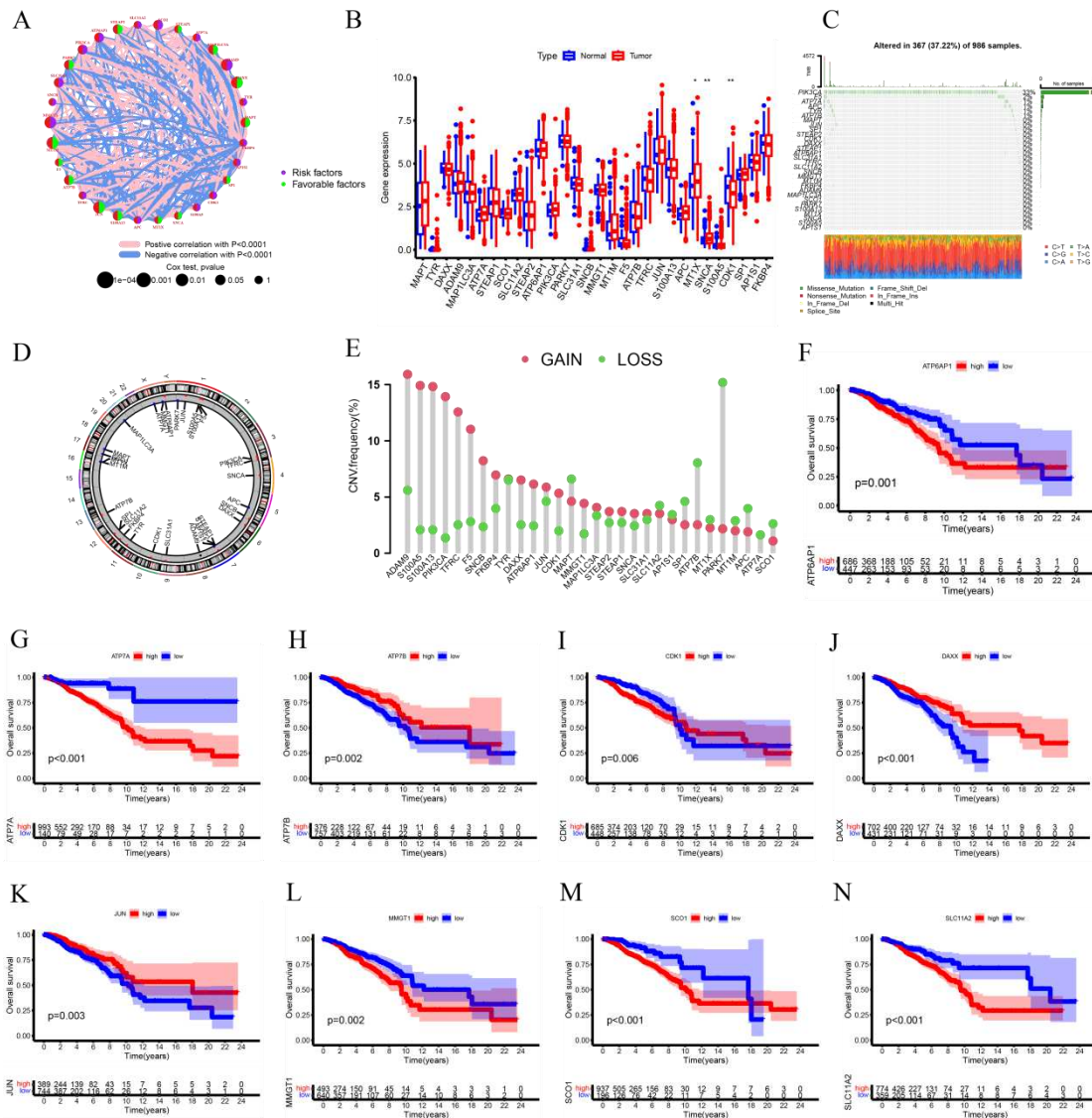


Figure 1. Expression profiling and clinical significance of 30 CMRGs in TCGA breast cancer cohort. (A) Differential expression of 30 CMRGs between breast cancer tissues and normal breast tissues (* $p < 0.05$; *** $p < 0.001$). (B) Mutation frequency landscape of CMRGs in 986 breast cancer patients. (C) Chromosomal distribution of copy number variation (CNV) alterations in CMRGs. (D) Frequency distribution of CNV gains, losses, and stable regions among CMRGs. (E-N) Kaplan-Meier survival analysis demonstrating the prognostic value of 10 representative CMRGs.

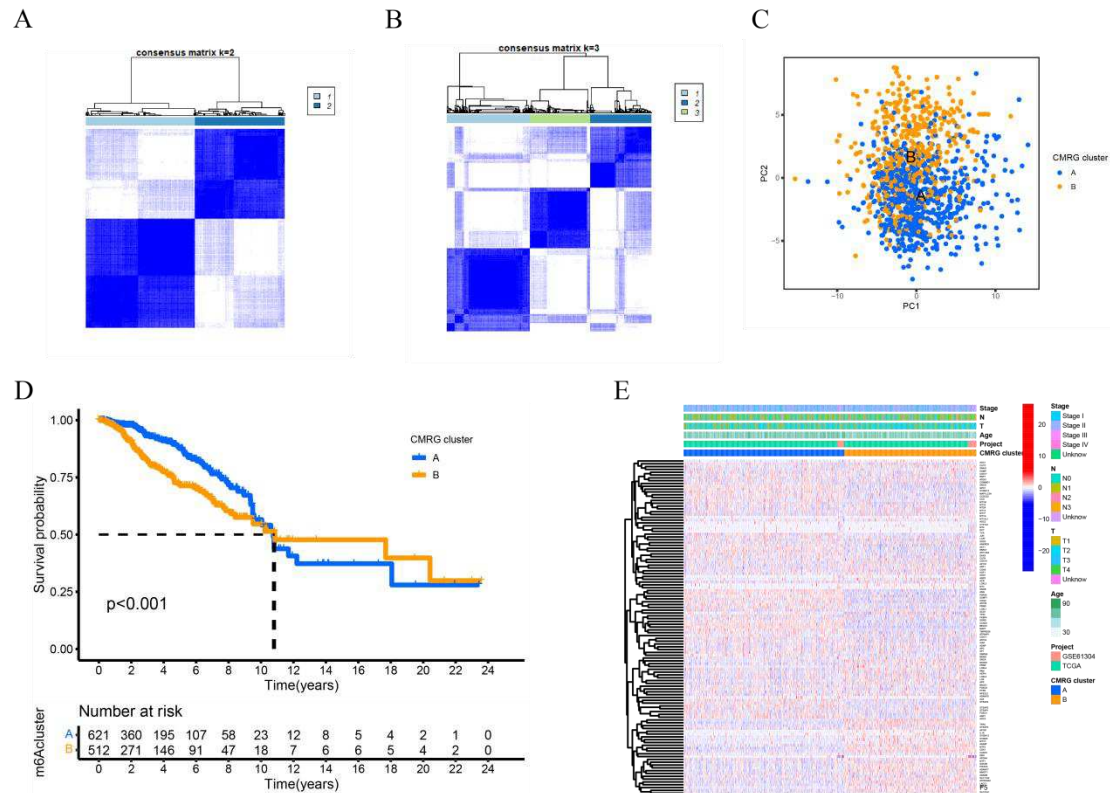


Figure 2. Molecular and clinical characterization of CMRG-based subtypes. (A) Protein-protein interaction network of CMRGs in breast cancer (red and blue lines indicate positive and negative correlations, respectively; color intensity reflects correlation strength). (B) Consensus clustering heatmap identifying two distinct molecular subtypes ($\kappa=2$). (C) Principal component analysis (PCA) plot showing significant transcriptomic differences between subtypes. (D) Kaplan-Meier curves illustrating distinct overall survival outcomes between subtypes. (E) Heatmap displaying subtype-specific CMRG expression patterns and associated clinicopathological features.

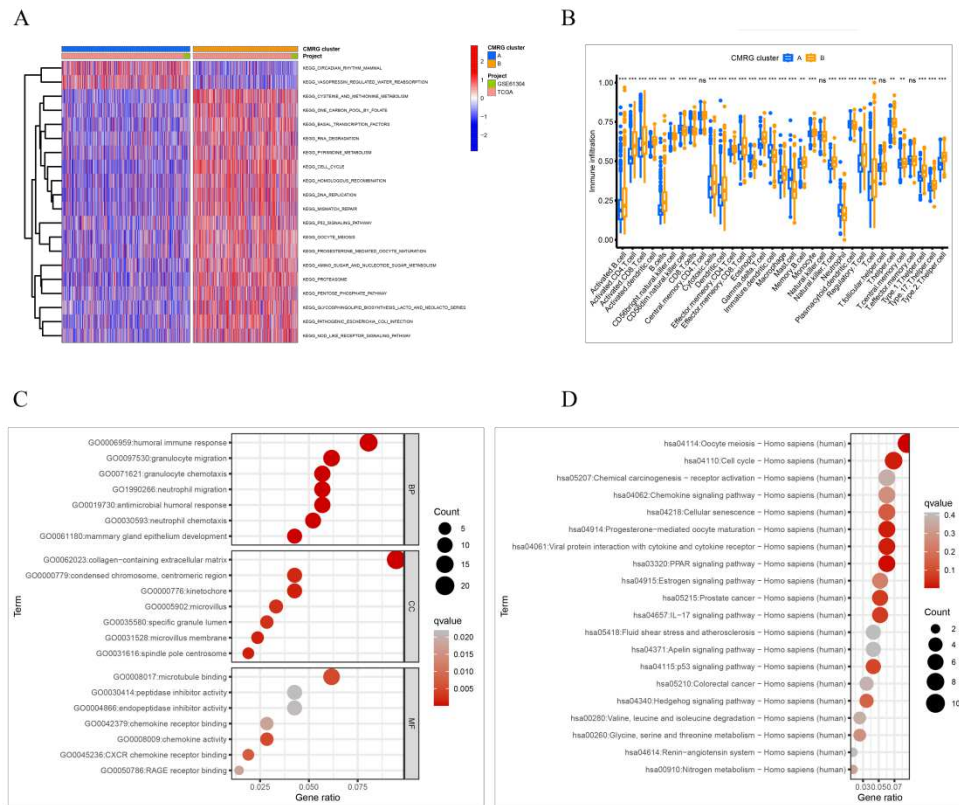


Figure 3. Tumor microenvironment analysis of CMRG subtypes. (A) Gene Set Variation Analysis (GSVA) of hallmark pathways in CMRG-defined subtypes (red: activated pathways; blue: suppressed pathways). (B) Differential immune cell infiltration patterns between CMRG subtypes. (C-D) Functional enrichment analysis of differentially expressed genes (DEGs) using Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

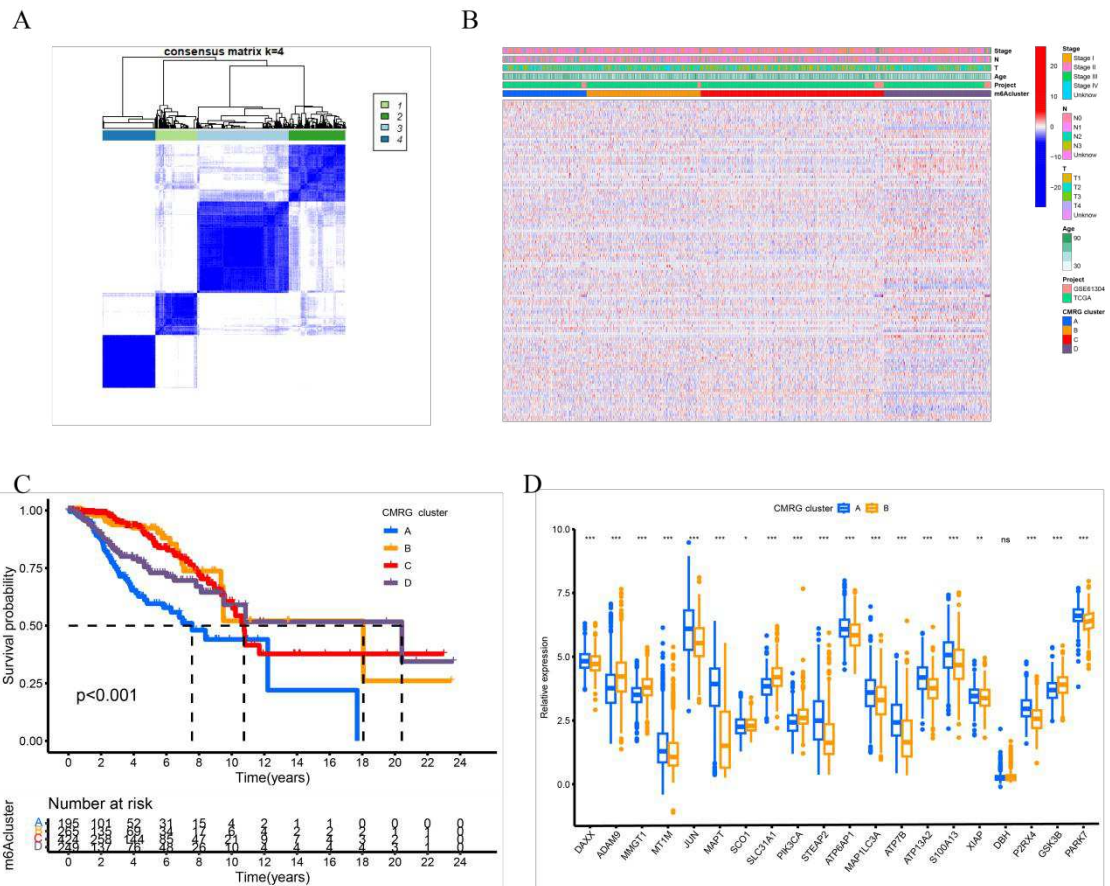


Figure 4. Molecular subtyping based on DEG profiles. (A) Consensus clustering heatmap defining four genomic subtypes ($\kappa=4$). (B) Clinicopathological characteristics across the four molecular subtypes. (C) Survival outcomes stratified by genomic subtype. (D) Expression patterns of 20 key CMRGs across the four subtypes ($*p < 0.05$, $**p < 0.01$, $***p < 0.001$).

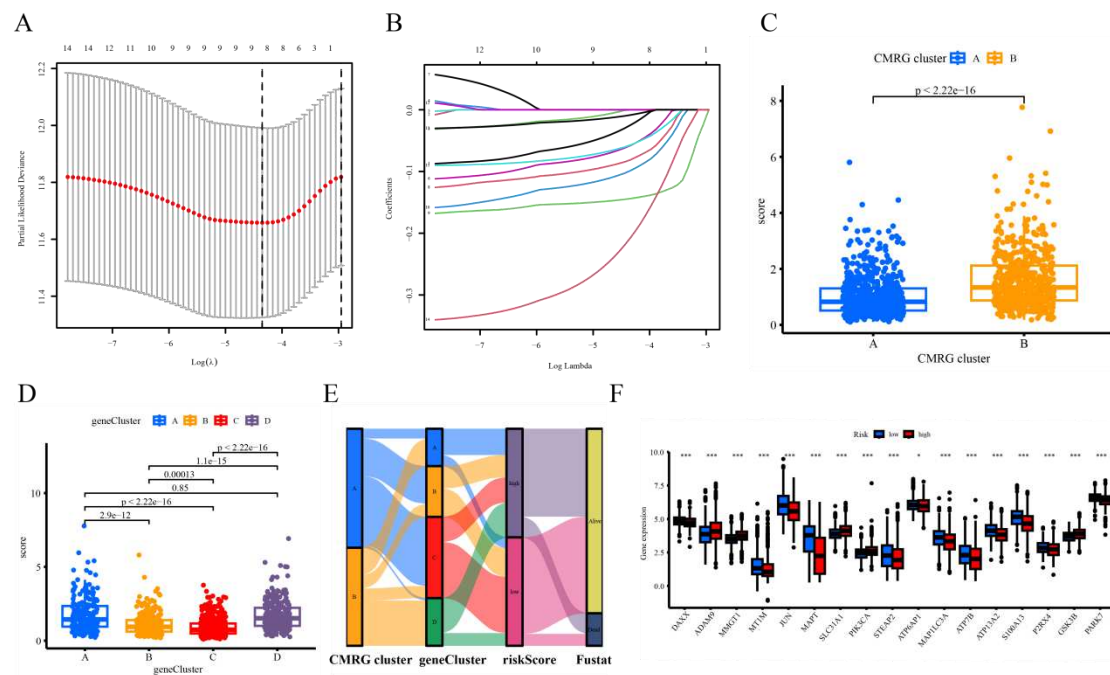


Figure 5. Development of CMRG prognostic signature. (A) Distribution of clinical subtypes, CMRG risk scores, and survival outcomes. (B) CMRG score variations between cuproptosis subtypes. (C) Prognostic gene (PRG) score differences among genomic subtypes. (D) LASSO regression cross-validation curve (red line) with standard deviation (gray area). (E) Parameter selection for LASSO regression model. (F) Differential expression of 17 signature CMRGs between risk groups.

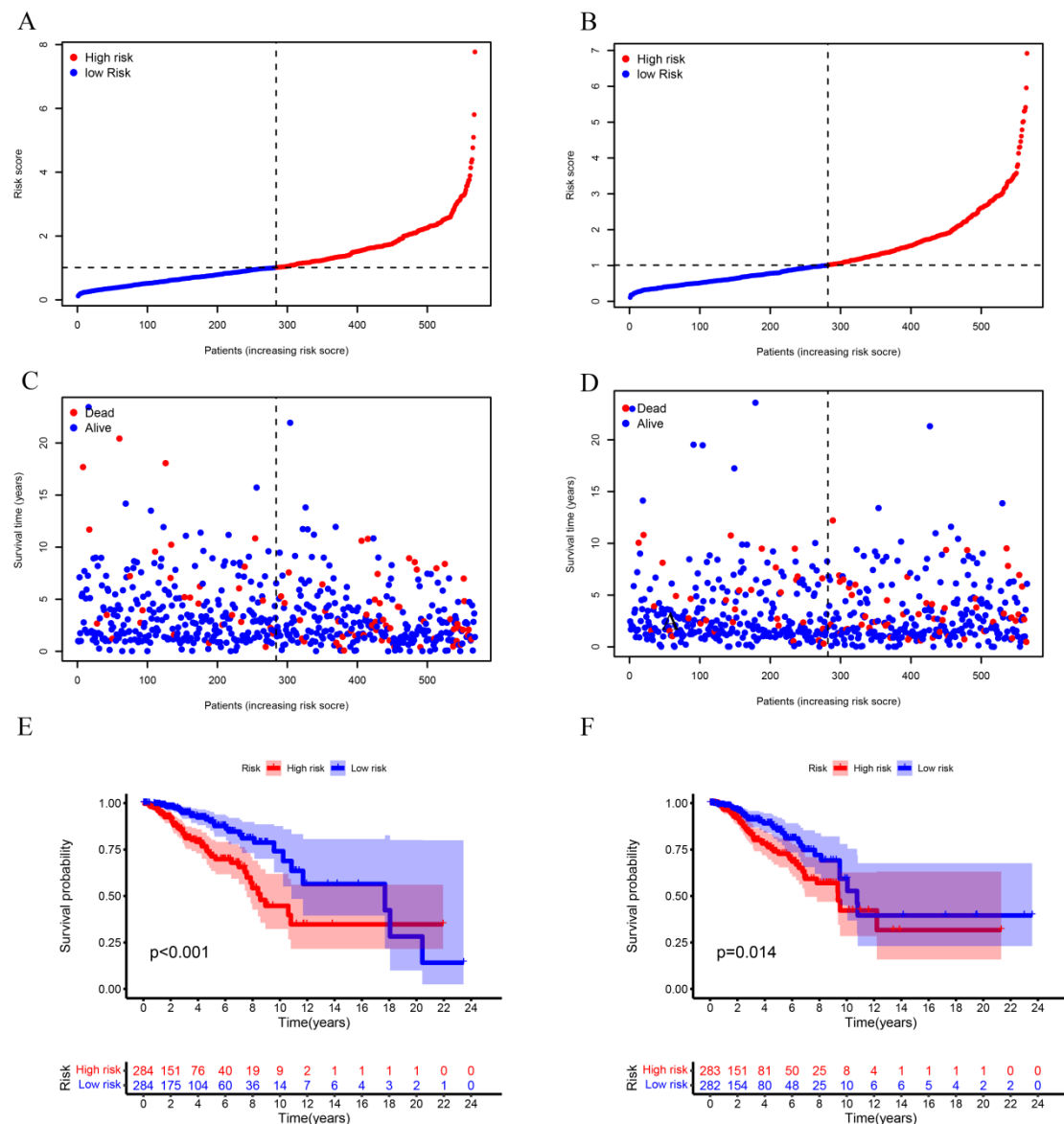


Figure 6. Risk stratification performance in the training and test cohorts. (A,C) Distribution of risk scores and survival status within the training cohort. (B,D) Distribution of risk scores and survival status within the test cohort. (E,F) Kaplan-Meier survival curves comparing high-risk and low-risk subgroups in the training and test cohorts.

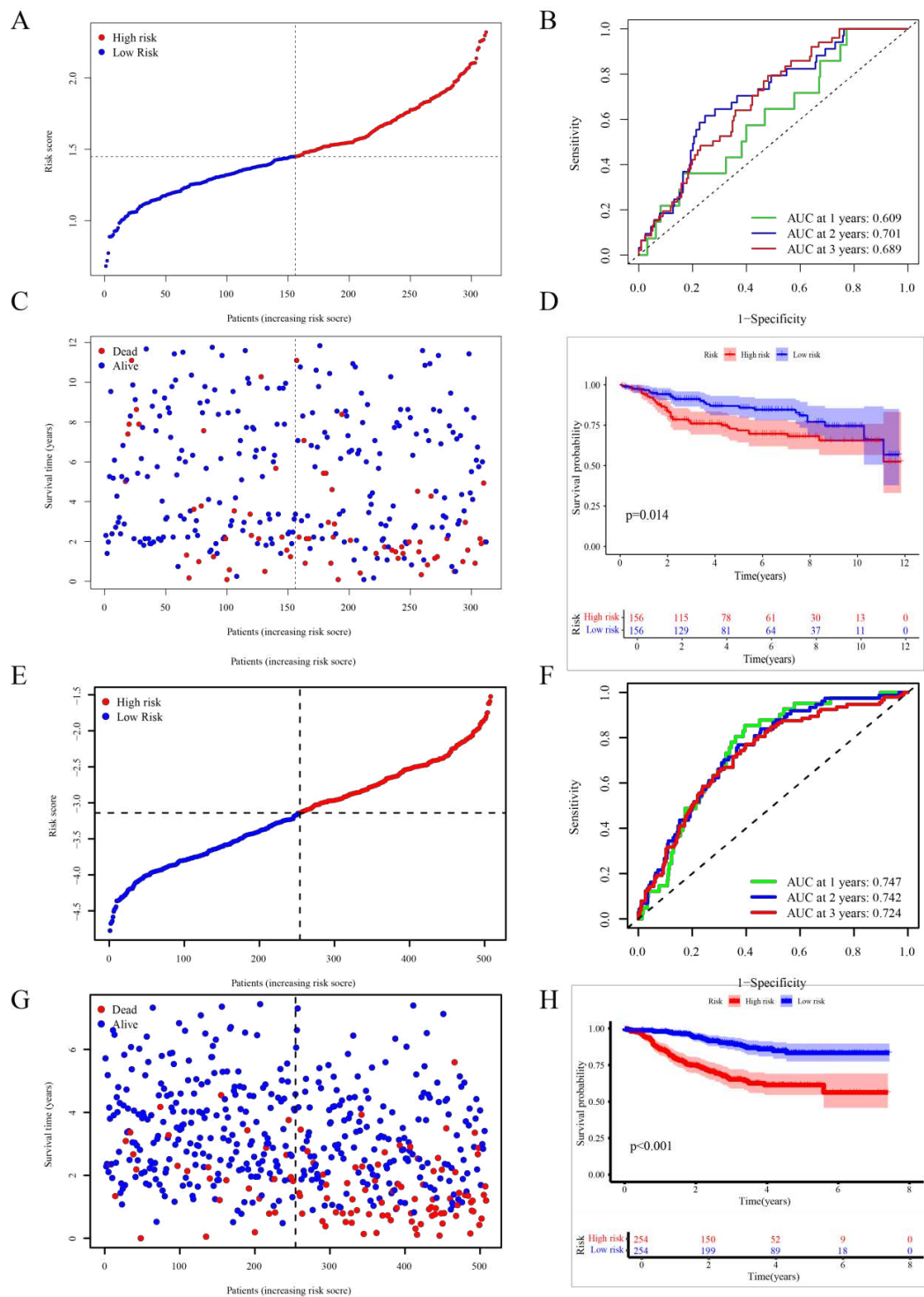


Figure 7. External validation of the CMRG prognostic signature in the GSE25066 and GSE58644 cohorts. (A, E) Risk score distribution along with survival status of breast cancer patients. (B, F) Time-dependent ROC curves assessing the predictive ability of the CMRG signature for survival. (C, G) Ranking of risk scores and corresponding patient survival status. (D, H) Kaplan-Meier survival curves illustrating overall survival differences between high-risk and low-risk subgroups.

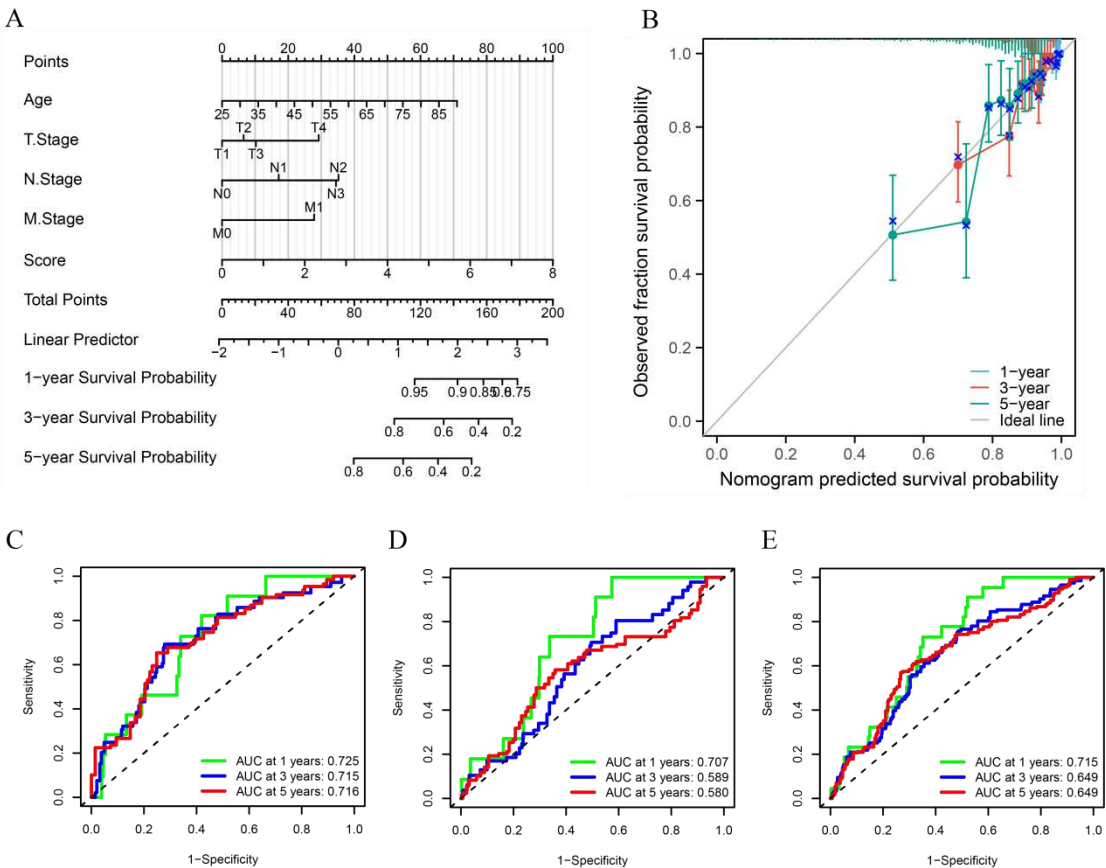


Figure 8. Construction and evaluation of prognostic nomogram. (A) Nomogram integrating CMRG risk score and clinical parameters for 1-, 3-, and 5-year survival prediction. (B) Calibration plot assessing nomogram accuracy. (C-E) Time-dependent ROC curves evaluating predictive performance in overall cohort, training and validation cohorts.

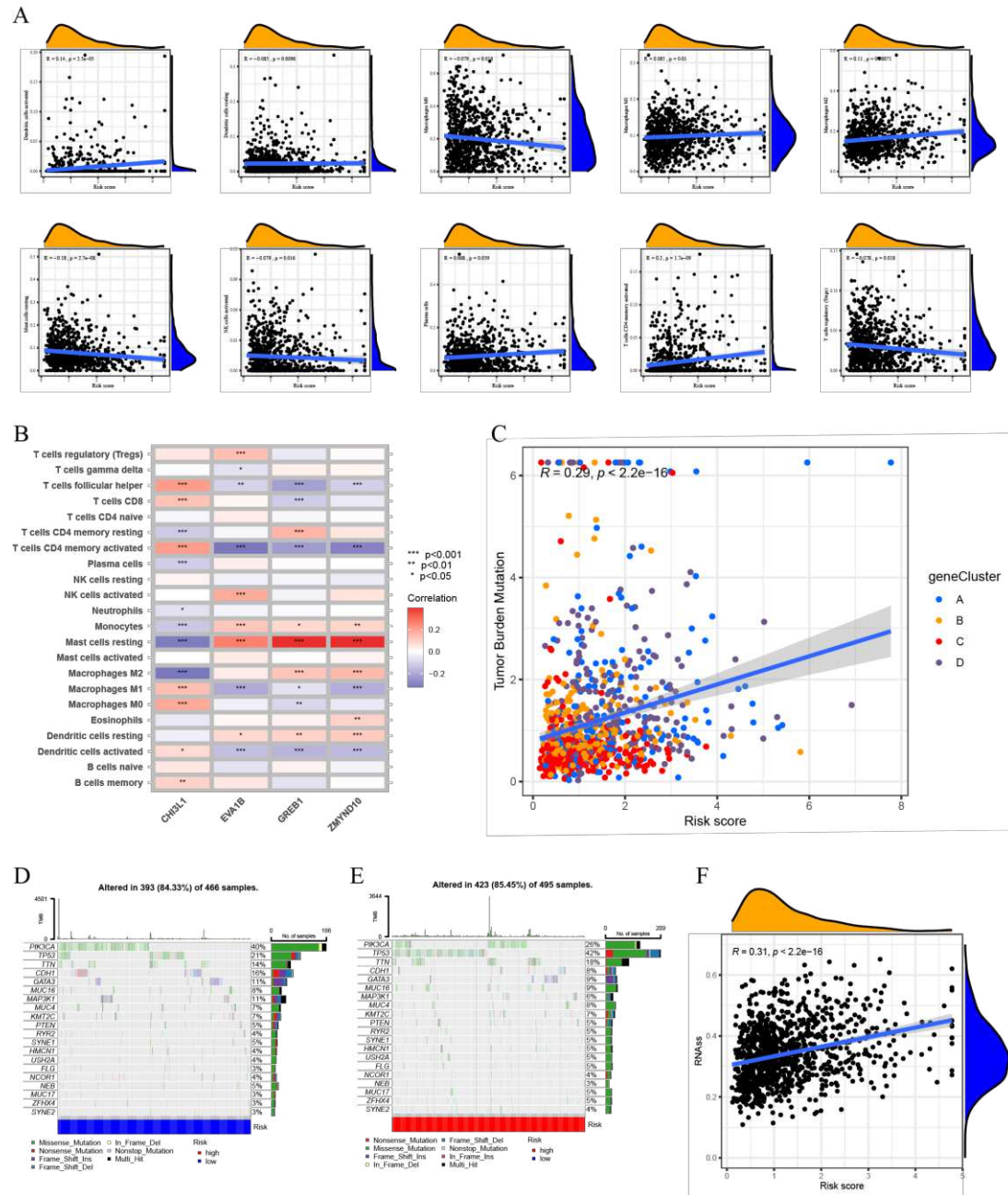


Figure 9. Comprehensive analysis of CMRG risk scores. (A) Heatmap illustrating the correlations between risk scores and immune cell infiltration levels. (B) Relationships between the signature genes and various immune cell populations. (C) Relationship between CMRG scores and tumor mutational burden (TMB). (D-E) Mutational landscape in high- and low-risk groups (columns represent individual patients; bar graphs show mutation frequencies and TMB levels). (F) Correlation between RNAss index and CMRG scores.

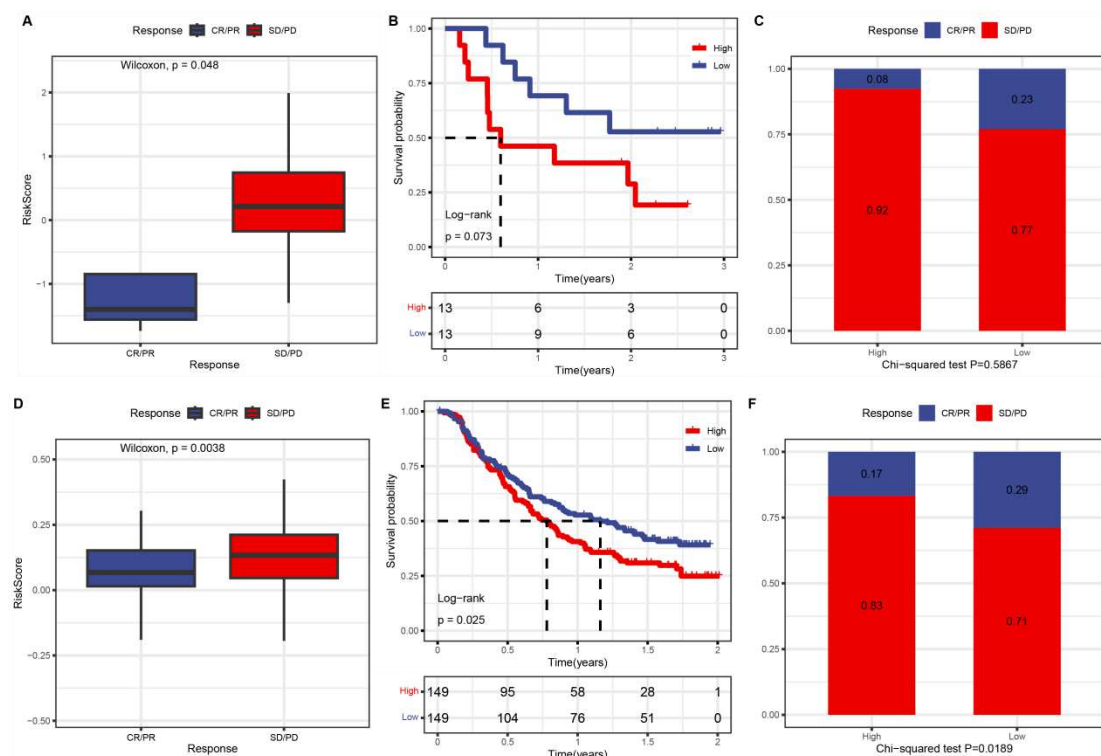


Figure 10. CMRG score as a predictor of immunotherapy response. (A) Comparison of CMRG risk scores between responders and non-responders in the GSE91061 immunotherapy cohort. (B) Kaplan-Meier survival curves of high-risk and low-risk subgroups in the GSE91061 cohort. (C) Clinical response rates of high-risk and low-risk subgroups in the GSE91061 cohort. (D) Comparison of CMRG risk scores between patients who responded and those who did not respond to immunotherapy in the IMvigor210 cohort. (E) Kaplan-Meier survival curves for the high-risk and low-risk subgroups within the IMvigor210 cohort. (F) Clinical response rates observed in the high-risk versus low-risk subgroups of the IMvigor210 cohort.

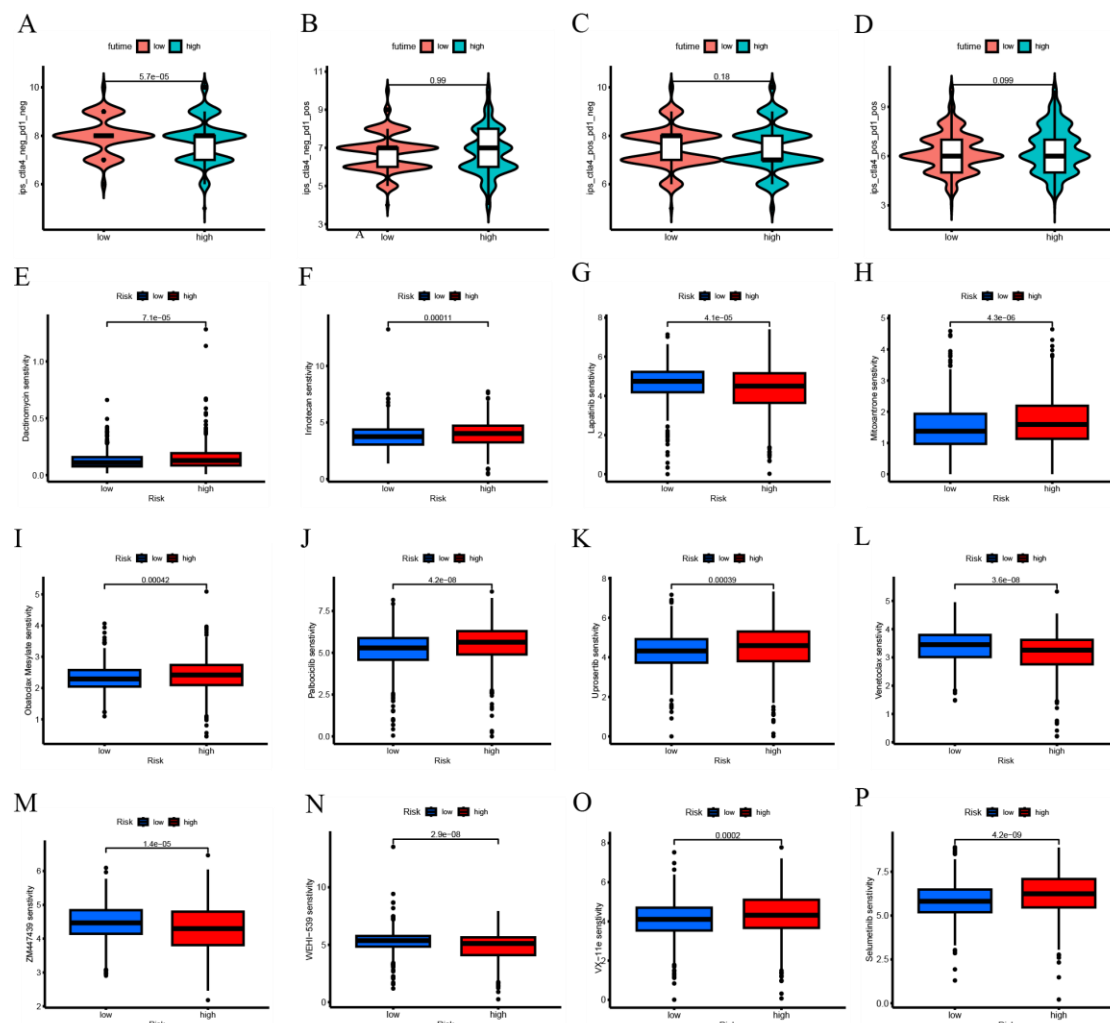


Figure 11. Therapeutic implications of CMRG risk stratification. (A-D) Relationship between CMRG scores and responses to immune checkpoint inhibitors (anti-PD-L1 and anti-CTLA-4). (E-P) Drug sensitivity patterns for standard chemotherapeutic drugs and targeted agents across different risk groups.

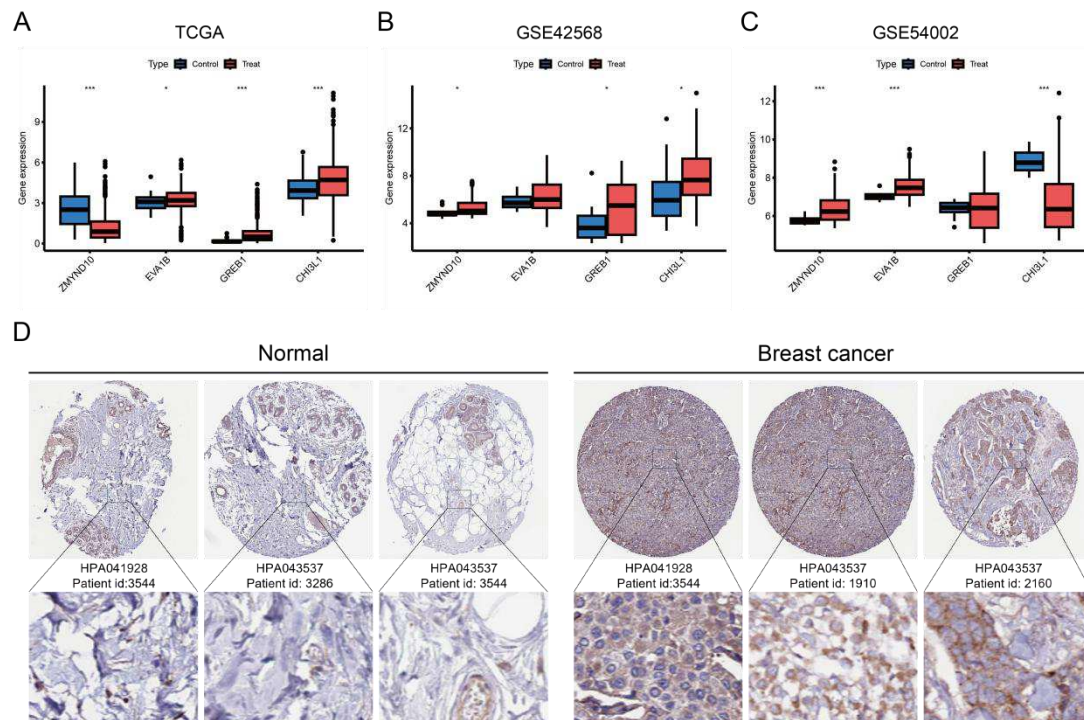


Figure 12. Validation of signature gene expression in TCGA, GEO, and HPA databases. (A) Expression levels of the four signature genes in the TCGA dataset. (B) Expression levels of the four signature genes in the GSE42568 dataset. (C) Expression levels of the four signature genes in the GSE54002 dataset. (D) Protein expression differences of EVA1B between normal controls and breast cancer tissues based on HPA database analysis.