Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative ceRNA Network Analysis Reveals Coordinated Immune Evasion, Angiogenesis, and Transcriptional Reprogramming in Chemoresistant Triple- Negative Breast Cancer

Yongtao Li

lyt1822@163.com

Xinjiang Medical University Affiliated Tumor Hospital

Miaomiao Zhang

Xinjiang Medical University Affiliated Tumor Hospital

Hao Lei

Xinjiang Medical University Affiliated Tumor Hospital

Wenting Xu

Xinjiang Medical University Affiliated Tumor Hospital

Rui Ma

Xinjiang Medical University Affiliated Tumor Hospital

Research Article

Keywords: ceRNA network, triple-negative breast cancer, chemoresistance, angiogenesis, immune evasion, transcriptional regulation

Posted Date: September 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10682433/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Triple-negative breast cancer (TNBC) frequently develops resistance to TAC chemotherapy. This study performed whole-transcriptome RNA sequencing of 15 TNBC specimens and applied a triple-algorithm machine learning consensus to dissect the ceRNA regulatory networks underlying chemoresistance.

Methods

RNA-seq data from nCR (resistant), pCR (sensitive), and normal tissues were analyzed by differential expression, WGCNA, ceRNA network construction, and molecular docking. A Lasso/SVM-RFE/Random Forest consensus was used for feature selection.

Results

Differential expression revealed that resistant tumors were characterized by significant downregulation of immune-related pathways and upregulation of angiogenesis and transcriptional regulation pathways (Fig. 1D). WGCNA identified four modules significantly associated with the nCR phenotype: the purple module (immune-related), cyan/red modules (angiogenesis-related), and yellow module (transcription factor-related) (Fig. 4C-E). CeRNA subnetwork analysis (Fig. 5A-E) further illustrated the regulatory flow among non-coding RNAs and these functional axes. Molecular docking predicted high binding affinities between epirubicin and resistance-associated proteins (e.g., *CYP1B1*, *PDCD1*), suggesting a possible drug-sequestration mechanism.

Conclusion

This integrative analysis reveals a coordinated ceRNA landscape where immune evasion, pathological angiogenesis, and transcriptional reprogramming converge to drive TNBC chemoresistance. The identified modules and their regulatory hubs provide hypothesis-generating targets for future therapeutic interventions.

1 Introduction

Triple-negative breast cancer (TNBC) remains the most aggressive subtype of breast cancer, which is characterized by the lack of expression of the human epidermal growth factor receptor 2 (HER2), progesterone receptor (PR), and estrogen receptor (ER)[1]. TNBC has a disproportionately poorer survival rate than other subtypes because of its high histological grade and propensity for early visceral metastasis[1]. Systemic chemotherapy, notably the TAC regimen (docetaxel, epirubicin, and cyclophosphamide), continues to be the standard of treatment in the absence of effective endocrine or

HER2-targeted therapies[1, 2]. However, the quick development of chemoresistance sometimes compromises clinical effectiveness, resulting in therapy failure[1–3]. There is a significant information vacuum in clinical oncology about the exact biological mechanisms behind this adaptive resistance.

Recent transcriptomic discoveries have identified circular RNAs (circRNAs) as essential regulators of tumor growth and treatment response[3]. CircRNAs have a covalently closed loop structure created by back-splicing, which gives them remarkable durability against exonuclease destruction in contrast to linear transcripts. Their promise as reliable diagnostic indicators and therapeutic targets is highlighted by their stability and tissue-specific expression patterns. CircRNAs act as molecular sponges for microRNAs (miRNAs) within the competitive endogenous RNA (ceRNA) framework, influencing the production of downstream messenger RNAs (mRNAs) implicated in carcinogenic pathways[4]. The majority of TNBC research still uses fragmented, pairwise interaction models, which are unable to fully represent the systemic complexity of the chemoresistant transcriptome, despite recent developments.

Moreover, chemoresistance is closely associated with the dynamic remodeling of the tumor microenvironment (TME) rather than being an independent biological event[3]. The survival of resistant cells has been found to be supported by both immune evasion and pathological angiogenesis. In addition to facilitating the flow of nutrients, aberrant angiogenic signaling, which is fueled by *CYP1B1* and *DLL4*, also forms a physical barrier that restricts the transport of drugs. Concurrently, resistant clones are able to evade T-cell-mediated monitoring due to the immunosuppressive characteristics of the TME, which are typified by the activation of checkpoints such *PDCD1 (PD-1)* and *CTLA4*[3]. Many TNBC patients have natural or acquired resistance to checkpoint inhibitors, despite the promise of immunotherapy. In the context of circRNA-mediated regulation, the relationship between these seemingly unrelated biological processes—angiogenesis, immune evasion, and transcriptional reprogramming—remains mainly unknown.

The absence of structural validation connecting predicted resistance genes to real pharmacological action is a major drawback of current transcriptome research. For example, resistance-associated proteins may directly link transcriptome dysregulation to clinical treatment failure through the metabolic inactivation or physical sequestration of medications such as epirubicin. An integrated strategy that moves from sequence-based prediction to structural biology validation is needed to address this.

An integrated strategy that moves from sequence-based prediction to structural biology validation is needed to address this. We therefore performed a multi-omics analysis of mRNA, lncRNA, miRNA, and circRNA profiles from 15 TNBC patients with divergent responses to neoadjuvant TAC therapy. To mitigate over-fitting and bias associated with limited sample size, we employed a triple-algorithm machine learning consensus (Lasso, SVM-RFE, Random Forest) for feature selection. Using WGCNA, we aimed to identify functional co-expression modules linked to the chemoresistant phenotype. Subsequently, we constructed ceRNA subnetworks and used Sankey diagrams to visualize the regulatory flow from non-coding RNAs to downstream functional axes. Finally, molecular docking was performed to assess potential drug-target interactions. Our computational analysis reveals a coordinated ceRNA

landscape in which immune-related pathways are downregulated while angiogenesis and transcriptional regulation pathways are upregulated. These three axes correspond to distinct WGCNA modules (purple for immunity, cyan/red for angiogenesis, yellow for transcription) and are integrated via a shared ceRNA network. This study provides a hypothesis-generating framework for understanding how multi-axis ceRNA regulation may contribute to TNBC chemoresistance.

2 Material and Methods

2.1 Clinical Specimen Acquisition and Cohort Stratification

The Ethical Committee of Xinjiang Medical University Affiliated Tumor Hospital formally approved this study (Approval No. K-2025078). 15 patients with triple-negative breast cancer (TNBC) who were undergoing the TAC neoadjuvant chemotherapy regimen (docetaxel, epirubicin, and cyclophosphamide) between 2020 and 2023 were prospectively included in a clinical cohort. Before starting therapy, fresh frozen specimens were obtained using a core needle biopsy. Patients were divided into three groups based on the pathological Miller-Payne (MP) grading system: the chemosensitive group (pCR, n = 6; MP grade 5), the chemoresistant group (nCR, n = 6; MP grade 1/2), and a normal control group made up of nearby non-tumor tissues (n = 3). Each subject provided written informed permission.

2.2 RNA Isolation and Multi-omics Library Preparation

TRIzol reagent (Invitrogen) was used to extract total RNA from frozen tissues. Only samples with a RIN > 7.0 were used for further library creation. RNA concentration and purity were measured using NanoDrop spectrophotometry, and structural integrity was confirmed using an Agilent Bioanalyzer 2100. The NEBNext Ultra II Directional RNA Library Prep Kit was used to create strand-specific RNA-seq libraries following ribosomal RNA depletion. The Illumina NovaSeq 6000 platform was used to perform high-throughput paired-end sequencing (PE150).

2.3 Transcriptome Data Processing and Genomic Alignment

Adapter sequences and low-quality bases (Phred score < 20) were removed from raw sequencing data in FASTQ format using the FASTX-Toolkit (v0.0.13). HISAT2 (v2.2.1) was used to align clean reads to the human reference genome (GRCh38), allowing for a maximum of four nucleotide mismatches[11]. Cleaned readings were matched to the human genome using Bowtie for miRNA analysis; matches to rRNAs and tRNAs were eliminated. The reads that did not align were gathered and submitted to new miRNA prediction using miRDeep, while the remaining reads were mapped against the miRBase database. Following conventional computation techniques, the levels of miRNA expression were then measured in transcripts per million (TPM). StringTie (v2.2.1) was used to assemble, quantify, and normalize mRNA and lncRNA transcripts using FPKM[1, 3, 12].

2.4 Computational Identification of circRNAs and lncRNAs

A multi-algorithm consensus process was used for circRNA identification in order to reduce false positives[13]. Candidates found by find_circ, circRNA_finder, CIRI2, and circexplorer2—at least two separate tools—were kept[14, 15]. The FcircSEC R package was used to rebuild full-length circRNA sequences, and Spliced Reads per Billion Mapping (SRPBM) was used to measure expression[16, 17]. A thorough screening process was used to identify lncRNAs. The coding potential of transcripts was assessed using CPC2, PLEK, CNCI, and CPAT[18–21]. Transcripts that were found to be non-coding by all four methods and had a length more than 200 nt and an intergenic distance greater than 1,000 bp from protein-coding genes were classified as high-confidence lncRNAs[20].

2.5 Differential Expression Analysis and Functional Annotation

The DESeq2 (v1.38.3) and edgeR R programs were used to conduct differential expression analysis between the nCR, pCR, and Normal cohorts[22, 23]. An adjusted P-value (FDR) < 0.05 and a linear fold change ≥ 1.5 (or ≤ 0.67) were considered statistically significant. Using the KOBAS 2.0 server, GO word and KEGG pathway enrichment analysis[24] was used to clarify the biological functions of differentially expressed genes (DEGs), with P-values adjusted using the Benjamini-Hochberg technique[25].

2.6 Feature Selection via Triple-Algorithm Consensus Framework

The high-confidence driver set was defined as the intersection of attributes that at least two of the three approaches prioritized[6, 8]. In order to reduce algorithm-specific biases, our cross-validation method made sure that candidates like *circBIRC6 (hsa_circ_0119838)*, which were found by both Lasso and Random Forest, were chosen based on their consistent performance across complementary mathematical models[5, 7, 9].

2.7 WGCNA and ceRNA Regulatory Network Construction

Weighted Gene Co-expression Network Analysis (WGCNA v1.72.1) was used to go beyond individual gene analysis and understand systems-level functional units[10]. The soft-thresholding power obtained from the scale-free topology fit index was used to build a scale-free co-expression network. 39 different modules of highly co-expressed genes were identified from the transcriptome. To find modules that were substantially connected with the chemoresistant (nCR) or chemosensitive (pCR) phenotypes, module-trait connections were measured. For the following creation of ceRNA networks, functional modules that were positively correlated with resistance (cyan and red) and sensitivity (purple and yellow) were given priority. Predicted miRNA-target binding sites (miRanda score ≥ 150, P < 0.05; and RNAhybrid) were integrated with expression patterns to create the ceRNA regulatory architecture. Only interactions that showed a contemporaneous positive correlation between ncRNAs and target mRNAs and a strong negative correlation between ncRNAs and miRNAs were kept.

2.8 Structural Validation through In Silico Molecular Docking

AutoDock Vina (v1.1.2) was used to assess the physical interaction stability between chemotherapeutic drugs and resistance-associated targets. Drug structures were obtained from PubChem, while three-dimensional structures for target proteins (such as *CYP1B1* and *PDCD1*) were obtained from the Protein Data Bank (PDB). Docking grids with an exhaustiveness of 20 were focused on anticipated binding pockets (60x60x60). Energy scores (kcal/mol) were used to quantify binding affinities, and PyMOL (v3.03) was used to display residue-specific interactions such as hydrophobic contacts and hydrogen bonds.

2.9 Statistical Analysis and Software Implementation

R (v4.2.3) and Python (v3.10) were used to construct computational analyses and visualizations, such as stacked bar charts and pattern diagrams. The Wald test or, when suitable, the hypergeometric tests were used to compare continuous variables. The Benjamini-Hochberg approach was used to limit the False Discovery Rate (FDR) by multiple comparison modifications.

3 Results

3.1 Overview of Sequencing Data and Transcriptomic Landscape

To describe the genetic landscape of chemoresistance, clinical specimens from 15 TNBC patients were subjected to multi-omics RNA sequencing. The chemoresistant (nCR), chemosensitive (pCR), and normal groups showed clear transcriptome separation using Principal Component Analysis (PCA), suggesting that the emergence of resistance is linked to discrete and repeatable molecular fingerprints (Fig. 1A). Throughout the three comparison cohorts, a wide range of resistance-associated genes were found using differential expression analysis. In particular, functional enrichment analysis using GO keywords showed that the nCR group's elevated genes were substantially engaged in angiogenesis, cell division, and extracellular matrix structure (Fig. 1D). On the other hand, genes linked to chemosensitivity were concentrated in interferon-gamma signaling and adaptive immune responses, indicating that a weakened immunological milieu is a symptom of treatment failure in TNBC. Quantitatively, differential expression analysis across the three comparisons (nCR vs. Normal, pCR vs. Normal, nCR vs. pCR) identified 2,310 differentially expressed genes (DEGs), 484 DElncRNAs, 194 DEmiRNAs, and 284 DEcircRNAs ($|FC| \geq 1.5$, $P \leq 0.05$). Intersection patterns of DEGs across comparisons are shown in Fig. 1B, and a heatmap illustrating the expression profiles of DEGs is shown in Fig. 1C. Notably, GO enrichment analysis of these DEGs (Fig. 1D) revealed that compared with the pCR group, the nCR group exhibited significant downregulation of immune-related pathways and upregulation of pathways involved in transcriptional regulation and angiogenesis. These observations suggest that chemoresistance in

TNBC may be driven by the convergence of three functional axes: immune evasion, pathological angiogenesis, and transcriptional reprogramming.

3.2 Expression Profiles of Non-coding RNAs

The non-coding RNA (ncRNA) landscape was systematically evaluated to identify potential regulatory drivers of chemoresistance.

The majority of discovered circRNAs had lengths centered between 200 and 1000 nt and typically contained 2 to 5 exons (Fig. 2A, C), and were predominantly derived from exonic regions (Fig. 2B).Compared to normal tissues, 2,066 circRNAs were uniquely and significantly elevated in the nCR group, displaying a distinct group-specific expression pattern that separated resistant from sensitive clones (Fig. 2D, E)[26–30]. Among these, several circRNAs derived from oncogene loci, such as *hsa_circ_0119838* (originating from the BIRC6 gene), showed consistent upregulation in resistant specimens. However, no single circRNA was universally overexpressed across all comparisons (Fig. 2E), indicating a complex, multi-factor regulatory landscape.

3.3 Expression Profiles of lncRNAs and miRNAs

In comparison to protein-coding mRNAs, a high number of unique lncRNAs with shorter transcript lengths and lower expression levels were predicted (Fig. 3A-C). A group of lncRNAs that were strongly linked to the chemoresistant phenotype were verified by differential expression analysis. These lncRNAs were then ranked according to their intergenic distance from recognized oncogenes (Fig. 3D).

In all three cohorts, differentially expressed miRNAs also showed unique, group-specific patterns (Fig. 3E, F). The coordinated dysregulation seen throughout these several non-coding RNA levels implies that a systemic, multi-layered regulatory network rather than single-gene alterations is responsible for resistance.

3.4 Identification of Co-expression Modules via WGCNA

Weighted Gene Co-expression Network Analysis (WGCNA) was used to the expression patterns of all identified transcripts in order to interpret the systems-level coordination of the TNBC transcriptome[10]. By choosing an ideal soft-thresholding power (**β**) based on the scale-free topology fit index ($\mathbf{R^2} > 0.9$), a scale-free co-expression network was created. The transcriptome was divided into 39 different modules by the resultant dendrogram (Fig. 4A, B).

Pivotal clusters linked to clinical characteristics were found using module-trait connection analysis (Fig. 4C). Based on their possible biological significance to the chemoresistant (nCR) phenotype, the cyan module (684 genes) and the red module (187 genes) were selected as candidates for additional investigation (**r** = 0.31, **P** = 0.3; and **r** = 0.5, **P** = 0.06, respectively). These modules were given priority because of their high gene density and functional enrichment in resistance-promoting pathways, even if

these associations did not approach formal statistical significance. On the other hand, resistance showed a negative connection with the purple module (326 genes). The correlation strengths, DEG counts across modules, and expression patterns of these four modules are shown in (Fig. 4C, D, E). Consistent with the GO findings in Fig. 1D, the purple module was enriched in immune-related processes and negatively correlated with resistance, while the cyan/red modules were enriched in angiogenesis pathways and the yellow module in transcriptional regulation, all positively correlated with the nCR phenotype (Fig. 4C-E).

3.5Construction and Functional Deconvolution of ceRNA Networks

3.5.1 Subnetwork Architecture of Functional Modules

To identify the circRNA/lncRNA-miRNA-mRNA regulatory axis, subnetworks were built for the modules that were given priority (Fig. 5A). To emphasize the key regulatory hubs, only the top ten highest-degree nodes were kept. Within each subnetwork, multiple circRNAs and lncRNAs acted as highly connected hubs bridging miRNAs to downstream mRNAs. For example, in the cyan/red subnetworks, angiogenesis-related effectors (e.g., *CYP1B1, DLL4*) were prominently targeted; in the purple subnetwork, immune checkpoints (e.g., *PDCD1, CTLA4*) were enriched; and in the yellow subnetwork, transcription factors (e.g., *SIX1, MEF2A*) dominated (Fig. 5A). The Sankey diagram (Fig. 5E) visually integrated these regulatory flows, illustrating how non-coding RNAs may coordinate the three functional axes.

3.5.2 Functional Enrichment and Sankey Integration

For every regulatory axis, GO enrichment analysis offered crucial functional context (Fig. 5B). The yellow module was mostly linked to transcription control, whereas the purple module was significantly enriched for immune-related activities. The tumor microenvironment and treatment failure were closely linked by the cyan and red modules' significant enrichment in angiogenesis-related pathways. Significant overlap between the purple module and 76 immune-associated genes (such *PDCD1* and *CTLA4*) and between the yellow module and 20 transcription factors (like *SIX1* and *MEF2A*) was confirmed by Venn diagrams (Fig. 5C, D)[31, 32]. The integrated regulatory flow from non-coding RNAs (including circRNAs and lncRNAs) to these prioritized functional axes is summarized in the Sankey diagram (Fig. 5E).

3.6Molecular Docking Prediction of Drug-Target Interactions

In silico molecular docking between important proteins expressed by the network mRNAs (such as *CYP1B1, PDCD1*, and *FN1*) and the chemotherapeutic drug epirubicin was carried out in order to explore the potential biological importance of the detected ceRNA axis (Fig. 6A). Several putative targets originating from the angiogenic and immunological modules showed strong binding affinities, according to the study (Fig. 6A). Significantly, epirubicin showed a substantial interaction with *PDCD1* at -9.49 kcal/mol and *CYP1B1* at -9.2 kcal/mol. Certain hydrogen bonds and hydrophobic contacts inside the

active sites were verified by visual examination of the interaction diagrams, indicating that these proteins may function as direct effectors of resistance (Fig. 6B). These high-affinity interactions suggest a possible drug-sequestration mechanism in which chemotherapeutic drugs are physically trapped by overexpressed resistance proteins, impeding their pharmacological potency and effective intracellular concentration.

4 Discussion

The intrinsic heterogeneity of triple-negative breast cancer (TNBC) and the rapid emergence of resistance to standard-of-care TAC neoadjuvant chemotherapy make clinical management of TNBC a difficult task[1, 3]. Although circular RNAs (circRNAs) have occasionally been identified as distinct actors in cancer in earlier research, our study goes beyond these disjointed discoveries by constructing a systematic, multi-layered regulatory architecture[4, 33]. By integrating whole-transcriptome sequencing with a triple-algorithm machine learning consensus and WGCNA, we identified three convergent functional axes — immune evasion, pathological angiogenesis, and transcriptional reprogramming — that are coordinately dysregulated in resistant tumors and linked through a shared ceRNA network[5–9].

4.1 The Drug-Sequestration Mechanism: A Structural Insight into Resistance

Having established the three functional axes through WGCNA module identification and their integration via the ceRNA network, we next sought to assess whether key proteins from these axes could physically interact with chemotherapeutic drugs. The molecular docking results raise the hypothesis of a potential drug-sequestration mechanism that requires experimental validation.In the past, ATP-binding cassette (ABC) transporters that actively pump medications out of the cell have been linked to chemoresistance. However, resistance-associated proteins, especially *CYP1B1*, may act as passive but effective "molecular traps," according to our in silico structure study, which provides a fresh mechanistic insight. The remarkably high binding affinity (-9.2 kcal/mol) between epirubicin and *CYP1B1* indicates that chemotherapeutic drugs may be physically sequestered inside the cytoplasm or particular organelles by overexpressed proteins in nCR tissues. This interaction may neutralize the drug's pro-apoptotic activity by preventing it from reaching nuclear targets such as topoisomerase II, without requiring active transport. TNBC cells have a multi-tiered defense mechanism thanks to this "sponge effect" at the protein level, which is similar to the "sponge effect" at the RNA level.

4.2 WGCNA Reveals Three Distinct Functional Modules Associated with Chemoresistance

To systematically dissect the transcriptomic architecture of chemoresistance, we performed WGCNA on all expressed genes. This analysis partitioned the transcriptome into 39 co-expression modules, among which four modules were closely associated with the nCR phenotype (Fig. 4C-E). Consistent with the GO findings from Fig. 1D, the purple module was significantly enriched in immune-related processes and

negatively correlated with resistance, whereas the cyan and red modules were positively correlated with resistance and strongly enriched in angiogenesis-related pathways. The yellow module, also positively correlated with the nCR phenotype, was predominantly enriched in transcriptional regulation. These results confirm that chemoresistance in TNBC is associated with three convergent functional axes: immune evasion (purple), pathological angiogenesis (cyan/red), and transcriptional reprogramming (yellow).

4.3 ceRNA Network Integrates the Three Functional Axes

Having identified the three functional modules, we next investigated whether they are interconnected through a shared ceRNA regulatory architecture. Using predicted miRNA-mRNA and ncRNA-miRNA interactions, we constructed circRNA/lncRNA-miRNA-mRNA subnetworks for each prioritized module (Fig. 5A). Multiple circRNAs and lncRNAs acted as highly connected hubs, bridging miRNAs to downstream mRNAs within each functional axis. More importantly, the Sankey diagram (Fig. 5E) visually integrates the regulatory flows from non-coding RNAs to the three functional axes, illustrating that immune evasion, angiogenesis, and transcriptional reprogramming are not isolated events but are coordinately regulated through a shared ceRNA network. This systems-level view suggests that ceRNA networks may serve as a molecular platform that integrates multiple resistance-promoting pathways in TNBC.

4.4 The Immune-Angiogenic Interplay: Co-regulating Pillars of Resistance

Our findings suggest a coordinated "excluded" or "suppressive" immune phenotype within the chemoresistant tumor microenvironment (TME)[3]. Pathological angiogenesis, driven by the cyan and red modules, often results in a disorganized vascular network that creates a physical and metabolic barrier. This structural abnormality not only promotes hypoxia but also significantly hinders the effective infiltration of cytotoxic T-cells into the tumor core[1]. Concurrently, the significant enrichment of immune-checkpoint pathways within the purple module—characterized by the upregulation of PDCD1 (PD-1) and CTLA4—renders the few infiltrating lymphocytes dysfunctional[3]. We hypothesize that the ceRNA network, via multiple ncRNA hubs (including circRNAs such as hsa_circ_0119838 and various lncRNAs), may function as a molecular bridge in this crosstalk, sponging specific miRNAs that would otherwise suppress both angiogenic factors and immune checkpoints[4, 33]. This hypothesis requires experimental validation. By co-regulating these two pillars, the ceRNA network may contribute to a hostile microenvironment that protects resistant clones from both chemotherapy and immune surveillance.

4.5 Transcriptional Plasticity and Cellular Adaptability

Through the yellow module, the ceRNA network may promote internal cellular adaptability via microenvironmental remodeling. This module contains high-confidence transcription factors including *MEF2A* and *SIX1*, which suggests a master regulatory layer that controls extensive transcriptional reprogramming[31, 32]. We suggest that TNBC cells are able to retain a high level of phenotypic flexibility

because of this regulatory axis. Such network-mediated plasticity may allow cells to dynamically change their metabolic profiles and activate alternative survival pathways, such as enhanced DNA repair and drug efflux, under the selective pressure of TAC agents (docetaxel, epirubicin, and cyclophosphamide), avoiding the apoptotic triggers of conventional chemotherapy.

4.6 Clinical Implications and Future Perspectives

The identified ceRNA network, particularly the crosstalk between the cyan/red modules (angiogenesis) and the purple module (immunity), offers potential targets for sensitizing resistant TNBC[33–35]. Non-coding RNAs such as circBIRC6, given their stability, may be candidates for liquid biopsy markers to monitor chemotherapy response in real time, but this requires further validation[33]. Additionally, the development of small-molecule inhibitors or PROTACs (Proteolysis Targeting Chimeras) specifically intended to disrupt the *CYP1B1*-drug interaction and release the "sequestered" chemotherapeutic agents to carry out their intended function is rationally supported by our molecular docking data. Combination strategies that pair TAC drugs with anti-angiogenic inhibitors (targeting the cyan/red modules) or with ncRNA-silencing oligonucleotides may offer synergistic approaches to overcome adaptive resistance[36].

4.7 Limitations

Although the multi-model consensus method is thorough, there are a few drawbacks that need to be noted[6, 8]. Several limitations of this study should be acknowledged. First, the sample size is limited (n = 15, with only 3 normal samples), which restricts statistical power and generalizability. Although a triple-algorithm consensus strategy was implemented to mitigate over-fitting, all differential expression and machine learning results are exploratory and require validation in larger, independent cohorts. Second, this study is purely computational and hypothesis-generating. No wet-lab experiments (e.g., qRT-PCR, luciferase reporter assays, or functional knockdown studies) were performed to validate the proposed ceRNA interactions or the drug-sequestration hypothesis. Third, the clinical cohort is derived from a single institution, which may introduce selection bias. Multi-center validation is needed. Fourth, the proposed regulatory network is based on predicted miRNA binding sites and expression correlations, which do not prove direct molecular interactions. Consequently, all mechanistic interpretations should be considered as hypotheses for future investigation rather than definitive conclusions[37–39].To corroborate these mechanistic theories, further wet-lab investigations are needed, such as dual-luciferase reporter assays to establish RNA interactions and in vitro/in vivo functional tests to demonstrate the role of *circBIRC6*. Furthermore, while several circRNAs (e.g., *hsa_circ_0119838*) were observed in the network, no single ncRNA was exclusively responsible for resistance, underscoring the complexity of the ceRNA landscape. Functional validation should therefore focus on network-level perturbations rather than on individual molecules.

4.8 Future Experimental Validation Roadmap

To validate the hypotheses generated by this computational study, the following experiments are recommended: (1) qRT-PCR validation of key non-coding RNAs (including selected circRNAs such as *hsa_circ_0119838* and relevant lncRNAs) and target mRNAs (*CYP1B1, PDCD1, DLL4*) in an independent,

larger TNBC cohort; (2) dual-luciferase reporter assays to confirm direct *circBIRC6-miRNA* and miRNA-mRNA interactions; (3) perturbation of specific modules (e.g., knockdown of cyan/red module-associated genes or overexpression of purple module immune checkpoints) in TNBC cell lines, followed by IC_{50} measurements for TAC drugs to establish causality;(4) Western blot or immunohistochemistry to confirm *CYP1B1* protein overexpression in chemoresistant versus chemosensitive tissues. These experiments will be essential to translate the current in silico findings into clinically actionable insights.

4.9 Conclusion: Integrated Mechanistic Insights

In conclusion, our work offers a thorough attempt to understand the systemic transcriptome rewiring underlying chemotherapy resistance in TNBC. We have produced a high-fidelity map of the non-coding RNA-mediated regulatory landscape by combining sophisticated computational modeling with multi-omics analysis of fifteen clinical cases.

4.9.1 Methodological Rigor: Stability through Consensus Modeling

The application of a triple-algorithm machine learning consensus framework is essential to our conclusions[5–9]. We were able to overcome the inherent volatility and small-sample bias that are frequently connected with clinical cohorts by combining Lasso regression, SVM-RFE, and Random Forest. This method provides robust feature selection and reduces over-fitting, offering a useful reference for biomarker discovery in resistant cancers.

4.9.2 Synthesis of the Tripartite Resistance Axis

Our study reveals a ceRNA network that operates in three convergent dimensions: (i) immune evasion (purple module), (ii) pathological angiogenesis (cyan/red modules), and (iii) transcriptional reprogramming (yellow module). These axes are linked through shared non-coding RNA regulators and are consistent with the GO enrichment from Fig. 1D and the module-trait correlations from Fig. 4C-E. The Sankey diagram (Fig. 5E) visually integrates these regulatory flows. Additionally, molecular docking predicts high binding affinities between epirubicin and key proteins from the immune (*PDCD1*) and angiogenic (*CYP1B1*) modules, suggesting a possible drug-sequestration mechanism. Together, our results support a coordinated, systems-level hypothesis in which multi-axis ceRNA regulation contributes to TNBC chemoresistance.

4.9.3 Clinical Impact and Future Perspectives

This work has immediate implications for precision oncology. Non-coding RNAs identified in this network, such as circBIRC6, may serve as candidate markers for liquid biopsy applications, but this requires further validation. [33]. Additionally, our molecular docking-based computational docking analysis offers a computational basis for future investigation[34, 35]. Antisense oligonucleotides or inhibitors that block the drug-protein binding pocket may be used to strategically disrupt the CYP1B1-

drug interaction or target key ncRNA hubs. This might "release" medications that have been sequestered and re-sensitize resistant TNBC cells to conventional chemotherapy.

4.9.4 Final Concluding Remarks

In the end, our work moves from straightforward sequence-based prediction to a multifaceted comprehension of pharmaceutical failure. We offer a comprehensive framework for comprehending how TNBC avoids therapeutic death by connecting transcriptome architecture with structural biology validation. The ceRNA network identified in this computational study, particularly the convergence of immune, angiogenic, and transcriptional modules, provides a hypothesis-generating framework for future investigation into adaptive chemoresistance and enhancing patient survival outcomes, even though further wet-lab validation and larger cohort studies are necessary to fully realize the clinical potential of these findings.

Abbreviations

ceRNA
competing endogenous RNA
circRNA
circular RNA
DEGs
differentially expressed genes
FPKM
fragments per kilobase of transcript per million mapped reads
GO
Gene Ontology
lncRNA
long non-coding RNA
miRNA
microRNA
MP
Miller-Payne
nCR
non-pathological complete response
pCR
pathological complete response
TAC
docetaxel/epirubicin/cyclophosphamide
TME
tumor microenvironment
TNBC

triple-negative breast cancer
TPM
transcripts per million
WGCNA
weighted gene co-expression network analysis

Declarations

5 Acknowledgements

The authors would like to thank the Xinjiang Key Laboratory of Oncology and the Xinjiang 'Tianshan Talent' Training Program and the Natural Science Youth Research Project of Xinjiang Medical University for their financial support.

6 Conflict of Interest

The authors declare that they have no competing interests.

7 Authors' contributions

YongTao Li: Conceptualization, Resources, Data Curation, Writing – Original Draft.

MiaoMiao Zhang: Methodology, Validation, Formal Analysis, Writing – Review & Editing, Visualization, Project Administration.

WenTing Xu: Writing – Review & Editing.

Yu Si: Writing – Review & Editing.

RuYu Pei: Writing – Review & Editing.

Tangnuer Buerliesi: Writing – Review & Editing.

Rui Ma: Writing – Review & Editing.

Bing Zhao: Supervision, Writing – Review & Editing.

8 Funding

This work was supported by the Xinjiang 'Tianshan Talent' Training Program - High-Level Medical and Health Talents Project [Grant Number: TSYC202401A044] and the Natural Science Youth Research Project of Xinjiang Medical University [Grant Number: 2026XYZR29]. The funders had no role in the study design; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

9Availability of data and materials

Availability of data and materials: The RNA-seq data generated in this study are publicly available under the BioProject accession number PRJCA047966 at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB), and can be accessed at: https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA047966.

10 Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Ethics Committee of Xinjiang Medical University Affiliated Tumor Hospital (Approval No. K-2025078; date of approval: 2025-07-24). Informed consent was obtained from all subjects involved in the study.

11 Patient consent for publication

Not applicable.

References

1. Mohanad M, Hamza HM, Bahnassy AA, Shaarawy S, Ahmed O, El-Mezayen HA, et al. Molecular profiling of breast cancer methylation pattern in triple negative versus non-triple negative breast cancer. Sci Rep. 2025;15(1):6894. 10.1038/s41598-025-90150-9.
2. Bianchini G, De Angelis C, Licata L, Gianni L. Treatment landscape of triple-negative breast cancer --- expanded options, evolving needs. Nat Rev Clin Oncol. 2022;19(2):91–113. 10.1038/s41571-021-00565-2.
3. Li Z, Qiu Y, Lu W, Jiang Y, Wang J. Immunotherapeutic interventions of Triple Negative Breast Cancer. J Transl Med. 2018;16(1):147. 10.1186/s12967-018-1514-7.
4. Qi X, Lin Y, Chen J, Shen B. Decoding competing endogenous RNA networks for cancer biomarker discovery. Brief Bioinform. 2020;21(2):441–57. 10.1093/bib/bbz006.
5. Liu Y, Yin Z, Wang Y, Chen H. Exploration and validation of key genes associated with early lymph node metastasis in thyroid carcinoma using weighted gene co-expression network analysis and machine learning. Front Endocrinol (Lausanne). 2023;14:1247709. 10.3389/fendo.2023.1247709.
6. Drotár P, Gazda J, Smékal Z. An experimental comparison of feature selection methods on two-class biomedical datasets. Comput Biol Med. 2015;66:1–10. 10.1016/j.compbiomed.2015.08.010.
7. Huang S, Cai N, Pacheco PP, Narrandes S, Wang Y, Xu W, et al. Applications of support vector machine (SVM) learning in cancer genomics. Cancer Genomics Proteom. 2018;15(1):41–51. 10.21873/cgp.20063.
8. Sarkar C, Cooley S, Srivastava J. Robust feature selection technique using rank aggregation. Appl Artif Intell. 2014;28(3):243–57. 10.1080/08839514.2014.883903.

9. Hu J, Szymczak S. A review on longitudinal data analysis with random forest. Brief Bioinform. 2023;24(2):bbad002. 10.1093/bib/bbad002.
10. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559. 10.1186/1471-2105-9-559.
11. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60. 10.1038/nmeth.3317.
12. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511–5. 10.1038/nbt.1621.
13. Hansen TB, Venø MT, Damgaard CK, Kjems J. Comparison of circular RNA prediction tools. Nucleic Acids Res. 2016;44(6):e58. 10.1093/nar/gkv1458.
14. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29(1):15–21. 10.1093/bioinformatics/bts635.
15. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. 2013. 10.48550/arXiv.1303.3997
16. Hossain MT, Peng Y, Feng S, Wei Y. FcircSEC: an R package for full length circRNA sequence extraction and classification. Int J Genomics. 2020;2020:9084901. 10.1155/2020/9084901.
17. Tan WL, Lim BT, Anene-Nzelu CG, Ackers-Johnson M, Dashi A, See K, et al. A landscape of circular RNA expression in the human heart. Cardiovasc Res. 2017;113(3):298–309. 10.1093/cvr/cvw250.
18. Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, et al. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Res. 2007;W345–9. 10.1093/nar/gkm391. 35(Web Server issue).
19. Wang G, Yin H, Li B, Yu C, Wang F, Xu X, et al. Characterization and identification of long non-coding RNAs based on feature relationship. Bioinformatics. 2019;35(17):2949–56. 10.1093/bioinformatics/btz008.
20. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, et al. Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. Nucleic Acids Res. 2013;41(17):e166. 10.1093/nar/gkt646.
21. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: coding-potential assessment tool using an alignment-free logistic regression model. Nucleic Acids Res. 2013;41(6):e74. 10.1093/nar/gkt006.
22. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. 10.1186/s13059-014-0550-8.
23. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139–40. 10.1093/bioinformatics/btp616.

24. Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Res. 2023;51(D1):D587–92. 10.1093/nar/gkac963. . PMID: 36300620; PMCID: PMC9825424.
25. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res. 2011;W316–22. 10.1093/nar/gkr483. 39(Web Server issue).
26. Ji P, Wu W, Chen S, Zheng Y, Zhou L, Zhang J, et al. Expanded expression landscape and prioritization of circular RNAs in mammals. Cell Rep. 2019;26(12):3444–e34605. 10.1016/j.celrep.2019.02.078.
27. Glažar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. RNA. 2014;20(11):1666–70. 10.1261/rna.043687.113.
28. Chen X, Han P, Zhou T, Guo X, Song X, Li Y. circRNADb: a comprehensive database for human circular RNAs with protein-coding annotations. Sci Rep. 2016;6:34985. 10.1038/srep34985.
29. Zheng LL, Li JH, Wu J, Sun WJ, Liu S, Wang ZL, et al. deepBase v2.0: identification, expression, evolution and function of small RNAs, LncRNAs and circular RNAs from deep-sequencing data. Nucleic Acids Res. 2016;44(D1):D196–202. 10.1093/nar/gkv1273.
30. Dong R, Ma XK, Li GW, Yang L. CIRCpedia v2: an updated database for comprehensive circular RNA annotation and expression comparison. Genomics Proteom Bioinf. 2018;16(4):226–33. 10.1016/j.gpb.2018.08.001.
31. Hu H, Miao YR, Jia LH, Yu QY, Zhang Q, Guo AY. AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. 2019;47(D1):D33–8. 10.1093/nar/gky822.
32. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT. The human transcription factors. Cell. 2018;172(4):650–65. 10.1016/j.cell.2018.01.029.
33. Guo LL, Song CH, Wang P, Dai LP, Zhang JY, Wang KJ. Competing endogenous RNA networks and gastric cancer. World J Gastroenterol. 2015;21(41):11680–7. 10.3748/wjg.v21.i41.11680.
34. Goodarzi M, Dejaegher B, Vander Heyden Y. Feature selection methods in QSAR studies. J AOAC Int. 2012;95(3):636–51. 10.5740/jaoacint.sge_goodarzi.
35. Cen L, Liu R, Liu W, Li Q, Cui H. Competing endogenous RNA networks in glioma. Front Genet. 2021;12:675498. 10.3389/fgene.2021.675498.
36. Hessen SH, Abdul-Kader HM, Khedr AE, Salem RK. Developing multiagent e-learning system-based machine learning and feature selection techniques. Comput Intell Neurosci. 2022;2022:2941840. 10.1155/2022/2941840.
37. Ma JG, Wang NJ, Yu WJ. Comparison of biological behavior between triple-negative breast cancer and non-triple-negative breast cancer. Nan Fang Yi Ke Da Xue Xue Bao. 2011;31(10):1729–32. Chinese. PMID: 22027778.
38. Nowacka-Zawisza M, Krajewska WM. Triple-negative breast cancer: molecular characteristics and potential therapeutic approaches. Postepy Hig Med Dosw (Online). 2013;67:1090-7. Polish.

10.5604/17322693.1077713

39. Nabi MG, Ahangar A, Wahid MA, Kuchay S. Clinicopathological comparison of triple negative breast cancers with non-triple negative breast cancers in a hospital in North India. Niger J Clin Pract. 2015;18(3):381–6. 10.4103/1119-3077.153248.

Figures

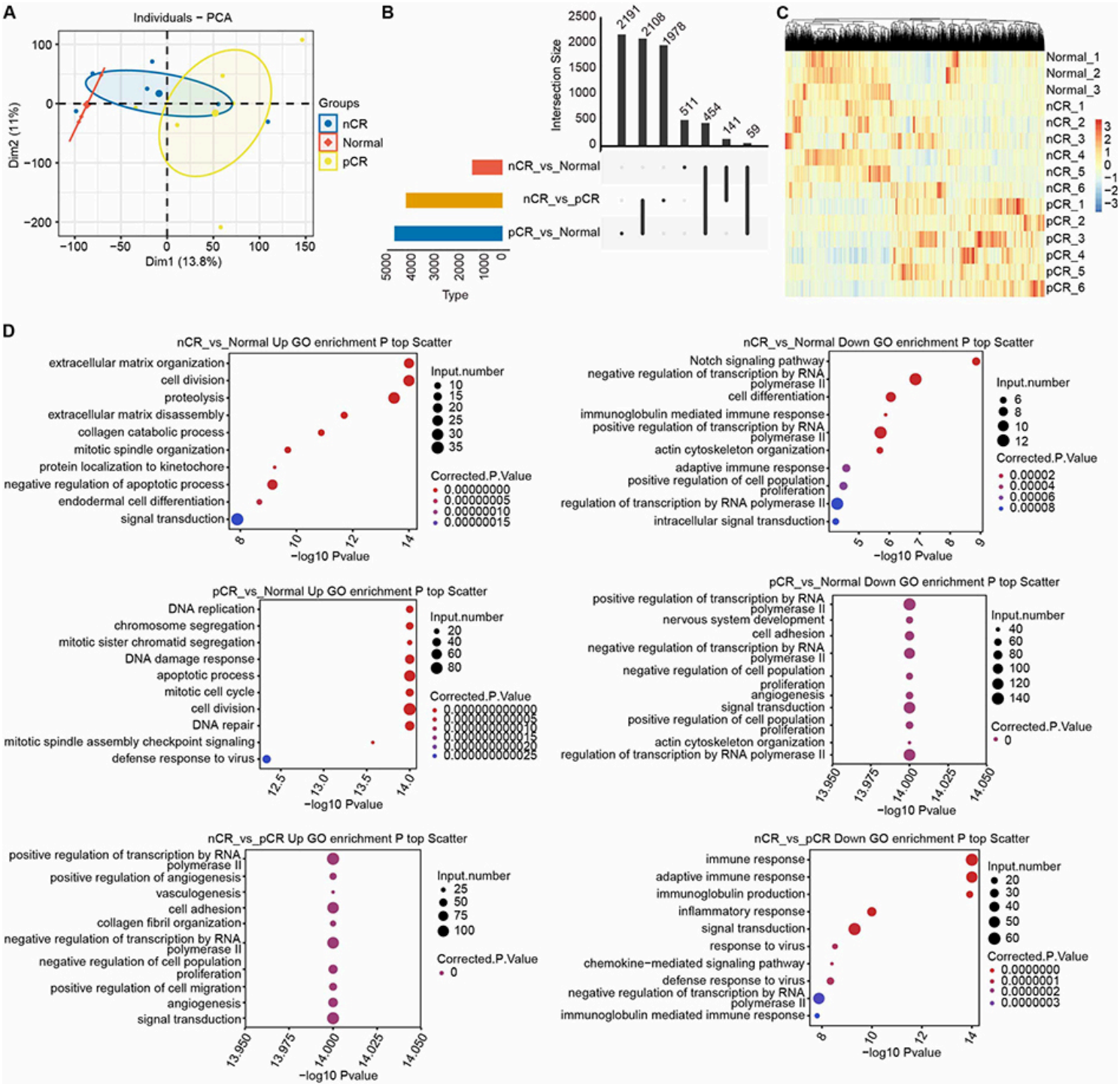


Figure 1

Transcriptomic profiling of chemoresistance in triple-negative breast cancer.

(A) Principal component analysis (PCA) based on FPKM values of all detected genes. Confidence ellipses show clear separation among the resistant (nCR), sensitive (pCR), and normal groups.

(B) UpSet plot illustrating the intersection patterns of differentially expressed genes (DEGs) from three comparisons: nCR vs Normal, pCR vs Normal, and nCR vs pCR (edge***R***, |FC| ≥1.5, ***P*** ≤0.05).

(C) Heatmap of DEGs. FPKM values were log2-transformed and median-centered per gene.

(D) Top 10 enriched GO biological process terms for DEGs. Compared with the pCR group, the nCR group showed significant downregulation of immune-related pathways and upregulation of transcriptional regulation and angiogenesis-related pathways.

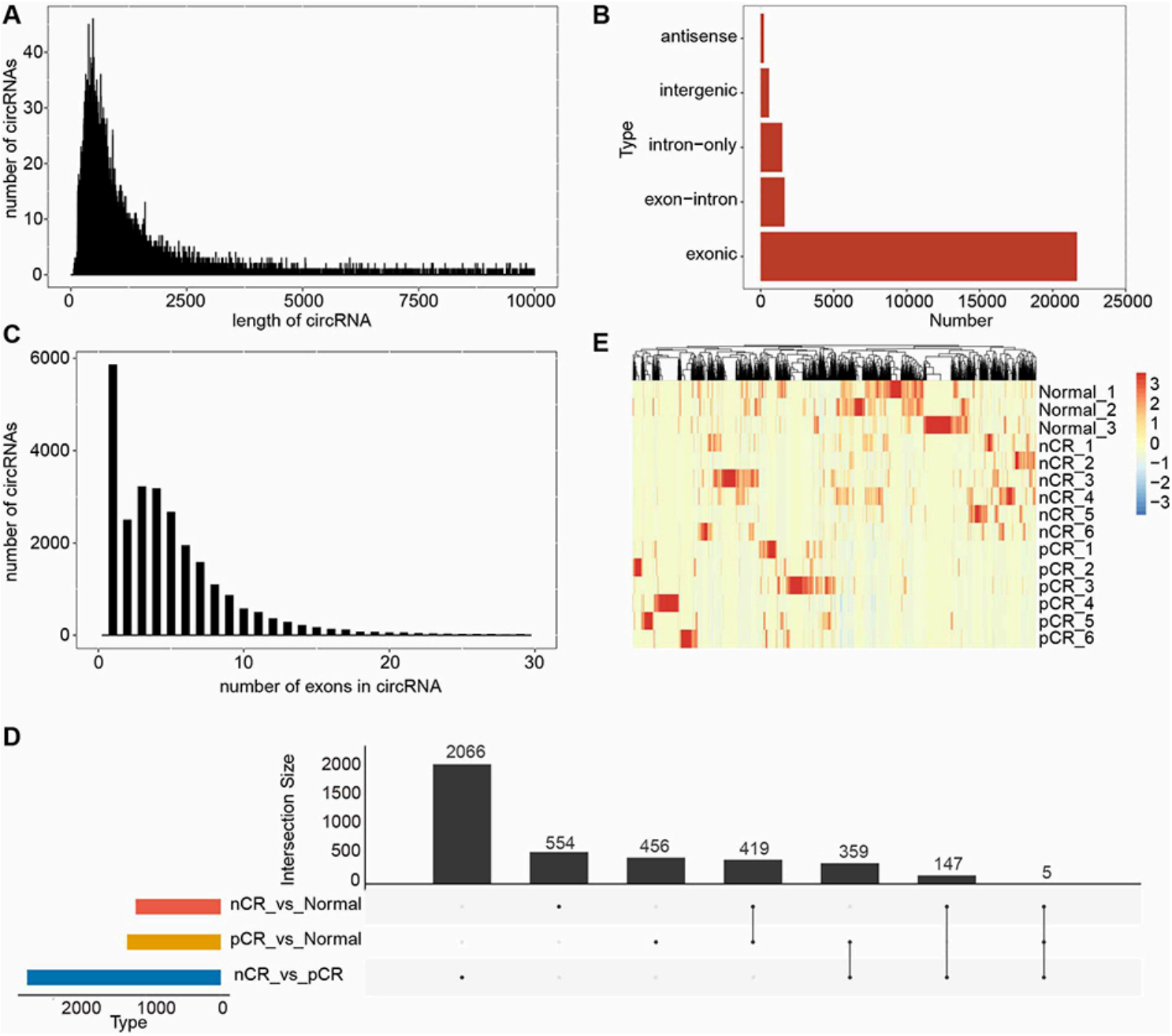


Figure 2

Analysis of circRNA expression between different groups.

(A)The bar chart displays length distribution statistics of circRNAs based on full-length sequences obtained through FcircSEC.

(B)The bar chart displays the distribution of circRNAs across different genomic regions.

(C)Bar plot showing exon count distribution in circRNAs.

(D)The UpSet diagram visualizes the intersection patterns of differentially expressed circRNAs from three comparative analyses: nCR_vs_Normal, pCR_vs_Normal, and nCR_vs_pCR. Differential expression analysis was performed using edgeR (FC: 1.5, Pvalue ≤0.05).

(E)The heatmap diagram showing the expression profile of DEcircRNAs.

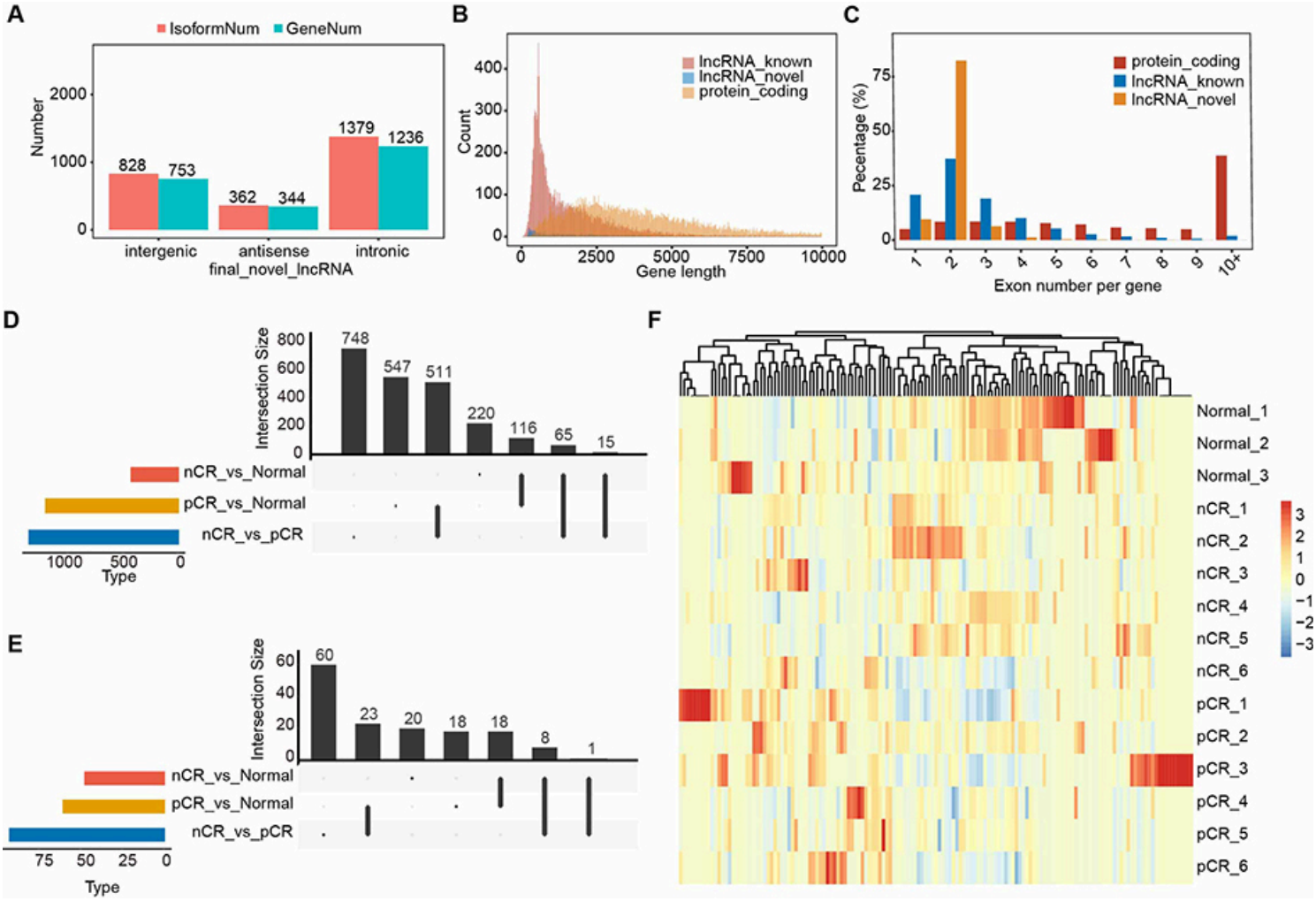


Figure 3

Expression profiles of DElncRNAs and DEmiRNAs in chemoresistance of triple-negative breast cancer

(A) Bar plot displaying the quantity of newly predicted lncRNAs.

(B) Bar chart comparing gene lengths between mRNAs and lncRNAs.

(C) Bar chart comparing gene number between mRNAs and lncRNAs.

(D) UpSet diagram visualizing the intersection patterns of differentially expressed lncRNAs from three comparative analyses: nCR_vs_Normal, pCR_vs_Normal, and nCR_vs_pCR. Differential expression analysis was performed using edgeR (FC: 1.5, ***P***≤0.05).

(E) UpSet diagram visualizing the intersection patterns of differentially expressed miRNAs from three comparative analyses: nCR_vs_Normal, pCR_vs_Normal, and nCR_vs_pCR. Differential expression analysis was performed using edgeR (FC: 1.5, ***P*** ≤0.05).

(F) Heatmap showing the expression profile of DEmiRNAs

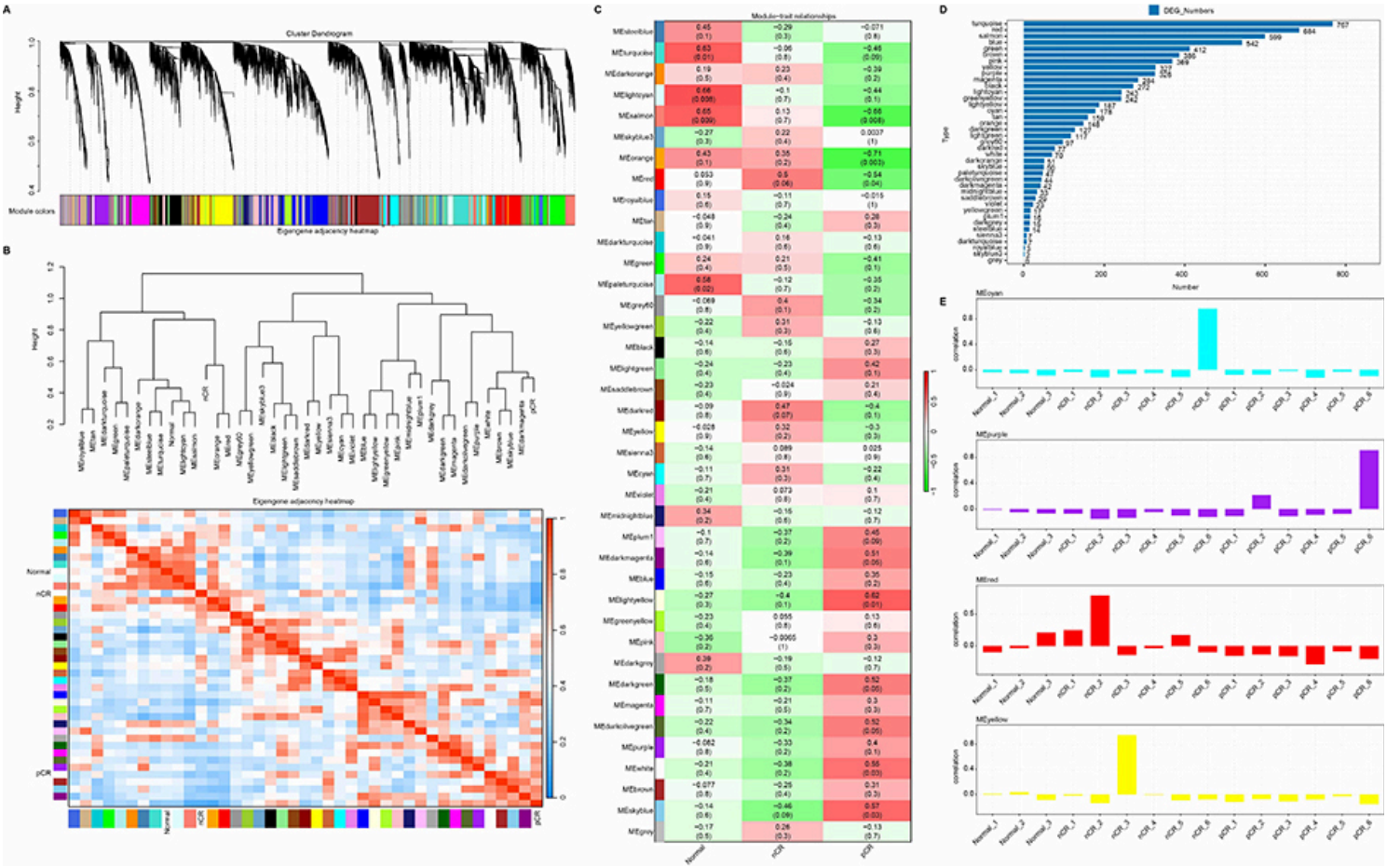


Figure 4

WGCNA was used to analyze all genes in Normal, pCR and nCR groups.

(A) Weighted gene co-expression network clustering plot showing the gene dendrogram and module colors.

(B) Hierarchical clustering of feature genes (top) and module-module correlation heatmap (bottom).

(C) Correlation heatmap between sample traits (Normal, pCR, nCR) and co-expression modules.

(D) Bar chart displaying the number of differentially expressed genes (DEGs) in each module.

(E) Expression patterns of cyan, purple, red, and yellow modules across three sample groups. Expression values are log2-transformed and median-centered.

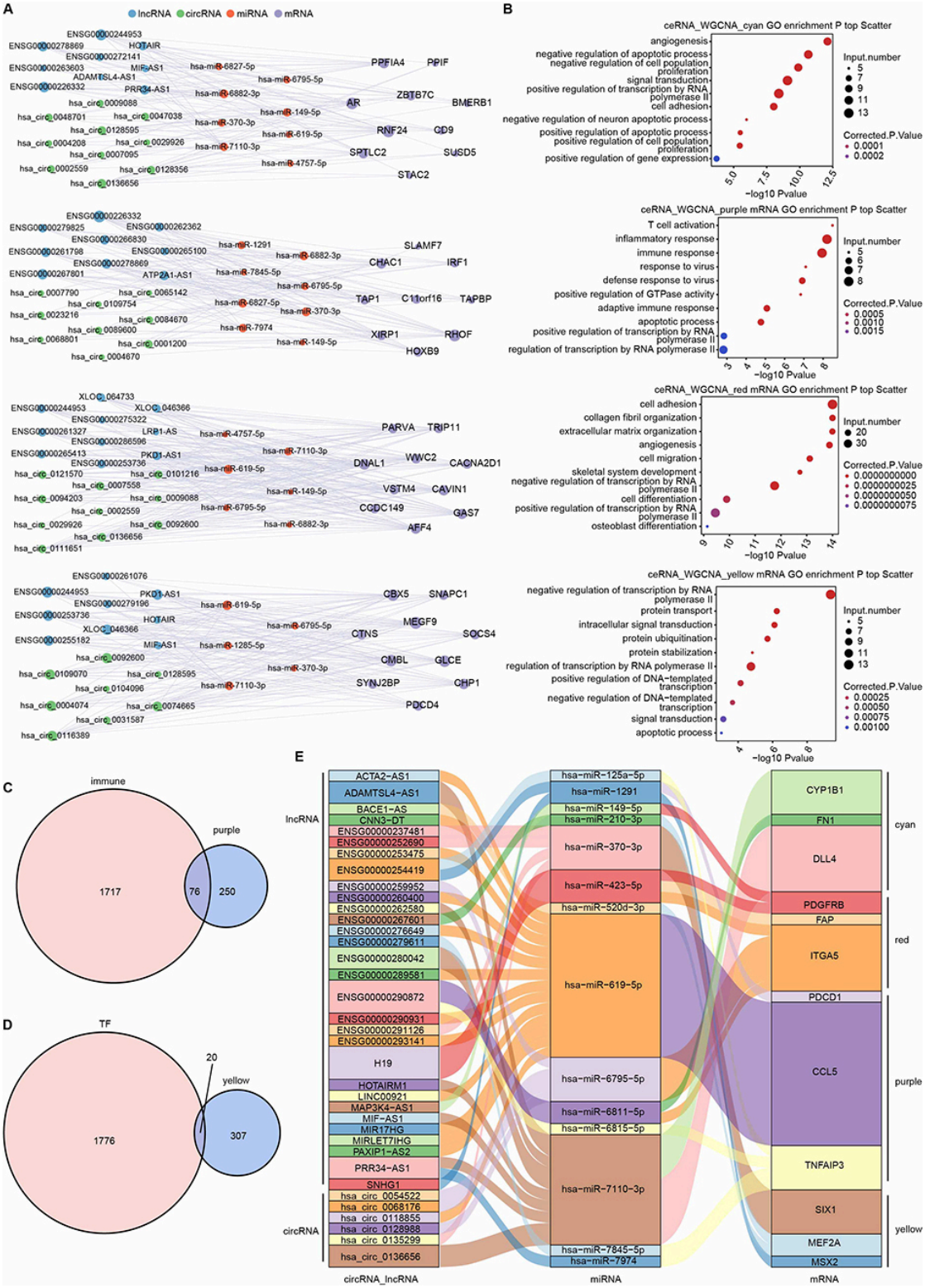


Figure 5

Construction and functional annotation of the ceRNA regulatory network.

(A) Subnetwork diagrams of circRNA/lncRNA-miRNA-mRNA ceRNA networks based on key WGCNA modules. Only the top 10 molecules with the highest connectivity in each subnetwork are shown, revealing complex regulatory relationships.

(B) Top 10 enriched GO biological process terms for genes in the four modules. The cyan/red modules are enriched in angiogenesis-related processes, the purple module in immune-related processes, and the yellow module in transcriptional regulation.

(C) Venn diagram showing the overlap between purple module genes and immune-related genes, yielding 76 immune-related candidate genes associated with chemoresistance.

(D) Venn diagram showing the overlap between yellow module genes and transcription factor-related genes, yielding 20 transcription factor-related candidate genes associated with chemoresistance.

(E) Sankey diagram illustrating the regulatory flow of angiogenesis-related (cyan/red modules), immune-related (purple module), and transcription factor-related (yellow module) genes in the ceRNA subnetwork. Only circRNAs and lncRNAs above expression thresholds were retained.

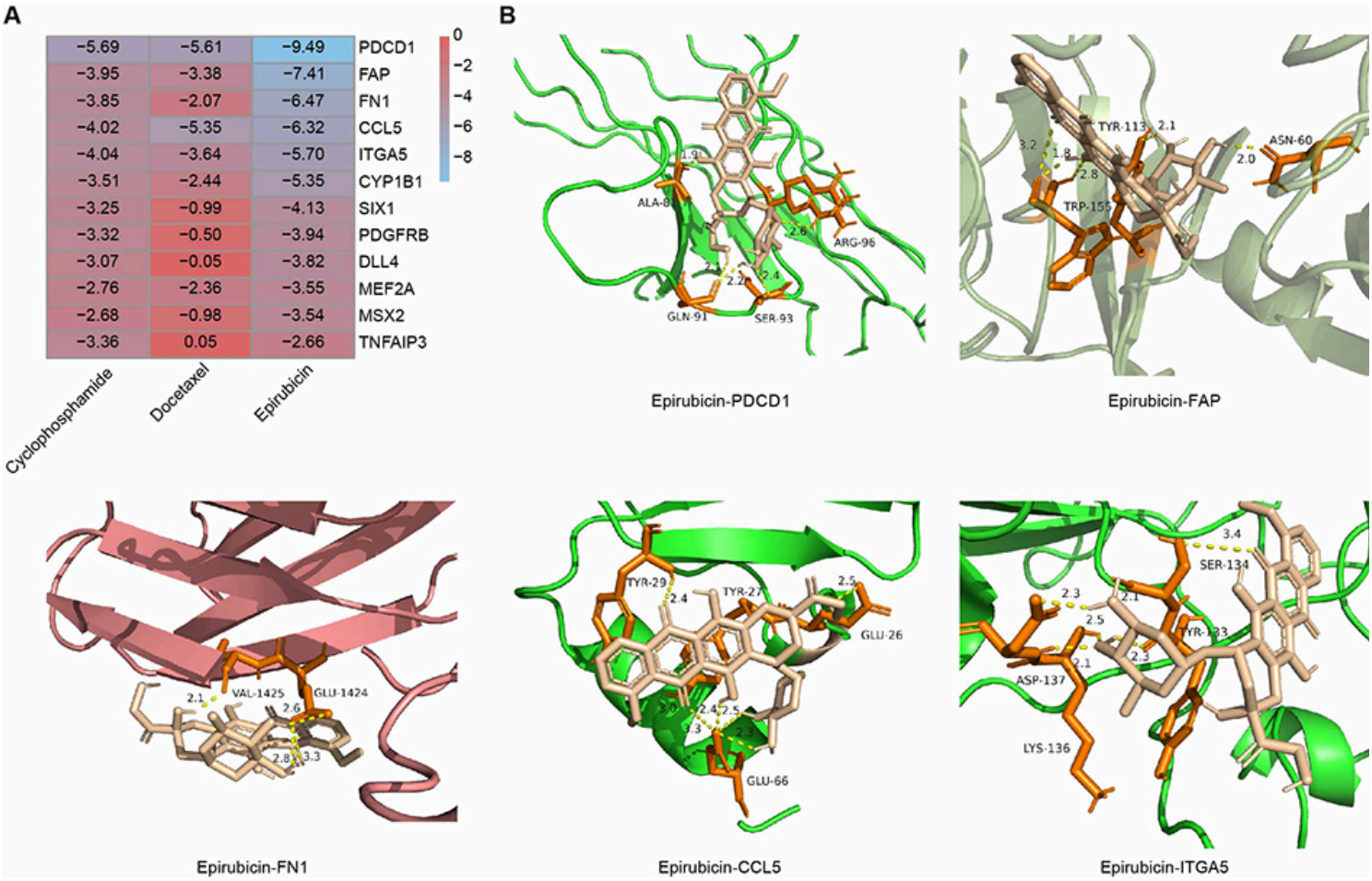


Figure 6

Molecular docking predicts drug-target interactions.

(A) Heatmap displaying Vina docking scores for epirubicin, docetaxel, and cyclophosphamide with top target proteins (*CYP1B1*, *PDCD1*, *FN1*, etc.).

(B) Detailed molecular docking diagrams illustrating interactions between epirubicin and top five target mRNAs. Hydrogen bonds are shown as dashed lines; hydrophobic contacts are indicated.