



OPEN

## Exploring ceRNA networks for key biomarkers in breast cancer subtypes and immune regulation

Alireza Shariatmadar Taleghani<sup>1</sup>, Yasaman Zohrab Beigi<sup>1</sup>, Fatemeh Zare-Mirakabad<sup>2</sup>✉ & Ali Masoudi-Nejad<sup>1</sup>✉

Breast cancer is a major global health concern, and recent researches have highlighted the critical roles of non-coding RNAs in both cancer and the immune system. The competing endogenous RNA hypothesis suggests that various types of RNA, including coding and non-coding RNAs, compete for microRNA targets, acting as molecular sponges. This study introduces the Pre\_CLM\_BCS pipeline to investigate the potential of long non-coding RNAs and circular RNAs as biomarkers in breast cancer subtypes. The pipeline identifies specific modules within each subtype that contain at least one long non-coding RNA or circular RNA exhibiting significantly distinct expression patterns when compared to other subtypes. The results reveal potential biomarker genes for each subtype, such as circ\_001845, circ\_001124, circ\_003925, circ\_000736, and circ\_003996 for the basal-like subtype, circ\_00306 and circ\_00128 for the luminal B subtype, circ\_000709 and NPHS1 for the normal-like subtype, CAMKV and circ\_001855 for the luminal A subtype, and circ\_00128 and circ\_00173 for the HER2+ subtype. Additionally, certain long non-coding RNAs and circular RNAs, including RGS5-AS1, C6orf223, HHLA3-AS1, circ\_000349, circ\_003996, circ\_003925, circ\_002665, circ\_001855, and DLEU1, are identified as potential regulators of T cell mechanisms, underscoring their importance in understanding breast cancer progression in various subtypes. This pipeline provides valuable insights into cancer and immune-related processes in breast cancer subtypes.

Breast cancer continues to hold its position as the prevailing type of cancer, constituting a significant portion of diagnoses among women globally, with approximately one in three cases being attributed to this malignancy. Statistical data from 2021 reveal that an estimated 281,550 new cases were diagnosed; accounting for 14.8% of all cancer diagnoses, and tragically, breast cancer was responsible for 43,600 deaths, representing 7.2% of the total cancer-related fatalities<sup>1</sup>. Characterized by its immunohistochemical properties, hormone receptors, and specific protein involvement, breast cancer is classified into five primary molecular subtypes: basal-like, luminal B, normal-like, luminal A, and HER2+<sup>2</sup>. These subtypes exhibit significant differences in characteristics<sup>3</sup>. Although substantial advancements have been made in clinical treatment, their efficacy could be further improved with a more precise understanding of the distinguishing features and molecular mechanisms associated with each breast cancer subtype<sup>4,5</sup>. This knowledge could significantly influence treatment outcomes and patient survival rates<sup>6</sup>.

Non-coding RNAs (ncRNAs) play pivotal roles in the regulating of gene expression<sup>7</sup>. The ncRNAs include long ncRNAs (lncRNAs), microRNAs (miRNAs), and circular RNAs (circRNAs). MiRNAs are short, single-stranded ncRNA molecules. They primarily modulate gene expression by binding to target mRNAs leading to their degradation or repression of translation<sup>8</sup>. LncRNAs are a broad class of ncRNAs exceeding 200 nucleotides in length. They regulate mRNA production by influencing transcription of protein-coding genes<sup>9</sup>. Moreover, lncRNAs may affect tumorigenesis by controlling key cancer-related genes<sup>10</sup>. CircRNAs are a relatively new class of endogenous small ncRNAs. Unlike most RNAs, they lack both a 5' cap and 3' poly-A tail, typically resulting from splicing errors<sup>11</sup>. Advances in RNA sequencing technologies and bioinformatics have accelerated research on circRNAs, highlighting their significant role in gene expression regulation, primarily by acting as miRNA sponges<sup>12</sup>. Similar to miRNAs and lncRNAs, circRNAs have been linked to various complex human diseases<sup>13,14</sup>, including several types of cancers<sup>15</sup>. Their dysregulation and potential functional roles make circRNAs promising candidates for further exploration and potential therapeutic interventions in cancer research. While the roles of miRNAs, lncRNAs, and circRNAs in cancer pathology and physiology are increasingly being elucidated, their

<sup>1</sup>Laboratory of Systems Biology and Bioinformatics (LBB), Institute of Biochemistry and Biophysics, University of Tehran, Tehran, Iran. <sup>2</sup>Department of Mathematics and Computer Science, Amirkabir University of Technology (Polytechnic Tehran), Tehran, Iran. ✉email: f.zare@aut.ac.ir; amasoudin@ut.ac.ir

interaction within competing endogenous RNAs (ceRNA) networks promises new insights into their regulatory mechanisms<sup>16</sup>. Some studies suggest that ceRNAs are pivotal to many vital biological processes and could serve as potential diagnostic markers, prognostic indicators, or therapeutic targets<sup>17,18</sup>. The regulatory influence of miRNAs goes beyond their capacity to recognize target sites on various RNA molecules<sup>19</sup>. It has been suggested that miRNAs can facilitate regulatory crosstalk among different components of the transcriptome<sup>20</sup>. This intricate miRNA-mediated regulation is modulated by other RNA molecules, including mRNAs and ncRNAs, which have been found to interact with miRNAs and exhibit significant expression across diverse biological conditions. This mechanism introduces an additional layer of post-transcriptional gene regulation, providing a complementary perspective on the functional relevance of the vast number of transcribed, yet un-translated RNAs. The concept of ceRNAs further enhances this post-transcriptional regulatory landscape, where ncRNAs gain new significance. The interplay mediated by miRNAs among different types of RNA molecules has been observed in numerous contexts<sup>21,22</sup>.

The ceRNAs hypothesis posits that a single miRNA can regulate multiple target RNAs, as long as they contain the specific miRNA response element. This hypothesis introduces the notion that RNAs can compete for a limited pool of miRNAs, serving as the foundation for constructing a ceRNA network<sup>23</sup>. Recently, the emergence of public databases has facilitated ceRNA network reconstruction, shedding light on the previously elusive mechanisms of tumorigenesis in cancers, such as breast cancer<sup>24</sup>.

Although previous studies have separately analyzed lncRNA-miRNA-mRNA<sup>25</sup> and circRNA-miRNA-mRNA<sup>26</sup> ceRNA networks, no research to date has integrated both circRNAs and lncRNAs into the ceRNA network for the basal-like, luminal B, normal-like, luminal A, and HER2+ breast cancer subtypes. Examining ceRNA networks across these subtypes, and combining circRNAs and lncRNAs, can significantly enhance our understanding of immune-related tumorigenesis in various breast cancer subtypes.

In this study, we present a novel pipeline, called Pre\_CLM\_BCS, designed to identify potential circRNAs, lncRNAs, and mRNAs as effective biomarkers for each breast cancer subtype. Our pipeline initiates by constructing a primary ceRNA network for each cancer subtype using gene expression data from patients. Subsequently, specific ceRNA networks are extracted for each breast cancer subtype from their respective primary ceRNA networks. Within these specific networks, modules containing at least one lncRNA or circRNA are identified and enriched with gene ontology terms and pathways. Finally, we employ survival analysis and differential gene expression to predict biomarkers from the enriched genes in the modules, which are further evaluated by support vector machine.

The predicted biomarkers for each breast cancer subtype according to the pipelines are as follows:

- For the basal-like subtype: circ\_001845, circ\_001124, circ\_003925, circ\_000736, and circ\_003996.
- For the luminal B subtype: circ\_00306 and circ\_00128.
- For the normal-like subtype: circ\_000709 and nephrotic syndrome 1 (NPHS1 as an mRNA).
- For the luminal A subtype: cam kinase like vesicle associated (CAMKV as an mRNA) and circ\_001855.
- For the HER2+ subtype: circ\_00128 and circ\_00173.

Additionally, we find the following RNAs in modules that have regulatory effects on T cell mechanisms, further emphasizing the significance of these RNAs in understanding and addressing breast cancer progression across various subtypes:

- For the basal-like subtype: circ\_003996, and circ\_003925.
- For the luminal B subtype: chromosome 6 open reading frame 223 (C6orf223 as an lncRNA), HHLA3 Anti-sense RNA 1 (HHLA3-AS1 as an lncRNA), and circ\_000349.
- For the normal-like subtype: deleted in lymphocytic leukemia 1 (DLEU1 as an lncRNA).
- For the luminal A subtype: circ\_002665 and circ\_001855.
- For the HER2+ subtype: regulator of G protein signaling 5 antisense RNA 1 (RGS5-AS1 as an lncRNA).

## Materials and methods

In this section, our main goal is to introduce the proposed pipeline, Pre\_CLM\_BCS, for predicting suitable circRNAs, lncRNAs, and mRNAs as biomarkers for five breast cancer subtypes: basal-like, luminal B, normal-like, luminal A, and HER2+. Figure 1 visually outlines our pipeline, and the following subsections offer in-depth explanations for each step of the process.

### Dataset

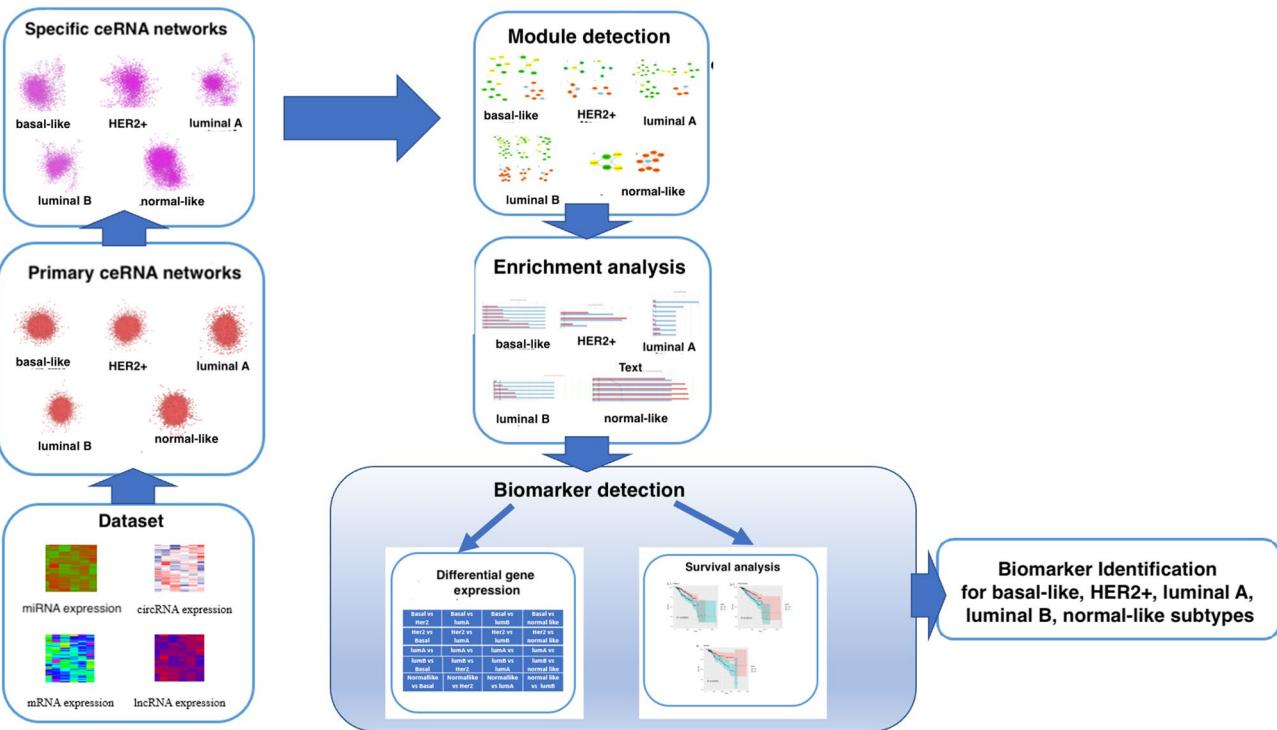
Breast cancer is traditionally classified into five subtypes as follows:

$$BCS = \{ \text{basal-like, luminal B, normal-like, luminal A, HER2+} \}.$$

In this study, we extract miRNA, mRNA, and lncRNA gene expression data, along with clinical data for breast cancer samples from The Cancer Genome Atlas (TCGA) database (as of August 2015).

We utilize the miRNA-seq dataset for miRNA gene expression and the RNA-SeqV2 level 3 dataset for mRNA and lncRNA expression profiles. To annotate the mRNA and lncRNA genes from the RNA-SeqV2 level 3 dataset, we use the BioMart tool available on the Ensembl website ([www.ensembl.org](http://www.ensembl.org)) as a genome browser<sup>27</sup>. Concurrently, circRNA expression data for breast cancer subtypes is sourced based on the findings of Asha et al.<sup>28</sup>.

Additionally, we collect clinical data, including immunohistochemical information, for breast cancer samples from TCGA database, in order to conduct survival analysis.



**Figure 1.** The Pre\_CLM\_BCS pipeline.

Considering the conventional categorization of breast cancer into five subtypes, our study samples are categorized accordingly. This results in the creation of five distinct datasets for mRNA, lncRNA, circRNA, and miRNA. For each cancer subtype, we establish a dataset denoted as  $\Delta^C$ , where  $C \in BCS$ .

Each dataset encapsulates the gene expression of mRNAs, lncRNAs, circRNAs, and miRNAs for samples within the respective breast cancer subtypes. The data preprocessing for each dataset  $\Delta^C$  comprises three main stages:

1. Elimination of genes with a mean expression value below 0.5.
2. Exclusion of genes with uniform expression across all five breast cancer subtypes, indicating low variance.
3. Detection of outliers. To identify outliers, our pipeline utilizes hierarchical clustering of the samples based on their expression profiles within the R environment (version 3.6)<sup>29</sup>.

### Primary ceRNA networks

In the second step of the Pre\_CLM\_BCS pipeline, we construct a primary ceRNA network named  $P^C = (N^C, E^C)$  for each breast cancer subtype  $C \in BCS$  using gene expression data from dataset  $\Delta^C$ . Here,  $N^C$  represents the collection of mRNAs, lncRNAs, and circRNAs in the dataset as nodes. The set of edges (interactions) in this network is defined as  $E^C$ , where each edge between two RNAs from set  $N^C$  indicates their interaction<sup>30</sup>. To identify these interactions, we utilize the Pearson correlation coefficient (PCC) for the gene expression dataset  $\Delta^C$ .

Our approach for constructing a primary ceRNA network is rooted in the ceRNA regulatory model. According to this model, lncRNAs and circRNAs compete with miRNAs, resulting in opposing effects on mRNA expression<sup>31,32</sup>. Therefore, when the expression of a miRNA increases, there is a corresponding decrease in the levels of lncRNAs, mRNAs, and circRNAs. This competition weakens the positive correlation among miRNAs, lncRNAs, and mRNAs<sup>33,34</sup>.

Consequently, a simultaneous increase in mRNA expression along with either lncRNA or circRNA implies the potential downregulation of a specific miRNA. To explore interactions as edges in the primary ceRNA network  $P^C$  among lncRNAs and mRNAs, we utilize the miRTarBase and TarBase databases to access experimentally verified mRNA–miRNA interactions<sup>35,36</sup>. We also extract experimentally verified mRNA–miRNA interactions from the miRTarBase database.

To predict interactions as edges among the union of mRNAs and lncRNAs in the set  $N^C$ , following the ceRNA regulatory model, we compute the PCC on the gene expression data of the RNAs, extracting their interactions from the miRTarBase and TarBase databases. In other words, assuming there is a documented interaction between miRNA A and both mRNA B and lncRNA C in the databases, we predict an edge between mRNA B and lncRNA C if the gene expression data shows a negative PCC between miRNA A and mRNA B, as well as between miRNA A and lncRNA C. Simultaneously, there should be a positive PCC between the gene expression

data of mRNA B and lncRNA C. Therefore, in cases where lncRNAs and miRNAs share a common miRNA, as indicated in the mentioned databases, we forecast potential interaction between them.

Due to the absence of experimental interaction data between circRNAs and miRNAs, we predict interactions among circRNAs or between circRNAs and mRNAs solely based on the PCC. In other words, to infer the interactions between circRNAs and mRNAs, we calculate the PCC for each pairing of circRNAs with miRNAs, mRNAs with miRNAs, and circRNAs with mRNAs. We predict an interaction between a circRNA and an mRNA when there is a positive PCC between their gene expression data, as well as a negative PCC between the gene expression data of the circRNA and a miRNA, as well as between the mRNA and that miRNA.

For both the scenarios: negative PCC values between microRNAs with lncRNAs, mRNAs, and circRNAs, and positive PCC values among lncRNAs, mRNAs, and circRNAs pairings, a stringent threshold of an absolute PCC surpassing 0.4 and a *p* value beneath 0.05 is instated for statistical robustness. Meanwhile, we remove RNAs from the set  $N^C$  if they lack edges connecting them to other RNAs.

In culmination, the visual representation of the primary ceRNA network  $P^C$  for each  $C \in \text{BCS}$  can be achieved via Cytoscape<sup>37</sup>.

### Specific ceRNA networks

In the third step of pipeline, we aim to establish a specific ceRNA network for each breast cancer subtype to delve into the distinctive regulatory mechanisms inherent to each subtype. To achieve this, we first extract a common ceRNA network from the primary ceRNA networks. Subsequently, we derive the specific ceRNA network for each subtype,  $C$ , by subtracting the common ceRNA network from the primary ceRNA network  $P^C = (N^C, E^C)$ .

To achieve this, we start by constructing the common ceRNA network called  $c\text{CENT} = (N^{c\text{CENT}}, E^{c\text{CENT}})$  by amalgamating the primary ceRNA networks,  $P^{\text{basal-like}}$ ,  $P^{\text{luminal B}}$ ,  $P^{\text{normal-like}}$ ,  $P^{\text{luminal A}}$  and  $P^{\text{HER2+}}$  where

$$N^{c\text{CENT}} = N^{\text{basal-like}} \cap N^{\text{luminal B}} \cap N^{\text{normal-like}} \cap N^{\text{luminal A}} \cap N^{\text{HER2+}},$$

$$E^{c\text{CENT}} = E^{\text{basal-like}} \cap E^{\text{luminal B}} \cap E^{\text{normal-like}} \cap E^{\text{luminal A}} \cap E^{\text{HER2+}}.$$

Then, a specific ceRNA network named  $s\text{CENT}^C = (N_{s\text{CENT}}^C, E_{s\text{CENT}}^C)$  is defined for each breast cancer subtype  $C \in \text{BCS}$  as bellow:

$$N_{s\text{CENT}}^C = N^C - N^{c\text{CENT}},$$

$$E_{s\text{CENT}}^C = E^C - E^{c\text{CENT}},$$

where  $N^C$  and  $E^C$  represent the number nodes (RNAs) and edges (interactions) among RNAs in the corresponding primary ceRNA network of breast cancer subtype  $C$ , respectively.

### Module detection

In the fourth step of the Pre\_CLM\_BCS pipeline, we employ ClusterMaker Cytoscape application<sup>38</sup> to detect modules within the specific ceRNA networks. This application leverages the Markov cluster algorithm<sup>39</sup>. For each breast cancer subtype  $C \in \text{BCS}$ , the specific ceRNA network  $s\text{CENT}^C$  is input into the application to identify modules. We select a collection of modules named  $LC - MODUL^C$  that contains one lncRNA or circRNA for further analysis in breast cancer subtype  $C$ .

### Enrichment analysis

In the fifth step of the pipeline, potential Gene Ontology (GO) terms of the modules in the set  $LC - MODUL^C$  of each breast cancer subtype,  $C \in \text{BCS}$ , are determined through GO enrichment analysis conducted using the ToppFun web tool. GO terms with FDR-corrected *p* values less than 0.05 are considered significant<sup>40</sup>.

### Biomarker detection

In the sixth step of our pipeline, we apply two methodologies to extract biomarkers from the set  $LC - Module^C$ , focusing on breast cancer subtype  $C$ . In the first methodology, we utilize survival analysis to identify effective mRNAs in breast cancer subtype  $C$ . In the second one, we employ differential gene expression data to predict biomarkers by comparing breast cancer subtype  $C$  to the other subtypes separately.

### Survival analysis

For each breast cancer subtype, denoted as  $C \in \text{BCS}$ , we calculate the union of mRNA genes from the modules listed in  $LC - Module^C$ . Subsequently, we explore the relationship between the expression of these mRNA genes and patients' overall survival by employing the Kaplan–Meier (KM) estimation and Log-rank test. TCGA survival data matching the mRNA genes are retrieved from TCGA database for survival analysis. A total of 88, 100, 76, 252, and 42 samples are collected for basal-like, luminal B, normal-like, luminal A, and HER2+ subtypes, respectively. For each gene's survival analysis, patients are divided into two groups based on their expression values. Those with values exceeding the mean are categorized as the high-expression group, while those with values below the mean are classified as the low-expression group. We compute Kaplan–Meier curves and Log-rank test statistical results for both groups using the Survminer<sup>41</sup> and survival<sup>42</sup> packages in R.

### Differential gene expression analysis

For the analysis of differentially expressed genes (DEG), we create the set of genes named  $G^C$  from  $LC\_Module^C$  for each breast cancer subtype  $C \in BCS$ . Then, we conduct DEG analysis for each pair of subtypes  $C, C' \in BCS$ , where  $C \neq C'$ , as the between-subtype differentially expressed genes, denoted as  $BSDEG^{C,C'}$ . This analysis is performed using the edgeR package on the gene expression data of the genes in  $G^C$ <sup>43</sup>. Each gene with adjusted  $p$  values less than 0.05 (Benjamini–Hochberg) and an absolute logFC value greater than 0.5 is included in the set  $BSDEG^{C,C'}$ . Subsequently, we extract the set of biomarkers ( $Bio\_DEG^C$ ) for breast subtype  $C$ , as follows:

$$Bio\_DEG^C = \bigcap_{\substack{C' \in BCS \\ C' \neq C}} BSDEG^{C,C'}$$

To validate the predicted biomarkers in  $Bio\_DEG^C$  for each cancer subtype  $C$ , we employ a support vector machine (SVM) classifier, and implement fivefold cross-validation. The validation procedures using the SVM classifier and fivefold cross-validation are performed individually for each subtype against all other subtypes. This process is repeated for each of the five subtypes.

After running SVM on the breast cancer subtype  $C \in BCS$ , we calculate the receiver operating characteristic (ROC) curve and the area under the curve (AUC) to evaluate the efficiency of the biomarkers,  $Bio\_DEG^C$ , in distinguishing the breast cancer subtype from the others. The caret<sup>44</sup>, plotROC<sup>45</sup>, MLmetrics<sup>46</sup>, and ggplot<sup>47</sup> R packages are utilized to implement the SVM algorithm and to calculate and illustrate the ROC curves.

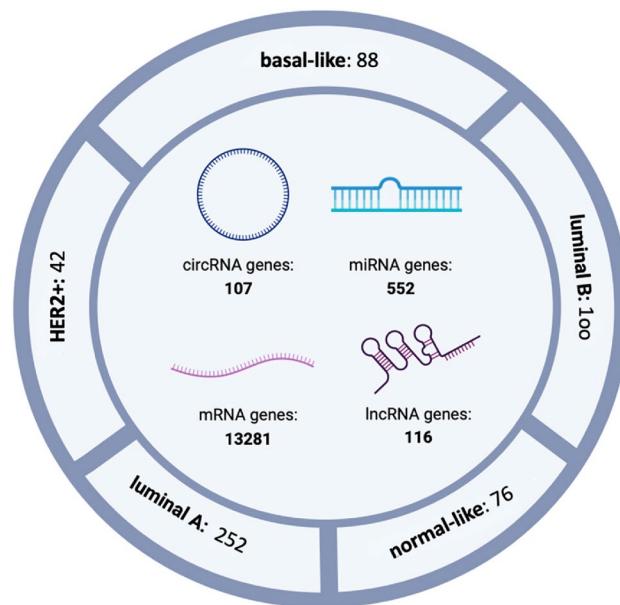
## Results

In this section, we offer a step-by-step representation of the results obtained from executing the pipeline outlined in Fig. 1 for biomarker detection to distinguish breast cancer subtypes.

### Dataset

Figure 2 illustrates the outcomes of the initial step in the Pre\_CLM\_BCS pipeline. During this phase, we extract gene expression of 30,480 genes, encompassing mRNA, miRNA, circRNA, and lncRNA. After the completion of the first and second data preprocessing stages within this step, we narrow down the initial gene pool to 14,056 genes. The figure also breaks down the retained genes by their respective types.

Furthermore, after the third data preprocessing step, we remove 105 out of 193 samples for the basal-like subtype, 119 out of 219 for luminal B, 67 out of 143 for the normal-like, 329 out of 581 for luminal A and 40 out of 82 for HER2+ subtypes of breast cancer as outlier data. Finally, for each subtype  $C \in BCS$ , we have a dataset as  $\Delta^C$  which includes gene expression data of 14,056 genes and  $n$  samples. The value of  $n$  varies depending on the cancer type. Specifically, for basal-like, luminal B, normal-like, luminal A, and HER2+ subtypes, we found 88, 100, 76, 252, and 42 samples, respectively. Supplementary Figure S1 exhibits dendograms illustrating the hierarchical clustering of samples conducted for outlier detection.

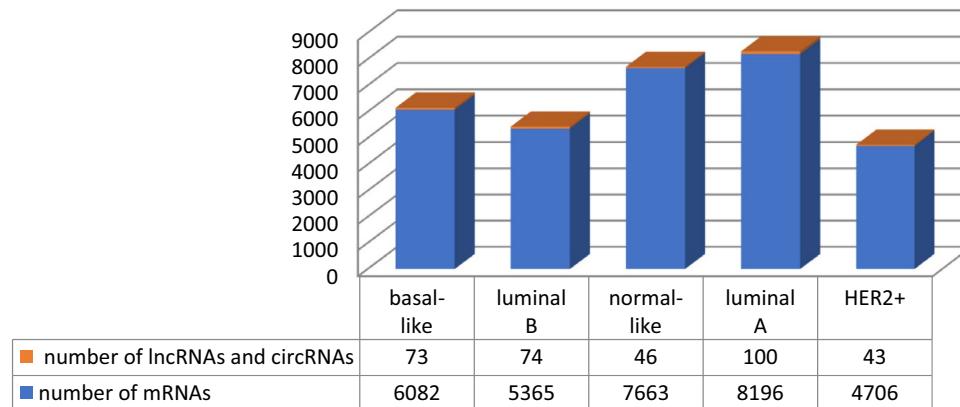


**Figure 2.** The number of genes and samples for breast cancer subtypes after preprocessing: basal-like, luminal B, normal-like, luminal A, and HER2+. Note: The picture was created using biorender software (biorender.com).

<b>C</b>	<b>Basal-like</b>	<b>Luminal B</b>	<b>Normal-like</b>	<b>Luminal A</b>	<b>HER2+</b>
$ N^C $	6155	5439	7709	8296	4749
$ E^C $	368,773	261,344	724,686	676,146	169,644

**Table 1.** The number of RNAs ( $|N^C|$ ) and the number of edges among RNAs ( $|E^C|$ ) for each subtype  $C \in BCS$  in the primary ceRNA network  $P^C = (N^C, E^C)$ .

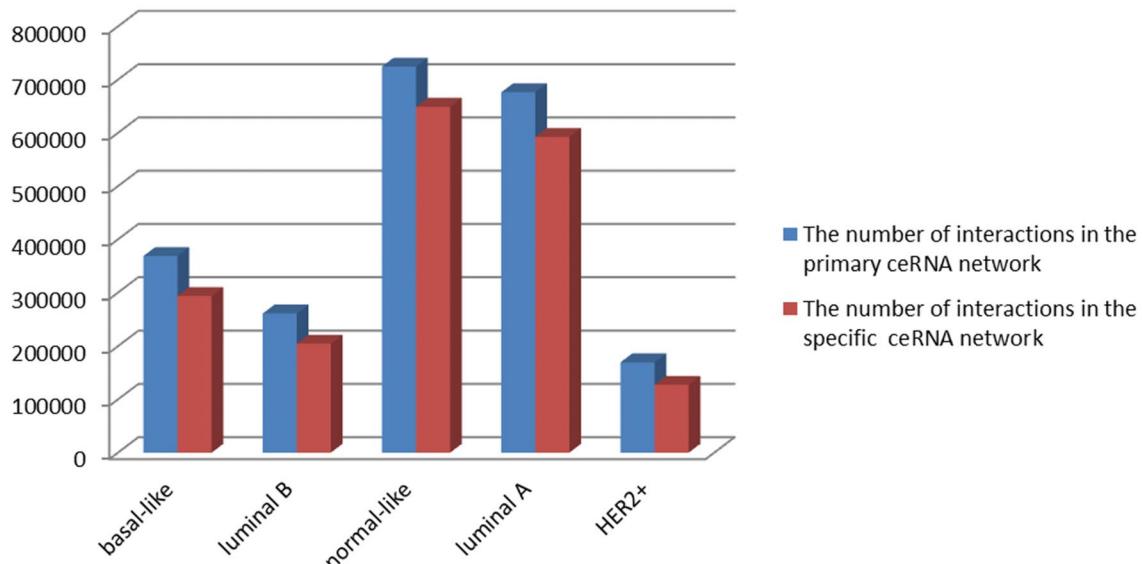
## RNA type distribution in the primary ceRNA network of each breast cancer subtype



**Figure 3.** RNA type distribution across breast cancer subtypes.

### Primary ceRNA networks

We define  $P^C = (N^C, E^C)$  to represent each subtype  $C \in BCS$  as the primary ceRNA network using the dataset  $\Delta^C$ . Table 1 shows the number of RNAs ( $|N^C|$ ) and the number of edges among RNAs ( $|E^C|$ ) for each subtype  $C \in BCS$ . Figure 3 illustrates the presence of mRNAs, lncRNAs and circRNAs in the primary ceRNA of each breast cancer subtype.



**Figure 4.** Comparison of the number of edges (interactions) in the specific and primary ceRNA networks for each breast cancer subtype.

### Specific ceRNA networks

For each breast cancer subtype  $C \in BCS$ , the specific ceRNA network named  $sCENT^C = (N_{sCENT}^C, E_{sCENT}^C)$  is derived by removing the common ceRNA network ( $cCENT$ ) from the primary ceRNA network  $P^C$ . After removing the  $cCENT$  network from the primary ceRNA network, we lose 8, 11, 2, 3 and 13 RNAs from the corresponding specific ceRNA networks for basal-like, luminal B, normal-like, luminal A and HER2+, respectively. Figure 4 compares the number of edges in  $sCENT^C$  ( $|E_{sCENT}^C|$ ) with the edges in the primary ceRNA network  $P^C$  for each breast cancer subtype  $C \in BCS$ .

### Module detection

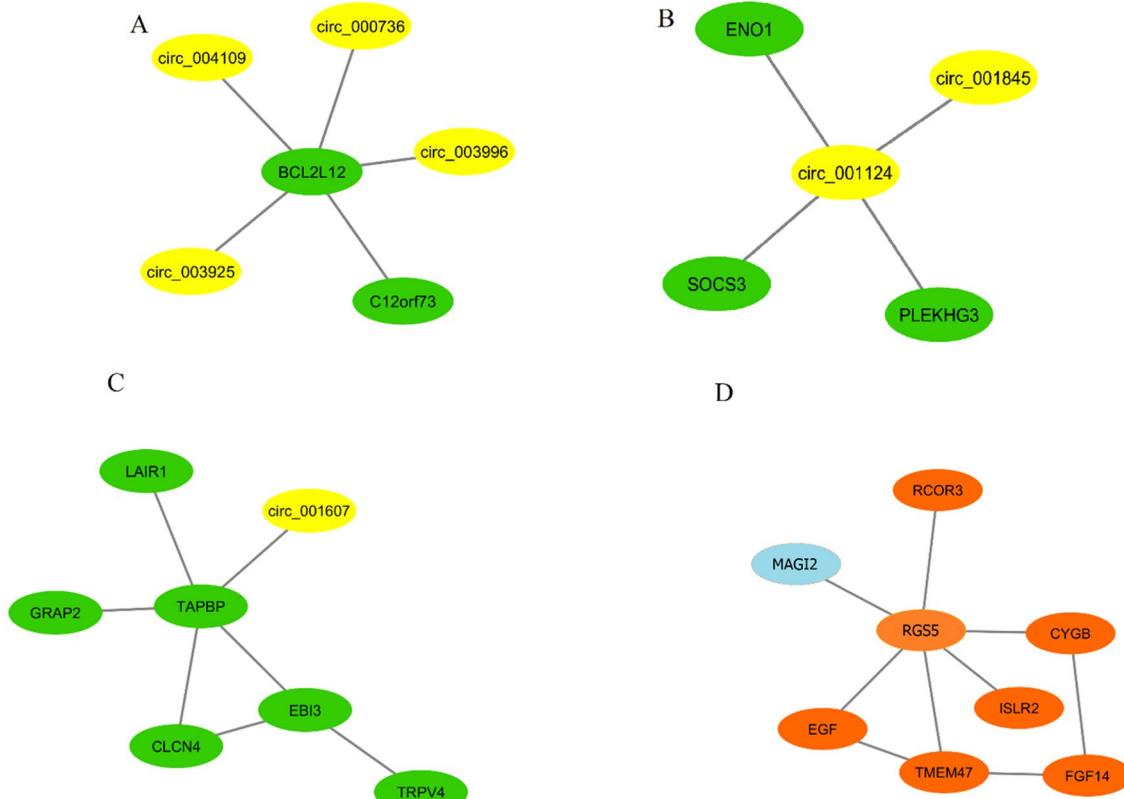
We select a set of modules as  $LC - MODUL^C$  including at least one lncRNA or circRNA. Figures 5, 6, 7, 8, and 9 depict these modules for basal-like, luminal B, normal-like, luminal A, and HER2+ subtypes, respectively. Notably, it is observed that in specific modules, such as module A in the basal-like subtype, module B in the luminal A subtype, and module A in the normal-like subtype, multiple circRNAs interacting with a single mRNA gene. This finding indicates that, despite lower expression levels of circRNAs compared to other RNA molecules, circRNAs play crucial roles in regulating specific mechanisms in different breast cancer subtypes<sup>48</sup>.

### Enrichment analysis

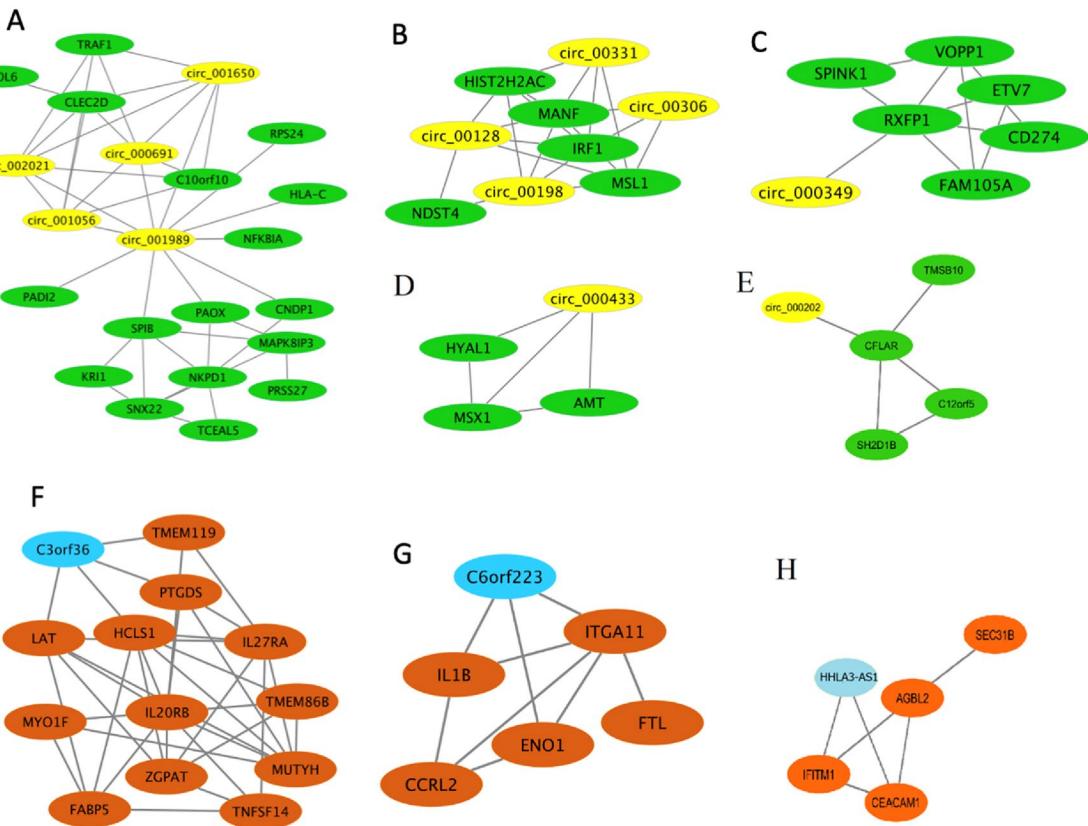
Enriched GO terms and pathways associated with the modules in  $LC - MODUL^C$ , for each subtype of cancer  $C$  are analyzed to identify their involvement in specific pathways and processes. The significant processes enriched by these modules in each breast cancer subtype can be categorized into three major groups: cancer-related processes, immune-related processes, and oncogenic signaling pathways in cancer.

Table 2 displays the cancer-related processes and oncogenic signaling pathways in cancer derived from the modules in  $LC - MODUL^C$  for each breast cancer subtype  $C$ , while Table 3 illustrates immune-related processes derived from modules for each breast cancer subtype  $C$ .

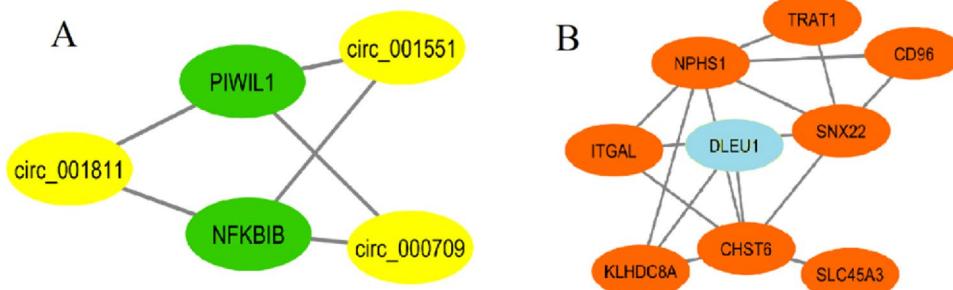
Notably, the enrichment analysis indicates that the immune response primarily involves T cell and B cell activation processes separately. Figure 10 shows the corresponding modules for T cell and B cell activation in each of the five subtypes. It is noteworthy that all five subtypes possess a distinct module for activating T cells, while only luminal B and normal-like subtypes have specific modules related to B cell activation, emphasizing the importance of T cells over B cells in breast cancer subtypes.



**Figure 5.** The basal-like subtype includes three distinct circRNA modules (A–C) and one lncRNA module (D). In A, B and C, the circRNAs and mRNAs are shown by yellow and green, respectively. In D, the lncRNA and mRNAs are shown by blue and brown, respectively.



**Figure 6.** The luminal B subtype includes five distinct circRNA modules (A–E) and three lncRNA modules (F–H). In (A–E), the circRNAs and mRNAs are shown by yellow and green, respectively. In (F–H), the lncRNAs and mRNAs are shown by blue and brown, respectively.



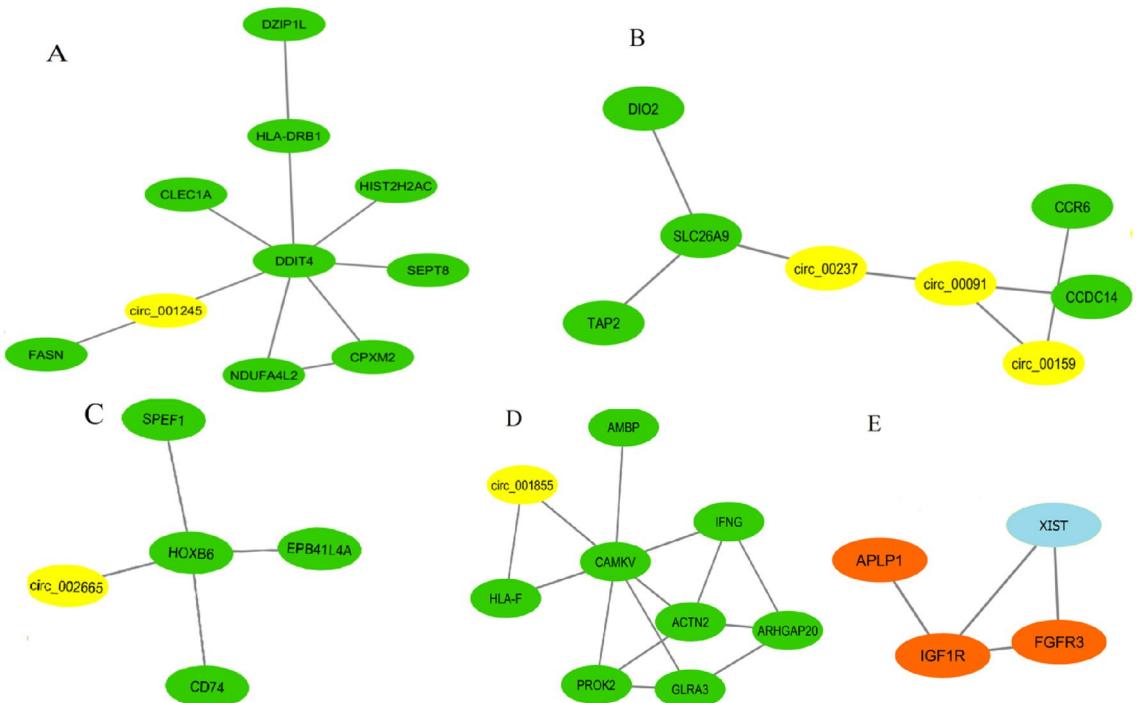
**Figure 7.** The normal-like subtype includes one circRNA module (A) and one lncRNA module (B). In (A), the circRNAs and mRNAs are shown by yellow and green, respectively. In (B), the lncRNAs and mRNAs are shown by blue and brown, respectively.

### Biomarker detection

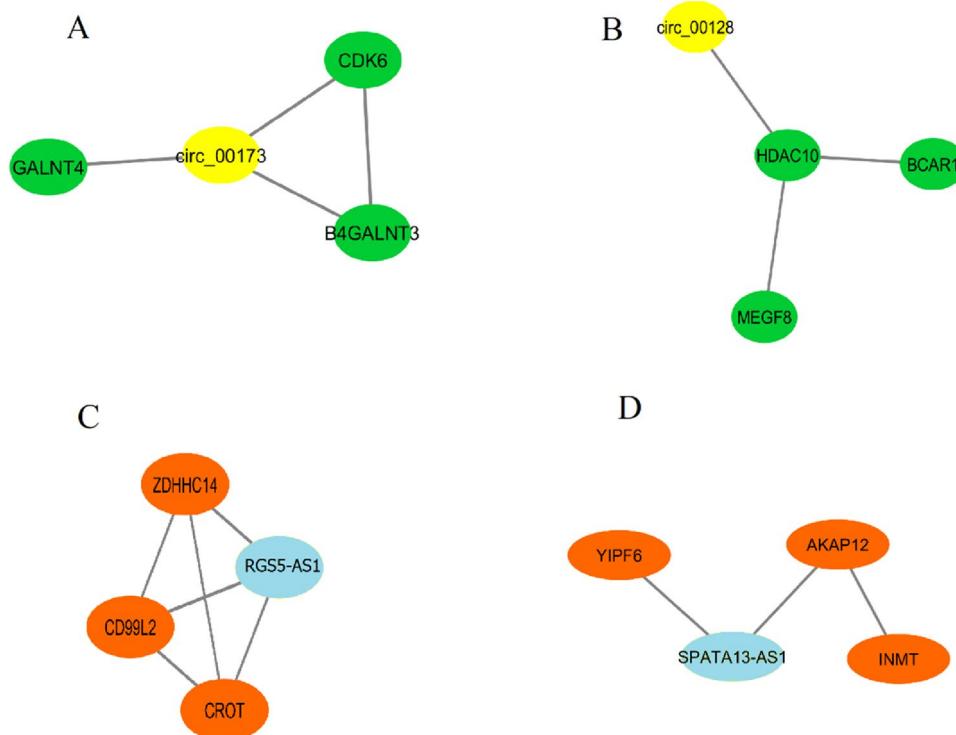
Through module enrichment, we establish the reliability of these modules for identifying biomarkers. In the following sections, we employ survival analysis and differential gene expression to discover biomarkers within these modules.

#### Survival analysis

We identify potential biomarkers from the mRNAs within the modules of  $LC - MODUL^C$  for each subtype  $C$  using survival analysis. Among the identified genes, three mRNAs, AGBL2, HIST2HAC, and NPHS1, show a significant association with the overall survival of breast cancer patients, with Log-rank test  $p$  values of  $< 0.05$ . These results suggest that AGBL2 and HIST2HAC are significant prognostic biomarkers for luminal B subtype patients, while NPHS1 holds prognostic value for normal-like subtype patients. Independent KM survival curves for AGBL2, HIST2HAC, and NPHS1 genes are depicted in Fig. 11. In the subsequent sub-section, the analysis



**Figure 8.** The luminal A subtype includes four distinct circRNA modules (A–D) and one lncRNA module (E). In (A–D), the circRNAs and mRNAs are shown by yellow and green, respectively. In (E), the lncRNA and mRNAs are shown by blue and brown, respectively.



**Figure 9.** The HER2+ subtype includes two distinct circRNA modules (A,B) and two lncRNA modules (C,D). In (A,B), the circRNAs and mRNAs are shown by yellow and green, respectively. In (C,D), the lncRNAs and mRNAs are shown by blue and brown, respectively.

Subtype C	Module's name	Type of RNAs in module	Go terms or pathways	FDR
Basal-like	D (Fig. 5)	lncRNA, mRNA	Breast cancer	0.03041
Luminal A	E (Fig. 8)	lncRNA, mRNA	Progesterone-mediated oocyte maturation	0.0497
Basal-like	A (Fig. 5)	circRNA, mRNA	Mammary gland development pathway—Involution (Stage 4 of 4)	0.01572
Luminal B	C (Fig. 6)	circRNA, mRNA	GO:0060658, nipple morphogenesis	0.01182

**Table 2.** Top pathways and GO enrichment terms of modules  $LC - MODUL^C$  related to breast cancer subtype C.

Subtype C	Module's Name	Type of RNAs in module	Go terms or Pathways	FDR
HER2+	C (Fig. 9)	lncRNA, mRNA	GO:2000409, positive regulation of T cell extravasation	0.03519
Luminal B	F (Fig. 6)	lncRNA, mRNA	B cell receptor signaling pathway	0.03837
Luminal B	G (Fig. 6)	lncRNA, mRNA	GO:0033083, regulation of immature T cell proliferation	0.04753
Luminal B	H (Fig. 6)	lncRNA, mRNA	GO:0042110, T cell activation	0.01656
Normal-like	B (Fig. 7)	lncRNA, mRNA	Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	0.04297
Normal-like	B (Fig. 7)	lncRNA, mRNA	Adaptive Immune System	0.04297
Normal-like	B (Fig. 7)	lncRNA, mRNA	T Helper Cell Surface Molecules	0.04297
Basal-like	A (Fig. 5)	circRNA, mRNA	T Helper-2 activation	0.03637
Luminal A	A (Fig. 8)	circRNA, mRNA	GO:0002819, regulation of adaptive immune response	0.03149
Luminal A	C (Fig. 8)	circRNA, mRNA	GO:0002286, T cell activation involved in immune response	0.02239
Luminal A	D (Fig. 8)	circRNA, mRNA	T cell activation	0.0419
Luminal B	B (Fig. 6)	circRNA, mRNA	GO:0002765, immune response-inhibiting signal transduction	0.04507
Luminal B	C (Fig. 6)	circRNA, mRNA	GO:1905399, regulation of activated CD4-positive, alpha-beta T cell apoptotic process	0.009291
Luminal B	E (Fig. 6)	circRNA, mRNA	CD40/CD40L signaling	0.07089
Normal-like	A (Fig. 7)	circRNA, mRNA	B cell receptor signaling pathway	0.02161

**Table 3.** Top pathways and GO enrichment terms of modules  $LC - MODUL^C$  related to the immune system in breast cancer subtype C.

of differentially expressed genes further confirms the potential of NPHS1 as a biomarker. The Log-rank test  $p$  value for all genes in the  $LC - MODUL^C$ , for each subtype  $C \in BCS$  can be found in Supplementary File S2.

#### Differential gene expression analysis

The common significantly differentially expressed genes between subtype C and the other subtypes are documented in the set  $Bio\_DEG^C$  as the potential genes for biomarkers. The genes expressed differently across the five subtypes are presented in Supplementary Files S3, S4, S5, S6, and S7. Table 4 displays the gene names for each subtype C in the  $Bio\_DEG^C$  set, along with the average values of absolute logFC and FDR for differentially expressed genes between subtype C and the other subtypes. Additionally, standard derivations of logFC and FDR of differentially expressed genes between subtype C and the other subtypes are provided in this table.

It is noteworthy that each of the five subtypes includes at least one circRNA gene considered as a biomarker. The negative values of logFC values shows all selected circRNA genes are down-regulated in the pairwise differential expression analysis between the five subtypes.

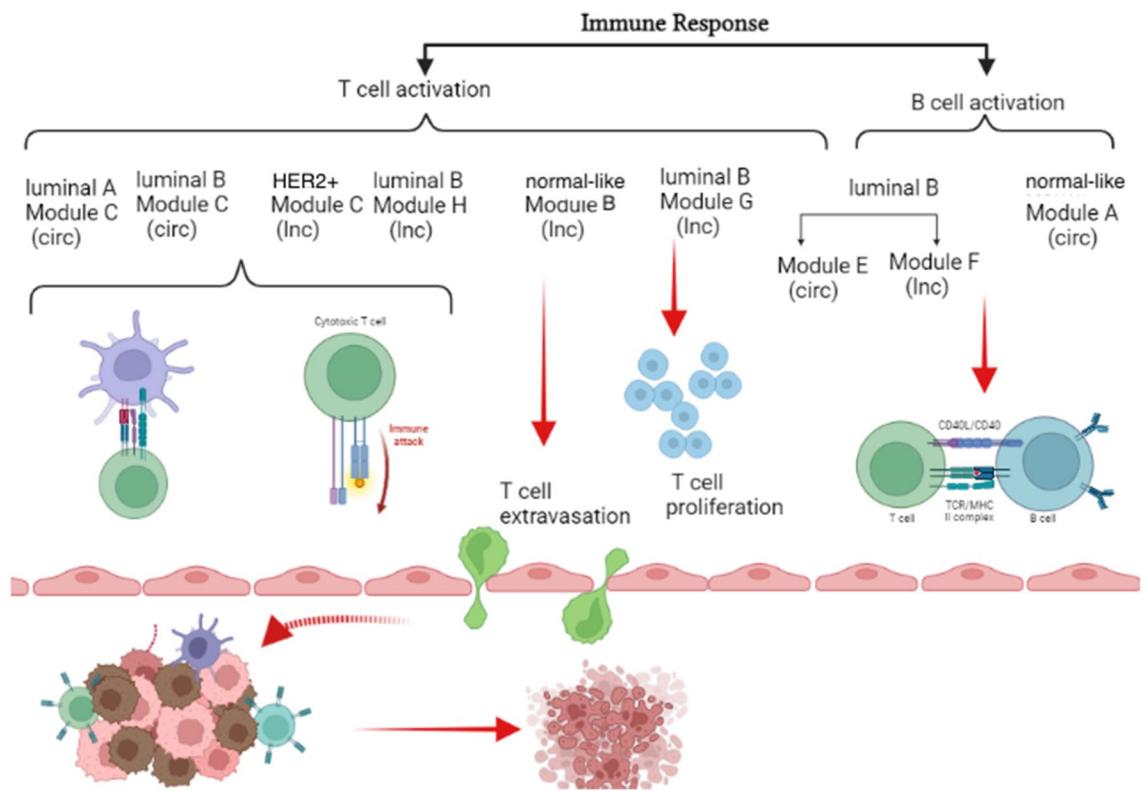
In the following, to evaluate the selected genes in Table 4 as effective biomarkers for each breast cancer subtype, we design a predictor for each subtype using selected genes as described in the last step of the pipeline. Figure 12 presents the corresponding ROC curves and AUC values for each subtype.

The validation results of the model on each breast cancer subtype demonstrate satisfactory performance. The average AUC values of fivefold for basal-like, HER2+, luminal A, luminal B, and normal-like are obtained 0.93, 0.92, 0.70, 0.82, and 0.86, respectively. Overall, these results indicate that the selected genes in the set  $Bio\_DEG^C$  for each subtype C serve as robust biomarker sets for distinguishing between breast cancer subtypes.

Among the biomarkers predicted in this stage, circRNAs and lncRNAs emerge as promising candidates for breast cancer subtype prediction.

## Discussion

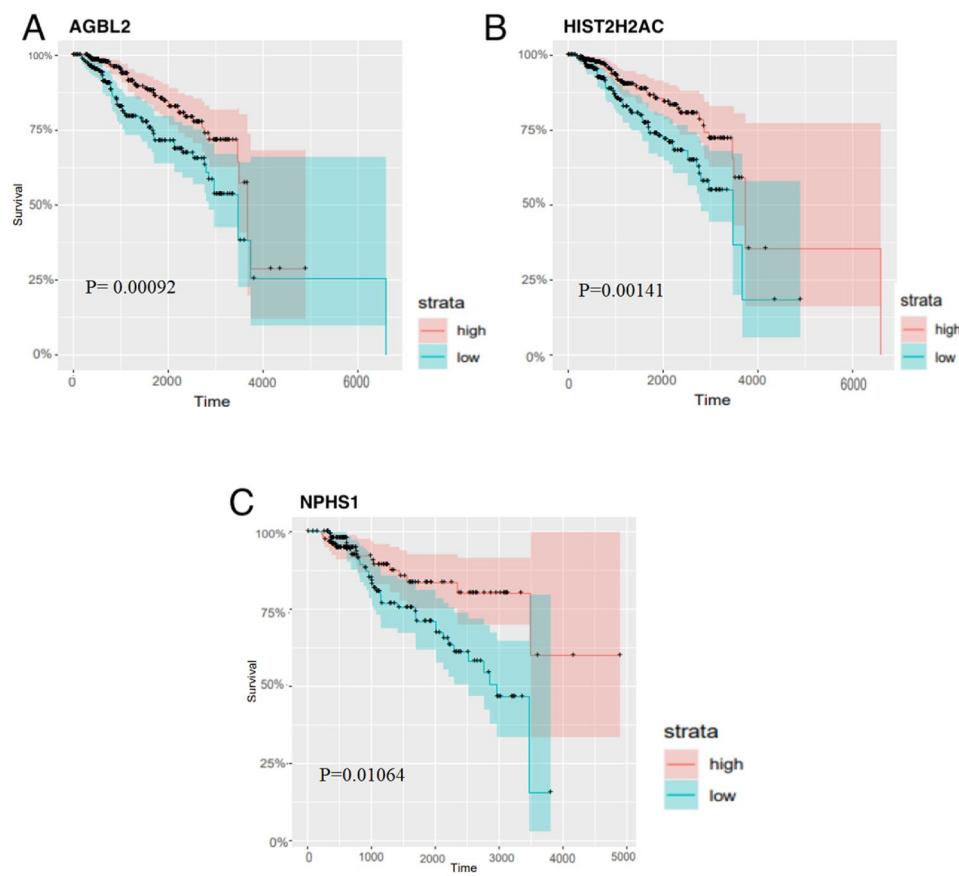
In the realm of immuno-oncology, a field at the intersection of immunology and cancer research, the development of potent immunotherapy treatments holds promise for various cancer types<sup>49</sup>. Understanding the immune system's intricacies is paramount to advancing these treatments. Breast cancer research has underscored the pivotal role of T cells in tumor progression, across various subtypes<sup>50–52</sup>. This section investigates T cell regulation



**Figure 10.** T cell and B cell activation processes in breast cancer subtypes. This figure represents the processes in which T cells and B cells are activated by the detected modules in breast cancer subtypes, with “lnc” and “circ” standing for lncRNA and circRNA, respectively. Note: The picture was created using biorender software (biorender.com).

and activation mechanisms within these subtypes, revealing at least one module in each subtype enriched in these processes, as follows:

- In the HER2+ subtype, module C in Fig. 9 comprises ZDHHC14 and CROT mRNA genes, along with the lncRNA RGS5-AS1. This module is associated with the positive activation of T cells, which is influenced by ZDHHC14 and CROT mRNA genes. These genes are subject to regulation by the lncRNA RGS5-AS1 through hsa-miR-20a-5p. Additionally, it's worth noting that ZDHHC14 is known as a tumor suppressor gene with relevance across various cancer types<sup>53</sup>.
- In the luminal B subtype, modules G and H in Fig. 6 are pivotal in regulating T cell proliferation and activation, while module C is associated with nipple morphogenesis and influences the T cell apoptotic process. These modules reveal the presence of hsa-miR-149-3p and hsa-miR-133a-3p miRNAs, which are shared between lncRNA C6orf223 and the mRNA molecules in module G, and between lncRNA HHLA3-AS1 and the mRNA molecules in module H. This discovery underscores the potential of these miRNAs as influential genes involved in suppressing various oncogenic pathways<sup>54</sup>. As a result of our analysis, we suggest that biologists consider lncRNAs C6orf223 (in module G) and HHLA3-AS1 (in module H), as well as circ\_000349 (in module C), as potential biomarkers for the luminal B subtype.
- In the basal-like subtype, module A (see Fig. 5) includes BCL2L12 mRNA linked to T Helper-2 (Th2) activation processes. This gene's role in modulating immune responses suggests it as a potential therapeutic target<sup>55</sup>. CircRNAs, circ\_003996, circ\_003925, circ\_004109, and circ\_000736, in this module further regulate BCL2L12 via hsa-miR-539-5p.
- In the luminal A subtype, circ\_002665 (module C in Fig. 8) and circ\_001855 (module D in Fig. 8) appear to play essential roles in regulating T-cell activation mechanisms. These circRNAs are involved in the regulation of HOXB6 in Module C and HLA-F and CAMKV in Module D. Previous studies have discussed the roles of HOXB6 in breast cancer<sup>56</sup>, and the dysregulation of HLA-F and CAMKV has been reported in various cancers<sup>57–59</sup>. This analysis underscores the precise involvement of HOXB6, HLA-F, and CAMKV in tumor progression through T cell activation processes in the luminal A subtype.
- In the normal-like subtype, module B (Fig. 7), which is enriched in T Helper Cell Surface Molecules, includes DLEU1. DLEU1 has been identified as an up-regulated lncRNA in breast cancer tissues and cells, particularly in tumors with high malignancy<sup>60,61</sup>. This lncRNA interacts with NPHS1 mRNA. Based on our analysis, NPHS1 exhibits significant prognostic value and shows differential expression between normal-like subtype and the other four breast cancer subtypes. This underscores NPHS1 as a potential dependable immune molecular marker for normal-like subtype. Interestingly, three shared miRNAs, namely hsa-miR-378a-5p,

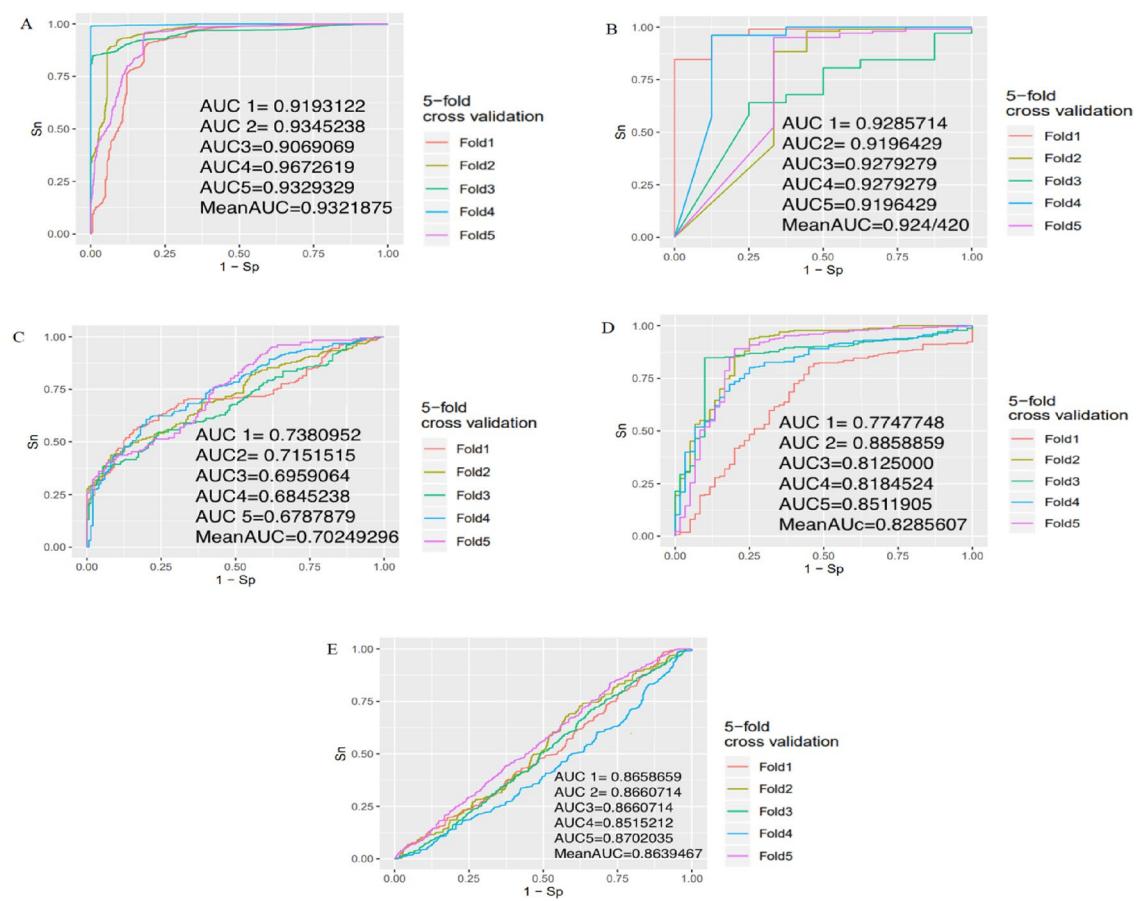


**Figure 11.** Kaplan–Meier survival curves for candidate mRNAs with lowest Log-rank test  $p$  values for (A, B) luminal B subtype and (C) normal-like subtype. Patients are divided into high-risk and low-risk groups based on the gene expression of these genes. The  $p$  value of the Log-rank test statistic has also been shown in each diagram.

Subtype C	<i>Bio_DEG<sup>C</sup></i>	avg_ logFC	std_ logFC	avg_FDR	std_FDR
Basal-like	circ_001845	0.909275	0.131957831	0.0234475	0.025229302
	circ_001124	0.855975	0.232192239	0.024875	0.025912079
	circ_003925	0.76815	0.001815673	0.0158925	0.023758302
	circ_000736	0.710225	0.107696406	0.028525	0.025866758
	circ_003996	0.762175	0.001776467	0.0158875	0.023748417
Luminal B	circ_00306	0.928925	0.214181813	0.0132075	0.025267296
	circ_00128	1.0188	0.126878236	0.0096	0.013684541
Normal-like	circ_000709	0.80905	0.012791794	0.009645	0.007470718
	NPHS1	0.717425	0.177538191	0.0218425	0.021153169
Luminal A	CAMKV	0.93575	0.467635994	0.0137	0.019086994
	circ_001855	0.5703	0.014347822	0.0448	0.02248733
HER2+	circ_00128	1.79915	0.523564581	0.003755	0.007230682
	circ_00173	1.85335	0.351477837	0.0001825	0.000279568

**Table 4.** The common significantly differentially expressed genes between subtype C and the other subtypes are documented in the set *Bio\_DEG<sup>C</sup>* as the potential genes for biomarkers.

hsa-miR-125a-3p, and hsa-miR-150-5p, are found to exist between DLEU1 and NPHS1. These miRNAs are known as anti-apoptotic agents in breast cancer, further highlighting their potential roles in immune regulation within the normal-like subtype<sup>62–64</sup>.



**Figure 12.** The validation results of selected genes in the set *Bio\_DEG*<sup>C</sup> for subtype C using SVM which is (A) basal-like, (B) HER2+, (C) luminal A, (D) luminal B, and (E) normal-like. The AUC values for fivefold cross-validation are provided for each subtype, and the mean AUC values of fivefold cross-validation are also included.

Subtype	Biomarkers	
	Cancer-related processes	Immune-related processes
Basal-like	circ_001845	circ_003996
	circ_001124	circ_003925
	circ_003925	
	circ_000736	
	circ_003996	
Luminal B	circ_00306	C6orf223
	circ_00128	HLA3-AS1
		circ_000349
Normal-like	circ_000709	DLEU1
	NPHS1	
Luminal A	CAMKV	circ_002665
	circ_001855	circ_001855
HER2+	circ_00128	RGS5-AS1
	circ_00173	

**Table 5.** A compilation of potential biomarkers identified by our pipeline.

These findings emphasize the complex interplay between various RNA molecules and immune mechanisms across breast cancer subtypes, offering insights into potential therapeutic targets and diagnostic markers.

## Conclusion

In conclusion, breast cancer poses a significant global health challenge. Recent research highlights the critical roles of ncRNAs, specifically lncRNAs and circRNAs, in both cancer and immune system functions. The competing endogenous RNA hypothesis provides a valuable framework to understand the interplay between mRNAs and ncRNAs in the context of miRNA regulation.

This study employed an innovative pipeline named Pre\_CLM\_BCS to investigate the roles of ncRNAs in various breast cancer subtypes. By constructing subtype-specific ceRNA networks, we unveiled distinct modules associated with immune and cancer-related processes within each subtype. Table 5 reveals the results of our pipeline, which has allowed us to identify several potential biomarkers specific to different breast cancer subtypes. Additionally, our analysis has unveiled certain ncRNAs that appear to play pivotal roles in the progression of their respective breast cancer subtypes, possibly through their regulatory effects on T cell mechanisms.

In the realm of immuno-oncology, understanding T cell regulation and activation mechanisms is essential. This study revealed the involvement of specific RNA molecules in these processes across different breast cancer subtypes. These findings provide insights into potential therapeutic targets and diagnostic markers, shedding light on the intricate interplay between various RNA molecules and immune mechanisms in the context of breast cancer subtypes. This research contributes to the ongoing efforts to improve breast cancer diagnosis and treatment, ultimately reducing the global burden of this disease.

While we have developed a pipeline to investigate the competitive regulatory mechanisms of circRNAs and lncRNAs within the context of miRNAs competition for the detection of breast cancer subtypes, it's evident that each step of the pipeline has potential for improvement.

In our future research, we aim to explore the implementation of graph neural networks (GNNs)<sup>65</sup> to enhance the construction of specific ceRNA networks in the third step of the pipeline. GNNs have shown promise in improving the accuracy of network analysis and prediction, and we anticipate that leveraging this technology will lead to the discovery of more effective biomarkers.

Furthermore, we plan to investigate the integration of ceRNA networks with gene-protein signaling networks<sup>66–68</sup>. This combination promises to provide a more holistic view of the regulatory interactions within a biological system. By unifying coding and ncRNA interactions with protein signaling pathways, we can gain a deeper understanding of how these components mutually influence each other's functions, ultimately shedding light on complex biological processes and diseases. This integrated approach holds the potential to identify key regulatory nodes, uncover novel biomarkers, and elucidate the intricate interplay between ceRNA networks and protein signaling.

## Data availability

The study utilized miRNA-seq and RNA-SeqV2 level 3 data, which encompass mRNA and lncRNA expression data of breast cancer, sourced from TCGA Research Network (<http://cancergenome.nih.gov/>) using the National Cancer Institute (NCI) Genomic Data Commons (GDC) resource (<https://gdc.cancer.gov/>). CircRNA expression data for breast cancer was acquired based on the information provided by Asha et al. (<https://doi.org/10.18632/oncotarget.13134>)<sup>28</sup>.

Received: 6 July 2023; Accepted: 18 November 2023

Published online: 27 November 2023

## References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2021. *CA Cancer J. Clin.* **71**, 7–33 (2021).
2. Spitale, A., Mazzola, P., Soldini, D., Mazzucchelli, L. & Bordoni, A. Breast cancer classification according to immunohistochemical markers: Clinicopathologic features and short-term survival analysis in a population-based study from the South of Switzerland. *Ann. Oncol.* **20**, 628–635 (2009).
3. Zaha, D. C., Lazar, E. & Lazureanu, C. Clinicopathologic features and five years survival analysis in molecular subtypes of breast cancer. *Rom. J. Morphol. Embryol.* **51**, 85–89 (2010).
4. Zardavas, D., Irrthum, A., Swanton, C. & Piccart, M. Clinical management of breast cancer heterogeneity. *Nat. Rev. Clin. Oncol.* **12**, 381–394 (2015).
5. McDonald, E. S., Clark, A. S., Tchou, J., Zhang, P. & Freedman, G. M. Clinical diagnosis and management of breast cancer. *J. Nuclear Med.* **57**, 98–16S (2016).
6. Peppercorn, J., Perou, C. M. & Carey, L. A. Molecular subtypes in breast cancer evaluation and management: Divide and conquer. *Cancer Invest.* **26**, 1–10 (2008).
7. Brosnan, C. A. & Voinnet, O. The long and the short of noncoding RNAs. *Curr. Opin. Cell Biol.* **21**, 416–425 (2009).
8. Catalanotto, C., Cogoni, C. & Zardo, G. MicroRNA in control of gene expression: An overview of nuclear functions. *Int. J. Mol. Sci.* **17**, 1712 (2016).
9. Kung, J. T. Y., Colognori, D. & Lee, J. T. Long noncoding RNAs: Past, present, and future. *Genetics* **193**, 651–669 (2013).
10. Grixti, J. M. & Ayers, D. Long noncoding RNAs and their link to cancer. *Noncoding RNA Res.* **5**, 77–82 (2020).
11. Jeck, W. R. & Sharpless, N. E. Detecting and characterizing circular RNAs. *Nat. Biotechnol.* **32**, 453–461 (2014).
12. Memczak, S. *et al.* Circular RNAs are a large class of animal RNAs with regulatory potency. *Nature* **495**, 333–338 (2013).
13. Liu, W. *et al.* MPCLCDA: Predicting circRNA-disease associations by using automatically selected meta-path and contrastive learning. *Brief Bioinform.* **24**, bbad227 (2023).
14. Sun, F., Sun, J. & Zhao, Q. A deep learning method for predicting metabolite-disease associations via graph neural network. *Brief Bioinform.* **23**, bbac266 (2022).
15. Li, Z., Zhu, X. & Huang, S. Extracellular vesicle long non-coding RNAs and circular RNAs: Biology, functions and applications in cancer. *Cancer Lett.* **489**, 111–120 (2020).

16. Zhang, Y. *et al.* Comprehensive characterization of lncRNA–mRNA related ceRNA network across 12 major cancers. *Oncotarget* **7**, 64148 (2016).
17. Zhong, Y. *et al.* Circular RNAs function as ceRNAs to regulate and control human cancer progression. *Mol. Cancer* **17**, 1–11 (2018).
18. Yao, Y. *et al.* Integrated analysis of co-expression and ceRNA network identifies five lncRNAs as prognostic markers for breast cancer. *J. Cell Mol. Med.* **23**, 8410–8419 (2019).
19. Ahmadi, H. *et al.* HomoTarget: A new algorithm for prediction of microRNA targets in *Homo sapiens*. *Genomics* **101**, 94–100 (2013).
20. Motieghader, H., Kouhsar, M., Najafi, A., Sadeghi, B. & Masoudi-Nejad, A. mRNA–miRNA bipartite network reconstruction to predict prognostic module biomarkers in colorectal cancer stage differentiation. *Mol. Biosyst.* **13**, 2168–2180 (2017).
21. Ala, U. Competing endogenous RNAs, non-coding RNAs and diseases: An intertwined story. *Cells* **9**, 1574 (2020).
22. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: The Rosetta Stone of a hidden RNA language?. *Cell* **146**, 353–358 (2011).
23. Yang, C. *et al.* Competing endogenous RNA networks in human cancer: Hypothesis, validation, and perspectives. *Oncotarget* **7**, 13479 (2016).
24. Wang, Y. *et al.* The emerging function and mechanism of ceRNAs in cancer. *Trends Genet.* **32**, 211–224 (2016).
25. Fan, C.-N., Ma, L. & Liu, N. Systematic analysis of lncRNA–miRNA–mRNA competing endogenous RNA network identifies four-lncRNA signature as a prognostic biomarker for breast cancer. *J. Transl. Med.* **16**, 1–12 (2018).
26. Sang, M. *et al.* Identification of epithelial–mesenchymal transition-related circRNA–miRNA–mRNA ceRNA regulatory network in breast cancer. *Pathol. Res. Pract.* **216**, 153088 (2020).
27. Durinck, S. *et al.* BioMart and Bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).
28. Nair, A. A. *et al.* Circular RNAs and their associations with breast cancer subtypes. *Oncotarget* **7**, 80967 (2016).
29. Zhang, Q. & Zhang, Y. Hierarchical clustering of gene expression profiles with graphics hardware acceleration. *Pattern Recognit. Lett.* **27**, 676–681 (2006).
30. Kouhsar, M., Azimzadeh Jamalkandi, S., Moeini, A. & Masoudi-Nejad, A. Detection of novel biomarkers for early detection of non-muscle-invasive bladder cancer using competing endogenous RNA network analysis. *Sci. Rep.* **9**, 1–15 (2019).
31. Guo, L.-L. *et al.* Competing endogenous RNA networks and gastric cancer. *World J. Gastroenterol.* **21**, 11680 (2015).
32. Gao, L., Zhao, Y., Ma, X. & Zhang, L. Integrated analysis of lncRNA–miRNA–mRNA ceRNA network and the potential prognosis indicators in sarcomas. *BMC Med. Genom.* **14**, 1–11 (2021).
33. Mitra, A., Pfeifer, K. & Park, K.-S. Circular RNAs and competing endogenous RNA (ceRNA) networks. *Transl. Cancer Res.* **7**, S624 (2018).
34. Wang, Y. *et al.* Circular RNAs in human cancer. *Mol. Cancer* **16**, 1–8 (2017).
35. Huang, H.-Y. *et al.* miRTarBase 2020: Updates to the experimentally validated microRNA–target interaction database. *Nucleic Acids Res.* **48**, D148–D154 (2020).
36. Karagkouni, D. *et al.* DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA–gene interactions. *Nucleic Acids Res.* **46**, D239–D245 (2018).
37. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
38. Morris, J. H. *et al.* clusterMaker: A multi-algorithm clustering plugin for Cytoscape. *BMC Bioinform.* **12**, 1–14 (2011).
39. Enright, A. J., Van Dongen, S. & Ouzounis, C. A. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575–1584 (2002).
40. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37**, W305–W311 (2009).
41. Kassambara, A., Kosinski, M., Biecek, P. & Fabian, S. Package ‘survminer’. *Drawing Survival Curves using ‘ggplot2’(R package version 0.3.1)* (2017).
42. Therneau, T. M. & Lumley, T. Package ‘survival’. *R Top Doc* **128**, 28–33 (2015).
43. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
44. Kuhn, M. Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).
45. Sachs, M. C. plotROC: A tool for plotting ROC curves. *J. Stat. Softw.* **79**, 2 (2017).
46. Yan, Y. MLmetrics: Machine learning evaluation metrics. *R Package Version* **1**, 1 (2016).
47. Wickham, H. & Wickham, M. H. The ggplot package. *Google Scholar* (2007).
48. Chekulaeva, M. & Rajewsky, N. Roles of long noncoding RNAs and circular RNAs in translation. *Cold Spring Harb. Perspect. Biol.* **11**, a032680 (2019).
49. Finn, O. J. Immuno-oncology: Understanding the function and dysfunction of the immune system in cancer. *Ann. Oncol.* **23**, viii6–viii9 (2012).
50. Buckner, J. H. & Ziegler, S. F. Regulating the immune system: The induction of regulatory T cells in the periphery. *Arthritis Res. Ther.* **6**, 1–8 (2004).
51. Kos, K. & de Visser, K. E. The multifaceted role of regulatory T cells in breast cancer. *Annu. Rev. Cancer Biol.* **5**, 291 (2021).
52. Speiser, D. E. & Verdeil, G. More T cells versus better T cells in patients with breast cancer. *Cancer Discov.* **7**, 1062–1064 (2017).
53. Yeste-Velasco, M. *et al.* Identification of ZDHHC14 as a novel tumor suppressor gene commonly downregulated in human cancers. *Cancer Res.* **72**, 4858 (2012).
54. Weber, D., Amar, L., Gödde, D. & Prinz, C. Extensive screening of microRNA populations identifies hsa-miR-375 and hsa-miR-133a-3p as selective markers for human rectal and colon cancer. *Oncotarget* **9**, 27256–27267 (2018).
55. Chen, X. *et al.* Bcl2-like protein 12 is required for the aberrant T helper-2 polarization in the heart by enhancing interleukin-4 expression and compromising apoptotic machinery in CD4+ T cells. *Circulation* **138**, 2559–2568 (2018).
56. Lewis, M. T. Homeobox genes in mammary gland development and neoplasia. *Breast Cancer Res.* **2**, 1–12 (2000).
57. Lin, A. & Yan, W.-H. The emerging roles of human leukocyte antigen-F in immune modulation and viral infection. *Front. Immunol.* **10**, 964 (2019).
58. Wuerfel, F. M. *et al.* Hla-G and HLA-F protein isoform expression in breast cancer patients receiving neoadjuvant treatment. *Sci. Rep.* **10**, 1–10 (2020).
59. Sussman, R. T. *et al.* CAMKV is a candidate immunotherapeutic target in MYCN amplified neuroblastoma. *Front. Oncol.* **10**, 302 (2020).
60. Ma, H. N., Chen, H. J., Liu, J. Q. & Li, W. T. Long non-coding RNA DLEU1 promotes malignancy of breast cancer by acting as an indispensable coactivator for HIF-1 $\alpha$ -induced transcription of CKAP2. *Cell Death Dis.* **13**, 1–10 (2022).
61. Wang, C., Xie, X. X., Li, W. J. & Jiang, D. Q. LncRNA DLEU1/microRNA-300/RAB22A axis regulates migration and invasion of breast cancer cells. *Eur. Rev. Med. Pharmacol. Sci.* **23**, 10410–10421 (2019).
62. Zheng, S., Li, M., Miao, K. & Xu, H. lncRNA GAS5-promoted apoptosis in triple-negative breast cancer by targeting miR-378a-5p/SUFU signalling. *J. Cell Biochem.* **121**, 2225–2235 (2020).
63. Xu, X., Lv, Y., Yan, C., Yi, J. & Ling, R. Enforced expression of hsa-miR-125a-3p in breast cancer cells potentiates docetaxel sensitivity via modulation of BRCA1 signaling. *Biochem. Biophys. Res. Commun.* **479**, 893–900 (2016).

64. Carvalho, T. M. *et al.* MicroRNAs miR-142-5p, miR-150-5p, miR-320a-3p, and miR-4433b-5p in serum and tissue: Potential biomarkers in sporadic breast cancer. *Front. Genet.* **14**, 865472 (2022).
65. Liu, W. *et al.* NSRGRN: A network structure refinement method for gene regulatory network inference. *Brief Bioinform.* **24**, bbad129 (2023).
66. Li, X. *et al.* Caspase-1 and Gasdermin D afford the optimal targets with distinct switching strategies in NLRP1b inflammasome-induced cell death. *Research* **2022**, 9838341 (2022).
67. Li, X. *et al.* RIP1-dependent linear and nonlinear recruitments of caspase-8 and RIP3 respectively to necrosome specify distinct cell death outcomes. *Protein Cell* **12**, 858–876 (2021).
68. Xu, F. *et al.* Specificity and competition of mRNAs dominate droplet pattern in protein phase separation. *Phys. Rev. Res.* **5**, 23159 (2023).

### Author contributions

F.Z.M. and A.M.N. proposed the idea. A.S.T. performed the data analysis, wrote the first draft and drawn all figures and illustration of the manuscript. Y.Z.B. helped with biological discussion and edited the manuscript. A.M.N. and F.Z.M. conceived and supervised and edited the final draft of the work. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47816-z>.

**Correspondence** and requests for materials should be addressed to F.Z.-M. or A.M.-N.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023