Briefings in Bioinformatics, 00(00), 2020, 1–14

doi: 10.1093/bib/bbaa088
Problem solving protocol

ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer

Xuanyi Wang†, Zixuan Chai†, Guizhi Pan†, Youjin Hao, Bo Li, Ting Ye, Yinghong Li, Fei Long, Lixin Xia, Mingwei Liu

Corresponding authors: Mingwei Liu, Tel.: +86-23-68485240; Fax: +86-23-68485240; E-mail: liumingwei@cqmu.edu.cn; Lixin Xia, Tel.: +86-755-26913253; Fax: +86-755-86671901; E-mail: xialixin@126.com
†These authors contributed equally to this work.

Abstract

Effective and safe implementation of precision oncology for breast cancer is a vital strategy to improve patient outcomes, which relies on the application of reliable biomarkers. As 'liquid biopsy' and novel resource for biomarkers, exosomes provide a promising avenue for the diagnosis and treatment of breast cancer. Although several exosome-related databases have been developed, there is still lacking of an integrated database for exosome-based biomarker discovery. To this end, a comprehensive database ExoBCD (https://exobcd.liumwei.org) was constructed with the combination of robust analysis of four high-throughput datasets, transcriptome validation of 1191 TCGA cases and manual mining of 950 studies. In ExoBCD, approximately 20 900 annotation entries were integrated from 25 external sources and 306 exosomal molecules (49 potential biomarkers and 257 biologically interesting molecules). The latter could be divided into 3 molecule types, including 121 mRNAs, 172 miRNAs and 13 lncRNAs. Thus, the well-linked information about molecular characters, experimental biology, gene expression patterns, overall survival, functional evidence, tumour stage and clinical use were fully integrated. As a data-driven and literature-based paradigm proposed of biomarker discovery, this study also demonstrated the corroborative analysis and identified 36 promising molecules, as well as the most promising prognostic biomarkers, *IGF1R* and *FRS2*. Taken together, ExoBCD is the first well-corroborated knowledge base for exosomal studies of breast cancer. It not only lays a foundation for subsequent studies but also strengthens the studies of probing molecular mechanisms, discovering biomarkers and developing meaningful clinical use.

Key words: breast cancer; precision oncology; exosome; biomarker; database ExoBCD;

Xuanyi Wang is a postgraduate student in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Zixuan Chai is a postgraduate student in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Guizhi Pan is a postgraduate student in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Youjin, Hao is a professor of cell biology and bioinformatics in the College of Life Sciences, Chongqing Normal University, Chongqing, China.
Bo Li has a PhD student of cell biology and bioinformatics from the College of Life Sciences, Chongqing Normal University, Chongqing, China.
Ting Ye is an undergraduate student in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Yinghong Li is a PhD student of the Key Laboratory on Big Data for Bio Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, China.
Fei Long is a postgraduate student in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Mingwei Liu an associate professor in the Key Laboratory of Clinical Laboratory Diagnostics, College of Laboratory Medicine, Chongqing Medical University, Chongqing, China.
Lixin Xia is a professor at the Shenzhen University Health Science Center, Shenzhen University, Shenzhen, China.
Submitted: 13 December 2019; **Received (in revised form):** 8 March 2020

Introduction

Breast cancer now is the most prevalent type of cancer in both the developing and developed countries and is the biggest cause of mortality of women worldwide [1]. Therefore, early diagnosis and efficient and precision treatments are crucial for prolonging and improving patient's life [2]. Clinically, the effective precision oncology depends upon high-fidelity and technically suitable biomarkers [3]. Genomic screening approaches have been commonly applied to identify tumour-specific, overexpressed proteins or genetic mutations that may lead to better clinical outcomes [4]. However, there is still an extreme lack of specific and sensitive biomarkers for breast cancer diagnosis and therapy [5].

Exosomes, small membrane vesicles secreted by a variety of healthy and unhealthy cells, entrap biomolecules (proteins, lipids, mRNAs and miRNAs) and mediate intercellular communication as well as selectively transfer of components to perform various functions in physiological and pathological conditions [6]. Especially, the enrichment in human body fluid and high amount of high-sensitive and high-specific molecules make exosomes important in tumour growth and metastasis, tumour invasion and angiogenesis and immune escape [7]. Moreover, due to the availability of the dynamic monitoring of tumour status and the real-time evaluation of disease development, exosomes are expected as novel sources of biomarkers in liquid biopsy for precision oncology and provide a promising avenue for the advantageous diagnosis and treatment of breast cancer [8, 9]. Therefore, discovering reliable biomarkers based on comprehensive and integrated analysis of large-scale high-throughput data, clinical data and relevant literature is valuable and significant for the diagnosis and treatment of breast cancer.

Significant advances have been made in the field of exosomal researches that have improved our understanding of the development and spread of breast cancers and advance the discovery of biomarkers. Khan *et al.* [10] reported that the antiapoptotic survivin and its isoforms were significantly expressed in exosomes derived from the serum of breast cancer patients. This differential expression pattern makes them important diagnostic or prognostic biomarkers in breast cancer. Recently, miRNAs have been identified in exosomes in a stable form and could be potential biomarkers of malignancy of cancer. Melo *et al.* [11] reported that *miR-21* and *miR-10b* in MDA-MB-231 cancer exosomes could bind the 3′-UTR of *PTEN* or *HOXD10* in normal breast cell line MCF10A and promoted the development and transfer of breast cancer cells. A recent report showed that *miR-21* and *miR-1246* were selectively abundant in breast cancer exosomes and their expression levels were significantly increased in the plasma of breast cancer patients. Therefore, the authors suggested that they could serve as diagnostic biomarkers for breast cancer [12]. Tan *et al.* [13] found that exosomal *RN7SL1* could promote the proliferation and growth of cancer cells. Hepatocellular carcinoma (HCC) patients with higher *RN7SL1* concentrations have a lower survival rate, which made it a reliable biomarker for HCC diagnosis and prognosis.

Although great achievements have been made on exosomal researches of breast cancer, there are still lots of challenges. (i) Due to lacking systematic knowledge, a more detailed relevance is still unknown about the regulation mechanisms of tumour progression and the clinically aggressive phenotypes. (ii) Although databases ExoCarta [14], Vesiclepedia [15], EVpedia [16], exoRBase [17] and EVmiRNA [18] have been developed, the complete exosome-related knowledge such as tumour stage and subtype, gene expression pattern, multiple molecular pathway and survival-related profile are little available, which holds back the reveal of crucial molecules and discovery of biomarkers in breast cancer. (iii) It is a complex, time-consuming and challenging process for biologists or clinicians to mine crucial molecules or biomarkers from the multifarious literatures and data sources with corroborative analysis.

To overcome the above limits, a comprehensive and web-based database ExoBCD (https://exobcd.liumwei.org) was developed in this study. In this database, 3 types of 306 valuable exosomal molecules, including 49 potential biomarkers (PBs) and 257 biologically interesting molecules (BIMs), were identified and well displayed along with their detailed and well-evidenced information such as molecular characters, biological experiments, gene expression patterns, overall survival, surveyed literatures as well as tumour stage or subtype and clinical use. All of these were fully merged into 15 types of annotations and 20 907 entries. Basically and systematically, all these were corroborated directly or indirectly from 950 latest valuable publications, reliable predictive analysis of 4 high-quality high-throughput datasets and reliable validation against 1191 TCGA breast cancer cases, which makes some exosomal molecules (e.g. *IGF1R* and *FRS2*) outstanding as promising potential biomarkers in the diagnosis of breast cancer prognosis, especially *IGF1R*. Furthermore, we also provided a query interface to facilitate users in browsing and visualizing the annotation information for exosomal molecules. Taken together, database ExoBCD provides a comprehensive and extensive resource with friendly interface and statistical visualization based on the integration of the robust analysis of high-throughput expression data and strict investigation of literatures. We believe that ExoBCD will be helpful for exosome-related studies and the discovery of clinical biomarkers.

Methods

The flowchart of ExoBCD construction was shown in Figure 1A, including literature-based mining, transcriptome expression analysis and biological function investigations. And a data-driven and literature-based paradigm of biomarker discovery with corroborative analysis was proposed as shown in Figure 1B. The details were described in the following sections.

Literature survey on breast cancer

To guarantee the high quality of collected literature data, molecules in ExoBCD were manually collected and integrated as follows (Figure 2A). Firstly, the detailed literature searching was performed in PubMed (http://www.ncbi.nlm.nih.gov/pubmed) and Web of Science (https://www.webofknowledge.com) using the keywords 'breast malignant neoplasm', 'breast cancer', 'breast carcinoma', 'mammary carcinoma', 'mammary neoplasms', 'breast tumour', 'breast neoplasm' or 'mammary cancer' and 'exosome'. Secondly, each returned article was assessed by two independent curators and then integrated using a more stringent criterion (Supplementary Note 1.1) according to the standards proposed by the International Society for Extracellular Vesicles (ISEV) as minimal information for studies of extracellular vesicles. Finally, the data were manually curated to ExoBCD terms, tags and descriptions based on a series of terms, such as molecule type (mRNA, miRNA, and lncRNA), sample (e.g. serum and urine), cell line, exosome isolation and verification method, physiological functions, control case, clinical use, tumour stage and subtype, PBs/BIMs and so on.

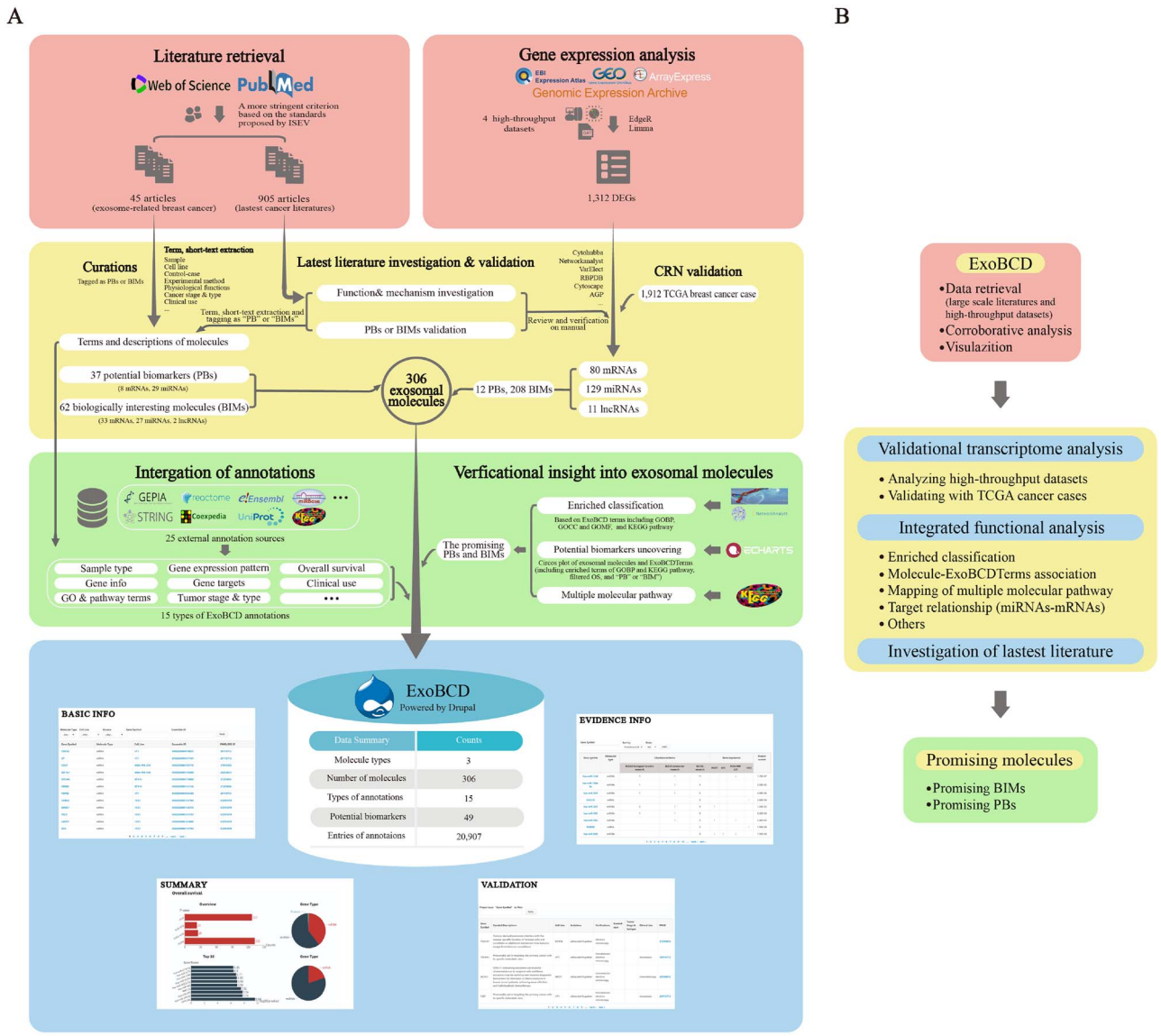

Figure 1. The flowchart of ExoBCD and the proposed paradigm of biomarker discovery. (A) The flowchart of ExoBCD construction based on the integration of large-scale literatures and high-throughput gene expression datasets. (B) The data-driven and literature-based paradigm schematically proposed biomarker discovery based on data retrieval, corroborative analysis and visualization, including validational transcriptome analysis, integrated functional analysis and investigation of the latest literatures.

High-throughput gene expression analysis

Data retrieving

Three miRNA microarray datasets [GSE60715 (10 samples), GSE60714 (four samples) and GSE66165 (three samples)] and one mRNA/lncRNA dataset [GSE93070 (four samples)] were retrieved from the Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) [19], ArrayExpress (https://www.ebi.ac.uk/arrayexpress) [20], Expression Atlas (https://www.ebi.ac.uk/gxa) [21] and Genomic Expression Archive (GEA, https://www.ddbj.nig.ac.jp/gea) [22] using built-in search engine and specialized tools [i.e. cBioPortal for Cancer Genomics (http://www.cbioportal.org) [23], EurOPDX Data Portal (https://dataportal.europdx.eu), Xena browser (https://xenabrowser.net), OmicsDI [24] (http://www.omicsdi.org), DataMed (https://datamed.org) [25] and Google Dataset Search (https://datasetsearch.research.google.com)] searching keywords including 'breast cancer', 'breast carcinoma', 'exosome', 'exosomes' and 'exosomal' (Supplementary Note 1.2, Figure 2B).

Data preprocessing and DEG identification

All the datasets were preprocessed to generate gene expression matrices using the limma [26] and edgeR [27] package (Supplementary Note 2.1). Differentially expressed genes (DEGs) were considered when FDR-adjusted $P < 0.05$ and absolute $\log_2$ fold change ≥ 1.5.

BIM and PB identification

For mRNA microarray data, DEGs were first validated in Cancer RNA-Seq Nexus (CRN, http://syslab4.nchu.edu.tw) ($P < 0.01$) [28] with 1191 TCGA breast cancer cases. Subsequently, the validated DEGs were used to identify the high-confidence hub genes that were identified by using 12 built-in centrality indices

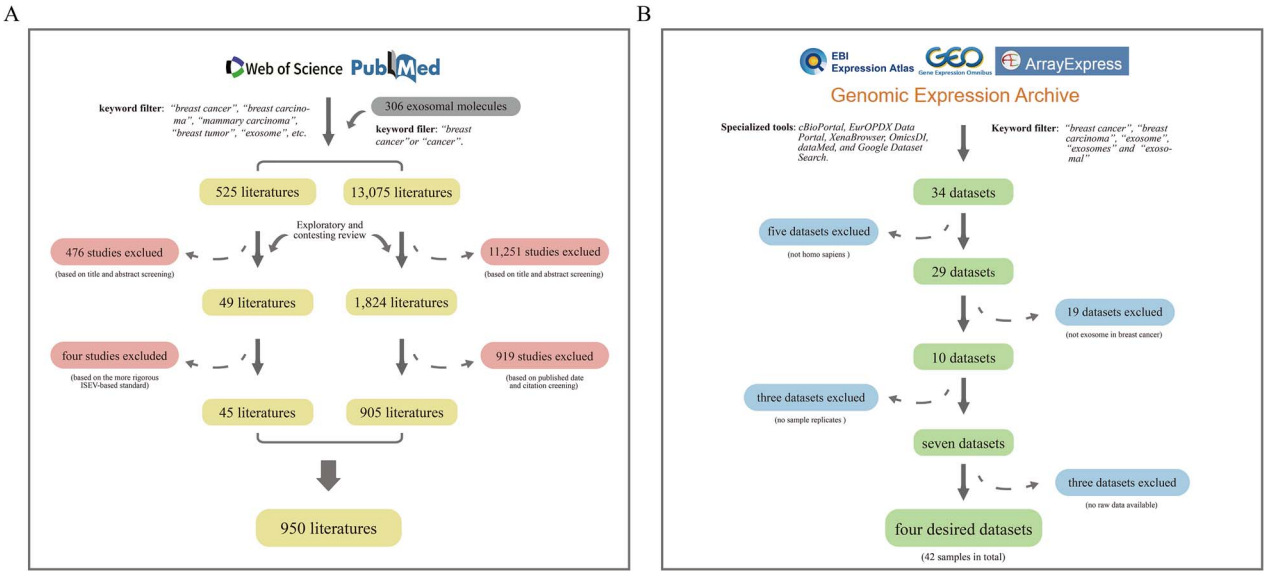

Figure 2. The pipeline of literature curation (A) plus dataset collection (B).

of CytoHubba [29]. GO and KEGG enrichment analysis was carried out by NetworkAnalyst [30]. The identified crucial hub genes with highlights on gene-phenotype evidences associated with breast cancer were considered as BIMs by VarElect [31] (Supplementary Notes 2.2, 2.3, 2.4). For miRNA microarray data, Wilcoxon signed-rank test was applied to screen BIMs when the breast cancer gene percentage (BCGP) values less than 0.05, which was numerically equivalent to the percentage of breast cancer-associated genes in the miRNA targets [32] (Supplementary Note 2.5, Supplementary Table S1).

To discover the BIMs from lncRNAs, all differentially expressed molecules (including mRNAs and lncRNAs) and filtered miRNAs (literature-originated exosomal biomarkers for breast cancer) were used to establish the competing endogenous RNA (ceRNA) network. The expression patterns of lncRNAs in the ceRNA network were also validated against the same TCGA cases mentioned above ($P < 0.001$) [28]. Moreover, potential targets of lncRNAs were searched in RNA-Binding Protein DataBase (RBPDB, http://rbpdb.ccbr.utoronto.ca) [33], and their biological functions in cancers were enriched by NetworkAnalyst [30] ($P < 0.05$). Notably, the latest literature review was also conducted to confirm the molecular function or mechanism, and the possibility of all molecules as BIMs or PBs, the described terms and short text in literatures and tags as 'PBs' or 'BIMs' were also extracted and merged as ExoBCD terms, tags and descriptions (Supplementary Note 2.6). All the molecules would be investigated according to the original studies to identify whether they were PBs/BIMs or not.

Functional and survival annotations for molecules

To better understand biological functions of identified molecules in breast cancer, the information of these molecules including molecular characters, GO and pathway terms, molecular interactions, expression patterns in different tissues and overall survival were annotated or analysed using Ensembl BioMart [34], miRBase [35], KEGG [36], GEPIA [37], Kaplan-Meier Plotter (http://kmplot.com/analysis) and so on. (Supplementary Tables S2 and S3).

Database construction

Based on literature mining and gene expression analysis, a database named ExoBCD with a friendly and easy-to-use web interface was developed by Drupal (https://www.drupal.org), in which molecules and their comprehensive information were integrated, such as enriched terms of GO and pathways, tumour stage and subtype, filtered OS (overall survival) and clinical use. To facilitate, explore and uncover exosomal biomarkers of breast cancer, a navigation function and search engine were provided. Users can easily browse and explore the target molecules based on research objectives but also can download the revealed roles of mRNAs in cancer data for a personalized analysis.

Verficational insights into exosomal molecules in ExoBCD

To better understand the roles of mRNAs and miRNAs in breast cancer, the enriched classification analysis of four terms from GOBP, GOCC, GOMF and KEGG pathway was performed using NetworkAnalyst [30] and the miRNA Enrichment Analysis and Annotation Tool (miEAA) [38] (Supplementary Note 3). The associations of exosomal molecules (121 mRNAs and 172 miRNAs) and ExoBCDTerms (including enriched terms of GOBP and KEGG pathway, filtered overall survival and 'BIMs/PBs') were plotted and visualized by Circos plot in ECharts (https://www.echartsjs.com/zh/index.html) to verify potential prognostic biomarkers. And to reveal the roles of mRNAs in cancer pathways, a manual map of multiple molecular pathways was reconstructed with KEGG Mapper. Additionally, the target correlation between mRNAs and miRNAs was revealed. After review of the latest cancer literatures, the promising PBs and BIMs were verified.

Results

Data summary

A total of 306 molecules (49 PBs and 257 BIMs), including 121 mRNAs, 172 miRNAs and 13 lncRNAs, were identified with the combination of literature mining and high-throughput expression analysis. Among them, 80 mRNAs, 129 miRNAs and

Table 1. Data summary of molecules and annotations in ExoBCD

ExoBCD term	Count	Category
Potential biomarkers	49	Molecule
Biologically interesting molecules	257	Molecule
Promising molecules	36	Molecule
GO term	3105	Annotation
KEGG pathway	1696	Annotation
Reactome pathway	579	Annotation
Wiki-pathway	712	Annotation
miRNA–mRNA interaction	10 492	Annotation
lncRNA–miRNA interaction	607	Annotation
lncRNA–protein interaction	526	Annotation
Overall survival	292	Annotation
Mutation	118	Annotation
Methylation	118	Annotation
Sequence	174	Annotation

11 lncRNAs from 4 high-throughput datasets were considered as BIMs or PBs. 37 PBs and 62 BIMs were retrieved from 525 literatures on breast cancer exosome, and they were all evidenced by experimental results. Furthermore, in ExoBCD, 15 types of annotations from 29 data sources were integrated with 20 907 entries. The detailed information was summarized in Table 1.

GO terms and pathway enrichment analysis

Gene ontology (GO) and pathway enrichment analysis were performed, and the results were shown in Figure 3A and B (Supplementary Table S4). About 121 mRNAs were mainly enriched in 4 terms, including signalling transductions (e.g. cGMP-PKG, MAPK, P53 and Ras signalling), pathways in cancer, cell cycle and proliferation and DNA damage response. This result suggested that these molecules were associated with cancer progression. For instance, three genes (*CDC42*, *FGFR3* and *TGFBR1*) were enriched in the MAPK signalling pathway, which is essential for growth factor-induced cell proliferation and differentiation. About 172 miRNAs were enriched into pathways associated with cancers, such as 'TGF beta signalling pathway', 'Wnt signalling pathway' and 'AMPK signalling pathway'. It's worth note that two miRNAs (*miR-639* and *miR-615-3p*) were predicted to target a set of genes with different functions in TGF beta- and P53-signalling pathways.

Identification of potential prognosis molecules

With association analysis by Circos plot, 6 mRNAs (*IGF1R*, *FRS2*, *FGFR2*, *SFN*, *PIK3R5* and *IL1RAP*) and 11 miRNAs (*miR-639*, *miR-615-3p*, *miR-595*, *miR-29b-3p*, *miR-149-3p*, *miR-130a-3p*, *miR-18a-5p*, *miR-195-5p*, *miR-223-3p*, *miR-124-3p* and *miR-17-5p*) were defined as potential prognostic biomarkers for breast cancers. As shown in Figure 3C and D, 17 molecules such as *IGF1R*, *FRS2*, *miR-18a-5p*, *miR-195-5p* and *miR-223-3p* showed good connections with filtered OS terms (all $P < 0.05$) and enriched cancer pathway terms such as MAPK, PI3K-Akt, P53 signalling pathway and so on. More connections were found between molecules and cancers, suggesting that the connected molecules (e.g. *IGF1R*, *FGFR2*, *SFN*, *miR-124-3p*, *miR-17-5p*, *miR-18a-5p* and *miR-195-5p*) were significantly correlated with prognosis (Supplementary Table S5).

Manual mapping of multiple molecular pathways

About 24 mRNAs were identified as key players in canonical pathways, such as P53, PI3K-Akt and MAPK signalling pathway, several other pathways involved in cell apoptosis, cell proliferation and angiogenesis (Figure 4). Among them, 10 mRNAs (e.g. *IGF1R*, *ERBB2*, *FGF9*, *FGFR2*, *ITGB1*, *CDC42*, *ARF6*, *ELK1*, *FGFR3* and *ITGB7*) are involved in the regulation of actin cytoskeleton, focal adhesion, endocytosis and adherent or tight junction. And five mRNAs (e.g. *IGF1R*, *ERBB2*, *FGF9*, *FGFR2* and *FZD5*) were predicted to play important roles in the invasion of breast cancer, gastric cancer and prostate cancer. Another 20 mRNAs (e.g. *IGF1R*, *ERBB2*, *FGF9*, *FGFR2*, *FGFR3*, *ELK1*, *CDC42*, *FRS2*, *NANOG*, *DUSP1*, *HTR7*, *TSC2*, *PKN2*, *LDHA*, *CHEK2*, *SFN*, *CACNB2*, *PIK3R5*, *BID* and *CD82*) are crucial in signalling transductions and proteoglycans in cancer and promote the development of cancer (Supplementary Table S5).

Target relationship of mRNAs and miRNAs

Based on 24 mRNAs above, target relationship analysis revealed that robust connections were detected between 16 mRNAs and 40 miRNAs (Figure 5A–C), while no connection was found for 8 mRNAs (*FGFR2*, *CACNB2*, *PIK3R5*, *ITGB7*, *BID*, *CD82*, *DUSP1* and *TSC2*), which were not targeted to any miRNA, but they were predicted to be important regulators and closely associated with breast cancer (Supplementary Table S5). Among the 16 mRNAs, the upstream regulators (miRNA) of *IGF1R* were up to 17 and followed by *FRS2*, *ARF6*, *FGFR3*, *FZD5* and *SFN*. It is worth note that *IGF1R* is a hub of the connection of 17 miRNAs and, of which, 5 miRNAs (e.g. *miR-340-5p*, *miR-583*, *miR-100*, *miR-145* and *miR-7109-5p*) also acted as bridges between *IGF1R* and 5 genes (*FRS2*, *ARF6*, *FGFR3*, *NANOG* and *ERBB2*), respectively. Interestingly, *IGF1R* targeted by eight miRNAs (e.g. *miR-122*, *miR-145*, *miR-223-3p*, *miR-340-5p*, *miR-372*, *miR-373*, *miR-486-5p* and *miR-503*) were considered as an exosomal biomarker for breast cancer [12, 39–43], and two exosomal miRNAs (*miR-100* and *miR-16*) targeted *IGF1R* were considered as important regulators and closely associated with breast cancer [44, 45]. And *FRS2* also has multiple target connections with seven miRNAs (*miR-101*, *miR-150-3p*, *miR-340-5p*, *miR-3654*, *miR-4443*, *miR-583* and *miR-8080*), such as *miR-101* and *miR-340-5p*, which were also considered as an exosomal biomarker for breast cancer [39, 42].

Taken together, based on the deeper analysis of molecular pathway mapping, target relationship and latest literature

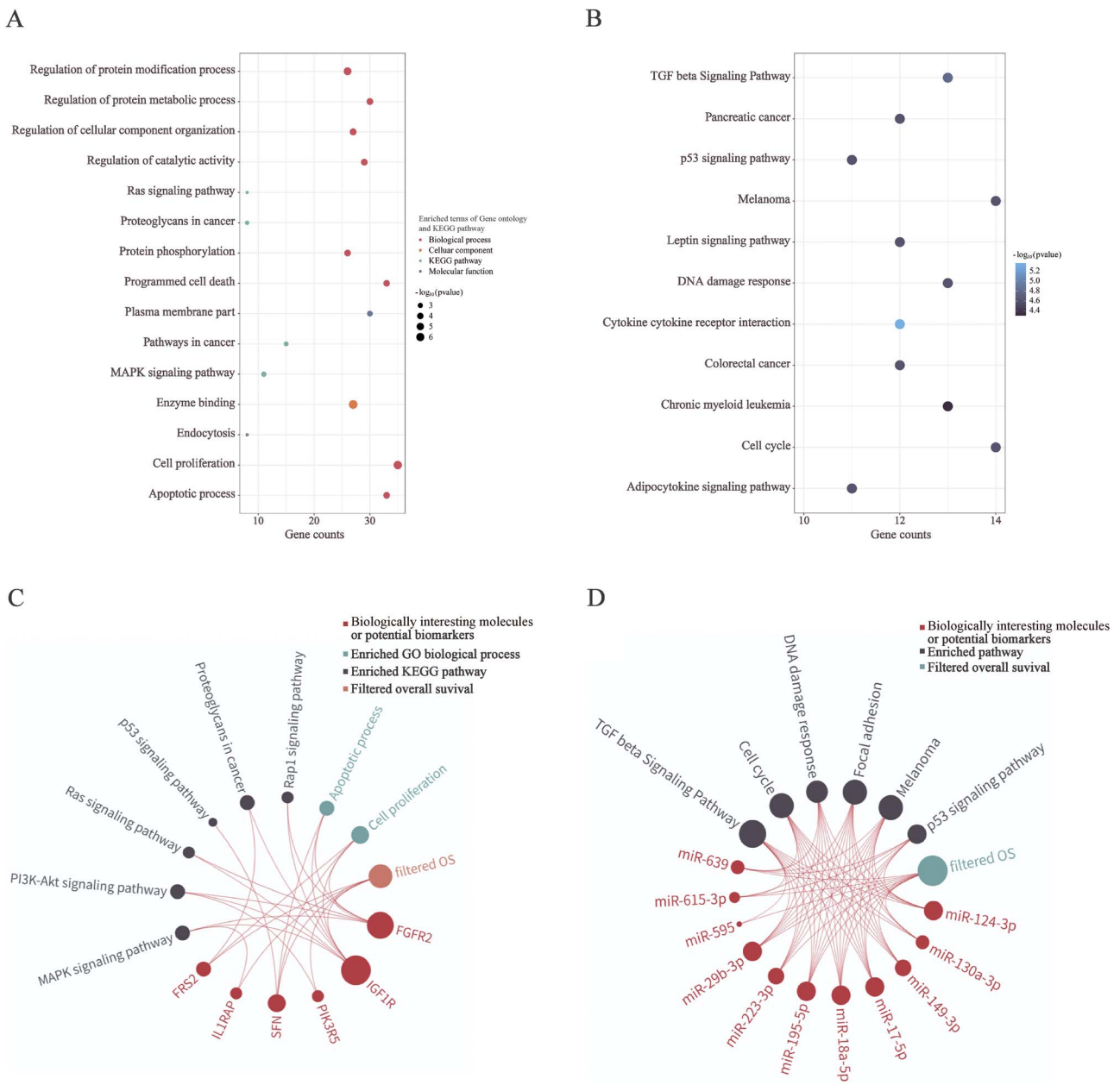

Figure 3. Classification of enriched exosome molecules in breast cancer and visualized by Circos plot. (A) Classification of enriched mRNAs. Annotation terms in biological process, cellular component, molecular function and KEGG pathway were marked in red, yellow, green and blue, respectively. Bubble size represents the value of -log10 of *P*-values of enrichment significance. (B) Classification of enriched miRNAs. Different colours represent the values of −log10 *P*-value of enrichment significance. (C and D) Circos plot of associations between enriched exosomal molecules and ExoBCD terms. Enriched GO terms in biological processes (BP) and KEGG pathways, filtered over survival (OS) ($P < 0.05$), were shown. They were presented by red, blue and yellow bubbles, respectively. Associations between molecules were linked. Line numbers represent the association strength. BIM, biologically interesting molecules; PM, potential biomarkers.

survey, 36 promising molecules (12 promising BIMs, 5 most promising BIMs, 17 promising PBs, and 2 most promising PBs) were uncovered (Supplementary Table S5), almost all of which are crucial for carcinogenesis, invasion, metastasis, development, recurrence and prognosis in breast cancer. Especially, two molecules (*IGF1R* and *FRS2*) were verified as the most promising PBs for breast cancer (Table 2).

Web interface

To facilitate users to explore breast cancer biomarkers from exosomes, a comprehensive ExoBCD with an easy-to-use web interface was constructed. Database navigation is from a set of menus, including HOME, BROWSE, SUMMARY, VALIDATION and DOWNLOAD.

On the HOME page, a navigation menu and data summary block were presented (Figure 6A). The information about all identified 306 molecules consisting of three molecule types and 15 types of annotations (over 20 900 entries) were displayed in the data summary block.

In the BROWSE page, 'Basic Info' and 'Evidence Info' options were created. In 'Basic Info' options, molecular information including gene symbol, type, source and cell source was integrated. With the input of the gene symbol or Ensemble ID

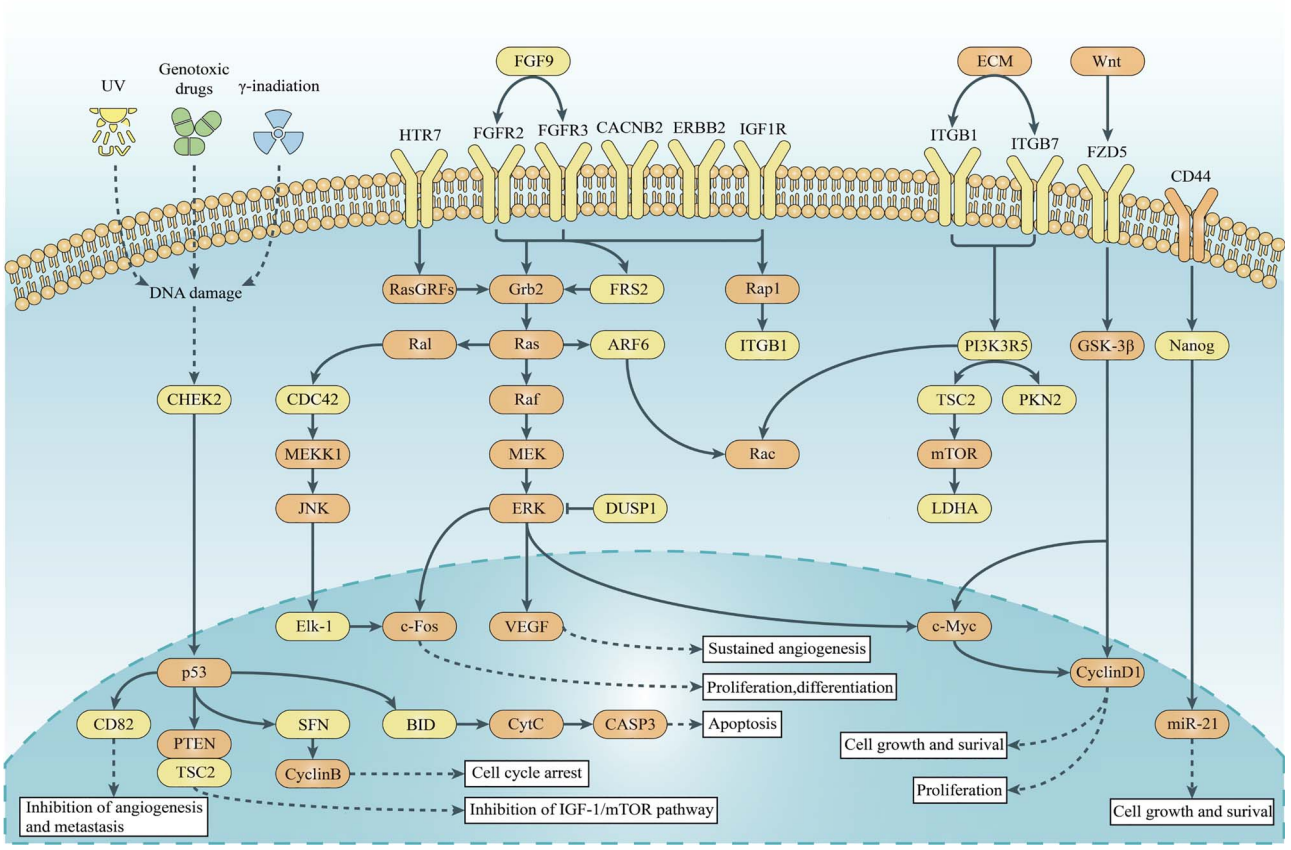

Figure 4. The manual map of multiple molecular pathways using KEGG mapper. High reliable mRNAs in breast cancer exosomes identified in this study were marked in yellow.

in the searching box, the search engine returns the detailed information. Moreover, users can also search for interest genes by filtering the results from the selected option. All available evidences including experimental evidences from public literatures associated with exosome and breast cancer (as well as other cancer types), high-throughput expression profiling in different cell lines, prognostic effect, and enriched functions, were available (Figure 6B).

In the SUMMARY page, result visualization of data processing was provided, including Kaplan-Meier chart of overall survival, bubble charts for KEGG pathways and GO term enrichment analysis, the regulatory network of ceRNAs and chord diagram of functional annotations (Figure 6C). In the regulatory network of ceRNAs, users can explore the correlation among lncRNA-miRNA-mRNA by clicking or dragging a node by mouse movement. Moreover, the regulatory information of lncRNAs in breast cancer genesis and development can be obtained from the ceRNA network, which is also helpful for biomarker prediction from exosomes.

In the VALIDATION page, the detailed experimental information, including the descriptions of case control, cell line, exosome isolation and verification method and condition information including function, tumour stage and subtype and clinical use, were integrated based on a literature review (Figure 6D).

In the DOWNLOAD page, literature information and functional annotations can be available in CSV, text or XML formats (Figure 6E).

Case study

ExoBCD aims to provide researchers a quick overview of PBs or BIMs linking exosome and breast cancer. Here, we use *miR-1246* and *miR-222* as cases to show how to effectively mine target molecules from the database.

First, users can use the BROWSE menu in the navigation bar to view the molecule information page. On the top, the searching box and the selected option as 'Basic Info' and 'Evidence Info' were available. When '*miR-1246*' was input in the searching box and 'Basic Info' was selected, the tabular information will be returned including 'Gene Symbol', 'Ensemble ID', 'Cell Line', 'PMID' and 'GEO ID'. By clicking the hyperlink for '*miR-1246*' in the table, users can view more detailed information such as 'KEGG pathways', 'potential target genes' and 'gene expression patterns' (Figure 7A). If users select the option of 'Evidence Info' and input gene symbol, some evidences for *miR-1246* as a potential biomarker will be displayed, in which descriptions of its roles in cancer development, high-throughput expression data, survival curve and the *P*-values, expression patterns in different cell lines, literature list and PubMed ID were included (Figure 7B).

Combining the information from 'Basic Info' and 'Evidence Info' query, it suggested that *miR-1246* was involved in the regulation of actin cytoskeleton, p53, Wnt/β-catenin signalling, etc., in lung, liver, cervical, oesophageal and pancreatic cancers [46–51]. *miR-1246* is also involved in promoting cell proliferation, invasion and drug resistance by targeting cyclin-G2 (CCNG2) in oral carcinomas and breast cancer [52, 53]. Statistical analysis showed that it is significantly associated with the prognosis of

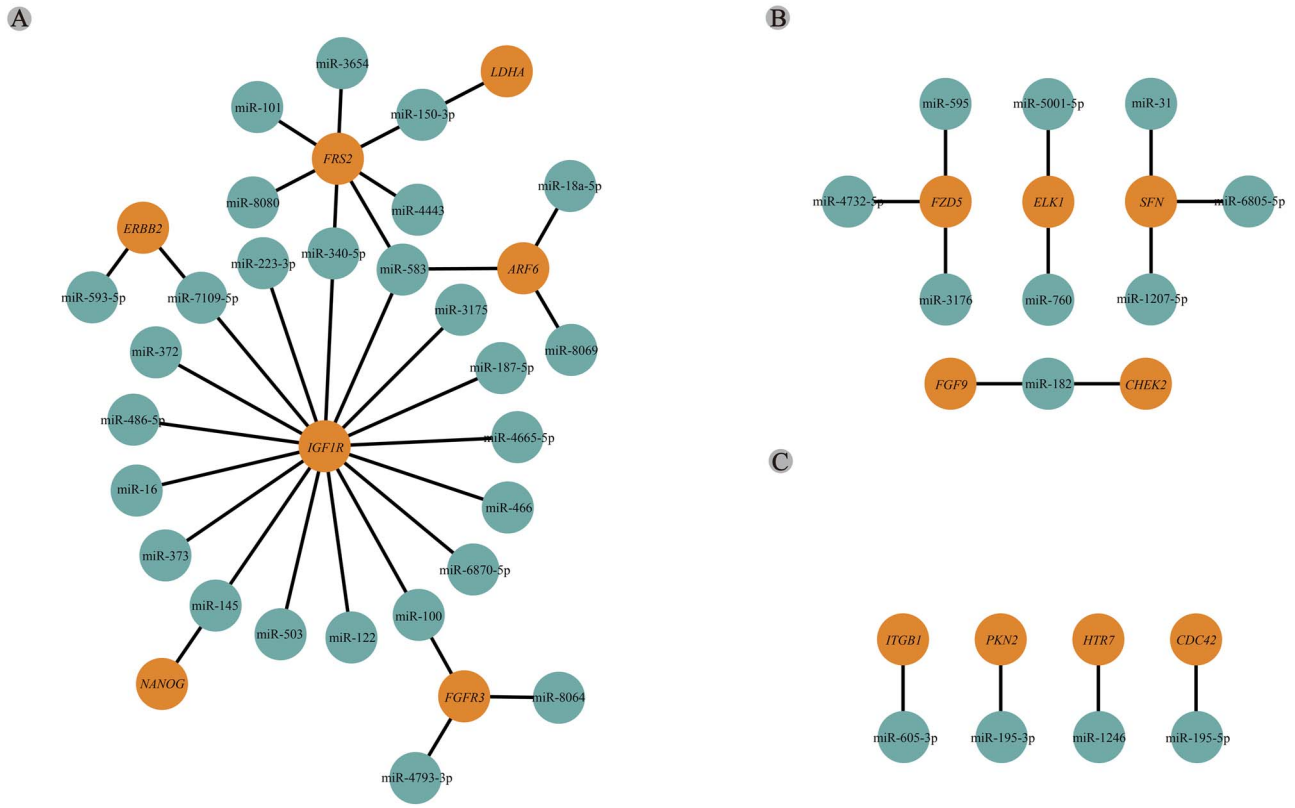

Figure 5. Target relationships of genes and miRNAs. Genes and miRNA were marked in orange and green bubbles, respectively.

breast cancer ($P < 0.001$). Therefore, we thought *miR-1246* should be a valuable biomarker of breast cancer, which was strongly supported by the original study on exosomes in breast cancer from using qRT-PCR and next-generation small RNA sequencing [12].

For another example, users can explore detailed information of *miR-222* on the VALIDATION page (Figure 7C), in which the function of *miR-222*, such as 'the transmission of chemoresistance by horizontal transfer of miRNA' and 'the migration and invasion in advanced breast cancer', was presented. Moreover, the clinical value of *miR-222* in breast cancer treatment can be available in the 'Tumor stage & subtype' and 'Clinical use' section. For example, in the 'Tumor stage & subtype' section, users can get some information about 'localized breast cancer' and 'neoadjuvant chemotherapy'. In addition, the detailed experimental information will provide a valuable reference for further studies.

Discussion

Exosomes are endosomal vehicles of intercellular signalling, vectors of cell therapy, biomarkers and prognosticators of disease. And presently, they have been considered as a promising 'liquid biopsy'. Although many researches on the exosomal biomarkers and related databases in breast cancer have emerged, there is still lacking an integrated and well-displayed knowledge base portal.

Compared with the ExoCarta, Vesiclepedia, EVpedia, exoRBase and EVmiRNA, we carried out an in-depth analysis and data integration to build a new and rich database for making available a huge increase of well-corroborated exosomal molecules in breast cancer from a systemic, knowledge-based perspective. Not only that, the adopted text extraction with a more rigorous ISEV-based standard and the coupled demonstration of well-linked molecules may power a great worth as a resource for researchers and clinicians studied on exosomal molecules in breast cancer. Besides, molecular characters and clinical data or clinical parameters such as tumour stage and subtype, clinical use and statistical estimation of prognosis were integrated and linked, which may enable a deeper understanding of molecular mechanisms and clinical characteristics and allow to make a better decision in the clinic [54].

We proposed a data-driven and literature-based paradigm with corroborative analysis, which can result in confirmable expects. In this study, each exosomal molecule as PB or BIM was required to pass the multiple rounds of selection: the statistical significance of differential gene expression, the large sample screening with the transcriptome profiling of cancerous and normal tissue samples on breast cancer from TCGA, the consistent identification by multiple enrichment (i.e. GO, KEGG and miRNA) and network-based analysis (e.g. hub gene, ceRNA, miRNA-mRNA targeting, multiple molecular pathways, etc.), as well as the reviewing by original reports. As one of the results, 47 of 91 exosomal molecules identified from GSE93070 were reported in original studies. And 41 of them may relate to the initiation and progression of breast cancer, and 22 revealed as potential breast cancer biomarkers (Supplementary Table S6). And all molecules in ExoBCD were verified as BIMs (257), PBs (49), promising BIMs (12), the most promising BIMs (5), promising PBs (17) and the most promising PBs (2) due to their molecular function, the closeness to breast cancer as well as the passed rounds of different selection.

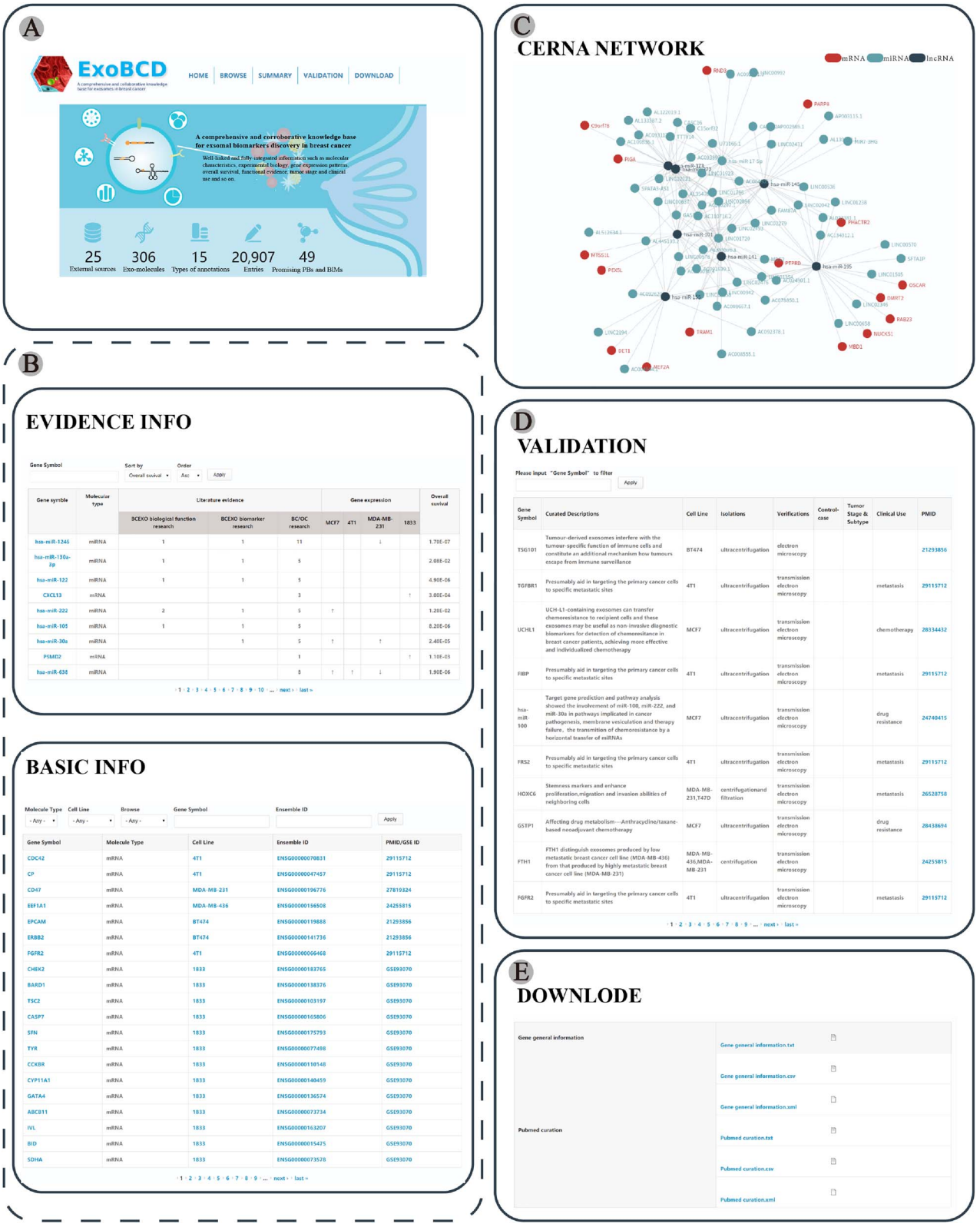

Figure 6. Illustration of the interface of database ExoBCD and relevant information of exosomal molecules. (A) Home page: five navigation menus and functional pages (HOME, BROWSE, SUMMARY, VALIDATION and 'DOWNLOAD') along with the whole data summary were presented. (B) BROWSE page: quick and precise search tool of 'Basic Info' and 'Evidence Info' for exosomal molecules was provided. (C) SUMMARY page: using the ceRNA network as a case, visualization of relevant information can be achieved by moving mouse or dragging. (D) VALIDATION page: experimental evidences and clinical information were integrated and presented. (E) DOWNLOAD page: literature information and functional annotations can be available in CSV, text or XML formats.

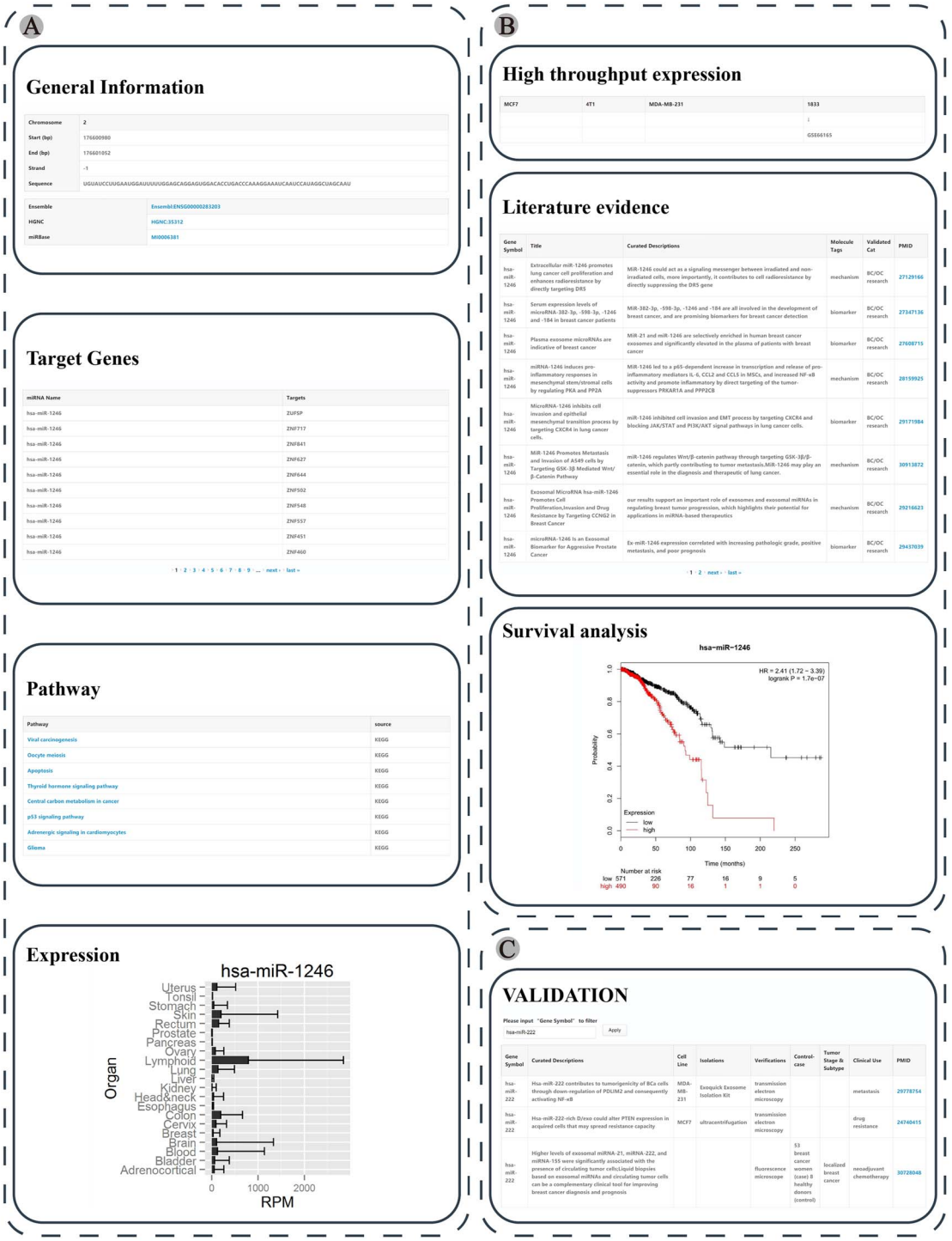

Figure 7. Demonstration of database ExoBCD using miR-1246 and miR-222 as a case. (A) Molecular characters, external database links, target genes and gene expression patterns in different tissues were obtained by the search engine in the BROWSE page using miR-1246 as a keyword. (B) Expression patterns of miR-1246 in cell line 1833, survival analysis plot and the coverage of cancer-related literatures were presented. (C) In the VALIDATION page, clinical and experimental information of miR-222 including curated description, isolation and validation method of exosomes, cell line description, clinical use and tumour stage and type were achieved.

Table 2. Summary of 12 promising PBs

Gene symbol	Signal pathway	Function	Diseases	Reference (PMID)	OS (*P*-value)
†††*IGF1R*[abc]	PI3K-Akt; MAPK; Ras; Rap1; mTOR; HIF-1; longevity regulating	Focal adhesion; adherens junction; endocytosis; EGFR tyrosine kinase inhibitor resistance; signalling pathways regulating pluripotency of stem cells	Breast cancer; hepatocellular carcinoma; melanoma; prostate cancer	29152592; 21574055	3.50E−05
†††*FRS2*[abc]	MAPK; PI3K-Akt; Ras; mTOR	Cell proliferation	Breast cancer	27533459; 29156384	1.60E−03
††*miR-18a-5*[c]	p53; TGF beta	Focal adhesion; cell cycle		29050253; 29988076	9.00E−04
††*miR-195-5p*[c]	p53; TGF beta	Focal adhesion; cell cycle	Melanoma	29050253; 26309570	1.80E−08
††*miR-223-3p*[c]	TGF beta	Focal adhesion; cell cycle		29805680; 31395736	6.00E−04
††*FGFR2*[bc]	MAPK; PI3K-Akt; Ras; Rap1	Endocytosis; EGFR tyrosine kinase inhibitor resistance; regulation of actin cytoskeleton; signalling pathways regulating pluripotency of stem cells	Gastric cancer; Prostate cancer	24043118; 23527311	3.10E−05
††*LDHA*[ab]	HIF-1			27598996; 19668225	9.10E−03
††*SFN*[abc]	p53	Cell cycle		19452229; 11423985	1.33E−02
††*miR-124-3p*[c]	p53; TGF beta	Focal adhesion; cell cycle	Melanoma	29050253; 27468577	4.05E−02
††*miR-130a-3p*[c]	TGF beta	Focal adhesion; cell cycle	Melanoma	29746865; 30510209	2.08E−02
††miR-17-5p[c]	p53; TGF beta	Focal adhesion; cell cycle	Melanoma	29050253; 30989460	2.00E−02
††miR-29b-3p[c]	p53; TGF beta	Focal adhesion; cell cycle	Melanoma	29050253; 31349873	6.50E−06

Notes: (i) 12 promising molecules, including 4 mRNA molecules without the prefix 'mi-' of their gene symbols and 7 miRNAs, were identified by Circos plot, manual mapping of multiple molecular pathways and target relationship. (ii) mRNAs targeted by multiple miRNAs have been marked with a superscript letter a ([a]), and those mapped in multiple molecular pathways have been marked with a superscript letter b ([b]). And mRNAs or miRNAs plotted in Circos have been marked with a superscript letter c ([c]). (iii) The results in the second, third and fourth columns were from the enrichment of KEGG mapping, including pathway, function and disease, respectively. The fifth and sixth columns contain literatures related to breast cancer and overall survival estimation, individually. (iv) All molecules were significant in overall survival (OS) analysis ($P < 0.05$). PMID (PubMed ID) of references with biomarker reporting in breast cancer and the significant OS value ($P < 0.05$) were underlined. (v) All of them were marked with asterisk, superscript or in red colour: potential biomarkers (PBs) are marked with one superscript dagger (†), which were molecules mined from original studies on exosomes in breast cancer or rigorously identified from high-throughput expression data. The promising PBs were marked with two superscript daggers (††), selected by the procedure of 'verification insight into exosomal molecules' (Figure 1) and reported as biomarkers for breast cancer in original studies. After a deeper investigation of the promising PBs, the most promising PBs were marked with three superscript daggers (†††) and set to font in red colour of their gene symbols. They are molecules targeted by multiple exosomal miRNAs in breast cancer, reported as biomarkers for breast cancer in original researches and playing key roles in the progress and prognosis of breast cancer. (vi) *IGF1R* and *FRS2*, the most promising PBs, being significant in prognosis and reported as biomarkers for breast cancers by multiple original studies, were hub molecules targeted by multiple key miRNAs in cancers being crucial for carcinogenesis and cancer development. More molecules and detailed information were shown in Supplementary Table S5.

As a result of the data-driven and literature-based paradigm proposed in this study, the promising molecules including 25 mRNAs and 11 miRNAs were uncovered (Supplementary Table S5). Especially, *IGF1R* and *FRS2* were verified as the most promising prognostic biomarkers according to multi-supported clues and investigation of the latest literatures. For instance, studies have shown that the insulin-like growth factor receptor (*IGF1R*) is a crucial member of the insulin signalling pathway and involved in the growth, survival and death of cancer cells. Clinically, more than 50% of breast cancers are triggered by *IGF1R* and subsequently activate other intermediate signalling proteins such as PI3K, RAS and MAPK. The compensatory cross talk between *IGF1R* and its receptor probably leads to poor therapy prediction outcomes [55]. As for *FRS2*, it is the fibroblast growth factor receptor substrate 2 which was found to play an important role in regulation of tumour cell differentiation, proliferation and tumorigenesis [56]. Meanwhile, being a key adaptor protein, *FRS2* was considered as an essential 'conning centre' in FGFR signalling, which had important functions in RAS/MAPK/ERK and PI3K/AKT/mTOR pathways [57]. For example, Manuvakhova *et al.* revealed the expression of the SNT-1/FRS2 phosphoserine binding domain inhibited activation of MAP kinase and PI3-kinase pathways. Additionally, *FRS2* was targeted by *miR-4653-3p*, and its overexpression was a potential high-risk factor for relapse in adjuvant treatment of HR+ breast cancer patients [58]. Besides, some exosomal molecules (e.g. *CDC42*, *ITGB1* and *ARF6*) may play key roles in the causes of breast cancer. Previous studies showed that *CDC42* was involved in the regulation of actin cytoskeleton

and triggers the actin-dependent and clathrin- and lipid raft-independent endocytosis in breast epithelial cells, which further promoted the entry of HPV pseudovirion and subsequently causes breast cancer [59, 60].

However, some limitations of this study should be noted. First, the statistical significance of involved molecules in ExoBCD could be affected due to the small samples. In this study, according to the proposed and data standard, 4 desired exosomal datasets, containing 42 samples in total, were collected availably from the public transcriptomic data repositories, but there were only 4 samples in the datasets of GSE93070 for the identification of mRNAs and lncRNAs (Supplementary Table S7). Although a rigorous process proposed as a data-driven and literature-based paradigm was employed for a satisfactory result (Supplementary Table S6), statistically significant risks still exist. So, we will integrate more exosomal datasets on breast cancer available in public repositories, timely, in the continued study. Second, the discovery of exosomal markers for breast cancer needs well-linked and better-graphed knowledge, which links to molecular mechanisms, tumour occurrence, cancer progress, clinical prognosis, etc. The clinical data such as tumour stage and subtype, clinical use and statistical estimation of prognosis have been integrated into ExoBCD, but still not enough. In the future, more combinations of cell-specific characteristics, multi-omics signatures, multiple molecular pathways, etc. will aid in the biomarker discovery and its precise implementation.

Conclusion

Based on a robust analysis of four high-throughput data and deep integration of 950 exosome-related literature mining, the first exosomal molecule database ExoBCD with high-quality annotations and evidentiary supports was constructed. In ExoBCD, not only a complete list of 49 PBs and 257 BIMs (306 molecules including 121 mRNAs, 172 miRNAs and 13 lncRNAs) was supplied but also correspondingly 4 information categories (molecular characters, descriptions of exosomal experiments, evidences for biomarkers and clinical use) and 15 detailed subcategories (~20 900 entries) with integration of 25 external sources, all of which were processed and well-linked. As a well-corroborated knowledge base, ExoBCD facilitates reliable biomarker mining from exosomes, which makes the reveal of 36 promising molecules (12 promising BIMs, 5 most promising BIMs, 17 promising PBs and 2 most promising PBs) with a data-driven and literature-based paradigm proposed in the study. Especially, two of the most promising PBs (*IGF1R* and *FRS2*) were verified as exosomal biomarkers for the prognostic of breast cancer. Taken together, this study is not only helpful for exosome-related studies on breast cancer but also serves as a valuable platform for the discovery of biomarkers in breast exosomes.

Key Points

- ExoBCD presents a visualized and systematic view of 306 exosomal molecules in breast cancer, which is the result of a robust analysis of 4 high-throughput datasets, validation against 1191 TCGA breast cancer cases and investigation of 950 related literatures, along with the integration of 29 annotation sources.
- Verification insight into exosomal molecules deduced 36 promising molecules (12 promising BIMs, 5 most promising BIMs, 17 promising PBs and 2 most promising PBs) and 2 of the most promising PBs (*IGF1R* and *FRS2*) for the prognosis of breast cancer according to a data-driven and literature-based paradigm proposed in this study, and case studies were described using the potential biomarkers (*miR-1246* and *miR-222*).
- ExoBCD is the first exosomal molecule database with a multi-faceted and well-corroborated annotative knowledge, including transcript-level expression analysis, clinic-level survival prediction and literature-level mining, which may work as an important and comprehensive resource for researchers and clinicians studying on exosomal molecules in breast cancer.

Supplementary Data

Supplementary data are available online at https://academic.oup.com/bib.

Acknowledgements

The authors are grateful for the help from Dr Lu Xie, Shanghai Center for Bioinformation Technology for the helpful suggestions.

Funding

Science Innovation Program of College of Laboratory Medicine, Chongqing Medical University (CX201704); Science and Technology Innovation Commission of Shenzhen (JCJY20170818094217688); Natural Science Foundation of Chongqing of China (cstc2019jcyj-msxmX0271, cstc2019jcyj-msxmX0527); Science and Technology Research Program of Chongqing Municipal Education Commission (KJQN201800523).

References

1. Bray F, Ferlay J, Soerjomataram I, *et al*. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;**68**:394–424.
2. Jameson JL, Longo DL. Precision medicine–personalized, problematic, and promising. *N Engl J Med* 2015;**372**:2229–34.
3. Vargas AJ, Harris CC. Biomarker development in the precision medicine era: lung cancer as a case study. *Nat Rev Cancer* 2016;**16**:525–37.
4. Beane J, Campbell JD, Lel J, *et al*. Genomic approaches to accelerate cancer interception. *Lancet Oncol* 2017;**18**: e494–502.
5. Loke SY, Lee ASG. The future of blood-based biomarkers for the early detection of breast cancer. *Eur J Cancer* 2018;**92**:54–68.
6. Zhang X, Yuan X, Shi H, *et al*. Exosomes in cancer: small particle, big player. *J Hematol Oncol* 2015;**8**:83.
7. Soung YH, Ford S, Zhang V, *et al*. Exosomes in cancer diagnostics. *Cancers (Basel)* 2017;**9**(1):8.
8. Xu R, Rai A, Chen M, *et al*. Extracellular vesicles in cancer - implications for future improvements in cancer care. *Nat Rev Clin Oncol* 2018;**15**:617–38.

9. Di Meo A, Bartlett J, Cheng Y, *et al*. Liquid biopsy: a step forward towards precision medicine in urologic malignancies. *Mol Cancer* 2017;**16**:80.
10. Khan S, Bennit HF, Turay D, *et al*. Early diagnostic value of survivin and its alternative splice variants in breast cancer. *BMC Cancer* 2014;**14**:176.
11. Melo SA, Sugimoto H, O'Connell JT, *et al*. Cancer exosomes perform cell-independent microRNA biogenesis and promote tumorigenesis. *Cancer Cell* 2014;**26**:707–21.
12. Hannafon BN, Trigoso YD, Calloway CL, *et al*. Plasma exosome microRNAs are indicative of breast cancer. *Breast Cancer Res* 2016;**18**:90.
13. Tan C, Cao J, Chen L, *et al*. Noncoding RNAs serve as diagnosis and prognosis biomarkers for hepatocellular carcinoma. *Clin Chem* 2019;**65**:905–15.
14. Keerthikumar S, Chisanga D, Ariyaratne D, *et al*. ExoCarta: a web-based compendium of Exosomal cargo. *J Mol Biol* 2016;**428**:688–92.
15. Pathan M, Fonseka P, Chitti SV, *et al*. Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. *Nucleic Acids Res* 2019;**47**: D516–9.
16. Kim DK, Lee J, Simpson RJ, *et al*. EVpedia: a community web resource for prokaryotic and eukaryotic extracellular vesicles research. *Semin Cell Dev Biol* 2015;**40**:4–7.
17. Li S, Li Y, Chen B, *et al*. exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. *Nucleic Acids Res* 2018;**46**:D106–12.
18. Liu T, Zhang Q, Zhang J, *et al*. EVmiRNA: a database of miRNA profiling in extracellular vesicles. *Nucleic Acids Res* 2019;**47**:D89–93.
19. Clough E, Barrett T. The gene expression omnibus database. *Methods Mol Biol* 2016;**1418**:93–110.
20. Parkinson H, Kapushesky M, Shojatalab M, *et al*. ArrayExpress–a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 2007;**35**:D747–50.
21. Papatheodorou I, Fonseca NA, Keays M, *et al*. Expression atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* 2018;**46**:D246–51.
22. Kodama Y, Mashima J, Kosuge T, *et al*. DDBJ update: the genomic expression archive (GEA) for functional genomics data. *Nucleic Acids Res* 2019;**47**:D69–73.
23. Cerami E, Gao J, Dogrusoz U, *et al*. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2012;**2**:401–4.
24. Perez-Riverol Y, Bai M, da Veiga LF, *et al*. Discovering and linking public omics data sets using the omics discovery index. *Nat Biotechnol* 2017;**35**:406–9.
25. Chen X, Gururaj AE, Ozyurt B, *et al*. DataMed - an open source discovery index for finding biomedical datasets. *J Am Med Inform Assoc* 2018;**25**:300–8.
26. Ritchie ME, Phipson B, Wu D, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;**43**:e47.
27. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;**26**:139–40.
28. Li JR, Sun CH, Li W, *et al*. Cancer RNA-Seq nexus: a database of phenotype-specific transcriptome profiling in cancer cells. *Nucleic Acids Res* 2016;**44**:D944–51.
29. Chin CH, Chen SH, Wu HH, *et al*. cytoHubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 2014;**8**(Suppl 4):S11.
30. Zhou G, Soufan O, Ewald J, *et al*. NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 2019;**47**:W234–41.
31. Stelzer G, Plaschkes I, Oz-Levi D, *et al*. VarElect: the phenotype-based variation prioritizer of the GeneCards suite. *BMC Genomics* 2016;**17**(Suppl 2):444.
32. Shen L, Lin Y, Sun Z, *et al*. Knowledge-guided bioinformatics model for identifying autism spectrum disorder diagnostic MicroRNA biomarkers. *Sci Rep* 2016;**6**:39663.
33. Cook KB, Kazan H, Zuberi K, *et al*. RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res* 2011;**39**:D301–8.
34. Kinsella RJ, Kahari A, Haider S, *et al*. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* 2011;**2011**:bar030.
35. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. *Nucleic Acids Res* 2019;**47**:D155–62.
36. Kanehisa M, Furumichi M, Tanabe M, *et al*. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 2017;**45**:D353–61.
37. Tang Z, Kang B, Li C, *et al*. GEPIA2: an enhanced web server for large-scale expression profiling and interactive analysis. *Nucleic Acids Res* 2019;**47**:W556–60.
38. Backes C, Khaleeq QT, Meese E, *et al*. miEAA: microRNA enrichment analysis and annotation. *Nucleic Acids Res* 2016;**44**:W110–6.
39. Sueta A, Yamamoto Y, Tomiguchi M, *et al*. Differential expression of exosomal miRNAs between breast cancer patients with and without recurrence. *Oncotarget* 2017;**8**:69934–44.
40. Bovy N, Blomme B, Freres P, *et al*. Endothelial exosomes contribute to the antitumor response during breast cancer neoadjuvant chemotherapy via microRNA transfer. *Oncotarget* 2015;**6**:10253–66.
41. Yoshikawa M, Iinuma H, Umemoto Y, *et al*. Exosome-encapsulated microRNA-223-3p as a minimally invasive biomarker for the early detection of invasive breast cancer. *Oncol Lett* 2018;**15**:9584–92.
42. Eichelser C, Stuckrath I, Muller V, *et al*. Increased serum levels of circulating exosomal microRNA-373 in receptor-negative breast cancer patients. *Oncotarget* 2014;**5**: 9650–63.
43. Gonzalez-Villasana V, Rashed MH, Gonzalez-Cantu Y, *et al*. Presence of circulating miR-145, miR-155, and miR-382 in exosomes isolated from serum of breast cancer patients and healthy donors. *Dis Markers* 2019;**2019**:6852917.
44. Chen WX, Liu XM, Lv MM, *et al*. Exosomes from drug-resistant breast cancer cells transmit chemoresistance by a horizontal transfer of microRNAs. *PLoS One* 2014;**9**:e95240.
45. Palma J, Yaddanapudi SC, Pigati L, *et al*. MicroRNAs are exported from malignant cells in customized particles. *Nucleic Acids Res* 2012;**40**:9125–38.
46. Zhang WC, Chin TM, Yang H, *et al*. Tumour-initiating cell-specific miR-1246 and miR-1290 expression converge to promote non-small cell lung cancer progression. *Nat Commun* 2016;**7**:11702.
47. Chai S, Ng KY, Tong M, *et al*. Octamer 4/microRNA-1246 signaling axis drives Wnt/beta-catenin activation in liver cancer stem cells. *Hepatology* 2016;**64**:2062–76.
48. Lazzarini R, Caffarini M, Delli Carpini G, *et al*. From 2646 to 15: differentially regulated microRNAs between progenitors from normal myometrium and leiomyoma. *Am J Obstet Gynecol* 2019;**S0002–9378(19)**:32775–9.

49. Yang F, Xiong H, Duan L, *et al*. MiR-1246 promotes metastasis and invasion of A549 cells by targeting GSK-3betaMediated Wnt/beta-catenin pathway. *Cancer Res Treat* 2019;**51**:1420–9.
50. Hoshino I, Yokota H, Ishige F, *et al*. Radiogenomics predicts the expression of microRNA-1246 in the serum of esophageal cancer patients. *Sci Rep* 2020;**10**:2532.
51. Hasegawa S, Eguchi H, Nagano H, *et al*. MicroRNA-1246 expression associated with CCNG2-mediated chemoresistance and stemness in pancreatic cancer. *Br J Cancer* 2014; **111**:1572–80.
52. Li XJ, Ren ZJ, Tang JH, *et al*. Exosomal MicroRNA MiR-1246 promotes cell proliferation, invasion and drug resistance by targeting CCNG2 in breast cancer. *Cell Physiol Biochem* 2017;**44**:1741–8.
53. Lin SS, Peng CY, Liao YW, *et al*. miR-1246 targets CCNG2 to enhance cancer Stemness and Chemoresistance in oral carcinomas. *Cancers (Basel)* 2018;**10**(8):272.
54. Taherian-Fard A, Srihari S, Ragan MA. Breast cancer classification: linking molecular mechanisms to disease prognosis. *Brief Bioinform* 2015;**16**:461–74.
55. Ekyalongo RC, Yee D. Revisiting the IGF-1R as a breast cancer target. *NPJ Precis Oncol* 2017;**1**:14.
56. Kouhara H, Hadari YR, Spivak-Kroizman T, *et al*. A lipid-anchored Grb2-binding protein that links FGF-receptor activation to the Ras/MAPK signaling pathway. *Cell* 1997; **89**:693–702.
57. Gotoh N. Regulation of growth factor signaling by FRS2 family docking/scaffold adaptor proteins. *Cancer Sci* 2008; **99**:1319–25.
58. Zhong X, Xie G, Zhang Z, *et al*. MiR-4653-3p and its target gene FRS2 are prognostic biomarkers for hormone receptor positive breast cancer patients receiving tamoxifen as adjuvant endocrine therapy. *Oncotarget* 2016;**7**:61166–82.
59. Khodabandehlou N, Mostafaei S, Etemadi A, *et al*. Human papilloma virus and breast cancer: the role of inflammation and viral expressed proteins. *BMC Cancer* 2019;**19**:61.
60. Schelhaas M, Shah B, Holzer M, *et al*. Entry of human papillomavirus type 16 by actin-dependent, clathrin- and lipid raft-independent endocytosis. *PLoS Pathog* 2012;**8**: e1002657.