ExomiRHub: a comprehensive database platform to integrate and analyze human extracellular miRNA transcriptome for discovering non-invasive biomarkers

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: Non-invasive biomarker, miRNA, Extracellular vesicles, Exosome, Cancer, Liquid biopsy, human disease, angiogenesis

Posted Date: February 10th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2566749/v1
Abstract

Background

Although studies reported that extracellular miRNAs have significant functions in regulating the development of human diseases, our understanding of their role in human diseases remains to be further addressed. Many extracellular miRNA expression data were deposited in public resources, which are heterogeneous and difficult to investigate due to the data generated from different high throughput platforms. To use these invaluable data for accelerating the discovery of non-invasive miRNA biomarkers, a comprehensive and user-friendly database platform is essential, especially for bench researchers who lack bioinformatics skills.

Methods

We integrated, standardized, and annotated human extracellular miRNA expression data and cancer-related miRNA transcriptome data from NCBI GEO and The Cancer Genome Atlas (TCGA), respectively. Moreover, we developed the ExomiRHub database platform that designed with comprehensive online analysis functions and tools to analyze these data or User's own data. These analysis functions and tools were designed to enable users to select samples, define groups and parameters for their own analysis.

Results

ExomiRHub includes 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, including 29,198 samples and 23 sample types. ExomiRHub further includes 16,012 miRNA transcriptome data of 156 cancer sub-types to enhance the usability of it in cancer research. To accelerate the identification of non-invasive miRNA biomarkers, ExomiRHub provides 25 online analytical and visualization functions to individually analyze these data. Moreover, ExomiRHub provides Web Service to enable users in conducting the analyses on their uploaded data. Furthermore, ExomiRHub provides four additional tools to evaluate the functions and targets of miRNAs and their variations. Finally, we used ExomiRHub and discovered non-invasive miRNA biomarkers associated with angiogenesis-related pathways for monitoring glioma progression.

Conclusion

The comprehensive data and functions of ExomiRHub can greatly accelerate the discovery of non-invasive miRNA biomarkers. It is freely accessible at the websites of http://hpcc.siat.ac.cn/exomirhub/ & http://www.biomedical-web.com/exomirhub/.
Background

MicroRNAs (miRNAs), a class of small non-coding RNA, which are widely expressed in tissues and circulates in various human bio-fluids to maintain cell homeostasis by negative gene regulation [1], have been confirmed to sever as a powerful diagnostic and therapeutic tool [2, 3]. Recently, miRNAs were found to be carried and secreted by extracellular vesicles such as exosomes or circulated in bio-fluids to mediate intercellular and intra-organ crosstalk between different tissues and thus regulate gene expression and the function of distant cells [4, 5]. Although studies suggested that extracellular miRNAs have significant functions in regulating the development of human diseases and can serve as promising and non-invasive biomarkers for disease diagnosis, progression and therapy [6, 7], our understanding of their role in human diseases remains to be further explored and addressed.

To explore the clinical significance of extracellular miRNA, several valuable databases have been designed to collect and provide the experimentally supported disease-extracellular miRNA association data from publications through manual curation of data [8–10] and analysis based on small amount of extracellular miRNA expression data [11–13]. For example, the databases of miRandola [8] and Vesiclepedia [9] were developed to provide extracellular miRNA through manual curation on published and unpublished studies. Different from these two databases, the CMEP [11], EVpedia (Kim et al., 2015) and EVmiRNA [13] database were developed to provide circulating microRNA expression profiling with the differential expression and KEGG functional pathway enrichment functions, while the ExoBCD [12] database aims to provide the exosomal miRNA association data of breast cancer through analysis on four high-throughput datasets and manual mining of publications. Recently, Li et al. developed a valuable database called CancerMIRNome [14] to integrate and visualize circulating and tissue miRNA expression data of human cancers from public resources, but it only collected the data including more than 1,000 samples and lacked various analytical functions of co-expression such as Weighted Gene Co-Expression Network Analysis (WGCNA) (Supplementary Table S1). Moreover, CancerMIRNome does not allow users to select specific samples and define their own groups of parameters for different analysis, or analyze the miRNA expression data uploaded by users (Supplementary Table S1).

Over the last ten years, a large number of human disease-related extracellular miRNA expression data have been accumulated in public resources such as National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [15, 16]. However, it is a great challenge for researchers without bioinformatics skills to analyze these invaluable miRNA expression data for discovering the clinical and biological significance of extracellular miRNA. There is an urgent need to develop a special online database platform, which can comprehensively collect, analyze, and visualize the miRNA expression data in customized grouping and setting manner. Such database platform can significantly accelerate our understanding of the significance of extracellular miRNA in human diseases.

In order to fill above gaps, we developed a comprehensive database platform named ExomiRHub (http://hpcc.siat.ac.cn/exomirhub/ & http://www.biomedical-web.com/exomirhub/) that aims to collect human extracellular miRNA transcriptomic data, and further provides interactive online analytical and
visualization functions to investigate these data individually for discovering significant extracellular miRNAs and to enhance understanding of their role in diseases. Currently, ExomiRHub provides 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, including 2,656 miRNAs, 29,198 samples, and 23 sample types. (Fig. 1A). To facilitate the application of this platform in the field of cancer research, ExomiRHub integrated 16,012 human miRNA transcriptomes of 156 cancer sub-types from The Cancer Genome Atlas (TCGA) projects [17] (Fig. 1A). To resolve the problems of data heterogeneity, we independently standardized and normalized the miRNA symbol, expression value and sample identifier in each dataset from a single study. In addition, we manually annotated each dataset and sample with rich biomedical information.

Moreover, we designed two web applications called ExomiRlyzer and TCGA-miRNAlyzer on ExomiRHub to provide four cutting-edge bioinformatics tool-kits, including Differential Expression Tool-kit, Co-expression Tool-kit, WGCNA Tool-kit, and Feature Selection Tool-kit (Fig. 1B). These four tool-kits provide a total of 25 analytical and visualization functions (Fig. 2A-D) to investigate the miRNA expression data for discovering the significant miRNAs. The analysis includes various aspects of differential expression and co-expression analysis, network construction, module detection, miRNA feature selection, calculations of topological properties, data simulation, Gene Ontology (GO) functional enrichment analysis etc. Moreover, these 25 functions were designed with the customized grouping and setting manner, and they can enable users to select specific samples, define groups and parameters for their own analysis (Fig. S1 and S2). Furthermore, all of the analytical functions were implemented with advanced visualizations and can support to generate publication-quality vector images in Portable Document Format (PDF) with user-friendly tables for further analysis and export (Fig. 2A-D).

To serve a wider community of research, we designed a Web Service application for users to upload their own miRNA expression data with sample information, and further to perform six, five, and ten analyses on the uploaded data in Differential Expression Tool-kit, Co-expression Tool-kit, and WGCNA Tool-kit, respectively (Fig. 1B-C and Fig. 2A-C). Moreover, four additional tools were implemented to predict and validate the potential functions and targets (including mRNA, lncRNA, and circRNA) of miRNAs and their variations (Fig. 1C). As a user-friendly database platform, ExomiRHub provides comprehensive data analytical and visualization functions for freely search, browse, analyze, and download data, as well as to submit new data for further integration (Fig. 1C). A detailed tutorial is available on the database “Help” webpage. In summary, we believe that ExomiRHub can serve as an important resource to improve our insights into the roles of miRNA in diseases, and accelerate the discovery and identification of non-invasive biomarkers.

Materials And Methods

Data collection, standardization, and annotation

We performed extensive search in NCBI GEO [15, 16] 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: (i) 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 bio-fluids such as plasma and serum; (ii) the expression data and sample information of the dataset could be freely downloaded and integrated. We further manually annotated each dataset with rich biomedical information, such as disease phenotype, grade, stage, drug, infection, metastasis, and genotype. Furthermore, the normalized expression data and sample metadata of the identified datasets were downloaded programmatically through the getGEO function in the R package GEOquery [18], 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, a systematic pipeline was designed to standardize and normalize 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 [19] and NCBI GEO [15, 16], respectively.

Given that different isolation methods and detection platforms have significant bias on the characterization and quantification of extracellular miRNA. In order to avoid the bias, we independently collected and standardized each dataset from a single 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 further comparison analysis. Although the datasets were generated from 14 different platforms, we classified the platforms to two categories, including next generation sequencing (NGS) and microarray. 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 one 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) [20], which generated from human tissues of cancer patient by miRNA sequencing and has been standardized by the same pipeline. The miRNA expression quantification data were further normalized by log2 transformation and merged into one expression matrix saved in a RDS file. In addition, the metadata of each sample were also downloaded from NCI GDC to annotate the sample with rich biospecimen and clinical information, such as sample type, cancer sub-type, 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, lifestyle, etc.

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

ExomiRlyzer and TCGA-miRlyzer was implemented to provide four helpful tool-kits, including Differential Expression Tool-Kit, Co-expression Tool-Kit, WGCNA Tool-Kit, and Feature Selection Tool-Kit. Moreover, 25 analytical and visualization functions were designed on the four tool-kits to individually perform comprehensive analysis and visualization on a single miRNA expression data, such as various aspects of differential expression, co-expression, WGCNA, feature selection, and GO function enrichment analysis (Figs. 1 & 2, Supplementary Table S1). To generate suitable, robust, and reliable hypothesis, all comparison analyses on the database are designed based on a single extracellular miRNA expression datasets from a single study, rather than different studies. The 25 functions were both implemented based on the R project (https://www.r-project.org/). The details of the R packages for the implementation and application of the 25 functions on the four tool-kits were described in Supplementary Table S2 and the “Help” webpage on ExomiRHub (http://hpcc.siat.ac.cn/exomirhub/help).

**Web Service**

We used R language to develop the Web Service application, which is a systematic pipeline to integrate and standardize the miRNA expression data and its sample metadata uploaded by researchers, and its further assigns a temporary identifier to the uploaded data, so that they can delete their uploaded data after completing the analysis in the database (Fig. 3). For example, the application can standardize the miRNA identifiers and expression values of read counts to RPM and further normalize it by log2 transformation. In addition, the human miRNA annotation data were downloaded and extracted from miRBase [19], which supports Web Service to convert different miRNA accessions and symbols to the primary/mature miRNA identifiers in miRBase. Moreover, Supplementary Table S1 describes the details of R package and resource used in the application implementation. Finally, this application was encapsulated by shell scripts and was deployed on our local server.

**miRNA Target Prediction and miRNA Mutation Evaluation**

It was suggested that miRNA can bind to the target of mRNA, circRNA, and IncRNA for regulating their biological functions [21], while the genomics variation on miRNA can impact its biological function in human diseases [22]. Therefore, we implemented the miRNA Target Prediction and miRNA Mutation Evaluation tools to predict the miRNA targets and evaluate the potential impacts of miRNA variation. To develop these two tools, we firstly extracted the sequences of human mRNA/IncRNA, and circRNA, miRNA seed region from the resources of GENCODE 2021 [23], circBase [24], and miRBase [19], respectively. Moreover, we installed miRanda [25] and wrote a shell script to encapsulate it with the extracted sequence data of miRNA, mRNA, IncRNA, and circRNA. To further implement the miRNA Mutation Evaluation tool, a R script was designed to calculate the gain or loss targets of miRNA for 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 firstly integrated the experimentally validated microRNA-target interaction data from the miRTarBase database [21]. Moreover, we wrote R scripts to implement miRNA Target Validation for systematically identifying experimental validated targets of miRNA. To predict the biological function of an interesting miRNA, we further implemented the miRNA Function Prediction tool based on the R packages of clusterProfiler[26], org.Hs.eg.db, pathview [27], topGO, and the GO [28] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [29] annotations of the miRNA validated targets in miRTarBase.

Data Storage And Web Implementation

ExomiRHub was implemented with a separated front-end and back-end framework. The methods of the framework built were described in detail previously [30, 31]. JAVA programming language was used to write the programs for data processing and application operation. In addition, MySQL is used to store and organize the association data for faster data data browsing and searching. We installed the R version 4.0.3 (https://www.r-project.org/) to support and run the analytical functions, severs and tools on ExomiRHub. The details of R packages and resources used in the implementation of them were detailed in Supplementary Table S2 and on the “Help” webpage (http://hpcc.siat.ac.cn/exomirhub/help). Finally, we deployed ExomiRHub on the Apache Tomcat server, and encourage users to access and utilize it publicly through the URL of http://hpcc.siat.ac.cn/exomirhub/.

Results

Data landscape and data access

ExomiRHub provides 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, including 2,656 miRNAs, 29,198 samples, and 23 sample types. (Fig. 1A). The statistic results shown that 80.63% (154/191) of the datasets provide miRNA expression profile from extracellular vesicle (including 145 datasets of exosome), while the rest 19.37% (37/191) provide circulating miRNA profiles of serum (13.09%, 25/191), plasma ( 4.71%, 9/191), and whole blood (1.57%, 3/191). These sample types include extracellular vesicles, exosomes, and circulating miRNA derived from whole blood, serum, plasma, urinary, ascites, cerebrospinal fluid, endometrial fluid, follicular fluid, pericardial fluid, Liquid milk, saliva, tissue fluid, umbilical cord blood, tracheal aspirate, faecal fluid, and the culture supernatant of cells with specific genotype and treatment. While the treatments include adjuvant chemotherapy, antibiotic, cell therapy, chemotherapy, immunotherapy, radiotherapy, target therapy, etc. The further statistical analysis indicated that the disease phenotypes included in ExomiRHub were associated with more than 23 body sites. The top six body sites with the largest numbers of the extracellular miRNA expression datasets are blood system, breast, brain, lung, intestine, and pancreas. Moreover, it was found that China, Japan, and USA are the top three countries with the largest contribution to the datasets, accounting for 25.65% (49/191), 24.08% (46/191) and 20.42% (39/191) respectively. Furthermore, we summarized the data features of body site, sample type,
isolation method, sample resource, extraction method, disease phenotype, genotype, and treatment on the “Home” and “Browse” webpages of the database platform.

Currently, about 63.87% (122/191) of the extracellular miRNA datasets are associated with 52 cancer sub-types. In order to enhance the usability of the database platform in the field of cancer research, ExomiRHub further integrated the human miRNA expression quantification data of 16,012 samples and 156 cancer sub-types from 42 TCGA projects (Fig. 1A). In addition, each sample from the TCGA projects was annotated with rich biospecimen and clinical data, including various demographics, diagnosis, progression, tumor microenvironment, and lifestyle, which facilitates users to browse and select specific samples for designing and defining the analytical and visual comparison.

In addition, ExomiRHub enables to quickly search and browse interesting extracellular miRNA expression dataset and further navigate to the ExomiRlyzer application for discover significant miRNA through user-definedly investigating on the dataset. All data in the ExomiRHub database platform can be freely downloaded for research, and it provides a variety of data formats for users to download, including CSV, TABLE, and JSON. Moreover, ExomiRHub encourages users to submit their novo human extracellular miRNA expression data to the database platform through the submit webpage (http://hpcc.siat.ac.cn/exomirhub/submit). Once the submitted data was approved by our submission review committee, these data will be included in the next version.

**Comprehensive Web Analysis And Visualization Function, Server, And Tool**

To understand the role of extracellular miRNA and further discover non-invasive biomarkers in human disease, the web applications of ExomiRlyzer and TCGA-miRNAlyzer have been designed and developed on ExomiRHub to provide four cutting-edge bioinformatics tool-kits, including Differential Expression Tool-kit, Co-expression Tool-Kit, WGCNA Tool-kit, and Feature Selection Tool-kit. These four tool-kits provide 25 analytical and visualization functions to integrated extracellular miRNA expression data and cancer related miRNA expression data. It involves various aspects of differential expression, co-expression, WGCNA, and GO function enrichment, COX Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Receiver Operating Characteristic (ROC), and Survival analysis and visualization. Due to the samples in each dataset were annotated with rich biomedical information, all 25 functions were designed with customized grouping and setting format, and allow users to select specific samples, define their own groups and parameters for their own comparison analysis (Fig. S1 and S2). We demonstrated with more details on the two web applications and four tool-kits with their analytical and visualization functions in Supplementary Table S1 and Fig. 2.

To serve a wider community of research, ExomiRHub provides the Web Service application, which is a systematic pipeline to integrate and standardize the miRNA expression data and its sample metadata uploaded by users, and its further assigns a temporary identifier to the uploaded data, so that they can perform the six, five, and ten further analyses and visualizations on their uploaded data in the Differential Expression, Co-expression and WGCNA tool-kits, respectively (Fig. 3 and Fig. 2A-C). To protect the security
of the data uploaded by users, this application enables them to delete their uploaded data at any time through its assigned temporary identifier, and we also clean up the uploaded data monthly (Fig. 3).

To increase the application efficiency of ExomiRHub in the community of miRNA research, ExomiRHub provides four additional tools to predict and validate the potential function and target (such as mRNA, lncRNA, and circRNA) of miRNA and its variation. The four tools include miRNA Function Prediction, miRNA Mutant Evaluation, miRNA Target Prediction, and miRNA Target Validation. A usage guide of the above analytical and visualization functions, servers, and tools were addressed on the “Help” webpage (http://hpcc.siat.ac.cn/exomirhub/help). Last but not least, the results from these applications are supported to generate publication-quality vector images in PDF format and tables for further analysis and download (Fig. 2A-D).

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

In order to discover and identify non-invasive biomarkers for diagnosis and monitoring progression of glioma, we performed comprehensive analysis on an exosomal miRNA dataset in ExomiRHub (ID: EMIR00000186), which provides miRNA expression profiles derived from plasma exosomes of healthy controls and glioblastoma patients. First, the WGCNA analysis results showed that plasma exosomal miRNAs could be used to distinguish glioblastoma patients from their healthy controls and obviously divided the patients into two subgroups named G1 and G2 (Fig. 4A). Subsequently, the hierarchical cluster analysis confirmed that the two subgroups have distinct exosomal miRNA expression profiles (Fig. 4B and Fig. S3). Moreover, we further performed WGCNA analysis on the two subgroups (Fig. S4) and identified four modules of eigengenes that associated with the two subgroups (Fig. 4C) and showed significant correlation with each other (Fig. 4D), including MEbrown, MEred, MEyellow, and MEturquoise. Further analysis results suggested that these four modules of eigengenes consistently enriched on the GO terms of regulation of angiogenesis and vasculature development (Fig. 4E-H), which GO terms have been proved to be closely related to the development, treatment and prognosis of glioma [32, 33].

To discover the difference of angiogenesis related pathway between G1 and G2, we conducted differential analysis and found 48 dysregulated exosomal miRNAs (|Log2(fold change)| >= 0.5 & p-value <= 0.05) between the two subgroups, including 21 down-regulated and 27 up-regulated (Fig. 5A & Supplementary Table S3). In line with the GO function enrichment analysis on the module eigengenes, we confirmed that these differential expression exosomal miRNAs enriched on the angiogenesis related pathways (Fig. 4I). For example, the dysregulation of hsa-miR-132-5p and hsa-miR-200a-5p associates with the positive regulation of blood vessel endothelial cell migration (Fig. 4I and Fig. 5A-C), and they showed significant correlation (Fig. 5D). Furthermore, the function of hsa-miR-132-5p and hsa-miR-200a-5p consistently enriched on the angiogenesis related pathway based on the GO annotations of their co-expression exosomal miRNAs (Fig. S5).

In order to further address the clinical significance of the two exosomal miRNAs in glioma, we performed COX regression and survival analysis on the glioma dataset integrated from TCGA on ExomiRHub. The
analysis results suggested that the down-regulation of angiogenic genes of hsa-mir-132 and hsa-mir-200a can significantly decrease the death event (Fig. 6A) and prolong the overall survival rate of glioma patients (Fig. 6B). In consistent with the above analysis results, the two miRNAs were up-regulated in high-grade glioma (HGG, WHO III & IV) compared with low-grade glioma (LGG, (WHO I & II)) (Fig. 6C&D) and showed significant correlation (Fig. 6E). Moreover, the ROC analysis result suggested that hsa-mir-132 (AUC = 0.98) and hsa-mir-200a (AUC = 0.88) could be used as independent factors to distinguish HGG from LGG (Fig. 6F). Thus, these results imply that the plasma exosomal miRNAs of hsa-miR-132-5p and hsa-miR-200a-5p may regulate the glioma progression through regulation of angiogenesis, and they have potential as non-invasive biomarkers for diagnosis and monitoring the progression of glioma.

In addition, the differential expression and WGCNA analysis on the miRNA expression datasets of LGG and HGG patient tissues suggested that the two different grades patients have significantly different miRNA expression patterns, and the differential features were consistently enriched on the functional pathway of angiogenesis regulation (Fig. 7 and Fig. S6). Clinically, we know that glioblastoma (WHO Grade IV) can be differentiated from anaplastic glioma (WHO Grade III) by the presence of neoplastic vasculature, so our non-invasive markers are in accordance with the known pathological feature. Altogether, these findings consistently suggested that miRNA plays a critical role in the progression of LGG to HGG through regulating angiogenesis related functional pathways, but need to be further validated by further in vitro and in vivo experiments, and other large-scale clinical analysis.

**Discussion And Conclusion**

ExomiRHub aims to comprehensively integrate human extracellular miRNA expression data and provide user-friendly analytical and visualization tools to investigate these data for discovering and identifying non-invasive biomarkers in human disease. In terms of data scale, web design and functionalities, ExomiRHub is significantly different from other databases because it collects and provides the experimentally supported disease-related extracellular miRNA association data through manual curation from publications [8, 9], and also includes analysis of a small amount of extracellular miRNA expression data [11, 12], such as the CMEP database (Supplementary Table S1).

In addition, ExomiRHub has significant different features from CancerMIRNome database [14] in terms of host and visualization of human cancer-related miRNA expression data (Supplementary Table S1). For example, ExomiRHub provides the extracellular miRNA expression data of various human diseases across 23 sample types, while CancerMIRNome provides the data of plasma/whole blood samples from human cancer patients only (Supplementary Table S1). As a result, 81.15% (155/191) of the extracellular miRNA datasets and 34.1% (5,458/16,012) of the cancer related miRNA transcriptomes in ExomiRHub are not found in CancerMIRNome. Moreover, 19 of 25 analytical and visualization functions on ExomiRHub for investigating the miRNA expression data are not supported by CancerMIRNome (Supplementary Table S1), including ten, five, and one function(s) in WGCNA Tool-kit (Fig. 2C), Co-expression Tool-kit (Fig. 2B), and Differential Expression Tool-kit (Fig. 2A), respectively. In addition, different from the traditional databases such as CancerMIRNome, ExomiRHub is designed to allow users to select specific samples,
define their own groups and parameters for performing comparison analysis. Furthermore, ExomiRHub annotates each sample with more detailed biomedical annotations than that in CancerMIRNome, including comprehensive demographics, drug/agent treatment, genotype, immune infiltration, and lifestyle (Supplementary Table S1), which can significantly increase the range of application scenarios for ExomiRHub. For example, one can try to identify miRNA that can regulate drug response and tumor micro-environment. Last but not least, ExomiRHub provides the Web Service application to enable users in conducting all analytical and visualization functions by themselves with their own uploaded data, while this application is not provided by CancerMIRNome (Supplementary Table S1). Thus, as a novel and powerful resource, ExomiRHub can serve as an important resource for researchers to conduct in-depth investigation of extracellular miRNA in human disease.

As different isolation methods and detection platforms have significant impacts on the characterization and relative quantification of extracellular miRNA, we do not recommend users to download and integrate the datasets from different studies in ExomiRHub for further comparison analysis.

In the foreseeable future, more and more human extracellular miRNA expression data are expected to be accessible in the public domain. Therefore, we will continue to maintain and update the analytical and visualization features of ExomiRHub as the new data becomes available. Moreover, we plan to integrate other extracellular molecular profiles and add new analytical features to enhance the application efficiency of the database platform, such as lncRNA/circRNA/mRNA and protein/peptide. In addition, we also plan to integrate datasets that include both the miRNA and mRNA/lncRNA/circRNA expression data from NCBI GEO [15, 16] and TCGA [17], and further add the co-expression function of miRNA-mRNA/lncRNA/circRNA in Co-expression Tool-kit. Furthermore, we will integrate the experimentally validated exosomal miRNA-disease interactions from the databases of miRandola [8] and Vesiclepedia [9] by constructing their interaction network on ExomiRHub. We believe that all these additional data, analytical and visualization features will increase the application efficiency of ExomiRHub in the miRNA research community.

Abbreviations

GEO: Gene Expression Omnibus; GDC: Genomic Data Commons; GO: Gene Ontology; HGG: high-grade glioma; LASSO: Least Absolute Shrinkage and Selection Operator; LGG: low-grade glioma; miRNAs: MicroRNAs; NCBI: National Center for Biotechnology Information; NCI: National Cancer Institute; NGS: next generation sequencing; PDF: Portable Document Format; ROC: Receiver Operating Characteristic; RPM: Reads Per Million mapped reads; TCGA: The Cancer Genome Atlas; WGCNA: Weighted Gene Co-Expression Network Analysis

Declarations

Funding
This work was supported by the National Natural Science Foundation of China under Grant number 32100513; Guangdong Basic and Applied Basic Research Foundation, China under Grant number 2020A1515110528; Strategic Priority Project of CAS under Grant number XDB38050100; the Youth Innovation Promotion Association of CAS under Grant number Y2021101; the Shenzhen Basic Research Fund under Grant number RCYX2020071411473419; the KQTD Project under Grant number KQTD20200820113106007 and JSGG20210802154539016; and the CAS Key Lab Foundation under Grant number 2011DP173015.

Authors’ contributions

Wenliang Zhang: Project administration, Conceptualization, Supervision, Methodology, Data curation and validation, Funding acquisition, and Writing—Original draft, Review & editing; Yang Liu: Data curation and validation, Methodology, Writing—Review & editing; Zhuochao Min: Web implementation, Methodology, and Writing—Review & editing; Jing Mo: Web implementation, Visualization, and Maintenance server; YanJie Wei: Resources, Supervision, Methodology, Funding acquisition, and Writing—Review & editing; Godfrey Chi-Fung Chan: Resources, Supervision, Writing—Review & editing; Zhen Ju: Formal analysis, Maintenance server; Jianliang Chen: Resources; Hanguang Li: Data curation and validation; Weiling Liang: Data curation and validation. All authors reviewed the manuscript.

Availability

The ExomiRHub database platform is accessible at websites of http://hpcc.siat.ac.cn/exomirhub and http://www.biomedicalweb.com/exomirhub.

Ethics approval and consent to participate

Not Applicable.

Consent for publication

All authors have read the manuscript and are consentaneous for publication.

Competing Interests

The authors declare that they have no conflict of interest.

References


**Figures**
Figure 1

The overview of data contents (A), analysis functions of Differential Expression Tool-kit, Co-expression Tool-kit, WGCNA-Tool kit, Feature Selection Tool-kit (B), web servers and tools (C) on ExomiRHub. PCA: Principal Component Analysis; WGCNA: Weighted Gene Co-Expression Network Analysis; TCGA: The Cancer Genome Atlas.
Figure 2

The advanced analysis functions and their visualizations of Differential Expression Tool-kit (A), Co-expression Tool-kit (B), WGCNA-Tool kit (C), Feature Selection Tool-kit (D) on ExomiRHub. GO: Gene Ontology; LASSO: Least Absolute Shrinkage and Selection Operator; ROC: Receiver Operating Characteristic; WGCNA: Weighted Gene Co-Expression Network Analysis.
Figure 3

The web interface and user guide of the Web Service application on the ExomiRHub database platform.
Figure 4

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. The Gene Ontology (GO) function enrichment analysis result of module eigengenes of MEbrown (E), MEred (F), MEyellow (G), and MEturquoise (H). (I) The GO function enrichment analysis result 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 volcano plot (A) and box plots (B & C) show that the angiogenic exosomal miRNAs of hsa-miR-132-5p and hsa-miR-200a-5p were up-regulated in the G1 glioblastoma subgroup compared with the G2 glioblastoma subgroup, and they show significant correlation (D).
The results of COX regression (A) and survival analysis (B) show that the down-regulation of angiogenic genes of hsa-mir-132 and hsa-mir-200a significantly decreased the death event and prolonged the overall survival rate of glioma patients. Compared with low-grade glioma (LGG, WHO I & II), the two miRNAs were up-regulated in high-grade glioma (HGG, WHO III & IV) (C & D) and they show significant co-expression in glioma tissues (E). ROC analysis show that HGG and LGG could be distinguished by the expression levels of hsa-mir-132 and hsa-mir-200a.
Intracellular miRNA plays a critical role in the progression of LGG to HGG through regulating angiogenesis related functional pathways. (A) Sample dendrogram and trait heatmap show that the LGG and HGG tissues have different miRNA expression pattern. Module-trait relationship (B), module eigengene dendrogram and trait heatmap (C), miRNA eigengenes expression of MEturquoise (D) results suggest that MEturquoise plays a critical role in the progression of LGG to HGG based on the miRNA co-expression
network analysis using the TCGA-miRlyzer application. The GO functional enrichment analysis on the miRNA eigengenes of MEturquoise (E) and the differential expression miRNA between the LGG and HGG tissues (F) were consistently enriched on the angiogenesis related pathways.

**Supplementary Files**

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

- [SupplementaryFiguresandTables.pdf](#)