Molecular mechanism of activated T cells in breast cancer

Introduction: This study aimed to explore the effect of activated T cells on breast cancer (BC) cells and provide a theoretical basis for the interaction mechanism studies between BC and immune cells.

Methods: The microarray dataset GSE73527 was downloaded from the Gene Expression Omnibus database. The common differentially expressed mRNAs (co-DEMs) and the common differentially expressed long non-coding RNAs (co-DElncRNAs) were identified between MDA-MB-231 cells and MCF7 activated human T cells, respectively. The RNA–miRNA–lncRNA (ceRNA) network was constructed. Furthermore, the Kyoto encyclopedia of genes and genomes pathway and the gene ontology function analyses were performed on co-DEMs. The protein–protein interaction networks and modules were investigated.

Results: A total of 639 co-DEMs (such as interleukin-6 [IL6] and signal transducer and activator of transcription 1 [STAT1]) were detected in this study. Defense response to other organisms and herpes simplex infection were the most outstanding function and pathway assembled with co-DEMs, respectively. One protein–protein interaction network and three modules were further constructed. A total of 88 RNA–miRNA–lncRNA relationships such as BTN3A1–has-mir-20-b-5p–HCP5 were explored in the ceRNA network.

Conclusion: Activated T cells may play a crucial role in the defense response to other organisms and herpes simplex infection pathways by upregulating IL6 and STAT1, which further affected the progression of BC. The BTN3A1–has-miR-20b-5p–HCP5 relationship may be the potential interaction mechanism between BC and immune cells.

Keywords: breast cancer, differentially expressed mRNA, function and pathway analysis, protein–protein interaction network, mRNA–miRNAs–lncRNA regulation network

Introduction

Breast cancer (BC) is the most frequently diagnosed cancer in women worldwide, accounting for 25% of all cancer cases1 with 1.68 million new cases and 522,000 deaths in a year.2 BC is typically treated by surgery, which may be followed by chemotherapy, radiation therapy, or other immune-modulating treatments.3 Although increasingly aggressive treatments are employed in clinical applications, the survival rates of BC in developing countries remain poor.1,4

Malignant human tumors are often accompanied by infiltration of immune cells into the region of tumor cell proliferation.3 Activated CD8+ T cells have been shown to play a critical role in the killing of tumor cells, and elevated T-cell activation scores are positively correlated with longer patient survival in BC.6 Additionally, a previous study demonstrated a relationship between tumor-infiltrating lymphocytes (TILs) and BC.7 TILs, which include T cells and B cells, are part of the larger category of “tumor-infiltrating immune cells” that are present in variable proportions.8 A previous study showed that TILs are good...
prognostic indicators in highly proliferative breast tumors.9 The results of large cohorts indicated an association between the TILs in early stage of BC and high response rates to neoadjuvant chemotherapy.10 TILs are involved in the process of BC via gene expression or pathway regulation.11 Numerous studies have revealed the molecular pathways such as PI3K-Akt-mTOR governing the migration of T lymphocytes into tumor tissue.12,13 In a previous study, Noonepalle et al compared the immunophenotype by mRNA sequencing between MDA-MB-231 cells and MCF7-activated human T cells.14 Their results showed that indoleamine 2,3-dioxygenase 1 (IDO1) methylation regulated anti-immune responses in BC subtypes and can be used as a predictive biomarker for immunotherapy. However, genetic evaluation of TILs in a BC specimen is currently not routine because the genetic relevance remains unclear.

Based on the microarray data provided by Noonepalle et al,14 we performed bioinformatics analysis to identify the common differentially expressed mRNAs (co-DEMs) and common differentially expressed long non-coding RNA (co-DElncRNA) between MDA-MB-231 cells and MCF7-activated human T cells, respectively. Functional and pathway enrichment analysis was performed based on these co-DEMs. Next, the protein–protein interaction (PPI) network and modules were investigated. Furthermore, lncRNAs are reported to act as ceRNAs that regulate the expression of miRNA-targeted genes by sponging miRNAs.15 lncRNAs have also been identified as promising prognostic and metastatic markers in BC.16,17 Thus, an mRNA–miRNA–lncRNA interaction network was constructed. We explored the effect of activated T cells on the expression of BC cell lines and provide a theoretical basis for studying the interaction mechanism between BC and immune cells.

Materials and methods

Data resource and preprocessing

Gene expression profile data in the dataset GSE7352714 were downloaded from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/). This dataset was produced on the GPL11154 Illumina HiSeq 2000 platform (Homo sapiens). A total of 12 cell samples including 3 MCF7 cell samples (MCF7-control samples), 3 MCF7 samples co-cultured with normal peripheral blood mononuclear cells activated with anti-CD3/CD28 antibodies (MCF-case samples), 3 MDA-MB-231 cell samples (MDA-MB-231-control samples), and 3 MDA-MB-231 samples co-cultured with normal peripheral blood mononuclear cells activated with anti-CD3/CD28 antibodies (MDA-MB-231-case samples) were examined in this profile.

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables known as principal components (or sometimes, principal modes of variation).18 PCA is widely used in data processing and dimensionality reduction.19 After downloading the standardized data of GSE73527, pattern recognition of data and samples distribution observation were performed by PCA. Next, combined with the protein coding gene annotation information (V25) provided by the Gencode database (https://www.gencodegenes.org/),20 the mRNA expression matrix and the lncRNA expression matrix for subsequent analysis were extracted, and our analysis was mainly based on mRNA expression.

Investigation of DEMs

According to different cell types and processing methods, the samples were grouped and analyzed as follows: MCF7-case vs MCF7-control (MCF7 group) and MDA-MB-231-case vs MDA-MB-231-control (MDA-MB-231 group). The P-value of DEMs in each group was calculated using the empirical Bayes linear model in Linear Models for Microarray Data (limma, http://www.bioconductor.org/packages/release/bioc/html/limma.html) package21 of R (version: 3.32.5) software. The adjusted P-values (adjust. P-value) of DEMs were obtained by multiple checking and correction based on the Benjamini and Hochberg method.22 An adjust. P.value < 0.05 and a |log fold change (FC)| > 1 were selected as the thresholds for identifying DEMs.

The intersection results for all DEMs (co-DEMs) in both the MCF7 and MDA-MB-231 groups were visualized using a Venn diagram. The expression values of DEMs were visualized using a heat map with heatmap software23 (clustering distance: Pearson’s correlation, clustering methods: linkage clustering) in R package (version: 1.0.8, https://cran.r-project.org/web/packages/heatmap).

Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis

GO analysis is commonly used for functional studies of large-scale genomic or transcriptomic data.24 The KEGG pathway database25 contains information regarding how molecules or genes are networked. In this study, the clusterProfiler software26 (version: 3.4.4, https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html) in R package was used to identify over-represented GO categories in biological process and significant KEGG pathways. Next, the Benjamini
and Hochberg method corrected adjust $P$ value $<0.05$ was selected as the threshold for identification.

**PPI network and modules construction**

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, version: 10.0, http://www.string-db.org/) was used to predict PPIs in this study. Based on the information in the STRING database, the PPI network was constructed from up- and downregulated proteins, and then visualized using Cytoscape (version: 3.4.0) software. The network topological index (degree) is defined as the number of links incident upon a node. The degree was calculated using CytoNCA software (version 2.1.6, http://apps.cytoscape.org/apps/cytonca). The key node (hub protein) was determined by the degree score: a higher score indicated a more important location in the network.

A previous study demonstrated that a gene from the same module in a PPI network plays a regulatory role in biological function. The MCODE package (version 1.4.2, http://apps.cytoscape.org/apps/MCODE) plugin in Cytoscape software (degree cutoff=2, node score cutoff=0.2, K-core=2, Max. Depth=100) was used to analyze modules in this study. Additionally, KEGG pathway enrichment analysis was carried out on outstanding modules. An adjust $P$ value $<0.05$ was selected as the threshold for identification.

**Investigation of differentially expressed lncRNAs**

According to the method for screening DEMs between groups, the differentially expressed lncRNAs (DElncRNA) were also explored as the adjust $P$ value $<0.05$ and $|\log FC| >1$. Next, the intersection results for all DElncRNAs in both the MCF7 and MDA-MB-231 groups were visualized using a Venn diagram. The expression values of DElncRNAs were visualized as a heat map using pheatmap software (clustering distance: Pearson’s correlation, clustering methods: linkage clustering) in R package.

**miRNA associated lncRNA exploration**

The miRNA prediction of co-DEMs was performed using miRWalk 2.0 software (http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/). Based on the predicted and experimentally confirmed miRNA–gene regulatory relationship in the miRWalk database, the miRNA–co-DEM interactions were further explored.

The starBase v2.0 database provides the most comprehensive CLIP-Seq experimentally supported miRNA–lncRNA interaction networks. In this study, the union set of miRNA–lncRNA interactions in both the starBase (version: 2.0) database and InCeDB (http://gyanxet-beta.com/lncedb) database were used as the background for investigating miRNA regulatory factors of co-DElncRNAs.

**Results**

**Differentially expressed analyses**

PCA analysis was performed on the downloaded data (Figure 1). The results showed that samples from the same group were clustered together, particularly for the MCF7-control samples and the MDA-MB-231-control samples. After annotation, 18,984 mRNAs were obtained from the expression matrix. A total of 2,781 DEMs (1,571 upregulated and 1,210 downregulated) were obtained between MCF7-case samples and MCF7-control samples in the MCF7 group. A total of 846 upregulated and 762 downregulated DEMs were detected between MDA-MB-231-case samples and MDA-MB-231-control samples in the MDA-MB-231 group.
Furthermore, 438 upregulated co-DEMs (including IDO1) and 201 downregulated co-DEMs in the intersection of two groups were identified (Figure 2A). The clustering heat map for these co-DEMs is shown in Figure 2B, indicating that co-DEMs could distinguish samples from different groups.

GO function and KEGG pathway enrichment analysis

GO categories in biological process functions enriched by co-DEMs are shown in Figure 3A. The results showed that these co-DEMs were mainly related to functions such as the defense response to other organisms (GO: 0098542, P: 3.68E-25, count: 72, gene: interleukin 6 [IL6], STAT1, 2′-5′-oligoadenylate synthetase 1 [OAS1], interferon regulatory factor 7 [IRF7], 2′-5′-oligoadenylate synthetase 2 [OAS2], etc), response to virus (GO: 0009615, P: 1.88E-28, count: 54, gene: interferon gamma inducible protein 16 [IFI16], OAS1, interferon regulatory factor 7 [IRF7], beta-2-microglobulin [B2M], interleukin 10 receptor subunit beta [IL10RB], etc), response to interferon-gamma (GO: 0034341, P: 1.88E-28, count: 54, gene: interferon gamma inducible protein 16 [IFI16], OAS1, interferon regulatory factor 7 [IRF7], beta-2-microglobulin [B2M], interleukin 10 receptor subunit beta [IL10RB], etc), and NOD-like receptor signaling pathway (hsa04621, P: 6.30E-15, gene: MEFV, pyrin innate immunity regulator [MEFV], guanylate binding protein 7 [GBP7], IFI16, OAS1, IRF7, etc.), influenza A (hsa05164, P: 1.00E-18, gene: OAS, IRF7, janus kinase 2 [JAK2], caspase 1 [CASP1], suppressor of cytokine signaling 3 [SOCS3], etc), and NOD-like receptor signaling pathway (hsa04621, P: 6.30E-15, gene: MEFV, pyrin innate immunity regulator [MEFV], guanylate binding protein 7 [GBP7], IFI16, OAS1, IRF7, etc.), as shown in Figure 3B.

PPI network analysis

To obtain more useful information from the co-DEMs mentioned above, a PPI network was constructed based on the relationships among protein interactions. The results revealed 499 nodes (348 upregulated and 151 downregulated proteins) and 4,769 interactions in the current PPI network. Among these nodes, IL6, STAT1, IRF7, OAS1, and OAS2 were the top five nodes according to the degree score and were considered as hub proteins in the network (Figure 4).

Modules and related pathway analysis

Three modules with the highest scores were selected for further investigation. The results showed that there were 57 upregulated proteins and 2 downregulated proteins in module A (Figure 4A); these proteins were mainly enriched in pathways such as herpes simplex infection and influenza A (Figure 5). There were 11 upregulated proteins and 18 downregulated proteins in module B (Figure 4B); these proteins were mainly enriched in pathways such as cell
Mechanism of activated T cells in BC

cycle and DNA replication (Figure 5). Furthermore, a total of 13 upregulated proteins and 17 downregulated proteins were found in module C (Figure 4C); these proteins were mainly enriched in pathways such as pyrimidine metabolism and DNA replication.

LncRNA investigation

After annotation, 11,023 lncRNAs were obtained. Differential expression analysis revealed 134 DeLncRNAs (103 upregulated and 31 downregulated) between MCF7-case samples and MCF7-control samples in the MCF7 group. There were 69 DeLncRNAs (30 upregulated and 39 downregulated) between MDA-MB-231-case samples and MDA-MB-231-control samples in the MDA-MB-231 group. A total of nine upregulated co-DElncRNAs and three downregulated co-DElncRNAs between the two groups were further obtained from these DeLncRNAs.

CeRNA regulatory network analysis

A total of 88 mRNA–miRNA–lncRNA regulation relationships including BTN3A1-has-mir-20b-5p-HCP5 and BTN3A1-has-mir-17–5 p-HCP5 were explored in ceRNA regulatory network analysis (Figure 6). Among these relationships, there were 2 upregulated lncRNAs (RP11-799D4.4 and HLA Complex P5 [HCP5]), 39 upregulated mRNAs (such as Kringle Containing Transmembrane Protein 1 [KREMEN1] and CD47 Molecule [CD47]), and 64 miRNAs (such as hsa-mir-20b-5p and hsa-mir-17–5 p) in the ceRNA
network. The results of network topological analysis showed that the most outstanding miRNAs regulating mRNA–lncRNA relationships were hsa-mir-20b-5p, hsa-mir-17–5p, hsa-mir-93–5p, and hsa-mir-106b-5p.

**Discussion**

BC is the most common invasive cancer in women worldwide. Although TILs are related to BC progression, the detailed molecular mechanism of TILs in BC remains unclear.

**Figure 4** Top three modules extracted from protein–protein interaction network.

**Notes:** (A) Module A; (B) module B; (C) module C; orange circular node represents the upregulated protein; blue diamond node represents the downregulated protein.

**Figure 5** Results of KEGG pathway enrichment analyses for proteins in three modules.

**Notes:** A darker color indicates a higher P-value; larger nodes indicate that a greater number of proteins are enriched.

**Abbreviation:** KEGG, Kyoto encyclopedia of genes and genomes.
Consistent with the findings by Noonepalle et al., our bioinformatics analysis also showed that IDO1 was co-upregulated between MDA-MB-231 cells and MCF7 cells-activated human T cells. Moreover, key co-DEMs including IL6 and STAT1 were detected in this study. GO function and KEGG pathway analysis showed that the defense response to other organisms and herpes simplex infection were the most outstanding functions and pathway assembled with co-DEMs, respectively. Furthermore, 88 mRNA–miRNA–lncRNA relationships including BTN3A1-hsa-mir-20b-5p-HCP5 and BTN3A1-hsa-mir-17-5p-HCP5 were found in the ceRNA network.
A recent study found that infiltration of tissue by immune cells contributes to BC development.\(^{38}\) The extent of T-cell infiltration into invasive BC has been reported to range from 1% to 45% of the total cellular mass.\(^{39}\) In rapidly proliferating tumors, the presence of T lymphocytes at tumor sites is a good prognostic indicator compared to non-immunogenic tumors.\(^{39}\) However, the exact composition of T lymphocyte infiltration varies greatly and may profoundly affect disease progression and overall patient survival.\(^{40}\) Furthermore, certain leukocytes, such as cytotoxic T lymphocytes, play a vital function in constraining tumor development.\(^{41}\) This is supported by epidemiologic data showing an increased incidence of viral-associated cancers.\(^{42}\) In the present study, co-DEMs between T-cell-activated samples and control samples were mainly enriched in immune response functions such as the defense response to other organisms and pathways such as herpes simplex infection. This result indicates that activated T cells affect BC cell lines via the defense response to other organism function and herpes simplex infection pathway. Interestingly, IL6 and SAT1 were two DEMs that were both outstanding in the defense response to other organism function and herpes simplex infection pathway. IL6 is secreted by T cells to stimulate immune responses during infection or other tissue damage, leading to inflammation.\(^{43}\) Kim et al found that normal mammary epithelial cells from healthy women release IL-6, but the expression of IL-6 was abolished in ductal infiltrating carcinomas.\(^{44}\) However, Garcia-Tuñón et al suggested that high expression of IL-6 and its receptors in breast tumors is related to enhanced cell proliferation in BC.\(^{45}\) IL-6 appears to have a dual function in the biology of in vitro BC cells. STAT1 is a member of the STAT family and plays an important role in orchestrating immune reactions.\(^{46}\) The antitumor function of STAT1 through its capacity to control the immune system and promote tumor immune surveillance has been widely examined.\(^{47}\) A previous study shows that interferon-γ restores BC sensitivity to the drug fulvestrant by regulating STAT1 and signaling to caspase-dependent apoptosis.\(^{48}\) In an animal study, Siebler et al showed that STAT1-deficient and T-bet-deficient mice were protected from T-cell-dependent liver injury, suggesting a close relationship between STAT1 and T-cell activation in disease.\(^{49}\) Thus, based on our results, we speculate that the activated T cells may play an important role in the defense response to other organism function and herpes simplex infection pathway by upregulating IL6 and STAT1, which further affect BC progression.

The expression of lncRNAs has recently been reported in tumorigenesis and plays a pivotal role in regulating the malignant behavior of cancers.\(^{50}\) HCP5 is a human endogenous retrovirus, but has now become an integral part of the human genome.\(^{51}\) Liang et al suggested that HCP5 could induce follicular thyroid carcinoma progression via an miRNA sponge.\(^{52}\) Furthermore, Venkatadri et al showed that modulating miRNAs with mimics or inhibitors validated the key roles of miR-542-3 p in MCF-7 and miR-122-5 p in MDA-MB-231 BC cell death.\(^{53}\) As an miRNA, miR-20b may underlie the metastatic heterogeneity of BC.\(^{54}\) Moreover, some genes such as early growth response-1 play an important role in miR-20b regulation during BC progression.\(^{55}\) Butyrophilin subfamily three member A1 (BTN3A1), a member of the butyrophilin gene family, plays a role in T-cell activation, activated T-cell proliferation, and immune response adaptation.\(^{56}\) It is widely known that in the ceRNA network of BC, mRNAs can communicate with each other by competing for a limited pool of miRNAs.\(^{57}\) A previous study showed that the 3’ untranslated region of the pseudogene CYP4Z2P promotes tumor angiogenesis in BC by acting as a ceRNA.\(^{58}\) In the present study, has-miR-20b-5p-BTN3A1-HCP5 showed an outstanding miRNA–mRNA–lncRNA regulation relationship in ceRNA regulation network. Thus, the has-miR-20b-5p-BTN3A1-HCP5 relationship may be involved in the interaction mechanism between BC and immune cells.

There were some limitations to this study. First, the sample size was small. Second, there was a lack of experimental verification because human or mouse tumor samples with activated T cells were difficult to obtain. Verification experiments with a large sample size are needed to confirm our results.

**Conclusion**

Activated T cells may play an important role in the defense response to other organism function and herpes simplex infection pathway by upregulating IL6 and STAT1, which further affect the progression of BC. Furthermore, the has-miR-20b-5p-BTN3A1-HCP5 relationship may be involved in the interaction mechanism between BC and immune cells.

**Acknowledgment**

This study was supported by the National Natural Science Foundation of China (No 11502146) and Shanghai Key Laboratory of Biliary Tract Disease Research Foundation (No 17DZ2260200).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


