ACCSCIENCE PUBLISHING

Gene & Protein in Disease

ORIGINAL RESEARCH ARTICLE

Identification of hub genes in breast cancer through genome-wide association and functional analyses of M2-like tumor-associated macrophages

Guang Yang[1]*, Zhaorui Hao[2], Xingrui Liu[2], and Zixi Tang[2]

[1]Hubei Provincial Demonstration Center for Experimental Medicine, School of Medicine, Jianghan University, Wuhan, Hubei, China

[2]School of Medicine, Jianghan University, Wuhan, Hubei, China

Abstract

Breast cancer remains a leading cause of cancer-related mortality among women, and immunosuppressive M2-like tumor-associated macrophages (TAMs) contribute to tumor progression and poor prognosis. This study investigated M2-like TAMs to identify key genes promoting breast cancer progression. Using weighted gene co-expression network analysis (WGCNA), which leveraged the opposing prognostic roles of M1/M2 macrophages, we identified 3,127 M2-associated module genes. Intersection with up-regulated differentially expressed genes from GSE42568 and the Cancer Genome Atlas Breast Invasive Carcinoma dataset yielded 85 pivotal genes. Functional enrichment analyses using Gene Ontology and Kyoto Encyclopedia of Genes and Genomes pathways linked these genes to tight junctions, Rap1 signaling, and PI3K–Akt pathways. Protein–protein interaction network analysis via CytoNCA and degree algorithms prioritized four significantly up-regulated hub genes: *FOXA1*, *AGR2*, *MUC1*, and *ERBB3*. A nomogram model demonstrated their prognostic value, while receiver operator characteristic analysis confirmed its diagnostic utility in distinguishing tumor from normal tissue. Hub gene expression positively correlated with the infiltration of mast cells, M2 macrophages, $CD4^+$ T cells, monocytes, B cells, and dendritic cells. Pan-cancer analysis revealed breast cancer-specific overexpression of *FOXA1* and *ERBB3*. Mendelian randomization using fixed-effect inverse variance weighting indicated causal associations between elevated expression of *FOXA1*, *MUC1*, and *ERBB3* ($\beta>0$) and increased breast cancer risk. Drug sensitivity analysis using the Genomics of Drug Sensitivity in Cancer database identified associations between lapatinib sensitivity and expression of *FOXA1*, *ERBB3*, and *MUC1*. This WGCNA-based approach defines *FOXA1*, *AGR2*, *MUC1*, and *ERBB3* as M2-like TAM-related hub genes with diagnostic, prognostic, and therapeutic relevance for breast cancer, particularly in informing drug sensitivity strategies.

Keywords: Breast invasive carcinoma; M2-like tumor-associated macrophages; Hub genes; Mendelian randomization; Genomics of drug sensitivity

***Corresponding author:**
Guang Yang
(yangguang@jhun.edu.cn)

Citation: Yang G, Hao Z, Liu X, Tang Z. Identification of hub genes in breast cancer through genome-wide association and functional analyses of M2-like tumor-associated macrophages. *Gene Protein Dis.* 2025;4(4):025250050. doi: 10.36922/GPD025250050

Received: June 21, 2025

Revised: August 18, 2025

Accepted: October 24, 2025

Published online: November 14, 2025

1. Introduction

Breast cancer is one of the most common malignant tumors worldwide, and it seriously endangers the life and health of women.[1] Although improvements in health awareness

and early diagnosis technologies have contributed to better prognoses, the 5-year survival rate for metastatic breast cancer remains <30%.[2] The hidden onset, recurrence and metastasis, poor prognosis, and lack of effective early diagnostic markers pose great challenges to the diagnosis and treatment of breast cancer. Therefore, there is an urgent need to identify new molecular biomarkers and therapeutic targets for breast cancer.

Bone marrow-derived cells that penetrate tumor tissues or aggregate in the solid tumor microenvironment (TME) are known as tumor-associated macrophages (TAMs).[3] As an important component of the TME, TAMs influence tumor growth, angiogenesis, immune regulation, metastasis, and chemotherapy resistance.[4,5] Under the influence of cytokines in the TME, macrophages differentiate into M1-like and M2-like phenotypes. M1-like macrophages are generally considered tumor-killing macrophages and primarily exert anti-tumor and immune-promoting functions. In contrast, M2-like macrophages are immunosuppressive and promote tumor progression.[6] Therefore, it is necessary to characterize the molecular features associated with M2-like macrophages and to determine the key regulatory factors involved in M2-like TAM polarization. With tumor progression, M1 macrophages gradually polarize into the M2 phenotype, and an increasing abundance of M2-like TAMs indicates poor prognosis, including in breast cancer.[7]

To provide new insights into the molecular characteristics of M2-like TAMs in breast invasive carcinoma (BRCA) patients, we identified groups with high and low M1 macrophage levels and groups with high and low M2 macrophage levels. Survival analysis showed a significant difference between M1-like and M2-like macrophages, and BRCA patients with low M2 and high M1 macrophage content had longer survival. Based on these survival differences, we conducted a series of analyses to identify biomarkers closely related to BRCA.

Mendelian randomization (MR) is a reliable analytical method widely used in recent years to infer potential causality between etiological factors and disease outcomes. MR uses genetic variants strongly associated with exposure factors as instrumental variables to deduce causal relationships between exposures (etiology) and outcomes (diseases). Typically, single-nucleotide polymorphisms (SNPs) are used as instrumental variables to assess causal relationships between exposure factors and outcomes.[8,9]

In this study, we first identified differentially expressed genes (DEGs) between normal and breast tumor tissues from The Cancer Genome Atlas BRCA (TCGA-BRCA) and GSE42568 datasets, respectively. Next, we screened pivotal module genes most positively related to M2 macrophages in BRCA using weighted gene co-expression network analysis (WGCNA), which substantially narrowed the range of hub genes. We then identified four hub genes, namely *FOXA1*, *AGR2*, *ERBB3*, and *MUC1*, as potential markers through protein–protein interaction (PPI) network analysis, which may contribute to BRCA and further elucidate the underlying mechanisms of breast cancer-associated risk genes. Finally, the expression patterns of these four hub genes in breast cancer and their causal relationships with BRCA were explored through functional enrichment, pan-cancer, and MR analyses. Meanwhile, drug sensitivity and IC_{50} associations for drugs targeting these hub genes were evaluated using the Genomics of Drug Sensitivity in Cancer (GDSC) database.

2. Materials and methods

2.1. Source data and preprocessing

A total of 1,226 cases with RNA-seq transcriptome profiling (High-Throughput Sequencing–Fragments Per Kilobase of transcript per Million mapped reads [HTSeq-FPKM]) and clinical information from TCGA-BRCA were obtained from the Genomic Data Commons TCGA data portal (https://portal.gdc.cancer.gov/). Two GEO cohorts (GSE42568, GSE58812) were downloaded from the Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/). Patients without survival information and RNA-sequencing data were excluded from the analysis. The FPKM values were transformed to TPM (transcripts per million) values for subsequent analyses.[10,11]

2.2. Macrophage-related survival analyses

The CIBERSORT algorithm computationally infers (deconvolves) the relative proportions of specific cell types within a mixed bulk tissue sample based on its gene expression profile. In essence, CIBERSORT applies a known reference signature to deconvolute bulk RNA expression data from complex tissues, thereby inferring their cellular composition. The relative abundance of macrophages in each sample from TCGA-BRCA and GSE58812 was calculated using the CIBERSORT algorithm[12] (perm ≥100, QN=TRUE). Kaplan–Meier survival analysis was performed using the R "survival" package to evaluate whether BRCA patient survival was correlated with M1 or M2 macrophage content. The "surv_cutpoint" function in the "survminer" package was used to determine the optimal cut-off values for high- and low-content M1 and M2 macrophage groups and to assess survival differences between the high- and low-M1 or high- and low-M2 groups.

2.3. DEGs identification

Identifying DEGs between normal and tumor tissues can uncover molecular drivers of cancer by pinpointing genes with significantly altered expression. Transcriptome data were obtained from the TCGA-BRCA cohort (1,113 tumor and 113 normal samples) and the GSE42568 dataset (104 tumor and 17 normal samples). All data were processed in R (version 4.3.1), including normalization and batch-effect correction. The R "limma" package was used for DEGs analysis. Volcano plots and heatmaps were generated using the "ggplot2" and "pheatmap" packages, respectively. DEGs were selected based on |log2FoldChange| > 1 and adjusted $p<0.05$.[13]

2.4. Acquisition of pivotal genes associated with M2-like macrophages in breast cancer

The WGCNA allows identification of highly correlated gene modules associated with diseases and can reveal candidate biomarkers and therapeutic targets; it has been widely applied in various biological fields, such as cancer research.[14] BRCA samples were grouped according to M1-like or M2-like macrophage levels, and TCGA-BRCA expression data were analyzed using the R "WGCNA" package to identify genes associated with M2-like macrophages. Samples were first clustered to evaluate overall sample relevance. The soft-thresholding power β was selected according to the minimum power required to achieve a high scale-free topology fitting index, and the minimum module size was set to 60 genes. The module most strongly associated with M2-like macrophages was identified through module–trait correlation analysis. The genes within this module were intersected with up-regulated DEGs between BRCA and normal tissues to obtain M2-like TAM-related genes. These intersecting genes were considered candidate pivotal genes for BRCA pathogenesis. Functional enrichment analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) was performed to better understand the potential biological processes and pathways involved in BRCA.[15]

2.5. Screening of hub genes in the PPI network

STRING (https://string-db.org) was used to predict and visualize the PPI network with a required minimum interaction score of 0.400 (medium confidence). Key genes in the PPI network were ranked using the "CytoNCA" toolkit and the degree algorithm in Cytoscape software (www.cytoscape.org). The top 10 genes were displayed, and the hub genes most strongly associated with BRCA were selected for subsequent analysis.

2.6. Nomogram scoring model construction

We used a nomogram scoring model to predict the risk of BRCA using the "rms" and "rmda" packages.[16,17] The risk of disease was predicted based on the risk score of hub genes. We then used the "ROC" package to construct receiver operator characteristic (ROC) curves to verify the predictive validity of the hub genes. The area under the ROC curve (AUC) was used to represent accuracy. If the AUC was between 0.8 and 1.0, it indicated excellent prediction accuracy.

2.7. Correlation, Gene Set Enrichment Analysis, and pan-cancer analysis

The correlation between hub gene expression and immune cell infiltration was evaluated using the CIBERSORT algorithm. Gene Set Enrichment Analysis (GSEA) was performed to identify functional pathways associated with hub genes. Pan-cancer analysis was then conducted to determine whether these M2-like TAM-associated hub genes were more highly expressed in BRCA than in other tumors.

2.8. Mendelian randomization study

Two-sample MR was used to evaluate the causal relationship between hub genes (exposures) and breast cancer risk (outcome) using SNPs as instrumental variables. From genome-wide association study (GWAS) data for the hub genes, we extracted SNPs strongly associated at genome-wide significance ($p<5 \times 10^{-6}$) to ensure independence of instrumental variables and exclude linkage disequilibrium effects.[18-20] The genetic distance was set to 10,000 kb, and the linkage disequilibrium parameter (r^2) threshold to 0.01. GWAS data included 89,677 individuals, comprising 42,892 controls and 46,785 cases of European ancestry. MR analysis was performed using the "TwoSampleMR" package, and MR-Egger regression, weighted median, fixed-effects inverse variance weighting (IVW), simple mode, and weighted mode were used to assess the relationship between hub gene levels and breast cancer risk. MR-Egger was used for sensitivity analysis. The intercept term in the causal estimation was assessed, and MR-Egger regression was used to evaluate directional pleiotropy. In addition, the IVW method and MR-Egger regression were applied to detect heterogeneity between individual genetic variation estimates.

2.9. Drug sensitivity analysis

The GDSC database (https://www.cancerrxgene.org/) contains more than 1,000 human cancer cell lines, 265 drug response datasets, and over 17,000 genomic markers. We

collected mRNA expression levels of hub genes and IC_{50} for 265 drugs across 51 BRCA-related cell lines and combined mRNA expression and drug susceptibility data. Pearson correlation analysis (p<0.03) was used to determine the correlation between hub gene mRNA expression and drug IC_{50} values, and to investigate whether hub gene expression was associated with drug sensitivity or resistance. The databases used in this study and their download links are listed in Table S1.

2.10. Statistical analysis

Survival differences between groups were evaluated using Kaplan–Meier curves and log-rank tests. Correlation coefficients were calculated using Pearson correlation analysis. Statistical analysis was performed using R software (version 4.3.1), and values are represented as mean ± standard deviation. p<0.05 was considered statistically significant.

3. Results

3.1. Screening M2-like TAM-related candidate hub genes

To elucidate the relationship between macrophages and BRCA prognosis, the proportions of M1 and M2 macrophages in TCGA-BRCA samples were estimated using the CIBERSORT algorithm. BRCA patients were then divided into high- and low-content M1 or M2 macrophage groups, respectively. Kaplan–Meier survival analysis showed a significant difference between the high-content and low-content M1 macrophage groups (p=0.032; Figure 1A). Furthermore, a more pronounced survival difference was observed between the high-content and low-content M2 macrophage groups, with patients in the high-content M2 macrophage group showing worse prognosis than those in the low-content group (p<0.001; Figure 1B). These results indicated that the prognostic effects of M1 and M2 macrophages were opposite, and that M2 macrophages played a more pivotal role in BRCA prognosis than M1 macrophages. Analysis of the GSE58812 datasets confirmed these results (Figure S1A–B).

Based on these observations, WGCNA was performed to identify the key module related to M2 macrophages, and the soft-threshold power β = 5 (scale-free R^2 = 0.92) was selected to construct a scale-free network. As shown in Figure 1C, the green module was positively correlated with M2 macrophages and negatively correlated with M1 macrophages. A total of 3,127 genes in the green module were selected for downstream analysis. In addition, 2,098 DEGs (935 up-regulated and 1,163 down-regulated) were identified in GSE42568 (Figure 1D), and 2,753 DEGs (1,078 up-regulated and 1,675 down-regulated) were identified in TCGA-BRCA (Figure 1E). Finally, 85 overlapping genes were identified by intersecting WGCNA-derived M2-like TAM-related genes with up-regulated DEGs associated with BRCA, and were considered candidate hub genes potentially involved in BRCA development and progression (Figure 1F).

3.2. GO and KEGG pathway enrichment analysis

The GO and KEGG analyses were conducted to further explore the potential roles of the 85 candidate hub genes. GO analysis showed that these overlapping genes were mainly enriched in cellular components such as the bicellular tight junction, tight junction, and apical junction complex (Figure 2A). KEGG pathway analysis indicated enrichment in tight junction, Ras-associated protein 1 (Rap1) signaling pathway, and phosphatidylinositol 3-kinase (PI3K)–Akt signaling pathway (Figure 2B). The Rap1 and PI3K–Akt signaling pathways are well known to be associated with proliferation, invasion, and metastasis in cancer, whereas the relationship between tight junctions and tumors has been less frequently reported. Breast cancer is a complex and heterogeneous disease, and proper adhesion between adjacent epithelial cells is essential for maintaining normal epithelial tissue structure and function. Increasing evidence suggests that dysregulation of cell–cell adhesion is associated with many cancers. One study has suggested that tight junctions may be involved in the development or progression of breast cancer.[21]

3.3. PPI network analysis of hub genes

The PPI network of candidate hub genes was constructed using the STRING online tool (Figure 2C). The top 10 up-regulated genes in breast cancer were then identified and visualized using the "CytoNCA" toolkit and degree algorithm. The genes *FOXA1*, *AGR2*, *MUC1*, *ERBB3*, *TFF1*, *EZR*, *GATA3*, *FGFR3*, *SPDEF* and *AGR3* were ranked (Figure 2D). A deeper red color indicated a higher relevance score and greater importance in BRCA. *FOXA1*, *AGR2*, *MUC1*, and *ERBB3* were the most prominent hub genes.

3.4. Construction of a nomogram model and receiver operator characteristic diagnostic curve

A nomogram model was constructed to predict BRCA risk (Figure 3A). Higher scores indicated higher disease risk. In the examples shown in Figure 3A, a total score of 100 for the four hub genes corresponds to a breast cancer risk exceeding 90%. Calibration analysis verified the reliability of the nomogram model (Figure 3B). ROC curves were then generated for the four hub genes (*FOXA1*, *AGR2*, *MUC1*, and *ERBB3*) to assess predictive performance; the AUC values were 0.877, 0.907, 0.865, and 0.930, respectively.

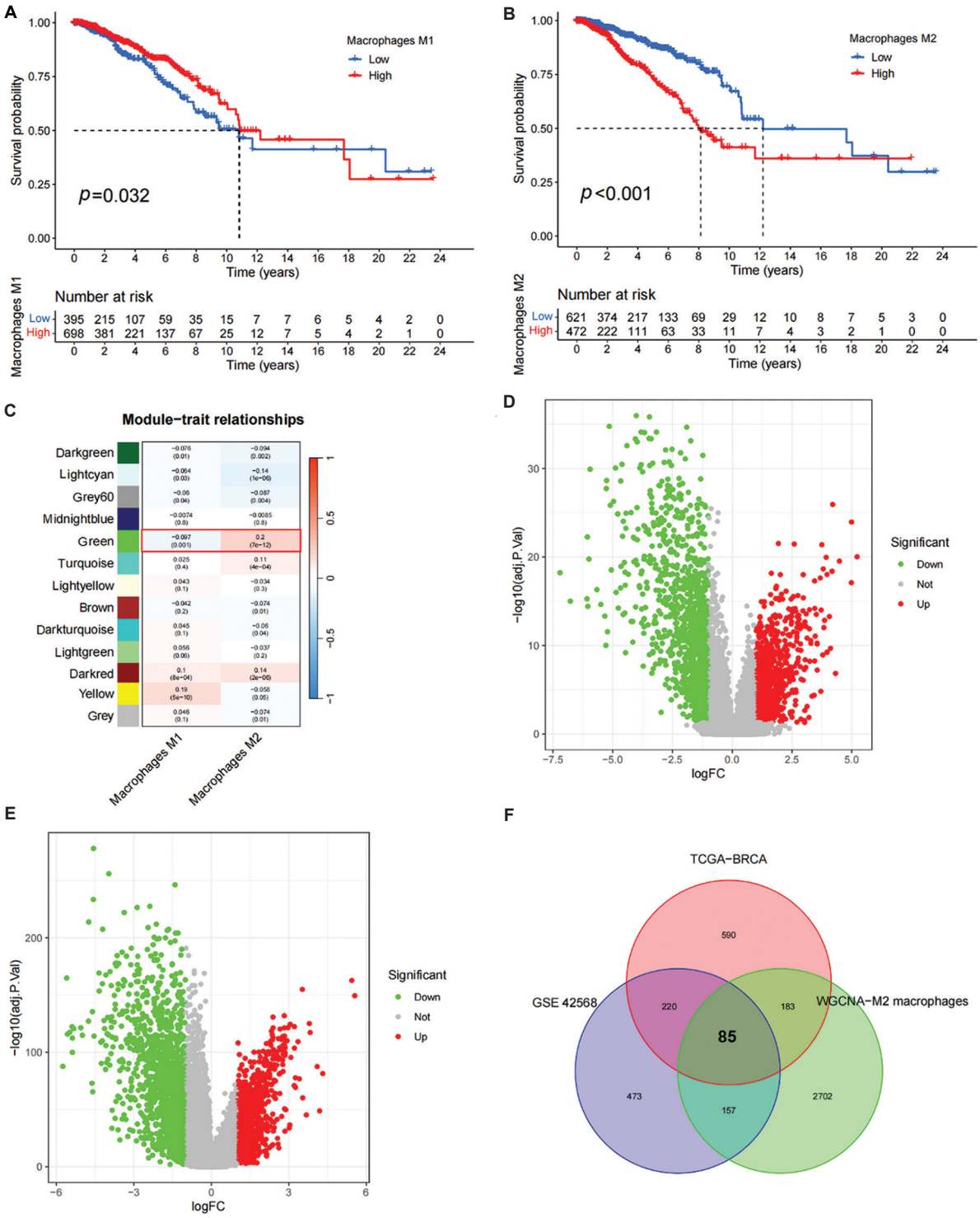

Figure 1. Survival analysis and screening of M2-like tumor-associated macrophage-related candidate hub genes. (A) Kaplan–Meier survival curves comparing high-content and low-content M1 macrophage groups. (B) Worse prognosis in the high-content M2 macrophage group. (C) Correlation analysis of modules and traits, with the green module most strongly associated with M2 macrophages. (D) Volcano plot of differentially expressed genes (DEGs) between normal and tumor tissues in GSE42568. Note: Red dots indicate up-regulated genes, green dots indicate down-regulated genes, and gray dots indicate nonsignificant genes. (E) Volcano plot of DEGs between normal and tumor tissues in TCGA-BRCA. (F) Venn diagram showing the intersection of weighted gene co-expression network analysis-derived M2-like TAM-related genes and up-regulated DEGs.
Notes: Red dots indicate up-regulated genes, green dots indicate down-regulated genes, and gray dots indicate nonsignificant genes.

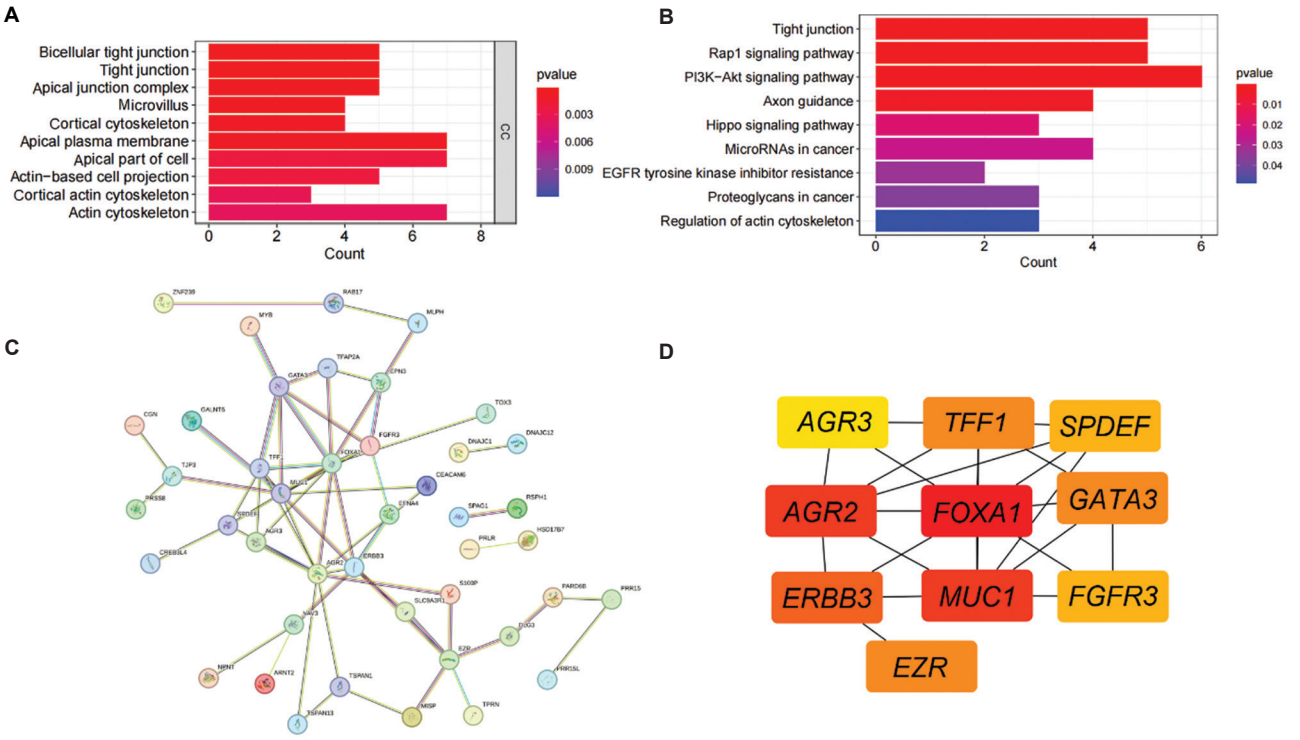

Figure 2. Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and protein–protein interaction (PPI) network analysis. (A) GO enrichment analysis of candidate hub genes. (B) KEGG pathway analysis of candidate hub genes. (C) PPI network of overlapping hub genes. (D) Core genes identified by the degree algorithm.

These results demonstrated that all four hub genes could effectively differentiate breast cancer from normal tissues (Figure 3C).

3.5. Functional analysis of M2-like tumor-associated macrophage-related hub genes

According to the correlation analysis between the four hub genes and M2-like macrophages, three of the four hub genes were closely related to M2-like macrophages. Among them, *FOXA1* and *ERBB3* showed the strongest correlations (Figure 4A). In the TCGA datasets, high *FOXA1* expression was associated with shorter overall survival (Figure 4B). In contrast, high *AGR2* expression was associated with better survival compared to low *AGR2* expression (Figure 4C).

Expression differences of the four hub genes between normal and breast tumor tissues were analyzed using TCGA-BRCA, and all four hub genes were significantly up-regulated in breast tumor tissues compared with normal tissues (Figure 4D). GSEA analysis showed that *FOXA1*, *AGR2*, and *ERBB3* were mainly enriched in tumor-related pathways, such as complement and coagulation cascades,[22] colorectal cancer, mTOR signaling pathway, and prostate cancer (Table 1). Complement and coagulation cascades are important components of the immune system. Platelets in the TME can promote tumor cell invasion and metastasis by activating the coagulation cascade.

In the pan-cancer analysis, the four hub genes were highly expressed in most tumor types, with particularly high expression of *FOXA1* and *ERBB3* in breast tumors (Figure S2A-D). Moreover, expression of the four hub genes was positively associated with age over 65 years in BRCA clinical data (Figure S2E-H).

Next, the relationship between hub gene expression and the TME was evaluated. CIBERSORT analysis showed that the four hub genes were significantly correlated with infiltration of immune cells, such as mast cells, M2 macrophages, $CD4^+$ T cells, monocytes, B cells, and dendritic cells (Figure S3A-D). In addition, *FOXA1*, *ERBB3*, and *AGR2* showed stronger correlations with M2 macrophages, consistent with the analysis shown in Figure 4A. Expression of all four hub genes was positively correlated with M2-like macrophage infiltration in BRCA, whereas the opposite pattern was observed for M1-like macrophage infiltration (Figure S4A-H).

3.6. Hub genes causally associated with the risk of invasive carcinoma

We assessed the causal association between the four hub genes and breast cancer. The SNPs associated with these

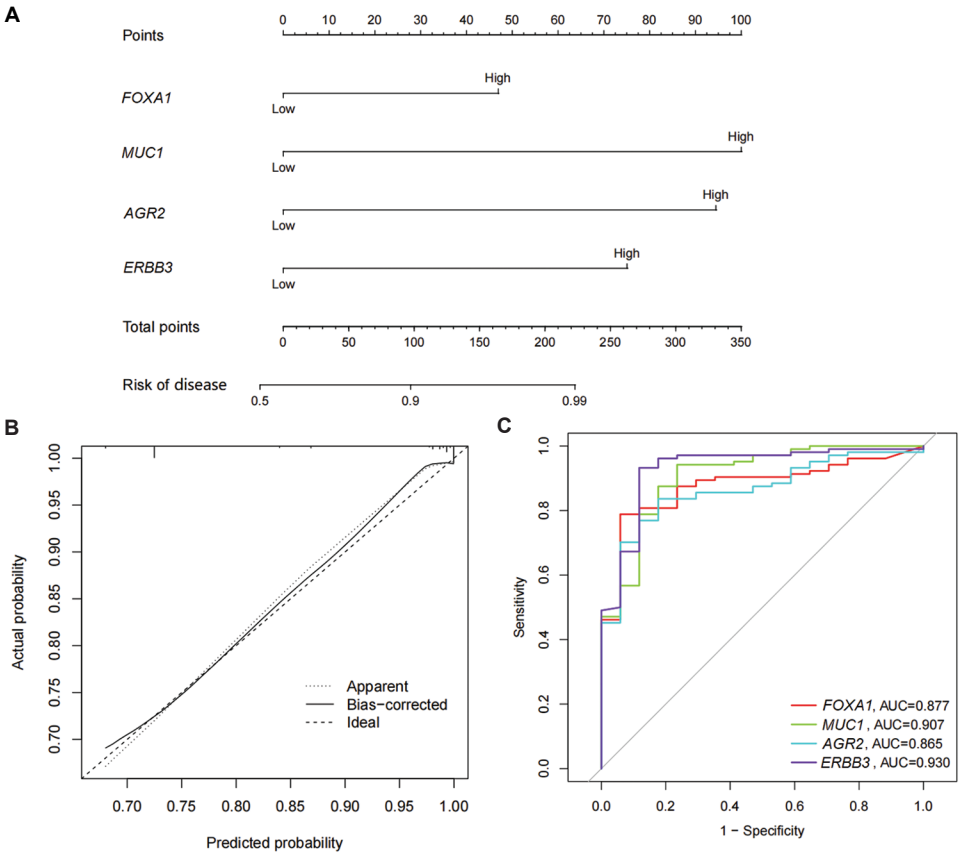

Figure 3. Nomogram model and receiver operator characteristic (ROC) curve analysis. (A) Nomogram model of four hub genes. (B) Calibration curve validating model performance. (C) ROC curves evaluating the predictive efficacy of each hub gene.
Abbreviation: AUC: Area under the ROC curve.

genes met the criteria for strong instrumental variables ($p<5\times10^{-6}$). Using the fixed-effects IVW method, we found that all four hub genes were associated with breast cancer risk (Table 2). A β-value greater than zero indicates that the exposure factor is a risk factor, and the risk of disease increases with increasing exposure. Except for *AGR2*, the expression of *FOXA1*, *ERBB3*, and *MUC1* was positively correlated with the occurrence of breast tumors ($\beta > 0$), which was consistent with the prognostic analysis results (Figure 4C). A strong causal association between *FOXA1* and breast cancer was observed (Figure 5A). The combined effect estimate of all SNPs was greater than zero, indicating that *FOXA1* is a risk factor and that breast cancer risk increases with higher *FOXA1* expression (Figure 5B). The weighted median and weighted mode methods also supported the results of the fixed-effects IVW method. The *FOXA1* funnel plot showed approximate symmetry around the central axis (Figure 5C), and no evidence of horizontal pleiotropy was observed at the MR-Egger intercept, indicating that pleiotropy did not bias the causal effect. In the leave-one-out analysis, MR analysis was repeated after removing each SNP, and the overall effect estimates remained consistent with the main result, indicating the robustness of the causal inference (Figure 5D).

3.7. Drug sensitivity analysis

To investigate whether the four hub genes influence drug sensitivity or resistance during chemotherapy, we integrated drug IC_{50} and gene expression data from GDSC BRCA cell lines (Table 3). Through this analysis, drugs most strongly correlated with the hub genes were identified.

The results showed that the IC_{50} values of A-443654, lapatinib, KIN001-102, CP724714, and ZSTK474 were negatively correlated with *FOXA1* expression, whereas the IC_{50} values of pyrimethamine and BX-795 were positively correlated with *FOXA1* expression, according to Pearson correlation analysis (Figure 6A). The IC_{50} values of lapatinib, A-770041, sorafenib, roscovitine, A-443654, MS-275, and paclitaxel were negatively correlated with *ERBB3* expression, indicating that *ERBB3* expression sensitizes cells to these drugs (Figure 6B). The IC_{50} values of saracatinib, imatinib,

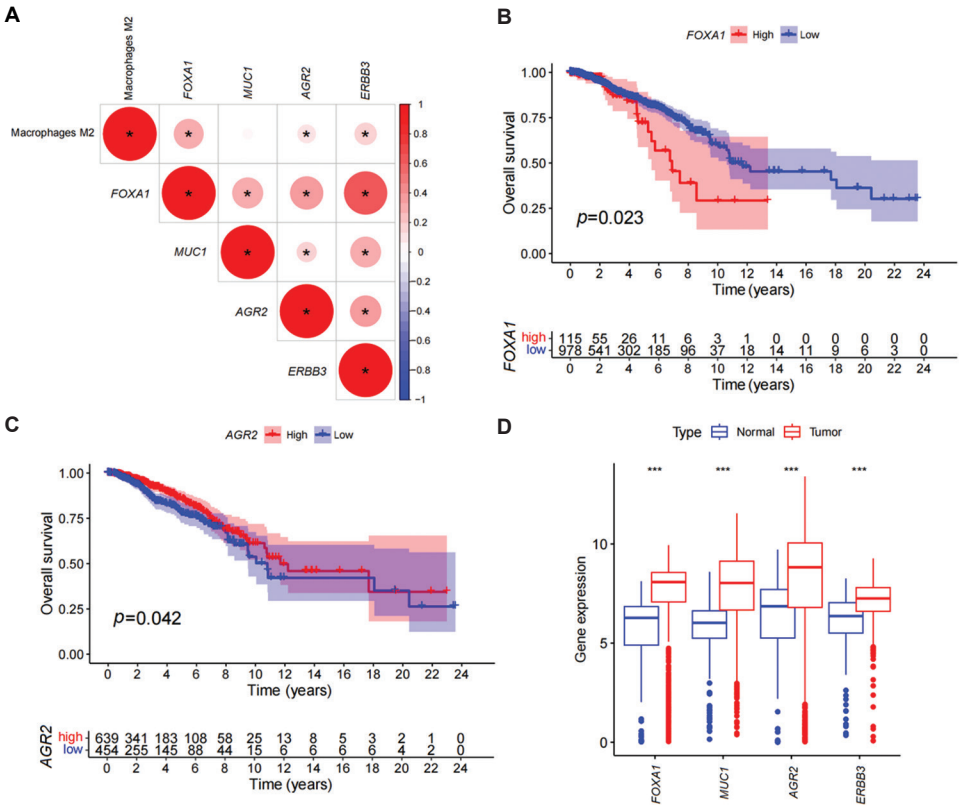

Figure 4. Correlation and survival analysis of M2-like tumor-associated macrophage-related hub genes. (A) Correlation between the four hub genes and M2-like macrophages. (B) Overall survival of TCGA-BRCA patients with high versus low *FOXA1* expression (Kaplan–Meier analysis). (C) Overall survival of TCGA-BRCA patients with high versus low *AGR2* expression (Kaplan–Meier analysis). (D) Expression of the four hub genes in normal versus breast tumor tissues.
Notes: Red: tumor samples; blue: normal samples. *p<0.05, **p<0.01 and ***p<0.001.

BMS-509744, and lapatinib were negatively correlated with *MUC1* expression, whereas olaparib, YK-4-279 were positively correlated with *MUC1* expression (Figure 6C). The IC_{50} values of KIN001-102, CP724714, ZSTK474, and Afatinib showed a negative correlation with *AGR2* expression. Conversely, TAE684, XMD8-85, AZ628, and Pyrimethamine exhibited a positive correlation (Figure 6D).

Interestingly, the expression of *FOXA1*, *ERBB3*, and *MUC1* was associated with increased sensitivity to lapatinib. Lapatinib was approved by the United States Food and Drug Administration in 2007 as a treatment for human epidermal growth factor receptor-2 (HER2)-positive advanced breast cancer. These findings suggest that combined expression of *FOXA1*, *ERBB3*, and *MUC1* may represent a potential marker for lapatinib responsiveness in clinical breast cancer patients.

In contrast, *AGR2* expression was closely related to resistance of TAE684, XMD8-85, AZ628, and pyrimethamine, indicating that higher *AGR2* expression may be linked to increased drug resistance when p<0.03.

4. Discussion

Metastatic breast cancer responds less effectively to immunotherapy than primary tumors, although certain immune-oncology targets, macrophages, and angiogenic features show preserved expression.[23] Stage IV (metastatic) breast cancer is treatable but not curable, and treatment goals include prolonging survival and improving quality of life.[24] Immunotherapy is a potentially effective approach for controlling tumor growth in patients with stage IV breast cancer.[25] Although encouraging immunotherapy outcomes have been reported in some clinical trials, several challenges remain. Therefore, personalized immunotherapy is considered a potential complementary therapy for breast cancer, particularly in combination with chemotherapy.[26] Considering that existing studies have proved the importance of TAMs in BRCA immunotherapy[27], we examined the reliability of M2-like TAMs as biomarkers and predictors of tumorigenesis, based on the significant prognostic differences observed between M1 and M2 macrophages in patients with breast cancer.

Table 1. Gene set enrichment analysis of four hub genes

Gene	KEGG pathway	Enrichment score	*p*-value
FOXA1	GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	0.64890	0.00064
	COMPLEMENT_AND_COAGULATION_CASCADES	0.46633	0.00027
	VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	0.51269	0.00147
	NUCLEOTIDE_EXCISION_REPAIR	0.49271	0.00267
	PROTEIN_EXPORT	0.57175	0.00729
AGR2	COMPLEMENT_AND_COAGULATION_CASCADES	0.44045	0.00459
	GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	0.59598	0.02040
	PRIMARY_IMMUNODEFICIENCY	−0.60826	0.02514
	SPLICEOSOME	0.32993	0.02020
ERBB3	PROSTATE_CANCER	0.42079	0.00012
	MTOR_SIGNALING_PATHWAY	0.50957	0.00023
	GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	0.67080	0.00039
	COLORECTAL_CANCER	0.42077	0.00204
	RNA_DEGRADATION	0.44905	0.00061
MUC1	VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	0.54677	0.00018
	FATTY_ACID_METABOLISM	0.51296	0.00144
	CIRCADIAN_RHYTHM_MAMMAL	0.69052	0.00805
	GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	0.51112	0.01853
	GALACTOSE_METABOLISM	0.48660	0.03024

Abbreviation: KEGG: Kyoto Encyclopedia of Genes and Genomes.

Table 2. Results of Mendelian randomization (MR) analysis

Gene	Exposure	Outcome	Method	nSNPs	β-value	*p*-value
FOXA1	prot-a-2611	Breast cancer id: ebi-a-GCST007236	MR-Egger	26	0.1462	0.9402
			Weighted median	26	0.9526	5.54×10^{-11}
			Inverse variance weighted (fixed effects)	26	0.6474	5.44×10^{-245}
			Simple mode	26	0.4266	0.1621
			Weighted mode	26	0.5861	9.18×10^{-7}
MUC1	prot-a-1967	Breast cancer id: ebi-a-GCST007236	MR-Egger	16	4.6470	0.0002
			Weighted median	16	1.8530	4.25×10^{-18}
			Inverse variance weighted (fixed effects)	16	0.7446	0
			Simple mode	16	1.6230	0.0040
			Weighted mode	16	1.6600	6.72×10^{-10}
AGR2	prot-a-58	Breast cancer id: ebi-a-GCST007236	MR-Egger	10	−1.3710	0.6640
			Weighted median	10	−0.6746	7.32×10^{-5}
			Inverse variance weighted (fixed effects)	10	−0.4915	6.48×10^{-53}
			Simple mode	10	−1.6350	0.0004
			Weighted mode	10	−0.6215	0.0061
ERBB3	prot-c-2617_56_35	Breast cancer id: ebi-a-GCST007236	MR-Egger	3	1.4770	0.9365
			Weighted median	3	0.3960	2.55×10^{-6}
			Inverse variance weighted (fixed effects)	3	0.03712	1.93×10^{-78}
			Simple mode	3	0.3527	0.0267
			Weighted mode	3	0.3025	0.0229

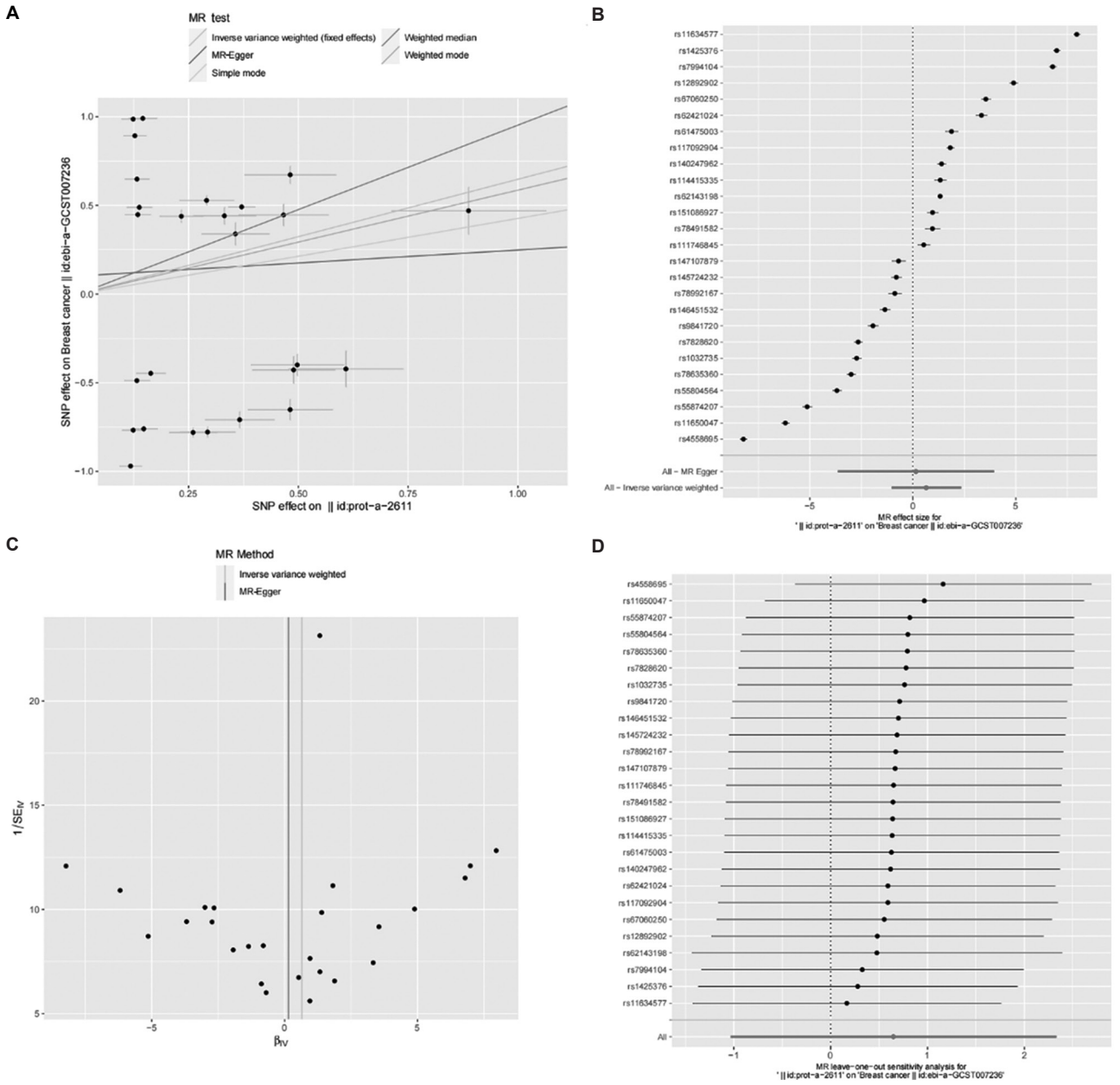

Figure 5. Mendelian randomization (MR) analysis of *FOXA1*. (A) Scatter plot showing the causal effect of *FOXA1* on breast cancer risk. (B) Forest plot showing the causal effect of individual single nucleatide polymorphism (SNP) on breast cancer risk. (C) Funnel plots assessing heterogeneity of MR estimates. (D) Leave-one-out plot evaluating the influence of individual SNPs on the causal estimate.

Tumor-associated macrophages are one of the most abundant immune cell populations in the TME.[28] M2-like TAMs are associated not only with poor prognosis in multiple cancers but also with the establishment of immunosuppressive TME.[27-30] M2-like macrophages secrete anti-inflammatory cytokines and growth factors and promote angiogenesis, supporting tumor cell proliferation.[31-33] Therefore, TAMs represent a promising therapeutic target, and identifying TAM-specific markers is critical. A TAM-related gene signature in breast cancer has been reported to be highly enriched in aggressive breast cancer subtypes,[34] and an M2-macrophage-related prognostic model has been developed in pancreatic ductal adenocarcinoma.[35] However, studies specifically investigating M2-like TAM-related hub genes in breast cancer remain limited.

To identify potential biomarkers for predicting prognosis in BRCA, we applied WGCNA and differential expression

Table 3. Results of drug sensitivity analysis

Gene	Drug	Pearson correlation coefficient	*p*-value
FOXA1	A-443654	−0.502	0.00736
	Lapatinib	−0.488	0.00883
	KIN001-102	−0.419	0.02051
	CP724714	−0.414	0.02173
	ZSTK474	−0.391	0.02821
	Pyrimethamine	0.634	0.02205
	BX-795	0.616	0.02705
ERBB3	Lapatinib	−0.633	0.00420
	A-770041	−0.589	0.00724
	Sorafenib	−0.530	0.01437
	Roscovitine	−0.493	0.02152
	A-443654	−0.490	0.02222
	MS-275	−0.474	0.02627
	Paclitaxel	−0.473	0.02655
MUC1	Saracatinib	−0.708	2.11×10^{-5}
	Imatinib	−0.402	0.01027
	BMS-509744	−0.347	0.02361
	Lapatinib	−0.334	0.02839
	Olaparib	0.491	0.01859
	YK 4-279	0.481	0.02153
AGR2	TAE684	0.709	0.00735
	XMD8-85	0.68	0.01109
	KIN001-102	−0.355	0.01299
	CP724714	−0.314	0.02230
	AZ628	0.625	0.02303
	Pyrimethamine	0.613	0.02678
	ZSTK474	−0.291	0.02974
	Afatinib	−0.291	0.02974

analysis to identify M2-like TAM-related hub genes. Firstly, we validated that the high M2-like TAM content was significantly associated with poorer prognosis in both the TCGA-BRCA and GSE58812 cohorts. Patients with high M2-like TAM content had significantly worse survival than those with low M2-like TAM content, indicating that M2-like TAMs are closely related to poor prognosis in BRCA. We then identified 85 candidate hub genes by intersecting M2-like macrophage modula genes with DEGs in BRCA. Next, four hub genes of breast cancer, including *FOXA1*, *ERBB3*, *AGR2*, and *MUC1*, were screened using the STRING and Cytoscape platforms. The nomogram model integrating these genes performed well in predicting breast cancer, and ROC curves confirmed that the four hub genes effectively distinguished breast cancer from normal tissues.

Functional analysis in TCGA-BRCA confirmed that the four hub genes were more highly expressed in tumors than in normal tissue, and high *FOXA1* expression was significantly associated with poor prognosis. GSEA analysis demonstrated that *FOXA1*, *AGR2*, and *ERBB3* were closely related to tumor-related pathways and functions. In pan-cancer analysis, these four hub genes were highly expressed in breast cancer patients over 65 years old. A study of 92 women over 65 years old reported that 77 women (83.6%) had malignant breast disease.[36] Mutations in *FOXA1* have been reported as a hallmark of estrogen receptor-positive (ER+) breast cancer.[37] *FOXA1* overexpression may be a prognostic factor for treatment resistance and a viable target for immunotherapy and chemotherapy sensitization in ER+ luminal breast cancer.[38] The positive correlation between *FOXA1* expression and M2 macrophage infiltration further supports its association with poorer prognosis. Previous research has also shown that *MUC1* can up-regulate M2 macrophage infiltration, and the cytoplasmic domain of *MUC1* plays an important role in promoting postpartum mammary tumors, suggesting a potential strategy for postpartum breast cancer prevention and treatment.[39]

We then explored the causal relationship between the expression levels of the four hub genes (*FOXA1*, *AGR2*, *MUC1*, and *ERBB3*) and breast cancer risk using two-sample MR based on large-scale GWAS data. This MR analysis suggested that *FOXA1*, *MUC1*, and *ERBB3* may be causally associated with increased breast cancer risk. MR reduces systematic bias caused by observational studies, such as confounding and reverse causality, in prospective randomized controlled trials.[40] To ensure that SNPs were not related to confounding factors between the four hub genes and breast cancer, only participants of the European ancestry were included. In addition, MR-Egger regression with an intercept term showed no evidence of directional pleiotropy, supporting the robustness of the causal inference.[41]

Finally, we explored whether hub gene expression influenced sensitivity or resistance to anticancer drugs based on the relationship between drug IC_{50} values and gene expression. *FOXA1*, *MUC1*, and *ERBB3* expression were associated with increased sensitivity to lapatinib. Lapatinib is a HER-2 tyrosine kinase inhibitor used clinically, often in combination with capecitabine, for the treatment of advanced or metastatic breast cancer. The findings may assist in supporting the clinical application of lapatinib.

While this study is constrained by its *in silico* nature, its primary value lies in guiding future translational research. The absence of wet-laboratory validation is not an endpoint but a starting point, identifying the most promising candidates

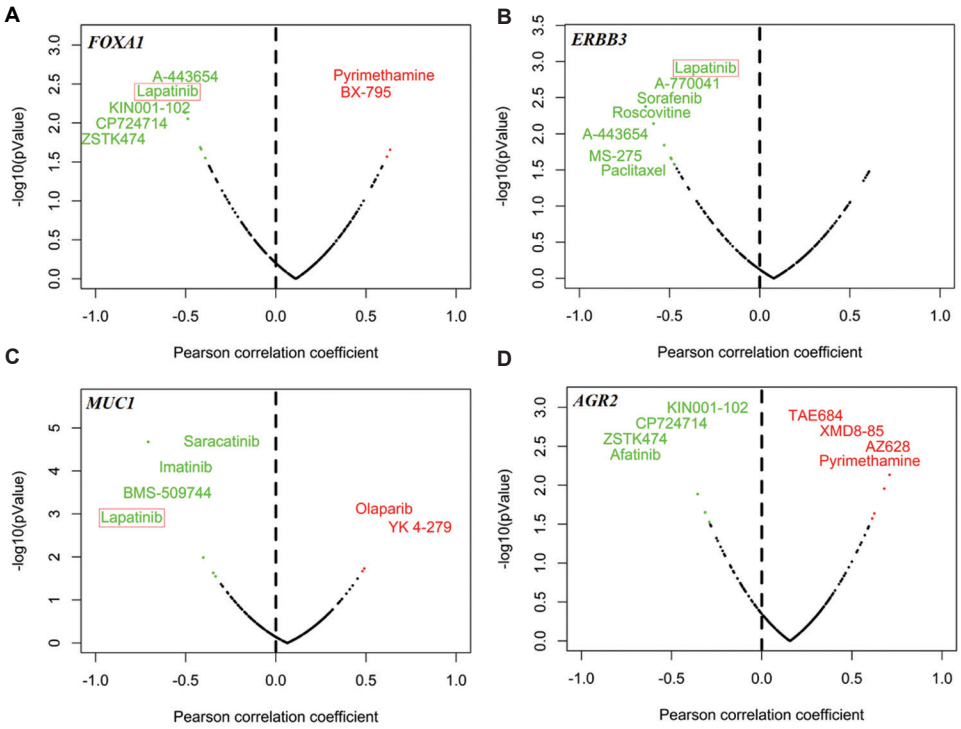

Figure 6. Pearson correlation analysis between hub gene mRNA expression and drug IC_{50} values. (A) *FOXA1*, (B) *ERBB3*, (C) *MUC1*, (D) *AGR2*.
Notes: Green indicates a negative correlation between drug IC_{50} and gene expression; red indicates a positive correlation.

for subsequent experimental investigation. Our integrated analysis serves as a filter, moving beyond correlation to suggest causality via MR and linking gene expression to clinically relevant drug responses. The genes identified here (*FOXA1*, *ERBB3*, *MUC1*) present a compelling rationale for future studies exploring their functional roles in modulating the tumor immune microenvironment and response to therapy, particularly to agents such as lapatinib. Thus, rather than a limitation, the computational design of this work represents a strength, offering a cost-effective basis to guide resource-intensive laboratory and clinical research toward the most viable targets.

5. Conclusion

Our study identified four M2-like TAM-related hub genes for predicting prognosis in breast cancer. These findings may contribute to improving prognostic assessment in breast cancer, and these genes may serve as potential biomarkers for clinical outcomes and immune responses.

Acknowledgments

None.

Funding

This work was supported by the Jianghan University Student Research Program [Grant Number 2024yb142].

Conflict of interest

The authors declare that they have no competing interests.

Author contributions

Conceptualization: Guang Yang
Formal Analysis: Guang Yang
Methodology: Guang Yang
Project administration: Zhaorui Hao
Supervision: Guang Yang
Validation: Zhaorui Hao, Xingrui Liu, Zixi Tang
Visualization: Zhaorui Hao, Xingrui Liu, Zixi Tang
Writing – Original Draft: Guang Yang
Writing – Review & Editing: Zhaorui Hao

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

All data were available in the TCGA database (https://portal.gdc.cancer.gov/), GEO database (https://www.ncbi.nlm.nih.gov/geo/), and GDSC database (https://www.cancerrxgene.org/). Questions regarding data can be directed to the corresponding author.

References

1. Sung H, Ferlay J, Siegel RL, *et al*. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71(3):209-249.

   doi: 10.3322/caac.21660

2. Hudson BI, Lippman ME. Comment on "The lingering mysteries of metastatic recurrence in breast cancer". *Br J Cancer*. 2023;128(3):484-485.

   doi: 10.1038/s41416-022-02012-0

3. Pathria P, Louis TL, Varner JA. Targeting tumor-associated macrophages in cancer. *Trends Immunol*. 2019;40(4): 310-327.

   doi: 10.1016/j.it.2019.02.003

4. Cassetta L, Pollard JW. Targeting macrophages: Therapeutic approaches in cancer. *Nat Rev Drug Discov*. 2018;17(12): 887-904.

   doi: 10.1038/nrd.2018.169

5. Mantovani A, Marchesi F, Malesci A, Laghi L, Allavena P. Tumour-associated macrophages as treatment targets in oncology. *Nat Rev Clin Oncol*. 2017;14(7):399-416.

   doi: 10.1038/nrclinonc.2016.217

6. Locati M, Curtale G, Mantovani A. Diversity, mechanisms, and significance of macrophage plasticity. *Annu Rev Pathol*. 2020;15:123-147.

   doi: 10.1146/annurev-pathmechdis-012418-012718

7. Chung W, Eum HH, Lee HO, *et al*. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun*. 2017;8:15081.

   doi: 10.1038/ncomms15081

8. Jansen H, Samani NJ, Schunkert H. Mendelian randomization studies in coronary artery disease. *Eur Heart J*. 2014;35(29):1917-1924.

   doi: 10.1093/eurheartj/ehu208

9. Emdin CA, Khera AV, Kathiresan S. Mendelian randomization. *JAMA*. 2017;318(19):1925-1926.

   doi: 10.1001/jama.2017.17219

10. Zhao S, Ye Z, Stanton R. Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA*. 2020;26(8):903-909.

    doi: 10.1261/rna.074922.120

11. Conesa A, Madrigal P, Tarazona S, *et al*. A survey of best practices for RNA-seq data analysis. *Genome Biol*. 2016;17:13.

    doi: 10.1186/s13059-016-0881-8

12. Newman AM, Steen CB, Liu CL, *et al*. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol*. 2019;37(7):773-782.

    doi: 10.1038/s41587-019-0114-2

13. Zheng X, Cao J, Wang H, *et al*. Effects of tauroursodeoxycholate on arsenic-induced hepatic injury in mice: A comparative transcriptomic analysis. *J Trace Elem Med Biol*. 2024;86:127512.

    doi: 10.1016/j.jtemb.2024.127512

14. Zheng H, Liu H, Ge Y, Wang X. Integrated single-cell and bulk RNA sequencing analysis identifies a cancer associated fibroblast-related signature for predicting prognosis and therapeutic responses in colorectal cancer. *Cancer Cell Int*. 2021;21(1):552.

    doi: 10.1186/s12935-021-02252-9

15. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;28(1):27-30.

    doi: 10.1093/nar/28.1.27

16. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. 2008;26(8):1364-1370.

    doi: 10.1200/JCO.2007.12.9791

17. Park SY. Nomogram: An analogue tool to deliver digital knowledge. *J Thorac Cardiovasc Surg*. 2018;155(4):1793.

    doi: 10.1016/j.jtcvs.2017.12.107

18. Matteini AM, Tanaka T, Karasik D, *et al*. GWAS analysis of handgrip and lower body strength in older adults in the CHARGE consortium. *Aging Cell*. 2016;15(5):792-800.

    doi: 10.1111/acel.12468.

19. Pegolo S, Cecchinato A, Savoia S, *et al*. Genome-wide association and pathway analysis of carcass and meat quality traits in Piemontese young bulls. *Animal*. 2020;14(2): 243-252.

    doi: 10.1017/S1751731119001812

20. Schooling CM, Johnson GD, Grassman J. Effects of blood lead on coronary artery disease and its risk factors: A Mendelian Randomization study. *Sci Rep***.** 2019;9(1):15995.

    doi: 10.1038/s41598-019-52482-1

21. Brennan K, Offiah G, McSherry EA, Hopkins AM. Tight junctions: A barrier to the initiation and progression of breast cancer? *J Biomed Biotechnol*. 2010;2010:460607.

    doi: 10.1155/2010/460607

22. Gong Z, He Y, Mi X, *et al*. Complement and coagulation cascades pathway-related signature as a predictor of immunotherapy in metastatic urothelial cancer. *Aging (Albany NY)*. 2023;15(18):9479-9498.

    doi: 10.18632/aging.205022

23. Trayes KP, Cokenakes SEH. Breast cancer treatment. *Am*

Fam Physician. 2021;104(2):171-178.

24. Lin M, Liang S, Jiang F, *et al*. 2003-2013, a valuable study: Autologous tumor lysate-pulsed dendritic cell immunotherapy with cytokine-induced killer cells improves survival in stage IV breast cancer. *Immunol Lett*. 2017;183:37-43.

    doi: 10.1016/j.imlet.2017.01.014

25. Barzaman K, Moradi-Kalbolandi S, Hosseinzadeh A, *et al*. Breast cancer immunotherapy: Current and novel approaches. *Int Immunopharmacol*. 2021;98:107886.

    doi: 10.1016/j.intimp.2021.107886

26. Gordon SR, Maute RL, Dulken BW, *et al*. PD-1 expression by tumour-associated macrophages inhibits phagocytosis and tumour immunity. *Nature*. 2017;545(7655):495-499.

    doi: 10.1038/nature22396

27. Cao X, Li B, Chen J, *et al*. Effect of cabazitaxel on macrophages improves CD47-targeted immunotherapy for triple-negative breast cancer. *J Immunother Cancer*. 2021;9(3):e002022.

    doi: 10.1136/jitc-2020-002022

28. Wang N, Wang S, Wang X, *et al*. Research trends in pharmacological modulation of tumor-associated macrophages. *Clin Transl Med*. 2021;11(1):e288.

    doi: 10.1002/ctm2.288

29. Xiang X, Wang J, Lu D, Xu X. Targeting tumor-associated macrophages to synergize tumor immunotherapy. *Signal Transduct Target Ther*. 2021;6(1):75.

    doi: 10.1038/s41392-021-00484-9

30. Litviakov N, Tsyganov M, Larionova I, *et al*. Expression of M2 macrophage markers YKL-39 and CCL18 in breast cancer is associated with the effect of neoadjuvant chemotherapy. *Cancer Chemother Pharmacol*. 2018;82(1):99-109.

    doi: 10.1007/s00280-018-3594-8

31. Muraoka D, Seo N, Hayashi T, *et al*. Antigen delivery targeted to tumor-associated macrophages overcomes tumor immune resistance. *J Clin Invest*. 2019;129(3): 1278-1294.

    doi: 10.1172/JCI97642

32. De Palma M, Biziato D, Petrova TV. Microenvironmental regulation of tumour angiogenesis. *Nat Rev Cancer*. 2017;17(8):457-474.

    doi: 10.1038/nrc.2017.51

33. Hara T, Chanoch-Myers R, Mathewson ND, *et al*. Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell*. 2021;39(6):779-792.e11.

    doi: 10.1016/j.ccell.2021.05.002

34. Qian BZ, Li J, Zhang H, *et al*. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature*. 2011;475(7355):222-225.

    doi: 10.1038/nature10138

35. Cassetta L, Fragkogianni S, Sims AH, *et al*. Human tumor-associated macrophage and monocyte transcriptional landscapes reveal cancer-specific reprogramming, biomarkers, and therapeutic targets. *Cancer Cell*. 2019;35(4):588-602.e10.

    doi: 10.1016/j.ccell.2019.02.009

36. Haddad A, Zoukar O, Daldoul A, *et al*. Breast diseases in women over the age of 65 in Monastir, Tunisia. *Pan Afr Med J*. 2018;31:67.

    doi: 10.11604/pamj.2018.31.67.16105

37. Arruabarrena-Aristorena A, Maag JLV, Kittane S, *et al*. FOXA1 mutations reveal distinct chromatin profiles and influence therapeutic response in breast cancer. *Cancer Cell*. 2020;38(4):534-550.e9.

    doi: 10.1016/j.ccell.2020.08.003

38. He Y, Wang L, Wei T, *et al*. FOXA1 overexpression suppresses interferon signaling and immune response in cancer. *J Clin Invest*. 2021;131(14):e147025.

    doi: 10.1172/JCI147025

39. Li Y, Pang Z, Dong X, *et al*. MUC1 induces M2 type macrophage influx during postpartum mammary gland involution and triggers breast cancer. *Oncotarget*. 2017;9(3):3446-3458.

    doi: 10.18632/oncotarget.23316

40. Gala H, Tomlinson I. The use of Mendelian randomisation to identify causal cancer risk factors: Promise and limitations. *J Pathol*. 2020;250(5):541-554.

    doi: 10.1002/path.5421

41. Hemani G, Zheng J, Elsworth B, *et al*. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408.

    doi: 10.7554/eLife.34408