Screening a Prognosis-related Target Gene in Patients with HER-2 Positive Breast Cancer

by Bioinformatics Analysis

Song Wang, Jian Deng, Xudong Gao, Hongying Lv, Yi Quan

Department of Breast Surgery, The Affiliated Hospital of Southwest Medical University,
Luzhou 646000, China.

Address all correspondence to:

Yi Quan
Department of Breast Surgery,
The Affiliated Hospital of Southwest Medical University,
Luzhou 646000, China.

Email: 415974602@qq.com

Running Title: Screening a Target Gene by Bioinformatics Analysis

Keywords: HER-2 positive breast cancer • Differentially-expressed genes • Bioinformatics analysis • Target genes • RRM2
Highlights of the Study

- Based on bioinformatics principles, this study identified RRM2 as a new biomarker and a therapeutic target for HER-2 positive breast cancer.

- RRM2 is associated with cell cycle regulation in the occurrence and development of HER-2 positive breast cancer.
Abstract

Objective: The objective of the present study was to determine a target gene and explore the molecular mechanisms involved in the pathogenesis of HER-2 positive breast cancer. Methods: Three RNA expression profiles obtained from the Gene Expression Omnibus (GEO) and the Cancer Genome Atlas (TCGA) were used to identify differentially-expressed genes (DEGs) using the R software. A Protein-Protein Interaction (PPI) network was then constructed, and hub genes were determined. Subsequently, the relationship between clinical parameters and hub genes was examined to screen for target genes. Next, DNA methylation and genomic alterations of the target gene were evaluated. To further explore potential molecular mechanisms, a functional enrichment analysis of genes co-expressed with the target gene was performed. Results: The differential expression analysis revealed 217 DEGs in HER-2 positive breast cancer samples compared to normal breast tissues. RRM2 was the only hub gene closely associated with lymphatic metastasis and the patients’ prognosis. Additionally, RRM2 was found to be consistently amplified and negatively associated with the level of methylation. Functional enrichment analysis showed that the co-expressed genes were mainly involved in cell cycle regulation. Conclusions: RRM2 was identified as a target gene associated with the initiation, progression and prognosis of HER-2 positive breast cancer, which may be considered as a new biomarker and therapeutic target.
**Introduction**

HER-2-positive breast cancer accounts for about 20% of all the breast cancer cases in China, and is closely associated with a highly aggressive clinical course, due to amplification of the HER-2 or overexpression of the protein. Trastuzumab plus chemotherapy has become the first-line systemic therapeutic regimen, which has been shown to significantly prolong the overall survival (OS) in breast cancer patients [1]. However, due to the development of primary or secondary resistance to targeted drugs, some patients still experience recurrence of the disease or metastasis after targeted therapy. Therefore, it is imperative to explore the underlying molecular mechanisms and identify new target genes involved in the pathogenesis of HER-2-positive breast cancer.

In recent years, high throughput sequencing technologies have provided oncologists with multiple levels of tumor data, including genomic, transcriptomic, proteomic and epigenetic data. Similarly, bioinformatic analysis has enabled us to investigate tumor progression and intrinsic mechanisms of carcinogenesis at the genome level for many types of cancer. The series of analyses performed on differential genes could facilitate the identification of biomarkers and drug targets for the diagnosis and treatment of cancers, allowing for the rapid evolution of precision oncology.

In this study, we carried out a comprehensive analysis of multiple authoritative databases using bioinformatics methods. Finally, RRM2 was identified as a target gene closely correlated with lymphatic metastasis and the prognosis of patients with HER-2-positive breast cancer. A list of the databases used in this study was displayed in supplementary Table 1. The study’s flowchart was shown in supplementary Fig. 1.
Materials and Methods

Data Sources and Preprocessing

Two independent microarray datasets (GSE29431 and GSE65194) based on the GPL570 platform of Affymetrix Human Genome U133 Plus 2.0 Array [HG-U133_Plus_2] were obtained from the GEO database. Data on breast cancer mRNA expression and corresponding clinical information were downloaded from the TCGA data portal. Respective characteristics of the 3 selected datasets were summarized in supplementary Table 2. Basic clinicopathological information of HER-2-positive breast cancer was presented in supplementary Table 3.

The probe ID was mapped to a gene symbol and the probe level expression values were averaged across the probes for genes with multiple probes. The array data was normalized and calibrated using the RMA algorithm.

Identification of DEGs

For GSE29431 and GSE65194, the LIMMA package was used for the detection of DEGs between HER-2-positive breast cancer and normal samples, while for mRNA expression data in TCGA, DEGs were identified using the DESeq2 package. The criteria for determining DEGs included a fold change of ≥ 2 and an adj p value (Benjamin-corrected p value) of < 0.01. The false discovery rate was controlled by correcting the p value with respect to the Benjamini–Hochberg procedure. The volcano plot was produced using the R software to display the results of differential gene expression analysis in three datasets. Finally, the common DEGs were identified by the Draw Venn Diagram tool.

PPI Network Construction and Modules Analysis

PPI analysis was carried out using the Search Tool for the Retrieval of Interacting Genes
(STRING) database with a combined score of $\geq 0.7$ as the threshold. Subsequently, the PPI network was visualized using the Cytoscape software. We used the Molecular Complex Detection (MCODE) and CytoHubba, two of Cytoscape plugin software, to perform module analysis of the PPI network and screen the hub genes, respectively. Moreover, degree cutoff=2, node score cutoff=0.2, k-core=2, and max.depth=100 were set as the cut-off criterion in modules analysis. The highest scores of the degree algorithm were retained as the criteria to screen out hub genes with high connectivity in the most significant module.

**Survival and Clinicopathological Parameters Analysis of Hub Genes**

We used the GEPIA database to perform an overall survival analysis of the previously identified hub genes. Using the clinical data (age, pathologic T, pathologic N and pathologic stage) of the 123 patients with HER-2 positive breast cancer, we further explored the correlation between hub genes and clinical parameters, and identified a target gene most closely related to the development of HER-2 positive breast cancer.

**Target Gene Validation Based on Multiple Databases**

In the present study, using the Oncomine, TIMER, UALCAN and Kaplan-Meier plotter database, we validated the overexpression and prognostic value of the target gene in HER-2 positive breast cancer.

**Target Gene Methylation Analysis**

We used UALCAN and cBioportal to assess the correlation between target gene mRNA expression and its methylation values. Moreover, MethSurv, a freely valuable platform, was used to assess the effects of target gene methylation on patient survival time.

**Genomic Alterations and Mutations Analysis of the Target Gene**
The cBioPortal was used to analyse the genomic alterations status of the target gene in breast cancer. Four breast cancer studies in the cBioPortal were utilized for this analysis. The four studies included Breast Invasive Carcinoma (TCGA, Cell 2015), Breast Invasive Carcinoma (TCGA, Firehose Legacy), Breast Invasive Carcinoma (TCGA, Nature 2012) and Breast Invasive Carcinoma (TCGA, PanCancer Atlas). The COSMIC database was used to examine the different types of mutations affecting the target gene in breast cancer.

*Functional Enrichment Analysis of Co-expressed Genes Correlated with the Target Gene in Breast Cancer*

We used the cBioPortal database to screen genes coexpressed with the target gene in breast cancer, with Spearman scores >0.5 as unique criteria. We then conducted an analysis of the Gene Ontology (GO) annotation and the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway using the clusterProfiler package [2]. A Benjamin-corrected p-value < 0.05 indicated significant enrichment of GO terms and KEGG pathways, and the GO annotation comprised the biological process, molecular function and cellular component.

**Results**

*Identification of DEGs*

The microarray datasets (GSE29431 and GSE65194) were standardized, and the results were illustrated in a box plot (Supplementary Fig.2). With a cutoff of adj $p$ value < 0.01 and fold change $\geq 2$, we eventually identified 503 DEGs (96 upregulated and 407 downregulated), 1,525 DEGs (809 upregulated and 716 downregulated) and 2,035 DEGs (1,000 upregulated and 1,035 downregulated) in GSE29431, GSE65194, and TCGA respectively, which were differentially expressed between HER-2 positive breast cancer samples and normal breast
tissues. The volcano plot showed the differentially expressed genes (Fig. 1A-C). The VENN plot delineated the common 217 DEGs in the three datasets, including 50 upregulated genes and 167 downregulated genes (Fig. 1D, E).

**PPI Network Construction and Module Analysis**

The 217 DEGs were input into the STRING database for PPI network analysis, which was then visualized by the Cytoscape software, and achieved a PPI network of 111 nodes and 310 edges, with combined scores $\geq 0.7$ (Fig. 2A). The most significant cluster 1 totally consisting of upregulated genes was identified using the MCODE plugin (Fig. 2B). Subsequently, with the highest connectivity degree=16, 13 genes selected from cluster 1 were considered as hub genes, including UBE2C, KIF20A, PTTG1, CENPF, CCNB1, BIRC5, TOP2A, CEP55, AURKA, RRM2, BUB1B, NUSAP1 and MELK. These genes were presented in supplementary Table 4.

**Survival and Clinicopathological Parameters Analysis of Hub Genes**

Using the GEPIA database, we assessed the prognostic value of hub genes in HER-2 positive breast cancer patients. The specific results of the survival analysis were shown in supplementary Table 5. Among them, only the upregulated NUSAP1 and RRM2 genes were associated with a poor prognosis based on the OS (Fig. 2C, D). Clinicopathological analysis of RRM2 and NUSAP1 using the $\chi^2$ test demonstrated that only RRM2 was linked with pathologic N ($p < 0.05$; Table 1). As shown in supplementary Table 6, there were no significant association of NUSAP1 expression with age ($p = 1.000$), pathologic T ($p = 0.618$), pathologic N (P=0.054) and pathologic stage ($p = 0.533$). These results suggested that RRM2 may be involved in the occurrence, development and prognosis of HER-2 positive breast cancer.
Validation of the Correlation Between RRM2 mRNA Expression and Prognosis

Using the Oncomine and TIMER databases, we found that RRM2 was overexpressed in multiple cancer types (Supplementary Fig. 3A, B), suggesting that it might be relevant to the oncogenesis of these tumors. From the Oncomine and UALCAN databases, RRM2 overexpression was confirmed in HER-2 positive breast cancer (Supplementary Fig. 3C, D). By using the Kaplan-Meier plotter database, we further corroborated the prognostic significance of RRM2 (Supplementary Fig. 3E). Consequently, RRM2 was confirmed to be associated with HER-2 positive breast cancer cinogenesis.

RRM2 Methylation Analysis

From the UALCAN and cBioportal databases, we observed that the RRM2 methylation levels in HER-2 positive breast cancer samples were apparently lower than those in normal breast tissues ($p < 0.05$), and showed a negative association with gene expression ($p < 0.05$, $r=-0.30$) (Fig. 3A, B). Using the MethSurv web tool, we found that hypermethylation was significantly correlated with a more favorable prognosis (Fig. 3C). We thus deduced that DNA hypomethylation was correlated with high gene expression levels, which may play an important role in the occurrence and development of HER-2 positive breast cancer.

Analysis of RRM2 Genomic Alterations and Mutations

Using 4 studies in the cBioPortal, we explored the genetic alterations of RRM2 in breast cancer patients. The results demonstrated that RRM2 was altered in 52 (1.4%) out of 3798 patients (Supplementary Table 7). As depicted in supplementary Fig. 4A and 4B, these alterations including amplification, deep deletion and mutation, were associated with a poor prognosis and a lower OS in breast cancer patients. Furthermore, the pie chart showed the
corresponding types of mutation of RRM2 in breast cancer patients from the COSMIC database, with missense and synonymous mutations being the most common types, and the mutations of the coding strand in RRM2 were as follows: G>A, A>T, G>T, A>G and C>T (Supplementary Fig. 4C, D).

Functional Enrichment Analysis of RRM2 Co-expressed Genes

According to the gene co-expression modules obtained from the cBioPortal database, a total of 292 genes with Spearman's Correlation scores >0.5 were identified as being co-expressed with RRM2. The top 10 co-expressed genes were listed in supplementary Table 8. For the GO analysis, the results pointed out that organelle fission was the principally enriched in biological processes (BP), the chromosomal region was the primarily enriched in cellular component (CC), whereas ATPase activity was the chiefly enriched in molecular functions (MF). Besides, KEGG pathway analysis indicated that these co-expressed genes were mainly enriched in the cell cycle. The results of the functional enrichment analysis were revealed in supplementary Fig. 5. As shown in Fig. 4, the corresponding genes were coloured red in the pathway mainly located in the S or G2/M phases of the cell cycle.

Discussion

HER-2 subtype breast cancer often presents with tumor recurrence and metastasis, and the patients' overall survival rate is lower. In recent years, targeted drug therapies have improved the overall survival for patients with HER-2 subtype breast cancer [1]. Despite the advent of this standardized initial therapy, some patients rapidly develop disease progression, which is predominantly due to the development of primary and secondary resistance to targeted drugs. Thus, further studies are necessary to explore the pathogenesis of breast cancer and search for
novel therapeutic targets.

In this study, we identified 217 DEGs in three datasets. With a highest connectivity degree of 16, 13 genes were considered as hub genes, including UBE2C, KIF20A, AURKA, BIRC5, PTTG1, CENPF, CCNB1, TOP2A, CEP55, BUB1B, NUSAP1, MELK and RRM2. Clinicopathological analysis suggested that only RRM2 was positively correlated with a poor prognosis and associated with lymph node metastasis in HER-2 positive breast cancer. Additionally, using the Oncomine and TIMER databases, we found that RRM2 was over-expressed in multiple tumors (Supplementary Fig. 3). Thus, we deduced that RRM2 may play the role of oncogene in diverse carcinomas, and is linked to the occurrence, development, diagnosis, treatment, and prognosis of HER-2-positive breast cancer.

High UBE2C expression contributes to estrogen-independent growth in HR+ breast cancer cells, which suggests that tamoxifen treatment combined with inhibitors targeting UBE2C could be a potential therapeutic strategy to overcome resistance in endocrine therapy [3]. KIF20A small interfering RNA (siRNA) has been reported to significantly inhibit the proliferation of pancreatic cancer cell lines [4]. Alisertib, a selective small-molecule AURKA inhibitor, was found to be able to promote preferential growth inhibition in AI-resistant cell lines [5]. Overexpression of BIRC5 has been shown to promote breast tumor proliferation by inducing genetic instability [6]. PTTG1 is correlated with the regulation of cancer stem cells in breast cancer [7]. Sun et al demonstrated that CENPF promoted breast cancer bone metastasis by activating the PI3K–AKT–mTOR signaling pathway [8]. CCNB1 silencing inhibits cell proliferation and promotes cell senescence via activation of the p53 signaling pathway in pancreatic cancer [9].
The PTEN/AKT signaling pathway affects the expression of TUBB3 and TOP2A, reducing cell growth and inducing apoptosis of human breast cancer MCF-7 cells through ATP and caspase-3 signaling pathways [10]. CEP55 have been shown to promote hepatocellular carcinoma cell migration and invasion by regulating the JAK2-STAT3-MMPs signaling pathway [11]. The oncogenic effect of BUB1B has been proven to be impaired by rapamycin via downregulation of the mTORC1 signaling pathway [12]. NUSAP1 is a potential prognostic marker for breast cancer patients that promotes the progression of breast cancer through the Wnt/β-catenin and epithelial-mesenchymal transition pathways [13]. Previous studies have demonstrated that MELK inhibition promotes cancer cell death through the p53 signaling pathway [14].

RRM2, a rate-limiting enzyme for DNA synthesis and repair, is involved in cell cycle regulation and over-expressed in various tumors, which is associated with cell proliferation, invasiveness and migration. This is consistent with the findings observed in the present study. Previous studies have shown that small molecular antagonists of the RRM2 gene significantly reduce tamoxifen-resistant cell proliferation and decreased tumor growth [15]. Moreover, suppression of RRM2 synthesis could enhance the tumor’s chemosensitivity to adriamycin, which has led to the recognition of RRM2 as a suitable biomarker [16]. However, the studies focused only on estrogen-negative breast cancer, and the cancer-forming pathways underlying this process are yet to be discussed. In this study, we aimed to find an effective target for the treatment of HER-2 positive breast cancer, comprising estrogen-positive breast cancer (luminal B subtype (positive HER-2 expression)). In this study, we elucidated the molecular mechanisms and signal pathways, and also performed further analysis of RRM2 genomic
alterations and mutations. We inferred that methylation of the RRM2 gene may also serve as a potential drug target.

In recent years, some current therapeutic regimens or drugs use RRM2 as the therapeutic target. GTI2040, as an RRM2 expression silencer, is complementary to a region of the coding sequence in the RRM2 mRNA and shows strong antitumor activity. Additionally, GTI2040 decreases RRM2 mRNA expression in human tumor cell lines and arrests cells in the early G1/S phase. GTI2040 in combination with weekly doses of gemcitabine has shown good treatment outcomes for patients with advanced solid tumors in a phase I study [17]. Another RRM2 inhibitor, GW8510, suppresses the survival of lung squamous cell carcinoma cells and reverses their resistance to gemcitabine by inducing DNA damage and cell apoptosis [18].

Subsequently, we found a significantly low level of RRM2 methylation and identified trends for the negative correlation between its methylation and expression levels, thus confirming that DNA hypomethylation was primarily responsible for the increased RRM2 expression in HER-2 positive breast cancer. Survival analysis revealed that increasing levels of methylation were related to poor prognosis, which indicates that level of methylation of RRM2 can be used as a prognostic predictor with breast carcinoma.

Based on these observations, we speculated that most RRM2-related genes are linked to the cell cycle, and promoted tumorigenesis and progression by regulating the G2/M phase transition in breast cancer. Previous studies have shown that multiple drugs can alter the proliferative activity of tumor cells by regulating the G2/M phase transition in the cell cycle. For example, CAPE could inhibit the cell cycle and eliminate breast cancer cells in the G2/M phase [19]. Moreover, cucurbitacin B has been shown to inhibit the proliferation of breast
cancer cells by inducing cell cycle arrest at the G2/M phase, leading to apoptosis [20]. Therefore, RRM2 inhibition by regulating the G2/M phase in HER-2 positive breast cancer could possibly serve as a novel therapeutic strategy.

**Conclusion**

In summary, this study demonstrated that RRM2 was overexpressed in HER-2 positive breast cancer and was associated with lymph node metastasis and the prognosis of these patients, indicating that RRM2 expression is closely related to tumorigenesis and cancer progression. Based on functional enrichment analysis, we speculate that RRM2 promotes tumorigenesis principally via cell cycle regulation. These findings suggest that RRM2 could be used as a potential therapeutic target against HER-2 positive breast cancer in the future. However, animal experiments and population-based studies are necessary to confirm the results obtained in this study.
References


7 Yoon CH, Kim MJ, Lee H, Kim RK, Lim EJ, Yoo KC, et al. PTTG1 oncogene promotes


Figures Legends

Fig. 1. The volcano plot and Venn diagram of DEGs in the three datasets. (A) Shows the volcano plot of DEGs for GSE29431; (B) Delineates the volcano plot of DEGs for GSE65194. (C) Outlines the volcano plot of DEGs for TCGA. (D) Venn diagram of common upregulated genes; (E) Venn diagram of common downregulated genes. In the volcano plot, each dot represents an individual gene (Red dots denote upregulated genes with logFC>2 and adj p value < 0.01, while blue dots represent downregulated genes with logFC<2 and adj p value < 0.01). FC, fold change; adj p-value, adjusted p-value; DEGs, differentially expressed genes.

Fig. 2. PPI network construction and Kaplan-Meier survival analysis. (A) PPI network was constructed from common DEGs; (B) The most significant cluster 1 was identified from the PPI network; The Red circle presented upregulated genes, while the black square presented downregulated genes. (C) Overall survival curves for NUSAP1; (D) Overall survival curves for RRM2.

Fig. 3. The associations between RRM2 DNA methylation and gene expression and survival. (A) RRM2 methylation levels analysis using the UALCAN database. (B) The relationships between RRM2 DNA methylation and gene expression using the cBioportal database. (C) Survival effect of RRM2 methylation on breast cancer patients using MethSurv.

Fig. 4. The position of RRM2-associated genes in the cell cycle. The corresponding genes in the pathway are marked with red circles.
Fig. 1
Fig. 2
Fig. 3
Fig. 4.