


Identification of Novel Biomarkers Associated With the Prognosis and Potential Pathogenesis of Breast Cancer via Integrated Bioinformatics Analysis

Technology in Cancer Research & Treatment
 Volume 20: 1-16
 © The Author(s) 2021
 Article reuse guidelines:
sagepub.com/journals-permissions
 DOI: 10.1177/1533033821992081
journals.sagepub.com/home/tct


Meng Wu, MD¹ , Qingdai Li, MD¹ , and Hongbing Wang, PhD² 

Abstract

Background: Breast cancer is the most commonly diagnosed malignancy and a major cause of cancer-related deaths in women globally. Identification of novel prognostic and pathogenesis biomarkers play a pivotal role in the management of the disease. **Methods:** Three data sets from the GEO database were used to identify differentially expressed genes (DEGs) in breast cancer. Gene Ontology (GO) enrichment and Kyoto Encyclopaedia of Genes and Genomes pathway analyses were performed to elucidate the functional roles of the DEGs. Besides, we investigated the translational and protein expression levels and survival data of the DEGs in patients with breast cancer from the Gene Expression Profiling Interactive Analysis (GEPIA), Oncomine, Human Protein Atlas, and Kaplan Meier plotter tool databases. The corresponding change in the expression level of microRNAs in the DEGs was also predicted using miRWalk and TargetScan, and the expression profiles were analyzed using OncoMiR. Finally, the expression of novel DEGs were validated in Chinese breast cancer tissues by RT-qPCR. **Results:** A total of 46 DEGs were identified, and GO analysis revealed that these genes were mainly associated with biological processes involved in fatty acid, lipid localization, and regulation of lipid metabolism. Two novel biomarkers, *ADH1A* and *IGSF10*, and 4 other genes (*APOD*, *KIT*, *RBP4*, and *SFRP1*) that were implicated in the prognosis and pathogenesis of breast cancer, exhibited low expression levels in breast cancer tissues. Besides, 14/25 microRNAs targeting 6 genes were first predicted to be associated with breast cancer prognosis. RT-qPCR results of *ADH1A* and *IGSF10* expression in Chinese breast cancer tissues were consistent with the database analysis and showed significant down-regulation. **Conclusion:** *ADH1A*, *IGSF10*, and the 14 microRNAs were found to be potential novel biomarkers for the diagnosis, treatment, and prognosis of breast cancer.

Keywords

breast cancer, biomarker, differentially expressed genes, microRNA, prognosis, bioinformatics analysis

Abbreviations

ADH1A, alcohol dehydrogenase 1A; BP, biological process; CC, cellular component; GEO, Gene Expression Omnibus; GEPIA, Gene Expression Profiling Interactive Analysis; GO, Gene Ontology; HPA, human protein atlas; KEGG, Kyoto Encyclopedia of Genes and Genomes; MF, molecular function; miRNA, microRNA; NAD, nicotinamide adenine dinucleotide; ORA, Over-Representation Analysis; *RBP4*, retinol-binding protein 4; TCGA, The Cancer Genome Atlas

Received: August 20, 2020; Revised: December 27, 2020; Accepted: January 13, 2021.

Introduction

Breast cancer is one of the most commonly diagnosed cancer, resulting in high rates of cancer-related mortality in women.¹ It has a high incidence rate, and the American Institute for Cancer Research reported that in 2018, there were about 2 million new breast cancer cases. Breast cancer biomarkers play a pivotal role in clinical practice since they can be used to guide early

¹ Department of Medical Oncology, The Affiliated Clinical College of Xuzhou Medical University, Xuzhou, Jiangsu, China

² Department of Medical Oncology, The Affiliated Hospital of Xuzhou Medical University, Xuzhou, Jiangsu, China

Corresponding Author:

Hongbing Wang, PhD, Department of Medical Oncology, The Affiliated Hospital of Xuzhou Medical University, 99 Huaihai West Road, Xuzhou, Jiangsu 221002, China.
 Email: hbw0316@163.com



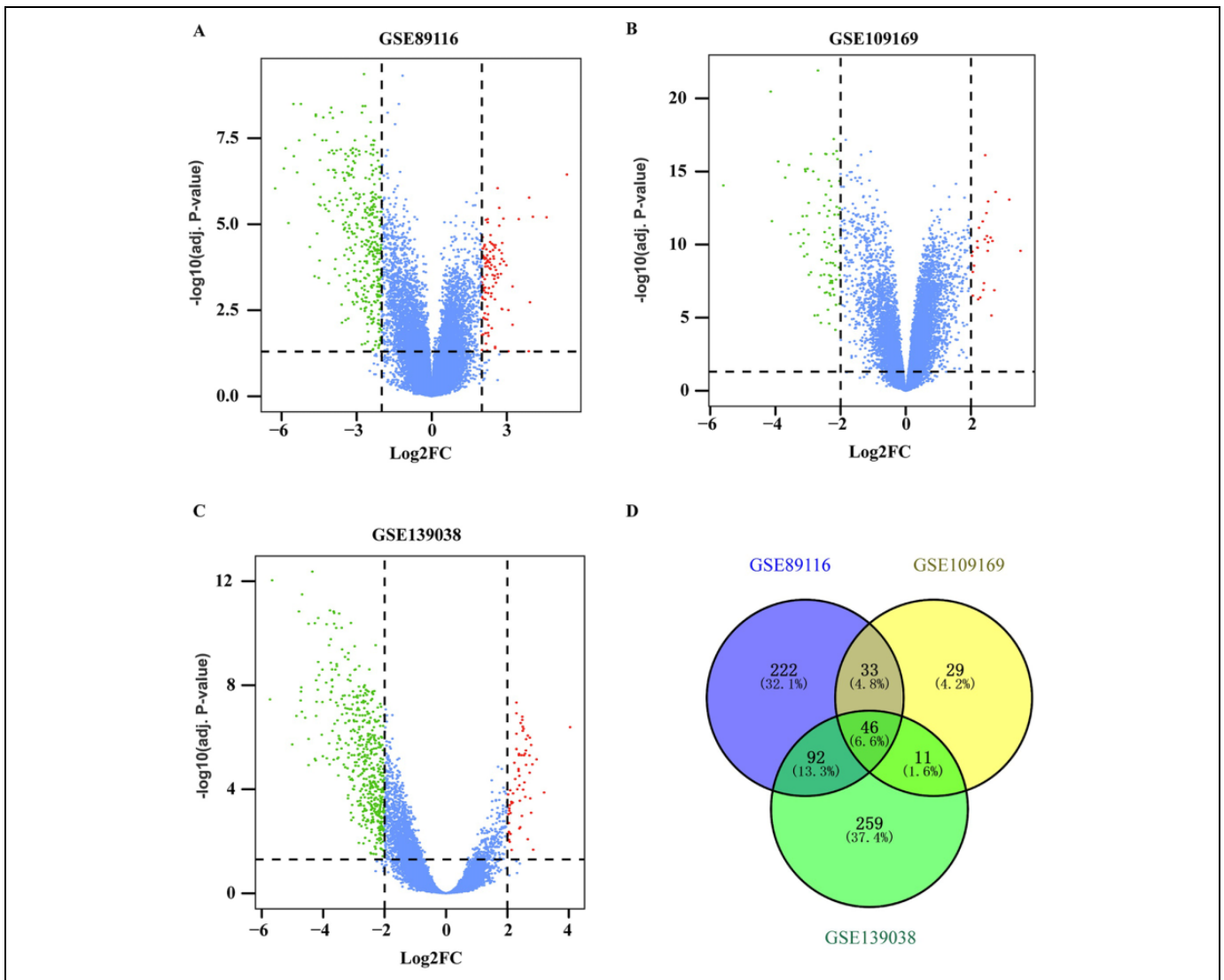


Figure 1. (A-C) Volcano plot of DEGs between breast cancer and normal breast tissues in each dataset. red dots, significantly upregulated genes in breast cancer tissues; green dots, significantly downregulated genes in breast cancer tissues; blue dots, not significantly expressed genes. $|\log_2FC| > 2$ and $\text{adj. } P < 0.05$ were considered to indicate a statistically significant difference. (D) Venn diagram of 46 overlapping DEGs from the GSE89116, GSE109169 and GSE139038 datasets. DEGs, differentially expressed genes.

diagnosis, individualized treatment, and prognosis. Although numerous risks and prognostic factors for breast cancer have been reported, and several biomarkers have also been identified, the mechanisms of breast cancer prognosis and pathogenesis remain to be fully elucidated. This renders the identification of novel biomarkers urgent.²

Comprehensive and in-depth cancer research entails the use of microarray and sequencing-based technologies to compare gene expression, for the accurate evaluation of the pathogenesis, and prediction of prognostic factors for various tumors.^{3,4} Recently, the advent of second-generation sequencing and protein profiling has markedly improved the efficiency of biological problem-solving, and these are currently widely used in cancer research.

This study aimed at identifying novel biomarkers associated with the prognosis and potential pathogenesis of breast

Table 1. A Total of 46 Overlapping DEGs Were Identified in the GSE89116, GSE109169, and GSE139038 Datasets, Including 5, Upregulated and 41 Downregulated.

The expression level of DEGs	Gene symbol
Upregulated	<i>NUSAP1, TOP2A, MMP1, MMP11, MMP13</i>
Downregulated	<i>ADH1A, IGSF10, KIT, RBP4, SFRP1, DMD, ADH1B, GPAM, CAV1, GSTM5, ABCA8, ACACB, CHRD1, PDK4, TIMP4, ADIPOQ, FIGF, ADHIC, PPARG, G0S2, SLC19A3, CFD, CD36, ABCA1, MAOA, TSHZ2, LPL, SRPX, TGFBR3, HLF, GPC3, ENPP2, FABP4, IGF1, DPT, MMRN1, OGN, MFAP4, APOD, FGF2, AKRIC2</i>

DEGs differentially expressed genes.

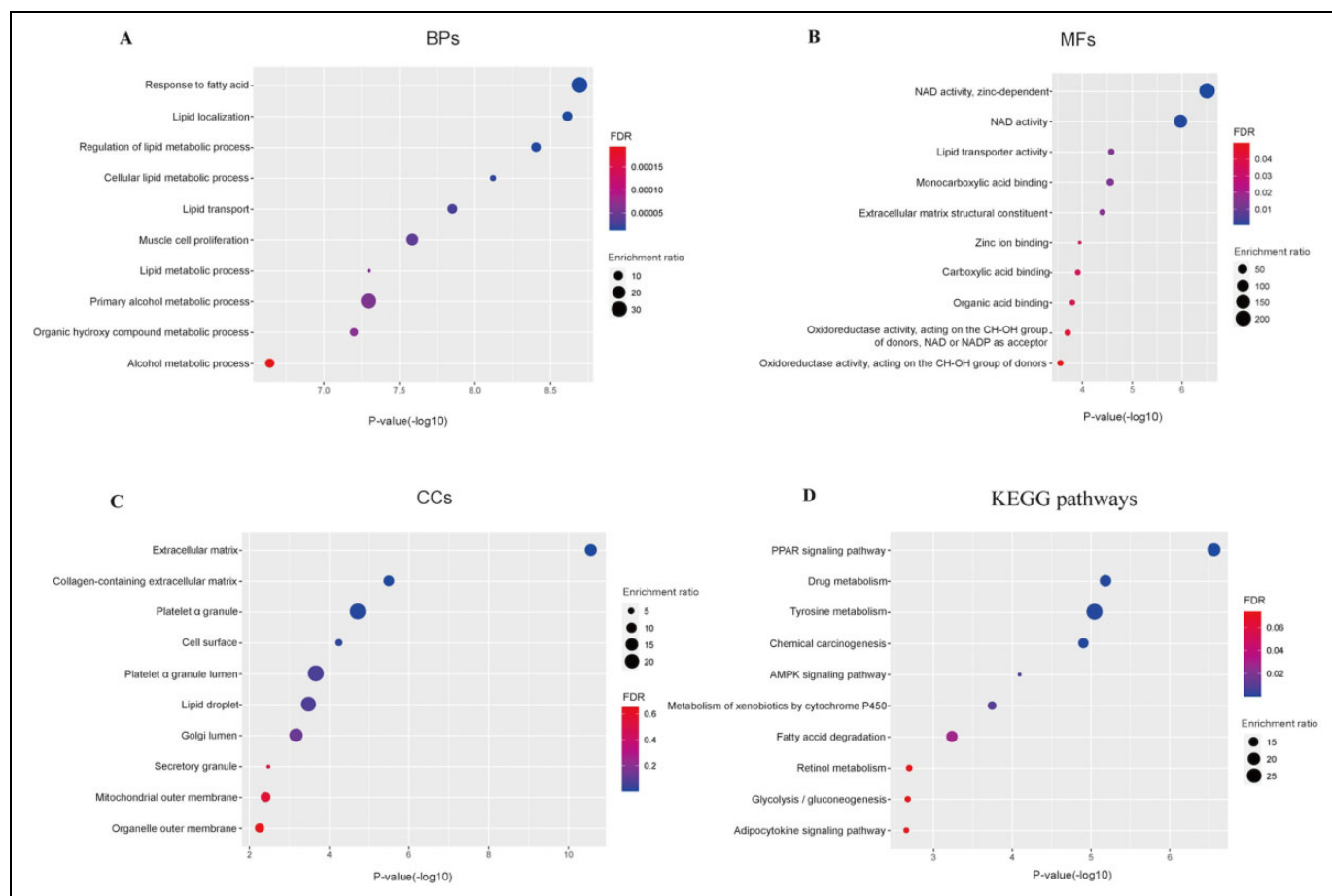


Figure 2. Enrichment analysis of 46 overlapping DEGs revealed top 10 geneSets in (A) BPs, (B) MFs, (C) CCs and (D) KEGG pathways by WebGestalt using the ggplot2 package in R language for visualization. DEGs, differentially expressed genes; BPs, Biological processes; MFs, Molecular Functions; CC, Cellular Components; KEGG, Kyoto Encyclopedia of Genes and Genomes; NAD, Alcohol Dehydrogenase; NADP, Nicotinamide Adenine Dinucleotide Phosphate.

cancer using integrated bioinformatics analysis. Briefly, the expression profiles of 3 genes, GSE89116, GSE109169, and GSE1390383, were downloaded from the Gene Expression Omnibus (GEO) database and analyzed using the GEO2 R online tool, followed by identification of the DEGs. The overlapping DEGs across the 3 datasets were analyzed using Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis.⁵ Significant DEGs were determined by analyzing the expression levels of the overlapping DEGs in breast tissues derived from breast cancer patients and normal subjects. The prognostic value of these factors was identified using Gene Expression Profiling Interactive Analysis (GEPIA),⁶ and translational and protein expression levels for the most significant DEGs were validated using OncoPrint and Human Protein Atlas (HPA), respectively.^{7,8} To identify the prognostic value of the most significantly expressed DEGs in breast cancer, Kaplan Meier plotter was used.⁹ miRWalk¹⁰ and TargetScan¹¹ were used to predict microRNA (miRNA) in the significantly expressed DEGs, and the intersecting miRNAs were verified using OncoPrint.¹²

These findings provide evidence of novel prognostic biomarkers for breast cancer.

Materials and Methods

Microarray Data

Expression profiles for the 3 genes, GSE89116, GSE109169, and GSE1390383, were downloaded from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>), a public functional genomics data repository that supports MIAME-compliant data submissions. In the database, tools are provided to help users query and download experimental and curated gene expression profiles. GSE89116 was sequenced on the GPL6947 platform (Illumina HumanHT-12 V3.0 expression BeadChip) from 9 breast cancer and 24 normal breast tissues, GSE109169 was generated on the GPL5175 platform, (HuEx-1_0-st) Affymetrix Human Exon 1.0 ST Array [transcript (gene) version] from 25 breast cancer and 25 normal breast tissues, and GSE139038 was sequenced on the GPL27630 platform [Print_1437 (Block_Column_Row IDs)] from 41 breast cancer and 18 normal breast tissues.¹³

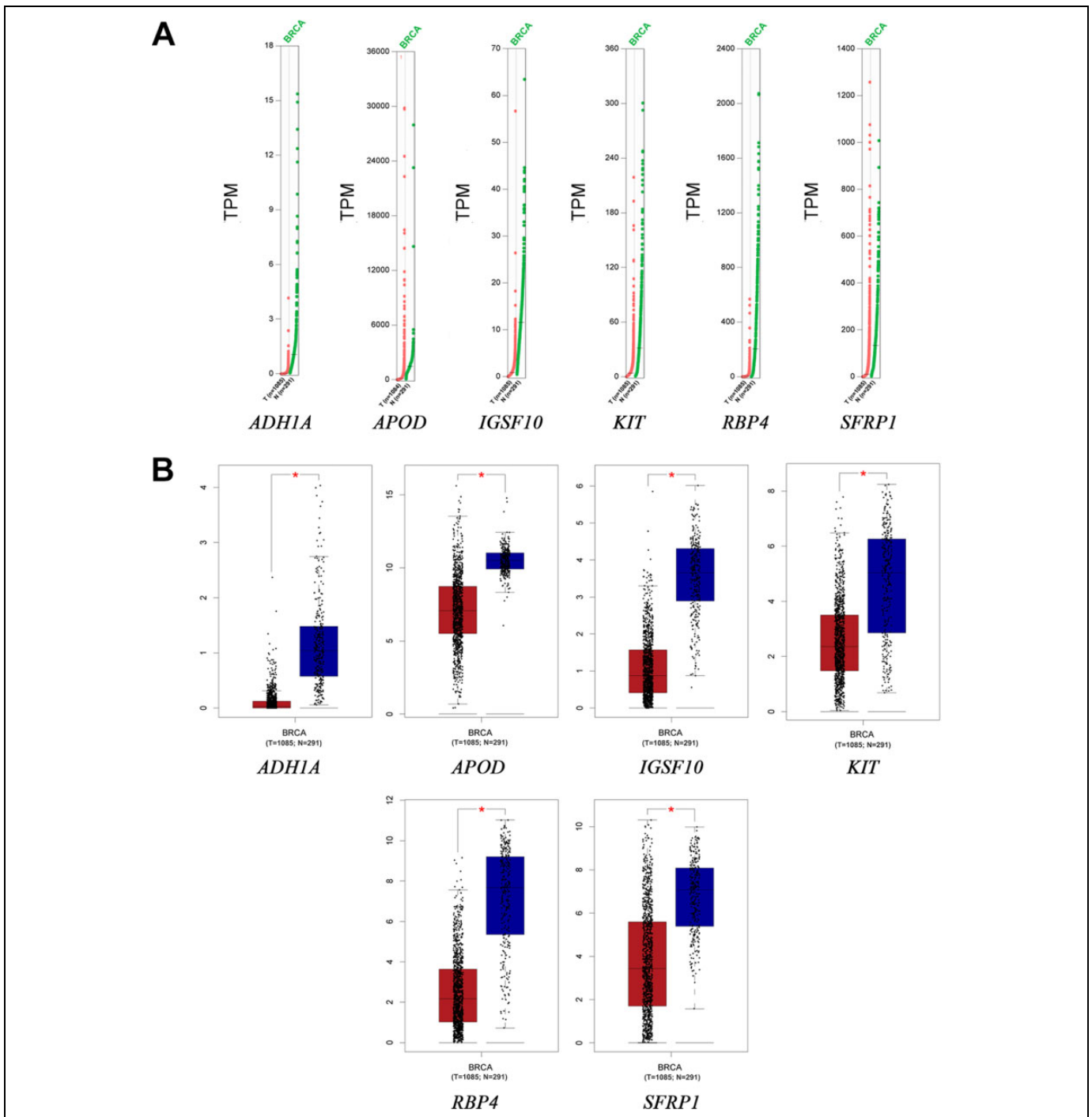


Figure 3. Expression level of ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1 between breast invasive carcinoma (BRCA) and normal breast tissues (GEPIA). (A) scatter diagram. (B) box plot. TPM, transcripts per million.

Identification of DEGs

DEGs were identified using GEO2 R (<http://www.ncbi.nlm.nih.gov/geo/geo2r>), an interactive online tool for comparing 2 or more datasets across experimental conditions. DEGs were identified using a classic t-test, as well as the standard of $|\log_{2}FC| > 2$, with $\text{adj. } P < 0.05$ was considered to be statistically significant.¹⁴ Overlapping DEGs across the 3 gene profiles

were identified using Venny 2.1 (<https://bioinfogp.cnb.csic.es/tools/venny/index.html>).

Enrichment Analysis

Overlapping DEGs were identified by GO enrichment and KEGG pathway analyses in WebGestalt (<http://www.webgestalt.org>),

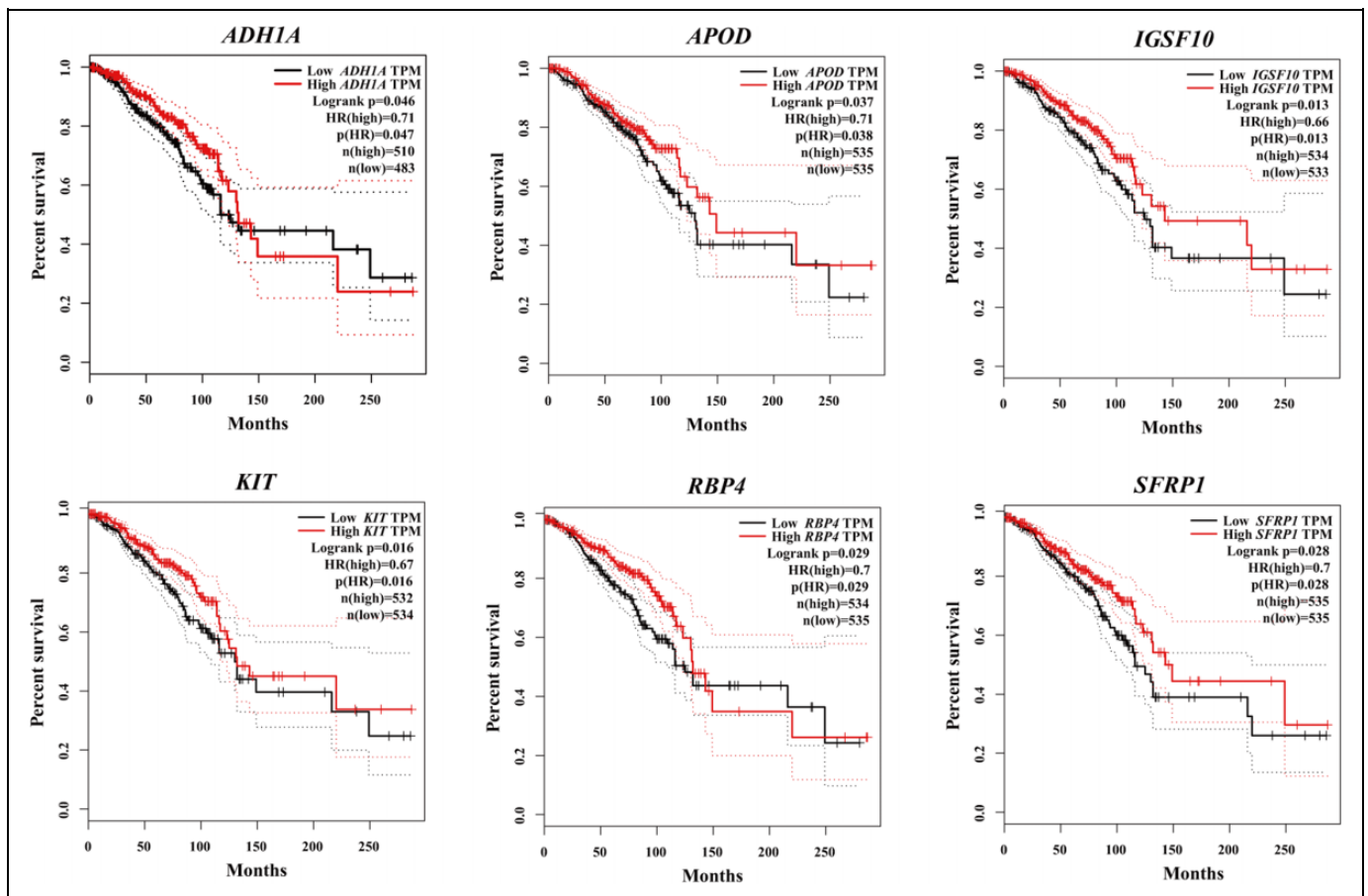


Figure 4. Prognostic value (overall survival) of the mRNA level of ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1 in breast cancer patients (GEPIA). TPM, transcripts per million.

using the Over-Representation Analysis (ORA) method. WebGestalt supports 3 well-established and complementary methods for enrichment analysis: ORA, Gene Set Enrichment Analysis, and Network Topology-based Analysis.⁵ GO enrichment targeted enriched biological process (BP), cellular component (CC), and molecular function (MF), with an FDR of <0.05 considered statistically significant.

Determination of Significant DEGs

To identify the most significant DEGs, the expression levels of overlapping DEGs in breast cancer tissues were compared with those from normal breast tissues using the standard fold change of >2 and $P < 0.01$. Prognostic values for breast cancer (Log rank $P < 0.05$) were determined using GEPIA (<http://gepia.cancer-pku.cn/>). This is a newly developed interactive web server analyzing RNA sequencing data for 9,736 tumors and 8,587 normal samples from The Cancer Genome Atlas (TCGA) and the GTEx projects, using a standard processing pipeline.⁶ The expression of the significantly expressed DEGs was analyzed across different pathological stages of breast cancer.

mRNA and Protein Expression Analysis of the Significantly Expressed DEGs

Analysis of the mRNA levels of DEGs in different cancer types was performed using OncoPrint gene expression array datasets (<https://www.oncoPrint.org/resource/login.html>, an online cancer microarray database). The resulting mRNA expression profiles from the significantly expressed DEGs across clinical cancer specimens were compared with those from normal controls, using a Student's t-test, $P < 0.01$, and fold change of 4. Furthermore, translational and protein expression levels of the significantly expressed DEGs were validated using OncoPrint and HPA, respectively. HPA (<https://www.proteinatlas.org/>) is a Swedish-based program initiated in 2003 that aids in the mapping of all human proteins in cells, tissues, and organs using the integration of various omics technologies, including antibody-based imaging, mass spectrometry-based proteomics, transcriptomics, and systems biology.⁷

Analysis of Prognostic Value

The identification of overall survival rates of the 6 significantly expressed DEGs in breast cancer was performed using the

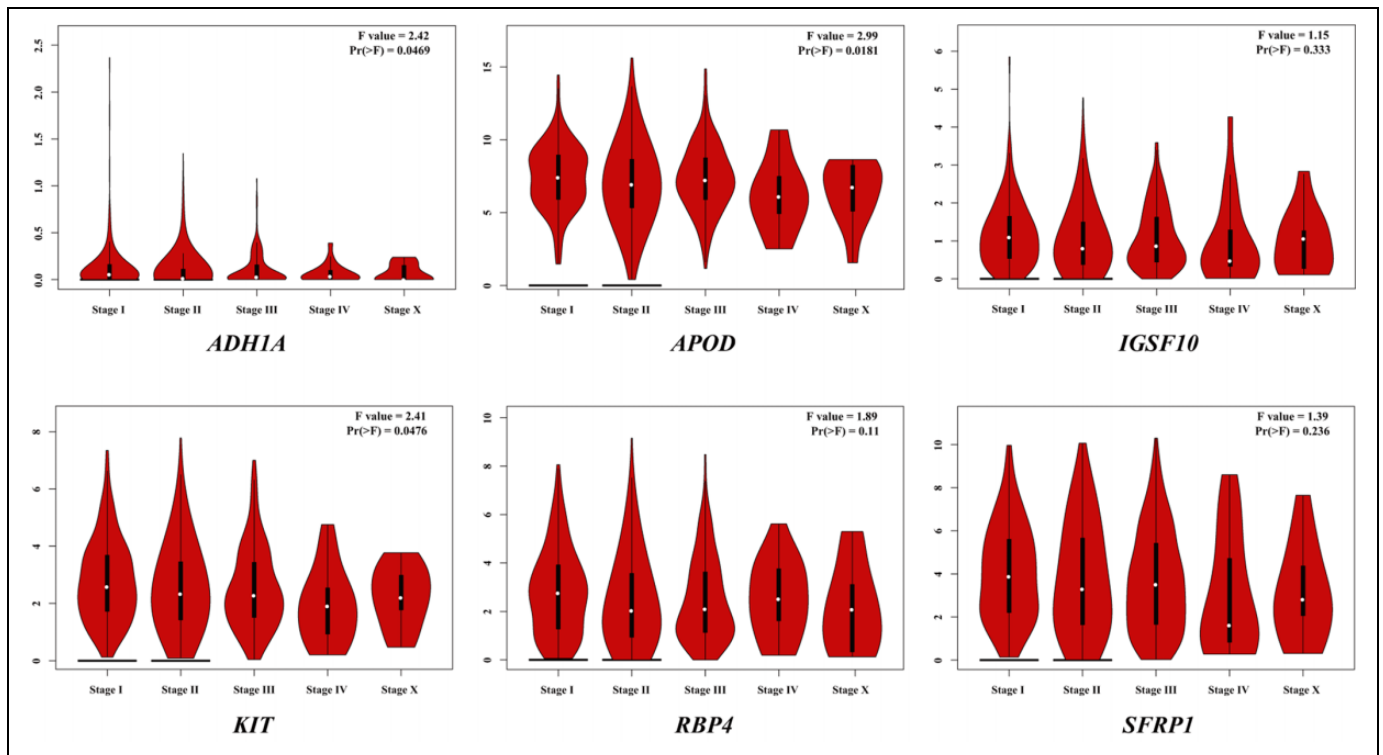


Figure 5. Correlation between the expression of the 6 most significant DEGs and tumor stage in breast cancer patients (gene expression profiling interactive analysis). DEGs, differentially expressed genes.

Kaplan Meier plotter (<https://kmplot.com/analysis>), an online tool used to assess the effect of 54 k genes on survival across 21 cancer types.⁹ A Log-rank $P < 0.05$ was considered to be statistically significant. The system comprised of a gene chip and RNA-seq data derived from GEO, EGA, and TCGA, while the primary purpose of the tool was meta-analysis-based discovery and validation of survival biomarkers.

miRNA Prediction

miRNAs in the 6 significantly expressed DEGs were predicted using online tools miRWalk (<http://mirwalk.umm.uni-heidelberg.de/>)¹⁰ and TargetScan (http://www.targetscan.org/vert_72/).¹¹ A Venn diagram was generated to describe the intersection. The resulting miRNA intersection was further analyzed using OncomiR (<http://www.oncomir.org/oncomir/index.html>),¹² and a t-test ($FDR < 0.05$) was performed to ascertain whether there were any significant differences in their expression levels, relative to those of corresponding target genes in breast cancer tissues. Finally, the overall survival rates of the miRNAs were analyzed using the Kaplan Meier plotter with a Log-rank $P < 0.05$.

Sample Collection

Ten samples from fresh tumor tissues and 10 samples from paired adjacent non-tumor tissues were collected from breast cancer patients after surgical resection at the Affiliated Hospital of Xuzhou Medical University, Xuzhou, China. This study

was carried out in accordance with the guidelines of the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the World Medical Association Declaration of Helsinki. The protocol was approved by the Affiliated Hospital of Xuzhou Medical University (Approval number: XYFY2019-KL095).

RNA Extraction and RT-qPCR

Total RNA was extracted using the Trizol reagent (Servicebio, China) according to the manufacturer's instructions. The cDNA was synthesized by using the Servicebio® RT First Strand cDNA Synthesis Kit (Servicebio, China). RT-qPCR was performed with 2×SYBR Green qPCR Master Mix (Low ROX) (Servicebio, China) on the ABI 7500 fast real-time PCR system. The amplification reaction procedure was as follows: 95°C for 10 min, followed by 95°C for 15 s and 60°C for 30 s for 40 cycles. GAPDH was applied as internal control for mRNA, and the relative expression level of mRNA was calculated by $2^{-\Delta\Delta ct}$ method. Primer sequences were listed in Table S1.

Results

Significantly Expressed DEGs

A total of 393 DEGs were obtained from the GSE89116 dataset. Among them, 97 were upregulated and 296 downregulated. A

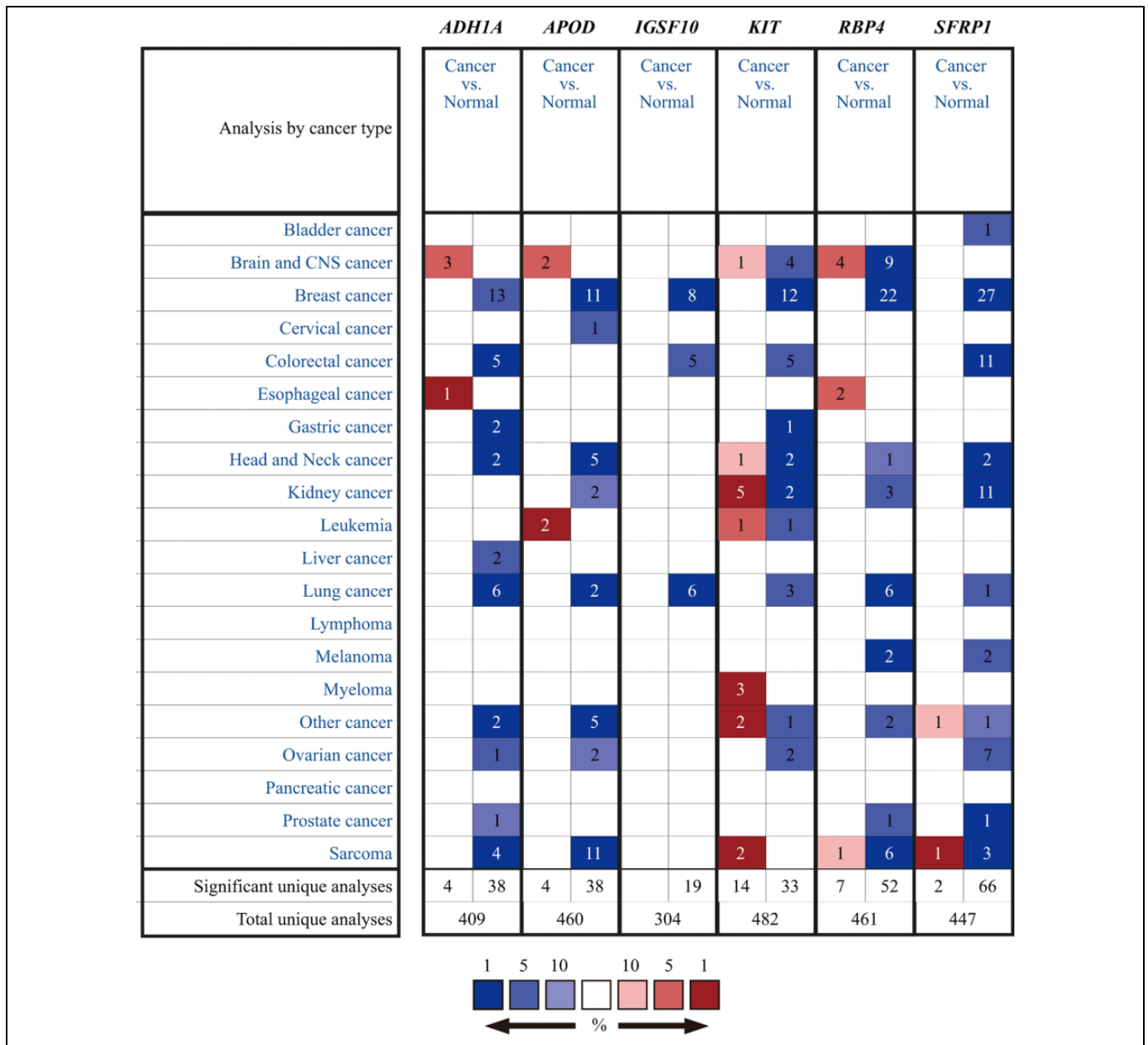


Figure 6. Transcription levels of ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1 in different types of cancer (Oncomine).

total of 119 DEGs were identified from GSE109169, 29 of which were upregulated and 90 downregulated. GSE139038 yielded 408 DEGs, 43 of which were upregulated and 365 downregulated. [$|\log_{2}FC| > 2$ and $adj. P < 0.05$; shown in volcano plots drawn in R software (ggpubr and ggthemes packages) (Figures 1A-C).

Besides, Venny 2.1 was used to identify 46 overlapping DEGs across the 3 datasets. Among them, 5 were upregulated and 41 downregulated (Figure 1D, Table 1).

Enrichment Analysis

Enrichment analysis revealed the top 10 gene sets in biological process (BP), molecular function (MF), cellular component

(CC), and KEGG pathways by WebGestalt using the ggplot2 package in the R language for visualization (Figure 2). The 46 DEGs were enriched in response to the following processes: Fatty acid, lipid localization, regulation of lipid metabolism, cellular lipid metabolism, lipid transport, muscle cell proliferation, lipid metabolism, primary alcohol metabolism, organic hydroxy compound metabolism and alcohol metabolism by BP analysis; nicotinamide adenine dinucleotide (NAD) activity, zinc-dependent, lipid transporter activity, monocarboxylic acid-binding, extracellular matrix structural constituent, zinc ion binding, carboxylic acid-binding, organic acid-binding, oxidoreductase activity, acting on the CH-OH group of donors, NAD or nicotinamide adenine dinucleotide phosphate as

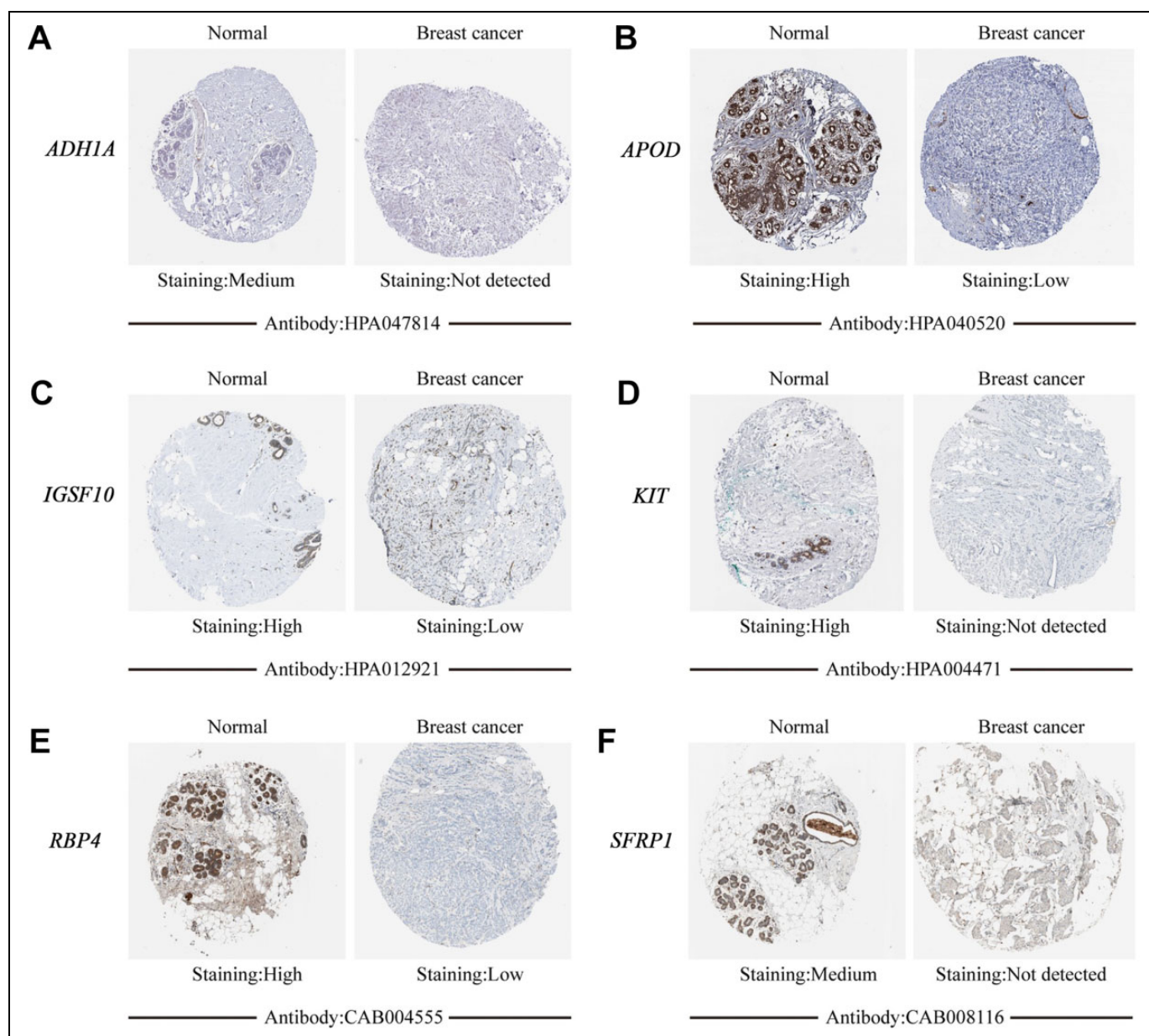


Figure 7. Representative immunohistochemical images of (A) *ADH1A*, (B) *APOD*, (C) *IGSF10*, (D) *KIT*, (E) *RBP4* and (F) *SFRP1* in breast cancer and normal breast tissues derived from the human protein atlas database.

acceptor, oxidoreductase activity, acting on CH-OH group of donors by MF analysis; extracellular matrix, collagen-containing extracellular matrix, platelet α granule and cell surface by CC analysis; and PPAR signaling pathway, drug metabolism, tyrosine metabolism, chemical carcinogenesis, AMPK signaling pathway, metabolism of xenobiotics by cytochrome P450 and fatty acid degradation by KEGG pathway analysis (FDR < 0.05).

Identification of the Significantly Expressed DEGs Through GEPIA

The most significant DEGs were identified by comparing the expression levels of the 46 DEGs between tissues from breast

cancer patients and those from normal subjects, followed by overall survival (OS) analysis using GEPIA. A total of 6 DEGs, *ADH1A*, *APOD*, *IGSF10*, *KIT*, *RBP4*, and *SFRP1*, were identified as the most significantly expressed, having met the expression threshold in the GEO and GEPIA databases (Figure 3). These DEGs were also associated with the prognosis of breast cancer following their low expression in tumor tissues (Figure 4). Among these DEGs, *ADH1A* and *IGSF10* were found to be novel biomarkers for breast cancer prognosis. Furthermore, the expression levels of *ADH1A* ($P = 0.047$), *APOD* ($P = 0.02$), and *KIT* ($P = 0.047$) were differentially expressed across tumor stages, whereas those of *IGSF10* ($P = 0.33$), *RBP4* ($P = 0.11$), and *SFRP1* ($P = 0.24$) were not (Figure 5).

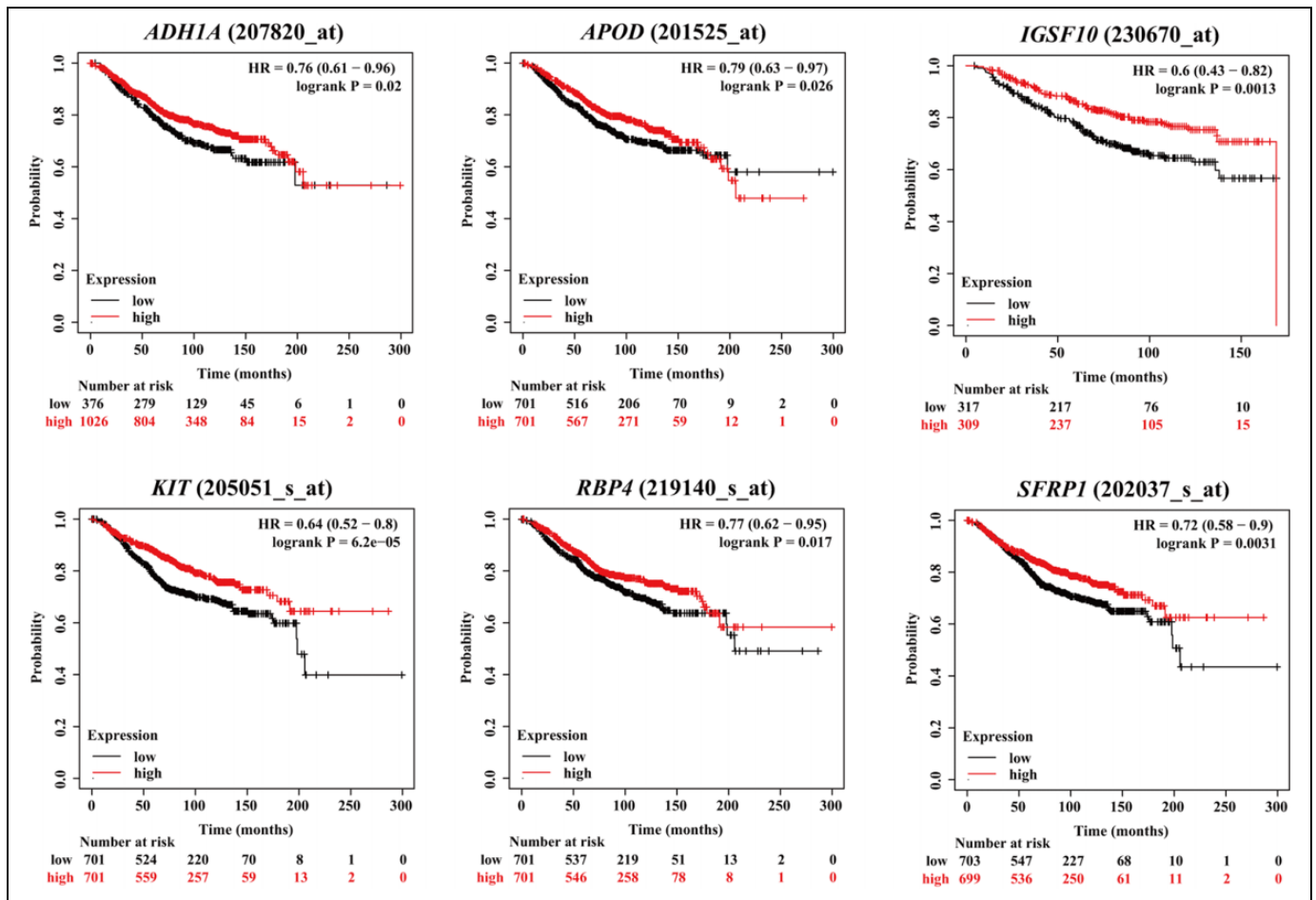


Figure 8. Prognostic value (overall survival) of mRNA level of ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1 in breast cancer patients (Kaplan-Meier plotter).

Profiles of mRNA and Protein Expression

Analysis of the expression levels in the 6 significantly expressed DEGs using Oncomine revealed lower levels in breast cancer when compared with normal tissues (Figure 6). Similarly, immunohistochemistry data from the HPA project, revealed that the proportions with low or not detected expression of ADH1A, APOD, IGSF10, KIT, RBP4, and SFRP1 were 11/12, 2/12, 2/12, 47/47, 20/23, and 9/12 in breast cancer samples, respectively, whereas the proportions with low or not detected expression of ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1 were 2/7, 0/2, 0/1, 2/7, 2/5 and 0/3 in normal breast tissues (Figure 7).

Prognostic Value of the 6 Most Significantly Expressed DEGs

Results from the Kaplan Meier plotter revealed a positive correlation between the expression of ADH1A ($P = 0.02$), APOD ($P = 0.026$), IGSF10 ($P = 0.0013$), KIT ($P = 6.2e-05$), RBP4 ($P = 0.017$), and SFRP1 ($P = 0.0031$) and good prognosis in breast cancer (Log rank $P < 0.05$). These results indicated that the 6 genes were prognostic factors for breast cancer (Figure 8).

miRNA Prediction and Prognostic Analysis

The miRWalk tool predicted 1,570, 1,417, 2,140, 2,000, 1,338 and 1,624 miRNAs corresponding to the 6 significantly expressed DEGs. Lower numbers of the significantly expressed DEGs (74, 71, 38, 35, 24 and 11 corresponding to ADH1A, APOD, IGSF10, KIT, RBP4 and SFRP1, respectively) were obtained using TargetScan. The analysis of the resulting Venn diagram revealed 32, 49, 26, 18, 12 and 5 intersecting miRNAs. Verification of expression generated the following 25 miRNAs across the 6 significantly expressed DEGs (Cytoscape software for visualization); hsa-miR-9-5p, hsa-miR-877-5p, hsa-miR-194-5p, hsa-miR-509-3p, hsa-miR-1301-3p, hsa-miR-17-5p, hsa-miR-93-5p, hsa-miR-182-5p, hsa-miR-153-3p, hsa-miR-106b-5p, hsa-miR-330-5p, hsa-miR-20b-5p, hsa-miR-493-5p, hsa-miR-23c, hsa-miR-18a-5p, hsa-miR-148a-3p, hsa-miR-193a-3p, hsa-miR-130b-3p, hsa-miR-148b-3p, hsa-miR-193b-3p, hsa-miR-454-3p, hsa-miR-34a-3p, hsa-miR-185-5p, hsa-miR-27a-3p and hsa-miR-128-3p (Figure 9, Table 2). Analysis of overall survival across the 25 miRNAs revealed that hsa-miR-9-5p ($P = 0.0046$), hsa-miR-877-5p ($P = 0.018$), hsa-miR-509-3p ($P = 0.01$), hsa-miR-1301-3p ($P = 0.036$), hsa-miR-17-5p ($P = 0.02$), hsa-miR-93-5p ($P = 0.036$), hsa-miR-106b-5p

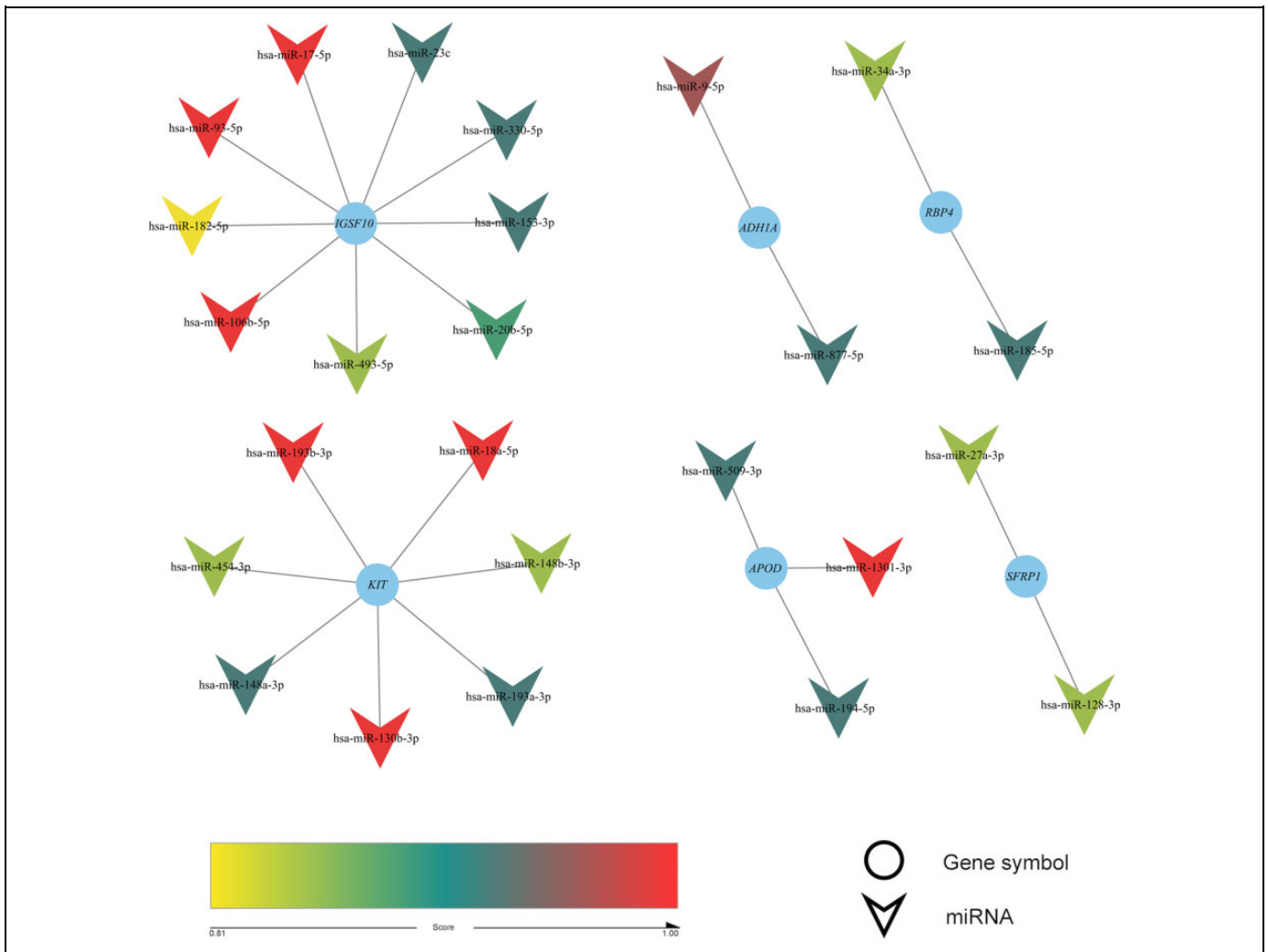


Figure 9. A total of 25 negative correlation pairs of miRNA-mRNA were predicted by miRWalk, targetscan and oncomir online tool. cytoscape software was used to construct the interaction network.

($P = 0.00055$), hsa-miR-493-5p ($P = 0.012$), hsa-miR-23c ($P = 0.003$), hsa-miR-18a-5p ($P = 0.00087$), hsa-miR-193a-3p ($P = 0.012$), hsa-miR-130b-3p ($P = 0.00024$), hsa-miR-148b-3p ($P = 0.04$), hsa-miR-193b-3p ($P = 0.018$), hsa-miR-454-3p ($P = 0.018$), hsa-miR-34a-3p ($P = 0.027$) and hsa-miR-185-5p ($P = 0.0051$) were associated with poor prognosis. Conversely, hsa-miR-182-5p ($P = 0.035$), hsa-miR-153-3p ($P = 0.0064$), hsa-miR-20b-5p ($P = 0.028$) and hsa-miR-148a-3p ($P = 0.0039$) were associated with good prognosis, whereas hsa-miR-194-5p ($P = 0.92$), hsa-miR-330-5p ($P = 0.86$), hsa-miR-27a-3p ($P = 0.17$) and hsa-miR-128-3p ($P = 0.78$) showed no significant association. Overall, 21 miRNAs were associated with breast cancer prognosis (Figure 10).

RT-qPCR Validation of the 2 DEGs in Chinese Breast Cancer Tissues

To further verify the expression of the 2 novel DEGs in breast cancer tissues, we detected their expression in 10 samples of

tumor tissues from Chinese breast cancer and 10 samples of adjacent non-tumor tissues. RT-qPCR results of *ADH1A* ($P < 0.01$) and *IGSF10* ($P < 0.01$) expression in Chinese breast cancer tissues were consistent with the database analysis and showed significant down-regulation. (Figure 11).

Discussion

Analysis of the 3 datasets from the GEO database revealed a total of 46 overlapping DEGs, out of which 5 and 41 were up- and downregulated, respectively. GO enrichment and KEGG pathway analysis of the DEGs further revealed the top 10 geneSets in BP, MF, and CC. Previous studies have described the importance of PPAR signaling pathway-related genes as predictors of neoadjuvant chemotherapy response and in early diagnosis of breast cancer.^{15,16} Particularly, oncogenic tyrosine kinase has been implicated in the induction of multiple types of cancer, including hematological malignancies, breast, prostate, colon, and lung cancer, as well as breast and pancreatic cancer progression.^{17,18} This signaling pathway may therefore be

Table 2. The miRNA Expression Levels of *ADH1A*, *APOD*, *IGSF10*, *KIT*, *RBP4*, and *SFRP1* Shows Opposite Expression Level or Their Corresponding Target Genes in Breast Cancer Tissue, as Determined by OncomiR (t-test, FDR < 0.05).

Gene symbol	miRNA name	t-test FDR	Upregulated in:	Tumor Log2 Mean expression	Normal Log2 Mean expression
<i>ADH1A</i>	hsa-miR-9-5p	1.41E-02	Breast cancer	9.82	9.04
	hsa-miR-877-5p	1.73E-08	Breast cancer	0.74	0.08
<i>APOD</i>	hsa-miR-194-5p	1.06E-03	Breast cancer	7.25	6.94
	hsa-miR-509-3p	1.01E-02	Breast cancer	1.98	1.66
<i>IGSF10</i>	hsa-miR-1301-3p	5.64E-26	Breast cancer	3.87	1.82
	hsa-miR-17-5p	8.25E-05	Breast cancer	8.2	7.62
	hsa-miR-93-5p	8.34E-13	Breast cancer	11.92	11.14
	hsa-miR-182-5p	1.28E-30	Breast cancer	15.11	12.97
	hsa-miR-153-3p	1.12E-02	Breast cancer	1.57	1.06
	hsa-miR-106b-5p	1.22E-15	Breast cancer	7.61	6.62
	hsa-miR-330-5p	1.71E-11	Breast cancer	4.95	4.2
	hsa-miR-20b-5p	1.90E-04	Breast cancer	4.17	3.28
	hsa-miR-493-5p	4.10E-14	Breast cancer	2.99	1.6
	hsa-miR-23c	6.28E-02	Breast cancer	0.14	0.04
<i>KIT</i>	hsa-miR-18a-5p	1.20E-05	Breast cancer	2.52	1.75
	hsa-miR-148a-3p	8.72E-12	Breast cancer	15.51	14.6
	hsa-miR-193a-3p	1.76E-03	Breast cancer	4.05	3.63
	hsa-miR-130b-3p	1.56E-20	Breast cancer	3.36	1.82
	hsa-miR-148b-3p	6.61E-24	Breast cancer	7.71	6.42
	hsa-miR-193b-3p	4.07E-03	Breast cancer	6.94	6.55
<i>RBP4</i>	hsa-miR-454-3p	1.31E-18	Breast cancer	3.2	1.61
	hsa-miR-34a-3p	3.15E-03	Breast cancer	0.26	0.09
<i>SFRP1</i>	hsa-miR-185-5p	3.49E-02	Breast cancer	5.59	5.32
	hsa-miR-27a-3p	1.72E-05	Breast cancer	9.96	9.54
	hsa-miR-128-3p	4.96E-11	Breast cancer	6.72	5.99

miRNA, microRNA.

considered an important mechanism for breast cancer development. Numerous animal studies have reported that several common environmental chemicals are mammary glands carcinogens. These chemicals have also been shown to activate related hormone response pathways and enhance the sensitivity of the mammary glands to carcinogenesis.¹⁹ AMPK is closely associated with the tumor-suppressive functions of LKB1 and P53, where it regulates cell survival signals, such as mTOR and Akt, leading to cell growth inhibition and cell cycle arrest. Besides, it participates in tumor drug resistance.²⁰ CYP2E1, on the other hand, is involved in the progression and metastasis of advanced breast cancer,²¹ whereas JAK/STAT3-regulated fatty acid β -oxidation is essential for self-renewal and drug resistance in breast cancer stem cells.²² The GO enrichment and KEGG pathway analysis results were consistent with findings from previous studies.

The present results indicated that *ADH1A*, *APOD*, *IGSF10*, *KIT*, *RBP4*, and *SFRP1* were the most significantly expressed DEGs, and their expression levels were found to be consistent with microarray data from the GEO database. Besides, these DEGs showed an association with breast cancer patient prognosis. The expression levels of *ADH1A*, *APOD*, and *KIT* were significantly different across different tumor stages, whereas Oncomine and HPA indicated that the 6 DEGs exhibited significantly lower expression in breast cancer patients, compared to normal subjects. Moreover, Kaplan Meier plotter results

revealed a positive correlation between the DEGs and breast cancer patient prognosis.

Previous studies have shown that alcohol dehydrogenase 1A (*ADH1A*) is positively correlated with the prognosis of hepatocellular carcinoma, as well as gastric and pancreatic cancer, confirming its potential as a biomarker.²³⁻²⁵ Besides, *ADH1A* participates in the occurrence and development of breast cancer via lipid metabolism and the PPAR pathway,²⁶ which is consistent with the results of our previous study, where PPAR signaling pathway and fatty acid degradation were identified using KEGG pathway analysis. In the present study, *ADH1A* was found to be a prognostic biomarker for breast cancer and has further been associated with tumor stage. *APOD* is a lipocalin that participates in various cellular processes, including cytoprotection, and is a biomarker positively correlated with the prognosis of breast and prostate cancer.²⁷⁻²⁹ *IGSF10* is the 10th member of the immunoglobulin superfamily and a biomarker positively correlated with the prognosis of lung cancer. Knocking out *IGSF10* is reported to activate the integrin β 1/FAK pathway, thereby affecting proliferation and adhesion in lung cancer cells.³⁰ These findings confirmed, for the first time, that *IGSF10* is positively correlated with breast cancer prognosis; and the underlying pathogenesis mechanism has been hypothesized to be the activation of the integrin β 1/FAK pathway with decreased expression of *IGSF10*.

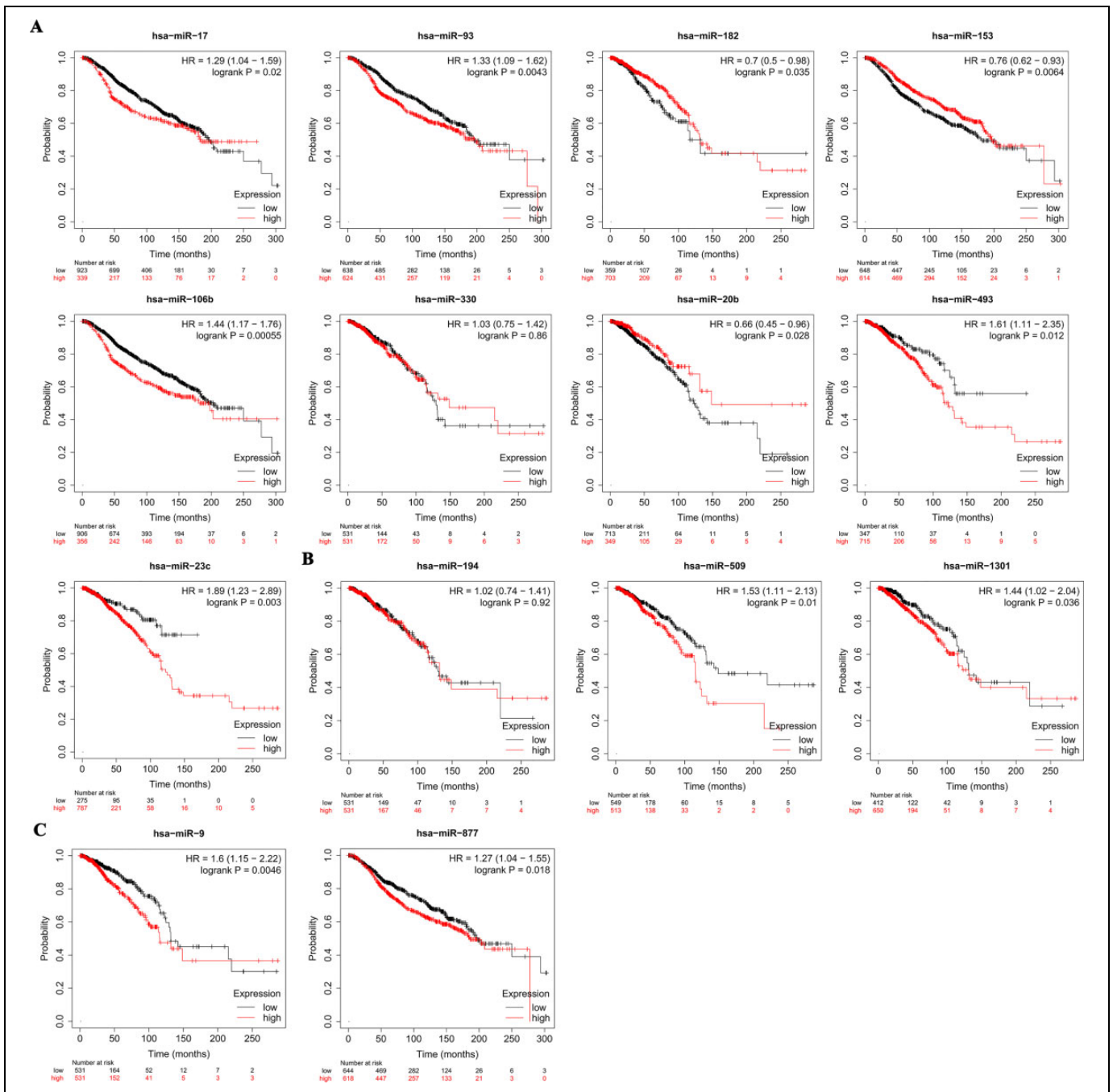


Figure 10. The prognostic value (overall survival) of the mRNA level of (A) IGSF10, (B) APOD, (C) ADH1A, (D) RBP4, (E) KIT, and (F) SFRP1 in breast cancer patients (Kaplan-Meier plotter).

The present findings revealed downregulation of retinol-binding protein 4 (*RBP4*) and a positive association between this downregulation and breast cancer prognosis. Previous studies have implicated increased *RBP4* levels with breast cancer risk.³¹ Particularly, metastatic breast cancer patients exhibit higher *RBP4* levels compared with healthy patients and non-metastatic breast cancer patients.³² This indicates that *RBP4* is a biomarker for poor prognosis, although other studies have reported an association between low expression levels and

breast cancer occurrence.¹⁶ Therefore, the association between the expression level and pathogenesis of *RBP4* in breast cancer remains unclear, necessitating further research. Functionally, *RBP4* participates in the pathogenesis of breast cancer through other molecules during lipid metabolism. *KIT*, a Receptor Tyrosine Kinase absent in 75% of breast-invasive ductal carcinoma, has previously been associated with lymph node metastasis.³³ Various tyrosine kinases, including *KIT*, *FLT3*, colony-stimulating factor 1 (*CSF-1*), and *RET* have been

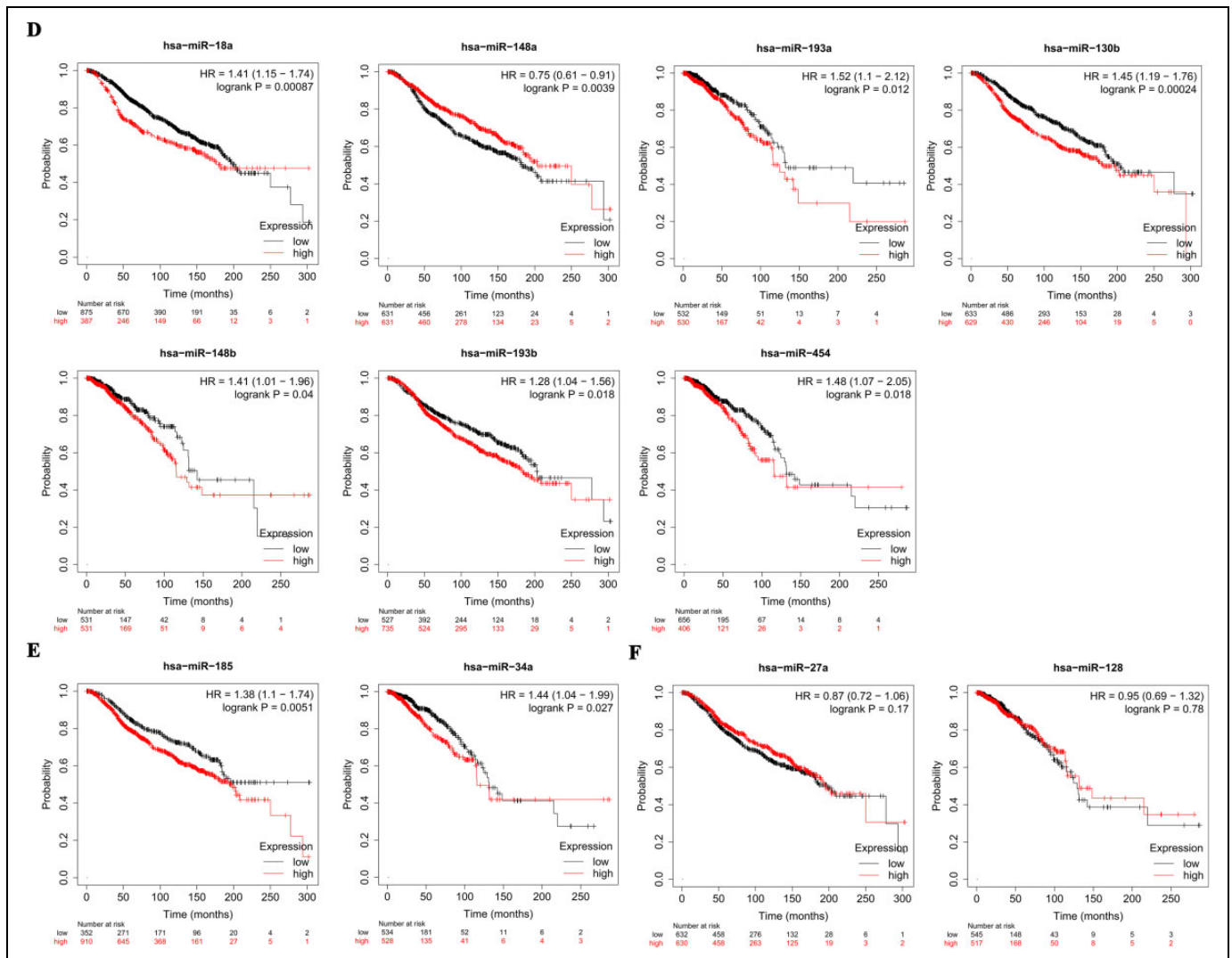


Figure 10. (Continued).

implicated in several malignant tumors, such as small cell lung cancer, gastrointestinal stromal tumor, breast cancer, and acute myeloid leukemia.³⁴ These studies indicate that a low *KIT* expression is associated with poor breast cancer prognosis.

The results of the present study revealed a low expression of *SFRP1* in breast cancer patients and were positively correlated with the disease prognosis. However, previous studies have shown that *SFRP1* is significantly overexpressed in triple-negative breast cancer, compared with other breast cancer subtypes, and its knockdown results in increased resistance to the standard chemotherapy for breast cancer.³⁵ These results indicate that breast cancer is a highly heterogeneous disease, and the same gene may exhibit different expression profiles across various subtypes. Therefore, individualized treatment is particularly important. miRNAs are a type of 18-25-nucleotide-long non-coding RNAs, which play crucial roles in human cancer. In the present study, miRWalk and TargetScan were used to predict corresponding miRNAs for the significantly expressed

DEGs. Their expression levels were verified using OncoMiR and their prognostic values were further predicted by Kaplan Meier plotter. Overall, 21 miRNAs associated with breast cancer prognosis were predicted, with 17 and 4 found to be negatively and positively associated with breast cancer prognosis, respectively.

Among these miRNAs, hsa-miR-9-5p, hsa-miR-18a-5p,³⁶ hsa-miR-17-5p,³⁷ hsa-miR-106b-5p,³⁸ hsa-miR-454-3p,³⁹ hsa-miR-182-5p⁴⁰ and hsa-miR-153-3p⁴¹ have previously been implicated in breast cancer prognosis, whereas hsa-miR-877-5p has been associated with the pathogenesis and treatment of hepatocellular carcinoma.⁴² On the other hand, hsa-miR-509-3p is associated with the prognosis of colorectal cancer⁴³ and hsa-miR-1301-3p with the recurrence of prostate cancer following radical prostatectomy.⁴⁴ hsa-miR-493-5p has been reported in adenovirus-infected children with a clear difference between infected and healthy people,⁴⁵ and hsa-miR-23c has been implicated in the migration and invasion of non-small cell

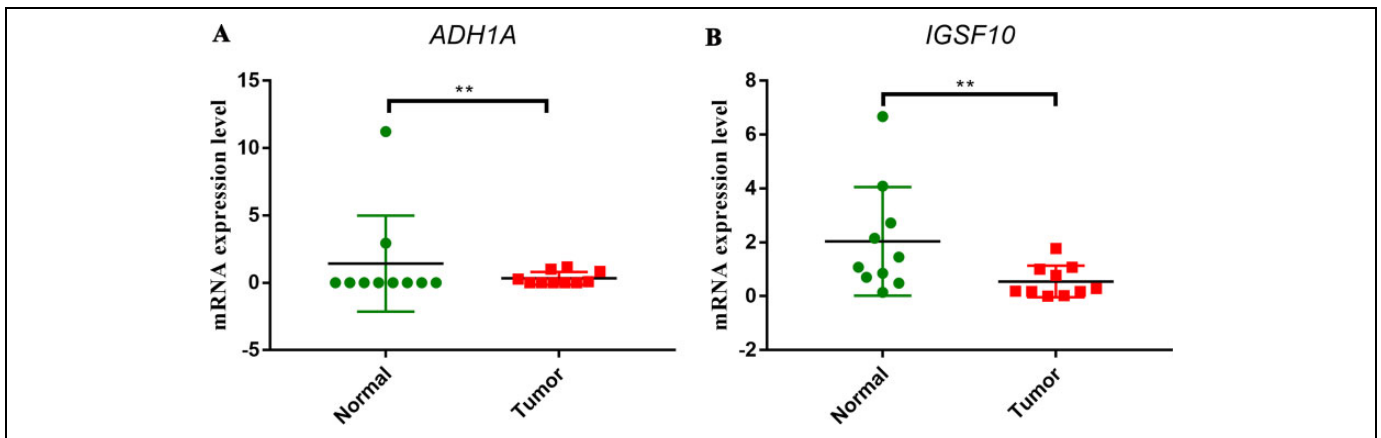


Figure 11. RT-qPCR validation of 2 DEGs in Chinese breast cancer tissues. (A) expression level of ADH1A gene, (B) expression level of IGSF10 gene. **P < 0.01.

lung cancer cells.⁴⁶ hsa-miR-193a-3p and hsa-miR-193b-3p are associated with precancerous lesions of gastric cancer,^{47,48} and hsa-miR-130b-3p is involved in the occurrence and development of prostate and liver cancer, and in metastatic adrenocortical carcinoma.⁴⁹⁻⁵¹ Moreover, studies have shown that hsa-miR-148a-3p and hsa-miR-148b-3p are downregulated and enriched in gastric cancer-related pathways,⁵² whereas hsa-miR-34a-3p is associated with metastatic conjunctival melanoma.⁵³ A defect in hsa-miR-185-5p is associated with the late-stage and low overall survival rates of bladder cancer,⁵⁴ whereas hsa-miR-20b-5p has been implicated in the prognosis of colorectal cancer.⁵⁵ Finally, hsa-miR-148a-3p has been linked to drug resistance and the invasiveness of esophageal squamous cell carcinoma. Although these studies describe the association between miRNAs with the pathogenesis and prognosis of other cancer types, their expression levels and roles in breast cancer remain unknown. However, the present results revealed that the 14 miRNAs were highly expressed in breast cancer patients and further associated with the prognosis of the disease. These miRNAs may provide new targets for the development of breast cancer treatment.

We further validated the expression of the 2 novel DEGs in Chinese breast cancer tissues. Our results indicated that the expression of *ADH1A* and *IGSF10* in Chinese breast cancer tissues was consistent with the database analysis and showed significant down-regulation.

The present study has some limitations. First, although new biomarkers associated with breast cancer are predicted, their mechanism of action remains unclear, necessitating further research. Secondly, the results need validation through experimental studies.

Conclusions

In the present study, 46 candidate DEGs are identified and their BP, MF and CC revealed. Further analysis of these genes reveals 6 significant DEGs that are downregulated in breast cancer patients, including 2 novel biomarkers (*ADH1A* and

IGSF10), and 4 other genes (*APOD*, *KIT*, *RBP4* and *SFRP1*) that are associated with the prognosis and potential pathogenesis of breast cancer. Besides, 25 miRNAs are predicted in the 6 genes, with 14 of these associated with breast cancer prognosis for the first time. Overall, the DEGs and miRNAs identified herein represent new biomarkers for the prediction, diagnosis and development of individualized prevention, treatment and prognosis approaches for breast cancer patients.

Ethics Approval and Consent to Participate

This study was carried out in accordance with the guidelines of the Ethics Committee of the Affiliated Hospital of Xuzhou Medical University, with written informed consent from all subjects. All subjects gave written informed consent in accordance with the World Medical Association Declaration of Helsinki. The protocol was approved by the Affiliated Hospital of Xuzhou Medical University (Approval number: XYFY2019-KL095).


Declaration of Conflicting Interests


The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.


Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Meng Wu, MD  <https://orcid.org/0000-0001-8972-8302>

Qingdai Li, MD  <https://orcid.org/0000-0001-8674-4331>

Hongbing Wang, PhD  <https://orcid.org/0000-0002-6480-5540>

Supplemental Material

Supplemental material for this article is available online.

References

- DeSantis CE, Ma J, Gaudet MM, et al. Breast cancer statistics, 2019. *CA Cancer J Clin.* 2019;69(6):438-451. doi:10.3322/caac.21583

2. Zhu QL, Xu WH, Tao MH. Biomarkers of the metabolic syndrome and breast cancer prognosis. *Cancers (Basel)*. 2010;2(2):721-739. doi:10.3390/cancers2020721
3. Tyanova S, Albrechtsen R, Kronqvist P, Cox J, Mann M, Geiger T. Proteomic maps of breast cancer subtypes. *Nat Commun*. 2016;7. doi:10.1038/ncomms10259
4. van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*. 2002;347(25):1999-2009. doi:10.1056/NEJMoa021967
5. Liao Y, Wang J, Jaehnig EJ, Shi Z, Zhang B. WebGestalt 2019: gene set analysis toolkit with revamped UIs and APIs. *Nucleic Acids Res*. 2019;47(W1):W199-W205. doi:10.1093/nar/gkz401
6. Tang Z, Li C, Kang B, Gao G, Li C, Zhang Z. GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses. *Nucleic Acids Res*. 2017;45(W1):W98-W102. doi:10.1093/nar/gkx247
7. Uhlen M, Zhang C, Lee S, et al. A pathology atlas of the human cancer transcriptome. *Science*. 2017;357(6352). doi:10.1126/science.aan2507
8. Wang L, Wang B, Quan Z. Identification of aberrantly methylated differentially expressed genes and gene ontology in prostate cancer. *Mol Med Rep*. 2020;21(2):744-758. doi:10.3892/mmr.2019.10876
9. Nagy A, Lanczky A, Menyhart O, Gyorffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci Rep*. 2018;8(1):9227. doi:10.1038/s41598-018-27521-y
10. Dweep H, Sticht C, Pandey P, Gretz N. miRWalk—database: prediction of possible miRNA binding sites by “walking” the genes of three genomes. *J Biomed Inform*. 2011;44(5):839-847. doi:10.1016/j.jbi.2011.05.002
11. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. *Elife*. 2015;4. doi:10.7554/eLife.05005
12. Wong NW, Chen Y, Chen S, Wang X. OncomiR: an online resource for exploring pan-cancer microRNA dysregulation. *Bioinformatics*. 2018;34(4):713-715. doi:10.1093/bioinformatics/btx627
13. Shen J, Yu S, Sun X, Yin M, Fei J, Zhou J. Identification of key biomarkers associated with development and prognosis in patients with ovarian carcinoma: evidence from bioinformatic analysis. *J Ovarian Res*. 2019;12(1):110. doi:10.1186/s13048-019-0578 -1
14. Chen X, Li D, Wang N, et al. Bioinformatic analysis suggests that UGT2B15 activates the HippoYAP signaling pathway leading to the pathogenesis of gastric cancer. *Oncol Rep*. 2018;40(4):1855-1862. doi:10.3892/or.2018.6604
15. Chen YZ, Xue JY, Chen CM, et al. PPAR signaling pathway may be an important predictor of breast cancer response to neoadjuvant chemotherapy. *Cancer Chemother Pharmacol*. 2012;70(5):637-644. doi:10.1007/s00280-012-1949-0
16. Sultan G, Zubair S, Tayubi IA, Dahms HU, Madar IH. Towards the early detection of ductal carcinoma (a common type of breast cancer) using biomarkers linked to the PPAR(gamma) signaling pathway. *Bioinformation*. 2019;15(11):799-805. doi:10.6026/97320630015799
17. Skorski T. Oncogenic tyrosine kinases and the DNA-damage response. *Nat Rev Cancer*. 2002;2(5):351-360. doi:10.1038/nrc799
18. Mulligan LM. RET revisited: expanding the oncogenic portfolio. *Nat Rev Cancer*. 2014;14(3):173-186. doi:10.1038/nrc3680
19. Rodgers KM, Udesky JO, Rudel RA, Brody JG. Environmental chemicals and breast cancer: an updated review of epidemiological literature informed by biological mechanisms. *Environ Res*. 2018;160:152-182. doi:10.1016/j.envres.2017.08.045
20. Wang Z, Wang N, Liu P, Xie X. AMPK and cancer. *Exp Suppl*. 2016;107:203-226. doi:10.1007/978-3-319-43589-3_9
21. Leung T, Rajendran R, Singh S, Garva R, Krstic-Demonacos M, Demonacos C. Cytochrome P450 2E1 (CYP2E1) regulates the response to oxidative stress and migration of breast cancer cells. *Breast Cancer Res*. 2013;15(6):R107. doi:10.1186/bcr3574
22. Wang T, Fahrman JF, Lee H, et al. JAK/STAT3-regulated fatty acid beta-oxidation is critical for breast cancer stem cell self-renewal and chemoresistance. *Cell Metab*. 2018;27(1):136-150 e5. doi:10.1016/j.cmet.2017.11.001
23. Gao Q, Zhu H, Dong L, et al. Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*. 2019;179(2):561-577 e22. doi:10.1016/j.cell.2019.08.052
24. Guo E, Wei H, Liao X, Xu Y, Li S, Zeng X. Prognostic value of alcohol dehydrogenase mRNA expression in gastric cancer. *Oncol Lett*. 2018;15(4):5505-5516. doi:10.3892/ol.2018.8007
25. Liao X, Huang R, Liu X, et al. Distinct prognostic values of alcohol dehydrogenase mRNA expression in pancreatic adenocarcinoma. *Onco Targets Ther*. 2017;10:3719-3732. doi:10.2147/OTT.S140221
26. Malvia S, Bagadi SAR, Pradhan D, et al. Study of gene expression profiles of breast cancers in Indian women. *Sci Rep*. 2019;9(1):10018. doi:10.1038/s41598-019-46261-1
27. Echevarria MI, Awasthi S, Cheng CH, et al. African American specific gene panel predictive of poor prostate cancer outcome. *J Urol*. 2019;202(2):247-255. doi:10.1097/JU.000000000000193
28. Soiland H, Janssen EA, Korner H, et al. Apolipoprotein D predicts adverse outcome in women >or=70 years with operable breast cancer. *Breast Cancer Res Treat*. 2009;113(3):519-528. doi:10.1007/s10549-008-9955-y
29. Li J, Liu C, Chen Y, et al. Tumor characterization in breast cancer identifies immune-relevant gene signatures associated with prognosis. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.01119
30. Ling B, Liao X, Huang Y, et al. Identification of prognostic markers of lung cancer through bioinformatics analysis and in vitro experiments. *Int J Oncol*. 2020;56(1):193-205. doi:10.3892/ijo.2019.4926
31. Jiao C, Cui L, Ma A, Li N, Si H. Elevated serum levels of retinol-binding protein 4 are associated with breast cancer risk: a case-control study. *PLoS One*. 2016;11(12). doi:10.1371/journal.pone.0167498
32. Papiernik D, Urbaniak A, Klopotoska D, et al. Retinol-binding protein 4 accelerates metastatic spread and increases impairment of blood flow in mouse mammary gland tumors. *Cancers (Basel)*. 2020;12(3). doi:10.3390/cancers12030623
33. Tsutsui S, Yasuda K, Suzuki K, et al. A loss of c-kit expression is associated with an advanced stage and poor prognosis in breast

- cancer. *Br J Cancer*. 2006;94(12):1874-1878. doi:10.1038/sj.bjc.6603183
34. Chow LQ, Eckhardt SG. Sunitinib: from rational design to clinical efficacy. *J Clin Oncol*. 2007;25(7):884-896. doi:10.1200/JCO.2006.06.3602
 35. Bernemann C, Hulsewig C, Ruckert C, et al. Influence of secreted frizzled receptor protein 1 (*SFRP1*) on neoadjuvant chemotherapy in triple negative breast cancer does not rely on WNT signaling. *Mol Cancer*. 2014;13(1):1-2. doi:10.1186/1476-4598-13-174
 36. Naorem LD, Muthaiyan M, Venkatesan A. Identification of dysregulated miRNAs in triple negative breast cancer: a meta-analysis approach. *J Cell Physiol*. 2019;234(7):11768-11779. doi:10.1002/jcp.27839
 37. Li Z, Peng Z, Gu S, et al. Global analysis of miRNA-mRNA interaction network in breast cancer with brain metastasis. *Anticancer Res*. 2017;37(8):4455-4468. doi:10.21873/anticancer.11841
 38. Kolacinska A, Morawiec J, Pawlowska Z, et al. Association of microRNA-93, 190, 200b and receptor status in core biopsies from stage III breast cancer patients. *DNA Cell Biol*. 2014;33(9):624-629. doi:10.1089/dna.2014.2419
 39. Jandrig B, Seitz S, Hinzmann B, et al. ST18 is a breast cancer tumor suppressor gene at human chromosome 8q11.2. *Oncogene*. 2004;23(57):9295-9302. doi:10.1038/sj.onc.1208131
 40. Uhr K, Prager-van der Smissen WJC, Heine AAJ, et al. MicroRNAs as possible indicators of drug sensitivity in breast cancer cell lines. *PLoS One*. 2019;14(5). doi:10.1371/journal.pone.0216400
 41. Sun L, Wang H, Jiang J, Bi X. [MiR-153-3p inhibits proliferation and migration of breast cancer cells via down-regulating ROCK1]. *Xi Bao Yu Fen Zi Mian Yi Xue Za Zhi*. 2020;36(2):138-144.
 42. Xiong DD, Dang YW, Lin P, et al. A circRNA-miRNA-mRNA network identification for exploring underlying pathogenesis and therapy strategy of hepatocellular carcinoma. *J Transl Med*. 2018;16(1):220. doi:10.1186/s12967-018-1593-5
 43. Wu XM, Yang HG, Zheng BA, Cao HF, Hu ZM, Wu WD. Functional genetic variations at the microRNA binding-site in the CD44 gene are associated with risk of colorectal cancer in Chinese populations. *PLoS One*. 2015;10(5). doi:10.1371/journal.pone.0127557
 44. Cheng B, He Q, Cheng Y, et al. A three-gene classifier associated with MicroRNA-mediated regulation predicts prostate cancer recurrence after radical prostatectomy. *Front Genet*. 2019;10. doi:10.3389/fgene.2019.01402
 45. Huang F, Zhang J, Yang D, et al. MicroRNA expression profile of whole blood is altered in adenovirus-infected pneumonia children. *Mediators Inflamm*. 2018;2018. doi:10.1155/2018/2320640
 46. Shangguan WJ, Liu HT, Que ZJ, Qian FF, Liu LS, Tian JH. TOB1-AS1 suppresses non-small cell lung cancer cell migration and invasion through a ceRNA network. *Exp Ther Med*. 2019;18(6):4249-4258. doi:10.3892/etm.2019.8103
 47. Wang XW, Wu Y, Wang D, Qin ZF. MicroRNA network analysis identifies key microRNAs and genes associated with precancerous lesions of gastric cancer. *Genet Mol Res*. 2014;13(4):8695-8703. doi:10.4238/2014.October.27.10
 48. Song B, Du J, Song DF, Ren JC, Feng Y. Dysregulation of NCAPG, KNL1, miR-148a-3p, miR-193b-3p, and miR-1179 may contribute to the progression of gastric cancer. *Biol Res*. 2018;51(1):44. doi:10.1186/s40659-018-0192-5
 49. Fort RS, Matho C, Oliveira-Rizzo C, Garat B, Sotelo-Silveira JR, Duhagon MA. An integrated view of the role of miR-130b/301b miRNA cluster in prostate cancer. *Exp Hematol Oncol*. 2018;7(1):10. doi:10.1186/s40164-018-0102-0
 50. Jiang Y, Gao H, Liu M, Mao Q. Sorting and biological characteristics analysis for side population cells in human primary hepatocellular carcinoma. *Am J Cancer Res*. 2016;6(9):1890-1905.
 51. Huang R, Liu Z, Tian T, et al. The construction and analysis of tumor-infiltrating immune cells and ceRNA networks in metastatic adrenal cortical carcinoma. *Biosci Rep*. 2020;40(3). doi:10.1042/BSR20200049
 52. Luo Y, Zhang C, Tang F, et al. Bioinformatics identification of potentially involved microRNAs in Tibetan with gastric cancer based on microRNA profiling. *Cancer Cell Int*. 2015;15(1):115. doi:10.1186/s12935-015-0266-1
 53. Mikkelsen LH, Andersen MK, Andreassen S, et al. Global microRNA profiling of metastatic conjunctival melanoma. *Melanoma Res*. 2019;29(5):465-473. doi:10.1097/CMR.0000000000000606
 54. Lin GB, Zhang CM, Chen XY, et al. Identification of circulating miRNAs as novel prognostic biomarkers for bladder cancer. *Math Biosci Eng*. 2019;17(1):834-844. doi:10.3934/mbe.2020044
 55. Ulivi P, Canale M, Passardi A, et al. Circulating plasma levels of miR-20b, miR-29b and miR-155 as predictors of bevacizumab efficacy in patients with metastatic colorectal cancer. *Int J Mol Sci*. 2018;19(1). doi:10.3390/ijms19010307