ExomiRHub: a database to explore human extracellular and intracellular microRNA transcriptomics data

Yang Liu  
Guangzhou Medical University

Zhuochao Min  
Tel Aviv University

Jing Mo  
Outstanding Biotechnology Co., Ltd.-Shenzhen

Zhen Ju  
Chinese Academy of Sciences

Jianliang Chen  
The University of Hong Kong-Shenzhen Hospital

Hanguang Li  
The University of Hong Kong-Shenzhen Hospital

Weiling Liang  
The University of Hong Kong-Shenzhen Hospital

Godfrey Chi-Fung Chan  
The University of Hong Kong-Shenzhen Hospital

Yanjie Wei  
Chinese Academy of Sciences

Wenliang Zhang (zhangwl25@mail3.sysu.edu.cn)  
Guangzhou Medical University

Research Article

Keywords: miRNA, Extracellular vesicle, Exosome, Cancer, Liquid biopsy

Posted Date: January 4th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-2566749/v2

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

Additional Declarations: The authors declare no competing interests.
Abstract

Extracellular microRNA (miRNA) expression data, generated by different laboratories, exhibit heterogeneity, posing challenges for researchers lacking bioinformatics expertise to explore these valuable data. To expedite the discovery of clinically relevant miRNA biomarkers, a user-friendly database is mandatory. Therefore, we formulated ExomiRHub, a database that incorporates 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, including 29,198 samples and 23 sample types. To enhance usability, ExomiRHub integrates 16,012 miRNA transcriptomes of 156 cancer subtypes from The Cancer Genome Atlas. Instead of mere collection, ExomiRHub standardizes and annotates the datasets and samples with rich annotations. In addition, it offers 25 analytical and visualization functions to interpret these datasets for identification of non-invasive miRNA biomarkers. These 25 functions empower users to select samples, define groups, and set parameters for personalized analyses. Moreover, ExomiRHub offers a web service enabling users to conduct analyses on their uploaded data. To further assist users, four additional tools are designed for evaluating the functions and targets of miRNAs and their variations. In a notable application of ExomiRHub, we identified non-invasive miRNA biomarkers associated with angiogenesis for monitoring glioma progression. This exemplifies how ExomiRHub can significantly expedite the discovery of non-invasive miRNA biomarkers. ExomiRHub is available at http://www.biomedical-web.com/exomirhub/.

Background

MicroRNAs (miRNAs), a class of small noncoding RNA, which are widely expressed in tissue cells and circulate in human biofluids to maintain cell homeostasis mainly by suppressing gene regulation [1]. miRNAs have been confirmed to serve as a powerful diagnostic and therapeutic tool [2,3]. Recently, miRNAs were found to be carried and secreted via extracellular vesicles such as exosomes or circulated in biofluids to mediate intercellular and intra-organ crosstalk between cells in different tissues, thus regulating the gene expression and function of distant tissue cells [4,5]. Although studies suggested that extracellular miRNA are associated with human diseases and can be used as a promising biomarker for disease diagnosis and monitoring [6,7], the exact mechanisms of action in human diseases remains elusive.

To investigate the clinical significance of extracellular miRNA, several valuable databases have been developed to collect experimentally supported disease-extracellular miRNA association data from publications through manual curation [8–11], and analysis of limited miRNA expression data [12–15]. For example, the databases of miRandola [8], Vesiclepedia [9], EVpedia [10], and ExoCarta [11] were created to offer extracellular miRNA curated studies, both published and unpublished. In contrary, CMEP [12], miREV [14], EVAtlas [15], and EVpedia [16] databases were designed to present circulating miRNA expression profiles, incorporating functions for differential expression and KEGG functional enrichment. Additionally, the ExoBCD [13] database focuses on providing exosomal miRNA association data in breast cancer, achieved through the analysis of four high-throughput datasets and manual literature mining. Notably, Li et al. have recently introduced the CancerMIRNome [17] database, which integrates and visualizes circulating and tissue miRNA expression data in human cancers. CancerMIRNome lacks various co-expression analytical functions such as weighted gene co-expression network analysis (WGCNA). Furthermore, CancerMIRNome does not permit users to select specific samples, define their own groups, and set parameters for different analyses, or analyze uploaded miRNA expression data.

Over the past decade, a substantial amount of human disease-related extracellular miRNA expression data has accumulated in National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [18,19]. However, these data are heterogeneous and pose a challenge for bench scientists without bioinformatics expertise to explore effectively. The identification of disease-related extracellular miRNA from these valuable datasets remains a significant hurdle. There is a pressing need for the development of a dedicated online resource capable of comprehensively collecting, analyzing, and visualizing miRNA expression data. Such a database has potential to greatly accelerate our understanding of the significance of extracellular miRNA in human diseases.

To tackle the aforementioned challenges, we introduced the ExomiRHub database (http://www.biomedical-web.com/exomirhub/). Its primary goal is to comprehensively integrate human extracellular and intracellular miRNA expression data, providing customizable analytical and visualization tools to facilitate the identification and discovery of non-invasive miRNA biomarkers. To cater for a broader research community, ExomiRHub offers a web service application allowing users to upload their miRNA expression data along with sample information, enabling them to perform comprehensive analyses on their own uploaded data. Additionally, ExomiRHub offers four extra tools for predicting and validating potential functions and targets of miRNAs and their variations. Designed for user-friendliness, ExomiRHub offers extensive data analytical and visualization functions for free searching, browsing, analysis,
downloading, and data submission. In summary, we envisage ExomiRHub as a vital resource in accelerating the identification and discovery of non-invasive miRNA biomarkers.

**Materials And Methods**

**Data collection and processing**

**Data collection, standardization, and annotation**

We performed extensive search in NCBI GEO for collecting dataset included human disease-related extracellular miRNA expression data. A total of 395 candidate datasets were screened through the searching strategy of “(((exosome [Title] OR exosomal [Title] OR extracellular vesicle [Title] OR plasma [Title] OR circulating [Title] OR serum [Title])) AND (miRNA [Title] OR microRNA [Title])) AND Homo sapiens [porgn:_txid9606])”. Moreover, we identified 191 human extracellular miRNA expression datasets from those candidate datasets according to two criteria: (1) the dataset including miRNA expression data from the extracellular vesicles and exosomes secreted by human tissues and cells, or circulating miRNA expression data from human biofluids such as plasma and serum; (2) the expression data and sample information of the dataset could be freely downloaded and integrated. We further manually curated each dataset and annotated it with rich biomedical information, including disease phenotype, grade, stage, drug, infection, metastasis, genotype etc. Furthermore, the sample metadata and normalized expression data of the identified datasets were downloaded programmatically through GEOquery [20], whereas the non-normalized expression data were downloaded manually.

It was found that the 191 datasets were generated from 14 different platforms, which leads to the heterogeneity among these datasets and further hinders their integration analysis and application. To remove the data heterogeneity, an in-house pipeline was designed to standardize the miRNA symbols/probes and sample identifiers in the datasets to the miRNA and GEO sample accession identifiers based on the annotation data downloaded from the miRBase Release 22.1 [21] and NCBI GEO, respectively.

Given that different isolation methods and detection platforms have significant bias on the characterization and quantification of extracellular miRNA. To avoid the bias, we independently collected and standardized each dataset from the same study, which used the same method and platform to capture, characterize, and quantify the extracellular miRNA, instead of integrating all datasets from different studies into a comprehensive dataset for comparison analysis. Although the datasets were generated from 14 different platforms, we classified the platforms into two categories, including next generation sequencing (NGS) and microarray. To resolve the problems of data heterogeneity, we independently standardized and normalized the miRNA symbol, expression value and sample identifier in each dataset from the same study. For the datasets from NGS, we transformed the read counts in the non-normalized expression data to Reads Per Million mapped reads (RPM) and further transformed by log2, while log2 transformation was applied to the original normalized expression data if it had not been applied by this transformation. For the rest of the datasets from microarray, we downloaded the original standardized expression data and apply log2 transformation to it, if the transformation does not apply it. If multiple probes match the same miRNA identifier, only the probe with the highest mean expression value was retained.

To develop the TCGA-miRlyzer web application, we created an in-house shell script to download the TCGA miRNA expression quantification data based on the curl API method provided by National Cancer Institute (NCI) Genomic Data Commons (GDC) [22], which are derived from human tissues of cancer patient through miRNA sequencing and have been standardized using the same pipeline. The miRNA expression quantification data were further normalized through log2 transformation and merged into a unified expression matrix saved in an RDS file. Furthermore, the metadata of each sample were obtained from NCI GDC to enrich the sample annotations with comprehensive biospecimen and clinical information. This information includes details such as sample type, cancer subtype, follow-up data, site of resection or biopsy, tissue or organ of origin, age, gender, race, therapy, various diagnosis and grading/staging information, tumor infiltration and necrosis information, progression or relapse, and lifestyle.

**Implementation of the ExomiRlyzer and TCGA-miRlyzer web applications**

ExomiRlyzer and TCGA-miRlyzer were implemented to offer four useful tool-kits, encompassing the differential expression tool-kit, co-expression tool-kit, WGCNA tool-kit, and feature selection tool-kit (**Figure 1**). Additionally, within these tool-kits, 25 analytical and visualization functions have been incorporated (**Figure 1**). These functions are designed to conduct comprehensive analysis and visualization on a individual miRNA expression data, covering various aspects of differential expression, co-expression, WGCNA,
feature selection, and gene ontology (GO) function enrichment analysis (Figure 2). The analytical and visualization functions were designed with a customized grouping and setting manner, and they can enable users to select specific samples, define groups, and set parameters for personalized analyses (Figure S1 and S2). In addition, all the functions were implemented with advanced visualizations and can support to generate publication-quality vector images in Portable Document Format (PDF) and user-friendly tables for further analysis and export. To formulate suitable, robust, and reliable hypotheses, all comparison analyses are structured on an individual miRNA expression data from the same study, rather than across different datasets and studies. The 25 functions were both implemented in R project (https://www.r-project.org/). Table S1 has described the details of R packages used to implement these functions on the four tool-kits.

Web service

The web service application was developed using R language, it serves as a pipeline, allowing users to upload and standardize their own miRNA expression data and sample metadata. This enables further customizable analysis of the uploaded data within the database. (Figure 3). For example, the application can standardize miRNA identifiers and expression values, converting read counts to RPM and subsequently normalizing them through log2 transformation. Additionally, human miRNA annotation data were downloaded and extracted from miRBase, supporting the web service to convert different miRNA accessions and symbols to primary/mature miRNA identifiers in miRBase. Table S1 has described R packages and resources utilized in the application implementation. The application was encapsulated using shell scripts and deployed to our local server.

miRNA target prediction and miRNA mutation evaluation

It has been suggested that miRNA can bind to the target of mRNA, circRNA, and IncRNA thereby regulating their biological functions [23], while the genomics variation in miRNA can impact its biological function in human diseases [24]. Consequently, we implemented miRNA target prediction and miRNA mutation evaluation tools to predict miRNA targets and evaluate the potential impacts of miRNA variation. To develop these tools, we initially extracted the sequences of human mRNA/lncRNA, circRNA, and miRNA seed region from the resources of GENCODE 2021 [25], circBase [26], and miRBase, respectively. Furthermore, we installed miRanda [27] and scripted a shell script to encapsulate it with the aforementioned extracted sequences. To further implement the miRNA mutation evaluation tool, an R script was designed to compute the gain or loss targets of miRNA, thereby evaluating the potential impact of its mutation.

miRNA function prediction and miRNA target validation

To develop the miRNA function prediction and miRNA target validation tools, we initially integrated experimentally validated miRNA-target interaction data from miRTarBase [23]. Subsequently, we authored R scripts to implement miRNA target validation, systematically identifying experimentally validated targets of miRNA. To predict the biological function of a specific miRNA of interest, we further implemented miRNA function prediction tool utilizing the clusterProfiler, org.Hs.eg.db, pathview, and topGO packages, in conjunction with the miRNA validated targets in miRTarBase.

Web implementation

The front-end and back-end framework utilized for the implementation of ExomiRHub has been previously detailed [28,29]. The JAVA programming language was employed for application operation and data processing. MySQL was used for storing and organizing association data, facilitating faster data browsing and searching. R 4.3.2 was installed to support and execute analytical functions, servers and tools on ExomiRHub. The specific R packages and resources employed are outlined in Table S1. Finally, ExomiRHub was deployed on the Apache Tomcat server and is available at http://www.biomedical-web.com/exomirhub/.
Database content and usage

Data landscape and access

ExomiRHub offers access to 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, including 2656 miRNAs, 29,198 samples, and 23 sample types (Figure 1A and Table 1). Statistically, 80.63% (154/191) of the datasets include miRNA expression profile from extracellular vesicle, including 145 datasets of exosome. The remaining 19.37% (37/191) present circulating miRNA profiles of serum (13.09%, 25/191), plasma (4.71%, 9/191), and whole blood (1.57%, 3/191). Sample types encompass extracellular vesicles, exosomes, and circulating miRNA derived from various sources, such as whole blood, serum, plasma, urinary, ascites, cerebrospinal fluid, endometrial fluid, follicular fluid, pericardial fluid, liquid milk, saliva, tissue fluid, umbilical cord blood, tracheal aspirate, fecal fluid, and the culture supernatant of cells with specific genotypes and treatments. The treatments span adjuvant chemotherapy, antibiotics, cell therapy, chemotherapy, immunotherapy, radiotherapy, target therapy, etc. Further statistical analysis indicates that the disease phenotypes included in ExomiRHub were associated with more than 23 body parts. Notably, the blood system, breast, and lung are the three body parts with the largest number of extracellular miRNA expression datasets. The top three contributing countries to the datasets are China, Japan, and United States, accounting for 25.65% (49/191), 24.08% (46/191), and 20.42% (39/191), respectively. Detailed data features are summarized on the “Home” webpage of the database, including body site, sample type, isolation method, sample resource, extraction method, disease phenotype, genotype, and treatment.

At present, approximately 63.87% (122/191) of the extracellular miRNA datasets are linked with 52 cancer subtypes. To augment the utility in the realm of cancer research, ExomiRHub has additionally incorporated human miRNA expression quantification data comprising 16,012 samples and spanning 156 cancer subtypes from 42 TCGA projects (Figure 1A). Each sample derived from the TCGA projects has been annotated with comprehensive biospecimen and clinical information, encompassing various demographics, diagnosis, progression, tumor microenvironment, and lifestyle. This annotation facilitates users in browsing and selecting specific samples for the purpose of designing and defining analytical and visual comparisons.

Moreover, ExomiRHub facilitates quick retrieval of interesting extracellular miRNA expression dataset. Users can then seamlessly navigate to the ExomiRlyzer application to discover significant extracellular miRNA through user-defined investigations on the dataset. All data in ExomiRHub is available in various formats, such as TABLE, CSV, and JSON, enabling users to freely download and analyze. Furthermore, ExomiRHub encourages users to submit their novel human extracellular miRNA expression data to the database through the submit webpage (http://www.biomedical-web.com/exomirhub/submit). Once the submitted data undergoes approval by our data submission review committee, they will be incorporated in the next version of the database.

Comprehensive web analytical and visualization function, server, and tool

To understand the role of extracellular miRNA and discover non-invasive biomarkers in human disease, ExomiRHub features the web applications of ExomiRlyzer and TCGA-miRNAlyzer. These applications are designed to offer four state-of-the-art bioinformatics tool-kits, including differential expression tool-kit, co-expression tool-kit, WGCNA tool-kit, and feature selection tool-kit. These four tool-kits collectively offer 25 analytical and visualization functions to analyze human extracellular miRNA expression data and cancer related miRNA expression data. The functions encompass various aspects of differential expression, co-expression, WGCNA, GO function enrichment, COX regression, receiver operating characteristic (ROC), least absolute shrinkage and selection operator (LASSO), and survival analysis. Given that each dataset's samples are annotated with comprehensive biomedical information, all 25 functions are structured with a customized grouping and setting manner. This allows users to select specific samples, define their own groups, and set parameters for personalized comparison analyses (Figure S1 and S2). Recognizing variations in isolation methods, detection platforms, and quality control processes among datasets from different studies, ExomiRHub restricts users to comparing and analyzing datasets exclusively within the context of the same study. This approach ensures the data quality and the reliability of analysis results. It is not recommended for users to download and integrate the datasets from different studies for further comparison analysis.

To serve a wider community of research, ExomiRHub offers a web service application, which enables users to upload and standardize the miRNA expression data and its sample metadata and assigned the uploaded data with a temporary identifier (Figure 3), so that they can further comprehensively analyze and visualize their uploaded data in the co-expression, differential expression, and WGCNA tool-
To enhance the utility of ExomiRHub in the miRNA research community, it offers four additional useful tools for predicting and validating the potential function and target (such as mRNA, IncRNA, and circRNA) of miRNA and its variation. The four tools include: (1) miRNA function prediction: predicting the biological function of miRNA based on the GO and KEGG annotations of its experimental validated targets in miRTarBase; (2) miRNA target prediction: predicting miRNA targets using miRanda, including mRNA, circRNA, and IncRNA; (3) miRNA mutant evaluation: predicting the gain or loss targets of miRNA after its mutation; (4) miRNA target validation: identifying the experimental validated miRNA targets in miRTarBase. A usage guide for these analytical and visualization functions, servers, and tools is available on the “Help” webpage (http://www.biomedical-web.com/exomirhub/help). Last but not least, the results from these applications can generate publication-quality vector figures in PDF and tables for further analysis and download (Figure 2).

**Case study: discover non-invasive miRNA biomarkers associated with angiogenesis for diagnosis and monitoring progression of glioma**

To discover and identify non-invasive biomarkers for the diagnosis and monitoring of glioma, we conducted a comprehensive analysis on an exosomal miRNA dataset in ExomiRHub (ID: EMIR00000186). This dataset provides miRNA expression profiles derived from plasma exosomes of healthy controls and glioblastoma patients. Firstly, WGCNA results demonstrated that plasma exosomal miRNAs could effectively distinguish patients with glioblastoma from healthy controls, dividing the patients into two subgroups named G1 and G2 (Figure 4A). Subsequently, hierarchical cluster analysis confirmed distinct exosomal miRNA expression profiles between the two subgroups (Figure 4B). Further analysis was performed on the two subgroups, identifying four modules of eigengenes (MEbrown, MEred, MEyellow, and MEturquoise) associated with the subgroups (Figure 4C). These modules showed significant correlation with each other (Figure 4D). Additional analysis revealed that these four modules of eigengenes consistently enriched in the GO terms related to the regulation of angiogenesis and vasculature development (Figure 4E–H). These GO terms are known to be closely associated with the development, treatment, and prognosis of glioma [30,31].

To elucidate the difference in angiogenesis related pathway between G1 and G2, we conducted a differential analysis and identified 48 dysregulated exosomal miRNAs (|log2(fold change)| ≥ 0.5 and P value ≤ 0.05) between the two subgroups, including 21 down-regulated and 27 up-regulated exosomal miRNAs (Figure 5A and Table S2). In line with the GO function enrichment analysis on the module eigengenes, we confirmed that these differential expression exosomal miRNAs were enriched in the angiogenesis related pathways (Figure 4I). For example, the dysregulation of hsa-miR-132-5p and hsa-miR-200a-5p was associated with the positive regulation of blood vessel endothelial cell migration (Figure 4I and Figure 5A–C), and these two miRNAs showed a significant correlation (Figure 5D). Furthermore, the function of hsa-miR-132-5p and hsa-miR-200a-5p consistently enriched in the angiogenesis related pathways based on the GO annotations of their co-expression exosomal miRNAs (Figure S3).

To further elucidate the clinical significance of the two exosomal miRNAs in glioma, we conducted COX regression and survival analysis on glioma datasets integrated from TCGA within ExomiRHub. The analysis results indicated that the down-regulation of angiogenic miRNAs hsa-miR-132 and hsa-miR-200a significantly decreased the death event (Figure 6A) and prolonged the overall survival probability of patients with glioma (Figure 6B). In consistent with the aforementioned results, the two miRNAs were up-regulated in high-grade glioma (HGG, WHO III and IV) compared with low-grade glioma (LGG, WHO I and II) (Figure 6C and D) and showed a significant correlation (Figure 6E). Moreover, the ROC analysis suggested that hsa-miR-132 (AUC = 0.98) and hsa-miR-200a (AUC = 0.88) could serve as independent factors to distinguish HGG from LGG (Figure 6F). Additionally, further analysis on an independent dataset from the Chinese Glioma Genome Atlas [32] also suggested that hsa-miR-132 and hsa-miR-200a were up-regulated in HGG and their up-regulation associated with the poor prognosis of glioma patients (Figure S4). Thus, these findings suggest that the plasma exosomal miRNAs, specifically hsa-miR-132-5p and hsa-miR-200a-5p, may play a role in regulating glioma progression through the modulation of angiogenesis, making them potential non-invasive biomarkers for the diagnosis and monitoring of glioma.

Furthermore, differential expression and WGCNA analysis on the miRNA expression datasets from tissues of LGG and HGG patient indicated significantly distinct miRNA expression patterns between the two grades of patients (Figure S5). The differential miRNAs were consistently enriched in the functional pathway of angiogenesis regulation (Figure S5). Clinically, it is known that glioblastoma (WHO Grade IV) can be differentiated from anaplastic glioma (WHO Grade III) by the presence of neoplastic vasculature. Therefore, our non-invasive markers align with the known pathological feature. In summary, these findings consistently suggest that miRNA plays an
important role in the progression of LGG to HGG through regulating angiogenesis related functional pathways. However, experimental validation in vitro and in vivo, as well as additional large-scale clinical analysis, are needed to confirm these observations.

**Discussion and perspectives**

ExomiRHub exhibits significant advantages in terms of data volume, coverage, and functionality when compared to other databases (Table 1). In contrast to databases like Vesiclepedia [9], ExoCarta [11], CMEP [12], miREV [14], EVAtlas [15], and EVpedi[a] [16], which compile and present experimentally supported disease-related extracellular miRNA and other biological molecular association data through manual curation of literature and analysis of public expression datasets, ExomiRHub concentrates on the collection and standardization of disease-related extracellular and intracellular miRNA expression datasets. Additionally, it offers extensive and customizable analytical and visualization functions for exploring these datasets, most of the capabilities that are not offered by these existing databases.

For example, when compared with CMEP, and EVAtlas, ExomiRHub exhibits a data set of extracellular miRNA that is 2.2 times larger than CMEP and 2.4 times larger than EVAtlas. Moreover, ExomiRHub provides a comprehensive suite of functionalities, including differential expression analysis, co-expression analysis, co-expression network analysis, and feature selection analysis, which features not offered by CMEP and EVAtlas. Additionally, ExomiRHub provides thousands miRNA transcriptomics data integrated from TCGA, a feature not offered by CMEP and EVAtlas. Additionally, ExomiRHub integrates advanced functionalities for miRNA function prediction, miRNA mutation evaluation, and miRNA target prediction and verification. These enhancements aim to encourage in-depth studies on the efficiency of miRNA regulation mechanisms. We believe these improvements further solidify the position of ExomiRHub as a valuable resource in the landscape of similar databases.

Compared to CancerMIRNome [17], ExomiRHub has significant advantages in various aspects (Table 1). For instance, ExomiRHub offers extracellular miRNA expression data for 112 human diseases from 23 sample types, whereas CancerMIRNome only offers data from plasma and whole blood samples of human cancer patients (Table 1). As a result, 81.15% (155/191) of extracellular miRNA datasets and 34.1% (5458/16,012) of cancer-related miRNA transcriptomes in ExomiRHub are not present in CancerMIRNome. Furthermore, ExomiRHub offers 25 analytical functions and four tools, with 21 of them not supported by CancerMIRNome (Table 1). Additionally, ExomiRHub offers more detailed biomedical annotations for each sample compared to CancerMIRNome, including comprehensive demographic information, drug/agent treatment, genotype, immune infiltration, and lifestyle (Table 1). This significantly broadens the application scenarios for ExomiRHub. For example, users can attempt to identify miRNA that can regulate drug response and tumor microenvironment.

In contrast to the existing databases, ExomiRHub is designed to allow users to select specific samples, define their own groups, and set parameters for comparison analysis. Additionally, ExomiRHub offers a web service application that enables users to perform all analytical and visualization functions with their own uploaded data, a feature not provided by the existing databases (Table 1). Therefore, as a novel and powerful resource, ExomiRHub serves as a crucial resource for users delving into the study of extracellular miRNA in human diseases.

In the foreseeable future, we anticipate an increasing availability of human miRNA expression data in the public domain. Accordingly, we commit to continuously maintaining and updating the analytical and visualization features of ExomiRHub to incorporate new data. Additionally, our future plans include the integration of other extracellular and intracellular biomolecular profiles and the introduction of new analytical features to enhance application efficiency. This expansion may involve the inclusion of profiles such as IncRNA, circRNA, mRNA, and protein/peptide data. Additionally, we intend to integrate multi-omics data from NCBI GEO [19] and TCGA [22], encompassing miRNA, mRNA, IncRNA, and circRNA, to introduce new functions for exploring interaction networks among these molecules. Furthermore, we plan to incorporate experimentally validated exosomal miRNA-disease interactions from databases such as miRandola [8] and Vesiclepedia [9] to construct comprehensive interaction networks. We are confident that these enhancements will significantly elevate the utility of ExomiRHub within the miRNA research community.

**Declarations**

**Data availability**

**CRediT author statement**

**Wenliang Zhang**: Conceptualization, Supervision, Methodology, Data curation, Formal analysis, Validation, Project administration, Funding acquisition, Writing–original draft, review and editing; **Yang Liu**: Formal analysis, Data curation, Methodology, Validation, Writing–original draft, review and editing; **Zhuochao Min**: Software, Formal analysis, Investigation, Methodology, Visualization, Writing–review and editing; **Jing Mo**: Software, Investigation, Methodology, Visualization; **Yanjie Wei**: Methodology, Resources, Funding acquisition, Supervision, Writing–review and editing; **Godfrey Chi-Fung Chan**: Supervision, Resources, Writing–review and editing; **Zhen Ju**: Formal analysis, Validation; **Jianliang Chen**: Resources; **Hanguang Li**: Data curation, Validation; **Weiling Liang**: Data curation, Validation. All authors reviewed the manuscript.

**Competing interests**

Jing Mo is a current employee of Outstanding Biotechnology Co., Ltd.-Shenzhen. Yang Liu and Wenliang Zhang is a current part-time consultant of Outstanding Biotechnology Co., Ltd.-Shenzhen. All the authors declare that they have no conflict of interest.

**Acknowledgments**

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32100513 and 62272449); Guangdong Basic and Applied Basic Research Foundation, China (Grant Nos. 2020A1515110528 and 2021B0101310002); Strategic Priority Project of Chinese Academy of Sciences, China (Grant No. XDB38050100); the Youth Innovation Promotion Association of Chinese Academy of Sciences, China (Grant No. Y2021101); the Shenzhen Basic Research Fund, China (Grant No. RCYX20200714114734194); the Shenzhen Science and Technology Program, China (Grant No. KQTD20200820113106007).

**ORCID**

0000-0001-7968-3403 (Yang Liu)
0000-0003-0017-6973 (Zhuochao Min)
0000-0002-2195-9080 (Jing Mo)
0000-0001-7720-1570 (Zhen Ju)
0000-0003-3780-3889 (Jianliang Chen)
0000-0001-9553-7465 (Hanguang Li)
0000-0003-0319-7876 (Weiling Liang)
0000-0001-5032-8985 (Godfrey Chi-Fung Chan)
0000-0002-4791-7540 (Yanjie Wei)
0000-0003-0454-6935 (Wenliang Zhang)

**References**


Tables

Table 1. The comparison of the data volume, coverage, and functionality of ExomiRHub with other databases.
<table>
<thead>
<tr>
<th>ases</th>
<th>ExomiRHub</th>
<th>CancerMIRNome</th>
<th>CMEP</th>
<th>miREV</th>
<th>EVAtlas</th>
<th>EVpedia</th>
<th>Vesiclepedia</th>
<th>ExoCarta</th>
</tr>
</thead>
<tbody>
<tr>
<td>cellular Acriptome</td>
<td>325,713</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>17,280</td>
<td>621,659</td>
<td>50,516</td>
<td></td>
</tr>
<tr>
<td>Number of entries</td>
<td>112</td>
<td>32 (Only cancer)</td>
<td>&lt; 66</td>
<td>9</td>
<td>&gt; 40</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of diseases</td>
<td>191 datasets</td>
<td>40 datasets</td>
<td>66 datasets</td>
<td>13 datasets</td>
<td>57 datasets</td>
<td>263 datasets, 6879 publication 1114</td>
<td>47 datasets, 3533 publications</td>
<td>-</td>
</tr>
<tr>
<td>Number of datasets or publications</td>
<td>29,198</td>
<td>28,633</td>
<td>10,419</td>
<td>428</td>
<td>2030</td>
<td>1114</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of samples</td>
<td>23</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>8</td>
<td>-</td>
<td>252</td>
<td></td>
</tr>
<tr>
<td>Number of sample types</td>
<td>miRNA</td>
<td>miRNA</td>
<td>miRNA</td>
<td>miRNA</td>
<td>miRNA, tRNA, piRNA, tRNA, snoRNA, snaRNA</td>
<td>mRNA, miRNA, protein, lipid, metabolite</td>
<td>miRNA, miRNA, protein, lipid, metabolite, DNA 56</td>
<td>organisms, such as human</td>
</tr>
<tr>
<td>Molecular type</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Organism</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>4 organisms, such as human</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Number of projects</td>
<td>156</td>
<td>33</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Number of cancer subtypes</td>
<td>16,012</td>
<td>10,554</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Number of samples</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Principal component analysis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Differential expression analysis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Differential expression Boxplot</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Volcano plot</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Heatmap</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td>√, GO</td>
<td>√, KEGG</td>
<td>√, KEGG</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>TCGA Pan-cancer analysis</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Co-expression Correlation</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Correlation plot</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Correlation boxplot</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Correlation boxplot</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Function</td>
<td>√, GO</td>
<td>√, KEGG</td>
<td>√, KEGG</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>miRNA-target correlation</td>
<td>x</td>
<td>√</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Scale</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Independence and mean connectivity</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Sample dendrogram and trait heatmap</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Dendrogram of module eigengenes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Cluster dendrogram among module Module eigengene expression pattern Network heatmap of top variable genes</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Dendrogram of module eigengenes and heatmap Heatmap of module-trait relationships Scatter plot miRNAs in module eigengenes Function</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>47 publications</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

Page 11/19
<table>
<thead>
<tr>
<th>Feature</th>
<th>COX</th>
<th>ROC</th>
<th>LASSO</th>
<th>Survival</th>
<th>miRNA function prediction</th>
<th>miRNA mutation evaluation</th>
<th>miRNA target prediction</th>
<th>miRNA target validation</th>
<th>Find suitable reference</th>
<th>miRNA related drug</th>
<th>Tissue and source specific expression</th>
<th>Network analysis</th>
<th>Sequence search</th>
<th>Uploaded data for analysis</th>
<th>Search</th>
<th>Browse</th>
<th>Download</th>
<th>Submission</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>×</td>
<td>√</td>
<td>×</td>
<td>√</td>
<td>√</td>
<td>×</td>
</tr>
</tbody>
</table>

Note: "-", the data and function feature is not provided or cannot be statistically calculated; "√", the data and function feature is provided by the database; "x", the data and function feature is not provided by the database; GO, gene ontology; KEGG: kyoto encyclopedia of genes and genomes; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; WGCNA, weighted gene co-expression network analysis.

**Figures**
Figure 1

The overview of ExomiRHub.

Figure 2

The advanced analysis functions and their visualizations of ExomiRHub.

Figure 3

The web interface and user guide of the web service application on the ExomiRHub database.
Extracellular miRNA regulates angiogenesis related functional pathways in glioblastoma.

**A.** Sample dendrogram and trait heatmap results of WGCNA analysis suggested that glioblastoma patients have different exosomal miRNA expression profiles from healthy controls, and these patients can be divided to two subgroups named G1 and G2. **B.** The G1 and G2 subgroups have distinct exosomal miRNA expression profiles. **C.** The module-trait relationships showing as a heatmap. **D.** Dendrogram of module eigengenes and their relationships of WGCNA analysis. **E.** The GO function enrichment of MEbrown module. **F.** The GO function enrichment of MERed module. **G.** The GO function enrichment of MEyellow module. **H.** The GO function enrichment of MEturquoise module. **I.** The GO function enrichment of the differential exosomal miRNA genes between the G1 and G2 subgroups. (The method and min module gene size for the WGCNA analysis was set to average and 30, respectively).
Figure 5

The angiogenic exosomal miRNAs, hsa-miR-132-5p and hsa-miR-200a-5p, exhibited up-regulation in the G1 glioblastoma subgroup as opposed to the G2 glioblastoma subgroup. Moreover, a notable correlation between these miRNAs was observed.

Figure 6

The angiogenic exosomal miRNAs, hsa-miR-132-5p and hsa-miR-200a-5p, associated with the progression of glioblastoma.

A. COX regression analysis suggested that down-regulation of hsa-mir-132 and hsa-mir-200a significantly decreased the death event and prolonged the overall survival rate of glioma patients. B. Survival analysis suggested that down-regulation of hsa-mir-132 and hsa-mir-200a significantly decreased the death event and prolonged the overall survival rate of glioma patients. C. hsa-mir-132 up-regulated in HGG. D. hsa-mir-200a up-regulated in HGG. E. hsa-mir-132 and hsa-mir-200a showed significant co-expression in glioblastoma tissues. F. ROC analysis revealed the capability to distinguish between HGG and LGG based on hsa-mir-132 and hsa-mir-200a. HGG, high-grade glioma; LGG, low-grade glioma; ****, P value ≤ 0.0001.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.tif
- FigureS2.tif
- FigureS3.tif
• FigureS4.tif
• FigureS5.tif
• TableS1.doc
• TableS2.doc