Computational and Structural Biotechnology Journal 23 (2024) 3104–3116

Contents lists available at ScienceDirect

Computational and Structural Biotechnology Journal

journal homepage: www.elsevier.com/locate/csbj

Database article

ExomiRHub: A comprehensive database for hosting and analyzing human disease-related extracellular microRNA transcriptomics data

Yang Liu [a,b,c,d,1], Zhuochao Min [e,f,1], Jing Mo [d,1], Zhen Ju [b], Jianliang Chen [c], Weiling Liang [c], Lantian Zhang [a], Hanguang Li [c], Godfrey Chi-Fung Chan [c,*], Yanjie Wei [b,g,**], Wenliang Zhang [a,b,c,d,***]

[a] The Key Laboratory of Advanced Interdisciplinary Studies, The First Affiliated Hospital of Guangzhou Medical University; GMU-GIBH Joint School of Life Sciences, The Guangdong-Hong Kong-Macao Joint Laboratory for Cell Fate Regulation and Diseases, Guangzhou Medical University, Guangzhou 510182, China
[b] Center for High Performance Computing, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[c] Department of Pediatrics, The University of Hong Kong-Shenzhen Hospital, Shenzhen 518053, China
[d] Department of Bioinformatics, Outstanding Biotechnology Co., Ltd.-Shenzhen, Shenzhen 518026, China
[e] School of Zoology, The George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 6997801, Israel
[f] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu 610054, China
[g] CAS Key Laboratory of Health Informatics, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China

ARTICLE INFO

Keywords:
MiRNA
Extracellular vesicle
Exosome
Cancer
Liquid biopsy

ABSTRACT

Extracellular microRNA (miRNA) expression data generated by different laboratories exhibit heterogeneity, which poses challenges for biologists without bioinformatics expertise. To address this, we introduce ExomiRHub (http://www.biomedical-web.com/exomirhub/), a user-friendly database designed for biologists. This database incorporates 191 human extracellular miRNA expression datasets associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, encompassing 29,198 and 23 sample types. ExomiRHub also integrates 16,012 miRNA transcriptomes of 156 cancer subtypes from The Cancer Genome Atlas. All the data in ExomiRHub were further standardized and curated with annotations. The platform offers 25 analytical functions, including differential expression, co-expression, Weighted Gene Co-Expression Network Analysis (WGCNA), feature selection, and functional enrichment, enabling users to select samples, define groups, and customize parameters for analyses. Moreover, ExomiRHub provides a web service that allows biologists to analyze their uploaded miRNA expression data. Four additional tools were developed to evaluate the functions and targets of miRNAs and miRNA variations. Through ExomiRHub, we identified extracellular miRNA biomarkers associated with angiogenesis for monitoring glioma progression, demonstrating its potential to significantly accelerate the discovery of extracellular miRNA biomarkers.

1. Introduction

MicroRNAs (miRNAs) are a class of small non-coding RNA expressed widely in tissue cells and circulating in human biofluids to maintain cellular homeostasis primarily by suppressing gene regulation [1]. These molecules serve as potent diagnostic and therapeutic targets [2,3]. Recently, miRNAs have been observed to be carried and secreted via extracellular vesicles such as exosomes or circulated in biofluids to mediate intercellular and intra-organ crosstalk, thus regulating gene expression and function of distant tissue cells [4,5]. Although extracellular miRNAs are associated with human diseases and hold promise as biomarkers for disease diagnosis and monitoring [6,7], their exact

* Corresponding author.
** Corresponding author at: Center for High Performance Computing, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China.
*** Corresponding author at: The Key Laboratory of Advanced Interdisciplinary Studies, The First Affiliated Hospital of Guangzhou Medical University; GMU-GIBH Joint School of Life Sciences, The Guangdong-Hong Kong-Macao Joint Laboratory for Cell Fate Regulation and Diseases, Guangzhou Medical University, Guangzhou 510182, China.
E-mail addresses: gcfchan@hku.hk (G.C.-F. Chan), yj.wei@siat.ac.cn (Y. Wei), zhangwl25@mail3.sysu.edu.cn (W. Zhang).
[1] Equal contribution.

https://doi.org/10.1016/j.csbj.2024.07.024
Received 15 April 2024; Received in revised form 30 July 2024; Accepted 31 July 2024
Available online 7 August 2024

mechanisms of action in human diseases remain poorly understood.

Several databases have been developed to collect experimentally supported disease-extracellular miRNA association data from publications through manual curation [8–11] and to analyze limited miRNA expression data [12–15]. For example, the miRandola (http://mirandola.iit.cnr.it/) [8], Vesiclepedia (http://www.microvesicles.org) [9], EVpedia (http://evpedia.info) [10], and ExoCarta (http://www.exocarta.org) [11] databases provide curated studies on extracellular miRNAs. Conversely, the CMEP (http://syslab5.nchu.edu.tw/CMEP) [12], miREV (https://www.physio.wzw.tum.de/mirev/), and EVAtlas (http://bioinfo.life.hust.edu.cn/EVAtlas) [15] databases present circulating miRNA expression profiles with functionalities for differential expression and functional enrichment. The ExoBCD (https://exobcd.liumwei.org) [13] database, focusing on exosomal miRNA data for breast cancer, was established through the analysis of four high-throughput datasets and manual literature mining. The latest version of miRTarBase (https://miRTarBase.cuhk.edu.cn/) integrates the expression of both extracellular and intracellular miRNAs [16]. However, databases that host and explore comprehensive human extracellular and intracellular miRNA transcriptomes are scarce. Li et al. introduced the CancerMIRNome (http://bioinfo.jialab-ucr.org/CancerMIRNome) [17] database, which integrates and visualizes circulating and tissue miRNA expression data for human cancers. However, CancerMIRNome is specifically designed for cancer and lacks advanced analytical functions such as Weighted Gene Co-Expression Network Analysis (WGCNA). Additionally, the analysis functions provided by CancerMIRNome are conventional and do not allow users to select specific samples, define their own groups, set parameters for various analyses, or analyze uploaded miRNA expression data.

Furthermore, a substantial amount of human disease-related extracellular miRNA expression data has been deposited to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) [18,19], which are heterogeneous and challenging for bench scientists without bioinformatics expertise to explore effectively. Identifying disease-related extracellular miRNAs from these datasets remains a significant challenge, highlighting the need for a user-friendly database capable of hosting and customizing human miRNA expression data to enhance understanding of the roles of extracellular miRNA in human diseases.

To address these challenges, we introduce ExomiRHub (http://www.biomedical-web.com/exomirhub/), a biologist-oriented database that integrates human extracellular miRNA expression data and provides extensive analytical and visualization functions for free search, browsing, analysis, downloading, and data submission. ExomiRHub also features a web application allowing users to upload their miRNA expression data along with sample information for comprehensive analyses. Furthermore, ExomiRHub offers four tools for predicting and validating the potential functions and targets of miRNAs and miRNA variations. We envision ExomiRHub as a crucial, accessible resource for accelerating the discovery of extracellular miRNA biomarkers.

2. Materials and methods

2.1. 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: (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 biofluids such as plasma and serum; (II) 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 biomedical information, including disease phenotype, grade, stage, drug, infection, metastasis, genotype etc. Furthermore, the sample metadata and normalized expression data of the datasets were downloaded programmatically through GEOquery [20], whereas the non-normalized expression data were downloaded manually. To remove the data heterogeneity, an in-house pipeline (Fig. S1) 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 assay platforms have significant bias on the characterization and quantification of extracellular miRNA. To avoid the bias, we independently integrated 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 different platforms, we classified the platforms into two categories, including next generation sequencing (NGS) and microarray. To overcome data heterogeneity, we independently standardized 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 (Fig. S1). For the rest of the datasets from microarray, we downloaded the original standardized expression data and applied log2 transformation to it, if the transformation were not applied (Fig. S1). 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 download The Cancer Genome Atlas (TCGA) miRNA expression quantification data based on the curl API method provided by National Cancer Institute (NCI) Genomic Data Commons (GDC) [22], which were 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 in an RDS file. Furthermore, sample metadata were obtained from NCI GDC to enrich annotations. The sample metadata include 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.

2.2. Implementation of the ExomiRlyzer and TCGA-miRlyzer web applications

ExomiRlyzer and TCGA-miRlyzer were implemented to offer four useful tool-kits, encompassing differential expression tool-kit, co-expression tool-kit, WGCNA tool-kit, and feature selection tool-kit (Fig. 1). Additionally, within these tool-kits, 25 analytical and visualization functions have been implemented (Fig. 2). The 25 functions were both implemented in R project (https://www.r-project.org/). The data processing for all functions utilized the R packages of stringr, dplyr, and jsonlite. The differential expression tool-kit was implemented using limma[23], factoextra, FactoMine, ggpubr, and reshape2. For the co-expression tool-kit, ggpubr and corrplot were employed. The WGCNA tool-kit was developed using the WGCNA package [24]. Feature selection was carried out with the R packages foreign, plotROC [25], ROCR [26], survival, and survminer. The GO functional enrichment analysis across these tool-kits was conducted using clusterProfiler [32], org.Hs.eg.db, pathview [27], Rgraphviz, and topGO. Table S1 described the

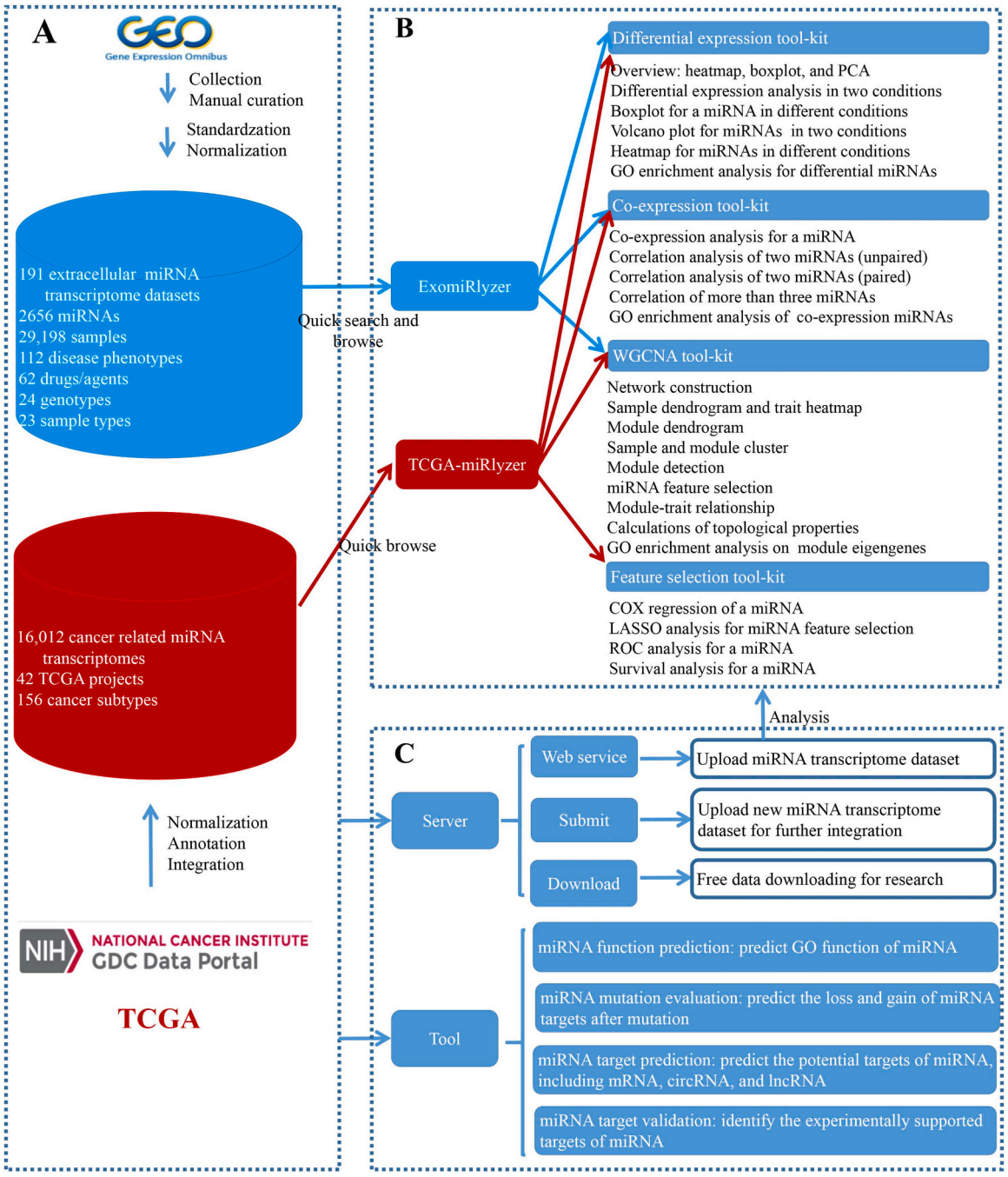

Fig. 1. The overview of ExomiRHub. A. Data content. B. The analysis functions of the differential expression, co-expression, WGCNA, feature selection tool-kits. C. Web servers and tools. PCA, principal component analysis; WGCNA, weighted gene co-expression network analysis; TCGA, The Cancer Genome Atlas.

details of R packages used to implement these functions on the four tool-kits.

2.3. 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 (Fig. S1). This enables further customizable analysis of the uploaded data within the database (Fig. 3). The application was encapsulated using shell scripts and deployed on our local server. To implement it, we extracted miRNA annotations from miRBase [21] and processed the data using the R packages of stringr, dplyr, and jsonlite. This application supports multiple miRNA nomenclatures, including accession numbers, IDs, and symbols for both primary and mature miRNAs from miRBase [21]. It is designed to handle various standardized forms of miRNA expression data, such as read counts, TPM, FPKM, RPM, TMM (Trimmed Mean of M-values), quantiles, and others. For example, the application can standardize miRNA identifiers and expression values, converting read counts to RPM and subsequently normalizing them through log2 transformation. Table S1 described R packages and resources utilized in the application implementation.

2.4. miRNA target prediction and miRNA mutation evaluation

It has been suggested that miRNA can bind to the target of mRNA,

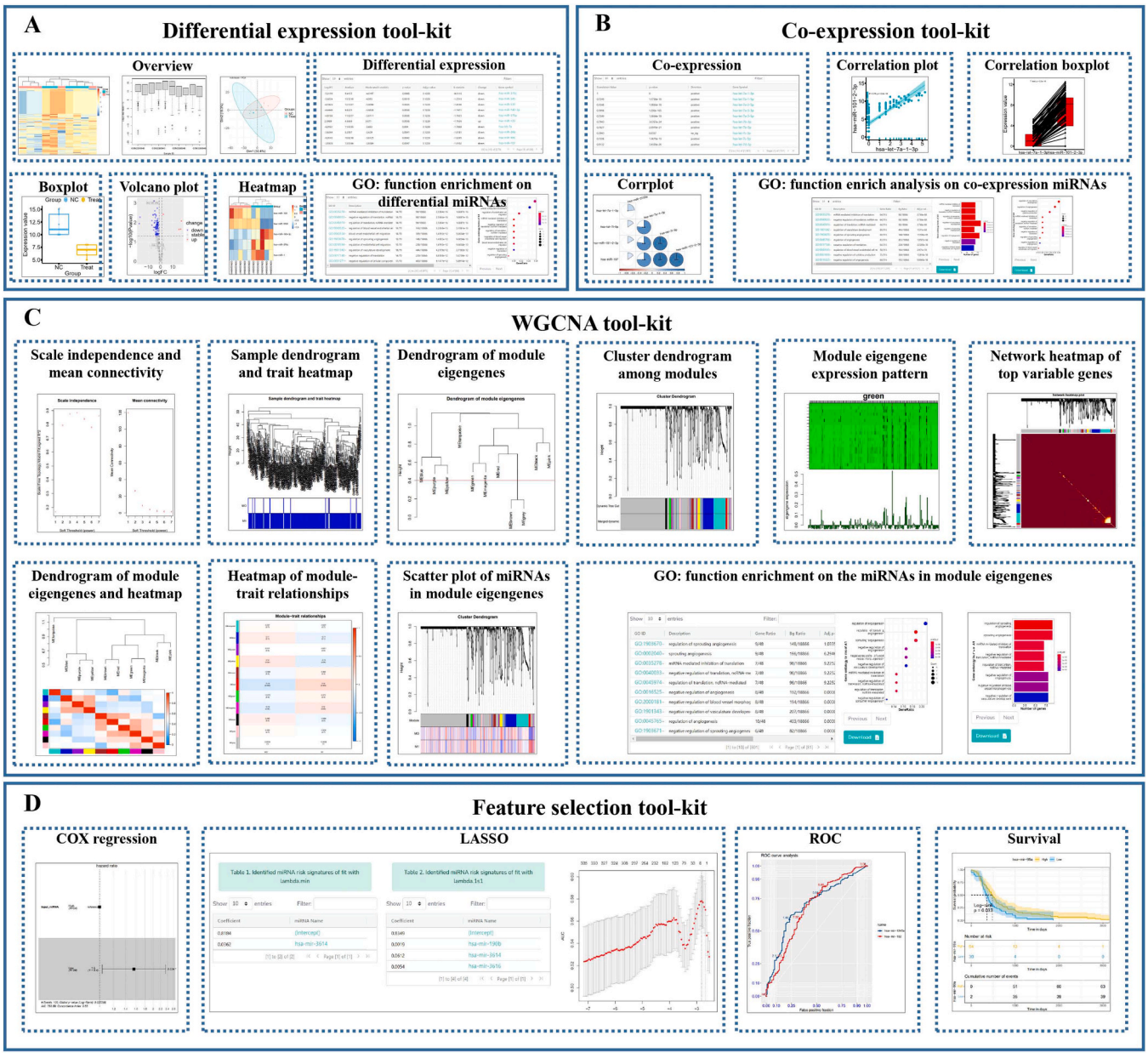

Fig. 2. The advanced analysis functions and their visualizations of ExomiRHub. A. Differential expression tool-kit. B. Co-expression tool-kit. C. WGCNA tool-kit. D. Feature selection tool-kit. GO, gene ontology; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic.

circRNA, and lncRNA thereby regulating their biological functions [16], while the genomics variation in miRNA can impact its biological function in human diseases [28]. 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 [29], circBase [30], and miRBase [21], respectively. Furthermore, we installed miRanda [31] 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. Similar to a miRNA target prediction tool, the miRNA mutation evaluation tool allows users to input sequences of mature miRNAs before and after mutation, set score and energy cutoffs, and specify target gene databases including mRNA 3′-UTR, lncRNA, and circRNA. It predicts reliable target genes for both pre- and post-mutation miRNA sequences using miRanda [31]. Subsequently, it calculates changes in target genes (including gained, lost, and retained) of post-mutation compared to pre-mutation miRNAs, along with their interaction details.

2.5. *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 [16]. Subsequently, we relied on 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 [32], org.Hs.eg.db, pathview [27], and topGO R packages, in conjunction with the miRNA validated targets in miRTarBase [16]. The R packages of stringr, dplyr, and jsonlite were used for data processing.

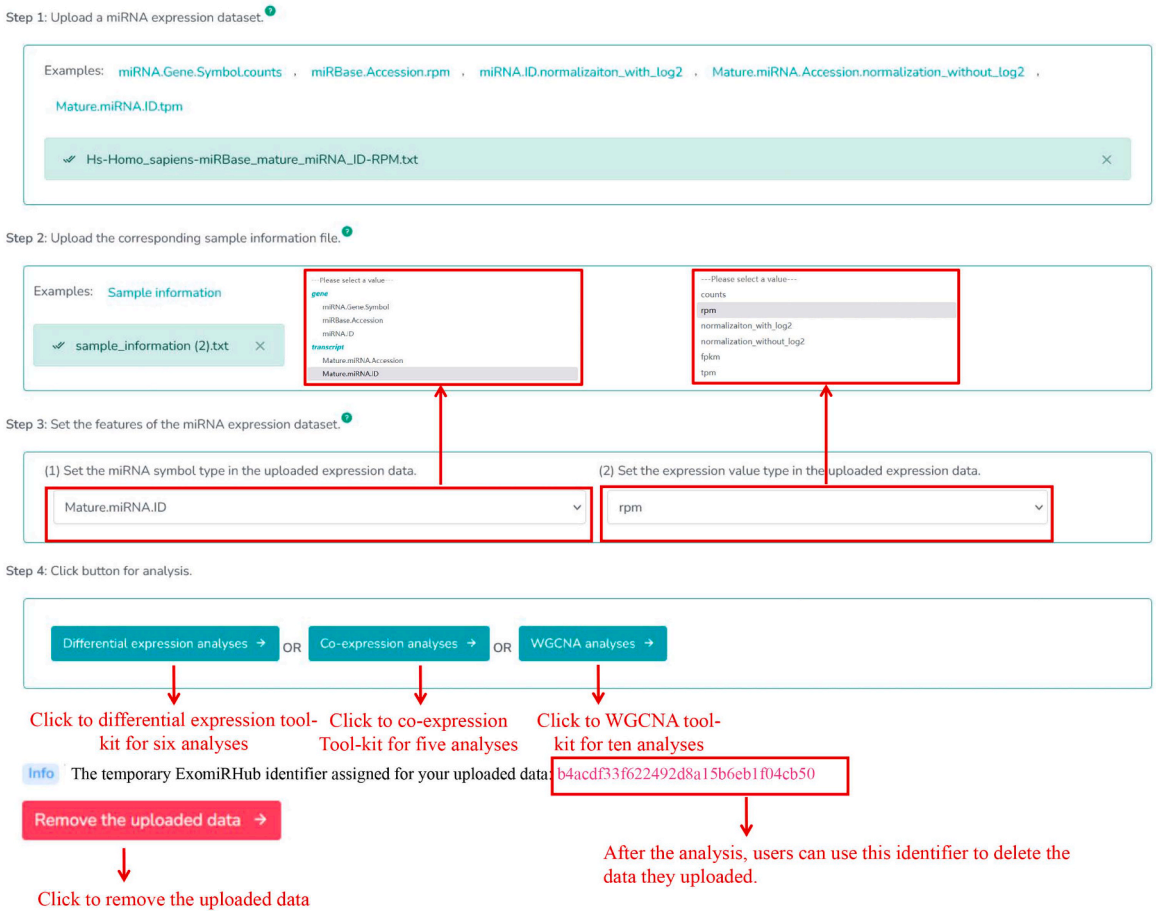

Fig. 3. The web interface and user guide of the web service application on the ExomiRHub database.

The miRNA function prediction and miRNA target validation tools enable users to input a set of miRNAs, which can be named using accession numbers for primary and mature miRNAs from the miRBase database, as well as by ID or symbol. Additionally, the miRNA function prediction tool offers parameters such as p-value, q-value, and GO type to highlight significant and specific GO terms. Data processing utilized the R packages stringr, dplyr, and jsonlite.

2.6. Web implementation

The front-end and back-end framework utilized for the implementation of ExomiRHub has been previously detailed [33,34]. 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 were outlined in Table S1. Finally, ExomiRHub was deployed on the Apache Tomcat server.

3. Results

3.1. Data landscape and access

ExomiRHub provides access to 191 human extracellular miRNA expression datasets (Table S2) associated with 112 disease phenotypes, 62 treatments, and 24 genotypes, comprising 2656 miRNAs, 29,198 samples, and 23 sample types (Fig. 1A and Table 1). Statistically, 80.63 % (154/191) of these datasets, including 145 datasets specifically from exosomes, are derived from extracellular vesicles. The remaining 19.37 % (37/191) include circulating miRNA profiles in serum (13.09 %, 25/191), plasma (4.71 %, 9/191), and whole blood (1.57 %, 3/191). Sample types include extracellular vesicles, exosomes, and circulating miRNAs from diverse sources such as whole blood, serum, plasma, urine, ascites, cerebrospinal fluid, endometrial fluid, follicular fluid, pericardial fluid, liquid milk, saliva, tissue fluid, umbilical cord blood, tracheal aspirate, fecal fluid, and cell culture supernatants under specific genotypes and treatments. These treatments range from adjuvant chemotherapy to targeted therapy. Further statistical analysis revealed that the disease phenotypes in ExomiRHub are associated with over 23 body sites, with the blood, breast, and lung containing the highest number of datasets. The top three contributing countries to the datasets are China, Japan, and the United States, which account for 25.65 % (49/191), 24.08 % (46/191), and 20.42 % (39/191) of the datasets, respectively. Detailed data features, including body site, sample type, isolation method, sample resource, extraction method, disease phenotype, genotype, and treatment, are accessible via the homepage of the database.

Currently, approximately 63.87 % (122/191) of the extracellular miRNA datasets are linked to 52 cancer subtypes, augmenting its utility in cancer research. ExomiRHub also integrates miRNA expression quantification data from 16,012 samples across 156 cancer subtypes from 42 TCGA projects (Fig. 1A), each annotated with extensive biospecimen and clinical information such as demographics, diagnosis, progression, tumor microenvironment, and lifestyle. This detailed annotation facilitates precise sample selection for analytical comparisons.

Moreover, ExomiRHub facilitates quick searches of interesting extracellular miRNA expression datasets. Users can seamlessly navigate the ExomiRlyzer application to identify significant extracellular miRNAs

Table 1
The comparison of the data volume, coverage, and functionality of ExomiRHub with other databases.

Databases		ExomiRHub	CancerMIRNome	CMEP	miREV	EVAtlas	EVpedia	Vesiclepedia	ExoCarta
Extracellular miRNA transcriptome data	Number of entries	325,713	-	-	-	-	17,280	621,659	50,516
	Number of diseases	112	32 (Only cancer)	< 66	9	> 40	-	-	-
	Number of datasets or publications	191 datasets	40 datasets	66 datasets	13 datasets	57 datasets	263 datasets, 6879 publication	47 datasets, 3533 publications	286 publications
	Number of samples	29,198	28,633	10,419	428	2030	1114	-	-
	Number of sample types	23	2	6	2	8	-	252	71
	Molecular type	miRNA	miRNA	miRNA	miRNA	miRNA, tRNA, piRNA, rRNA, snRNA, snoRNA	mRNA, miRNA, protein, lipid, metabolite	mRNA, miRNA, protein, lipid, metabolite, DNA	mRNA, miRNA, protein, lipid
	Organism	Human	Human	Human	Human	Human	4 organisms, such as human	56 organisms, such as human	10 organisms, such as human
TCGA miRNA transcriptomes	Number of projects	42	33	×	×	×	×	×	×
	Number of cancer subtypes	156	33	×	×	×	×	×	×
	Number of samples	16,012	10,554	×	×	×	×	×	×
Differential expression analysis functions	Principal component analysis	√	√	×	√	×	×	×	×
	Boxplot	√	√	×	×	√	×	×	×
	Differential expression	√	√	√	×	×	×	×	×
	Heatmap	√	×	√	×	×	×	×	×
	Volcano plot	√	√	×	×	×	×	×	×
	Function enrichment	√, GO	√, KEGG	√, KEGG	×	×	√, GO	×	×
	TCGA Pan-cancer analysis	×	√	×	×	√	×	×	×
Co-expression analysis function	Co-expression	√	×	×	×	×	×	×	×
	Correlation plot	√	×	×	×	×	×	×	×
	Correlation boxplot	√	×	×	×	×	×	×	×
	Corrplot	√	×	×	×	×	×	×	×
	Function enrichment	√, GO	×	×	×	×	×	×	×
	miRNA-target correlation	×	√	×	×	×	×	×	×
WGCNA analysis function	Scale Independence and mean connectivity	√	×	×	×	×	×	×	×
	Sample dendrogram and trait heatmap	√	×	×	×	×	×	×	×
	Dendrogram of module eigengenes	√	×	×	×	×	×	×	×
	Cluster dendrogram among module	√	×	×	×	×	×	×	×
	Module eigengene expression pattern	√	×	×	×	×	×	×	×
	Network heatmap of top variable genes	√	×	×	×	×	×	×	×
	Dendrogram of module eigengenes and heatmap	√	×	×	×	×	×	×	×
	Heatmap of module-trait relationships	√	×	×	×	×	×	×	×
	Scatter plot miRNAs in module eigengenes	√	×	×	×	×	×	×	×
	Function enrichment	√, GO	×	×	×	×	×	×	×
Feature selection analysis function	COX regression	√	×	×	×	×	×	×	×
	ROC	√	√	√	×	×	×	×	×
	LASSO	√	√	√	×	×	×	×	×
	Survival	√	√	×	×	×	×	×	×
Tool	miRNA function prediction	√	√	×	×	×	×	×	×

(*continued on next page*)

Table 1 (*continued*)

Databases		ExomiRHub	CancerMIRNome	CMEP	miREV	EVAtlas	EVpedia	Vesiclepedia	ExoCarta
	miRNA mutation evaluation	√	×	×	×	×	×	×	×
	miRNA target prediction	√	×	×	×	×	×	×	×
	miRNA target validation	√	×	×	×	×	×	×	×
	Find suitable reference miRNA	×	×	×	√	×	×	×	×
	miRNA related drug	×	×	×	×	√	×	×	×
	Tissue and source specific expression miRNA	×	×	×	×	√	√	×	×
	Network analysis	×	×	×	×	×	√	×	×
	Sequence search	×	×	×	×	×	√	×	×
Server	Uploaded data for analysis	√	×	×	×	×	√	×	×
	Search	√	√	√	×	√	√	√	√
	Browse	√	√	√	×	√	√	√	√
	Download	√	√	√	×	×	×	√	√
	Submission	√	×	×	×	×	×	√	√

Note: "-", the data and function feature is not provided or cannot be statistically calculated; "√", the data and function feature is provided by the database; "× ", 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.

through user-defined investigations of the dataset. Direct external links to the resources EVAtlas [15], miRandola [8], ExoCarta [11], and Vesiclepedia [9] have been added for the miRNAs in ExomiRHub. All data in the ExomiRHub are available in various formats, such as TABLE, CSV, and JSON, enabling users to freely download and analyze. Furthermore, ExomiRHub encourages users to submit novel human extracellular miRNA expression data to the database via the submit webpage (http://www.biomedical-web.com/exomirhub/submit). Upon approval by our review committee, these submissions are incorporated into subsequent database versions.

3.2. Comprehensive web analytical and visualization functions, servers, and tools

ExomiRHub facilitates the exploration of the role of extracellular miRNAs and the identification of non-invasive biomarkers for human diseases through web applications such as ExomiRlyzer and TCGA-miRNAlyzer. These applications provide access to four state-of-the-art bioinformatics tool-kits: differential expression, co-expression, WGCNA, and feature selection. Collectively, these tool-kits enable 25 analytical and visualization functions to analyze human extracellular miRNA expression data and cancer-related miRNA expression data. The functions include various aspects of differential expression, co-expression, WGCNA, Gene Ontology (GO) function enrichment, COX regression, receiver operating characteristic (ROC) analysis, least absolute shrinkage and selection operator (LASSO), and survival analysis (Fig. 2). With comprehensive biomedical information annotating each dataset sample, all 25 functions are customizable, allowing users to select specific samples, define their own groups, and set parameters for customized comparative analyses (Figs. S2 and S3). Recognizing variability in isolation methods, assay platforms, and quality control across studies, ExomiRHub restricts comparisons and analyses to datasets within the same study to ensure data quality and reliability.

To serve a broader research community, ExomiRHub includes a web service that permits users to upload and standardize miRNA expression data along with sample metadata, assigning a temporary identifier to each dataset (Fig. 3). This feature allows for extensive analysis and visualization using the co-expression, differential expression, and WGCNA tool-kits (Fig. 2A–C). For data security, users can delete their data anytime via the assigned temporary identifier. Uploaded data are regularly cleared monthly (Fig. 3).

To enhance the utility of ExomiRHub in the miRNA research community, it offers four additional valuable tools for predicting and validating potential functions and targets (such as mRNAs, lncRNAs, and circRNAs) of miRNAs and their variations. The four tools included (I) miRNA function prediction, predicting miRNA function based on the GO and KEGG annotations of its experimentally validated targets in miRTarBase; (II) miRNA target prediction, predicting miRNA targets using miRanda, including mRNA, circRNA, and lncRNA; (III) miRNA mutant evaluation, predicting the gain or loss targets of miRNA after its mutation; and (IV) miRNA target validation, identifying the experimentally 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). Finally, the results from these applications can generate publication-quality vector figures in PDF format and tables for further analysis and download (Fig. 2).

3.3. Case study: discover extracellular miRNA biomarkers associated with angiogenesis for diagnosis and monitoring the progression of glioma

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

To elucidate the difference in angiogenesis-related pathways between G1 and G2, we conducted a differential analysis and identified 48 dysregulated exosomal miRNAs ($|\log2(\text{fold change})| \geq 0.5$ and P value

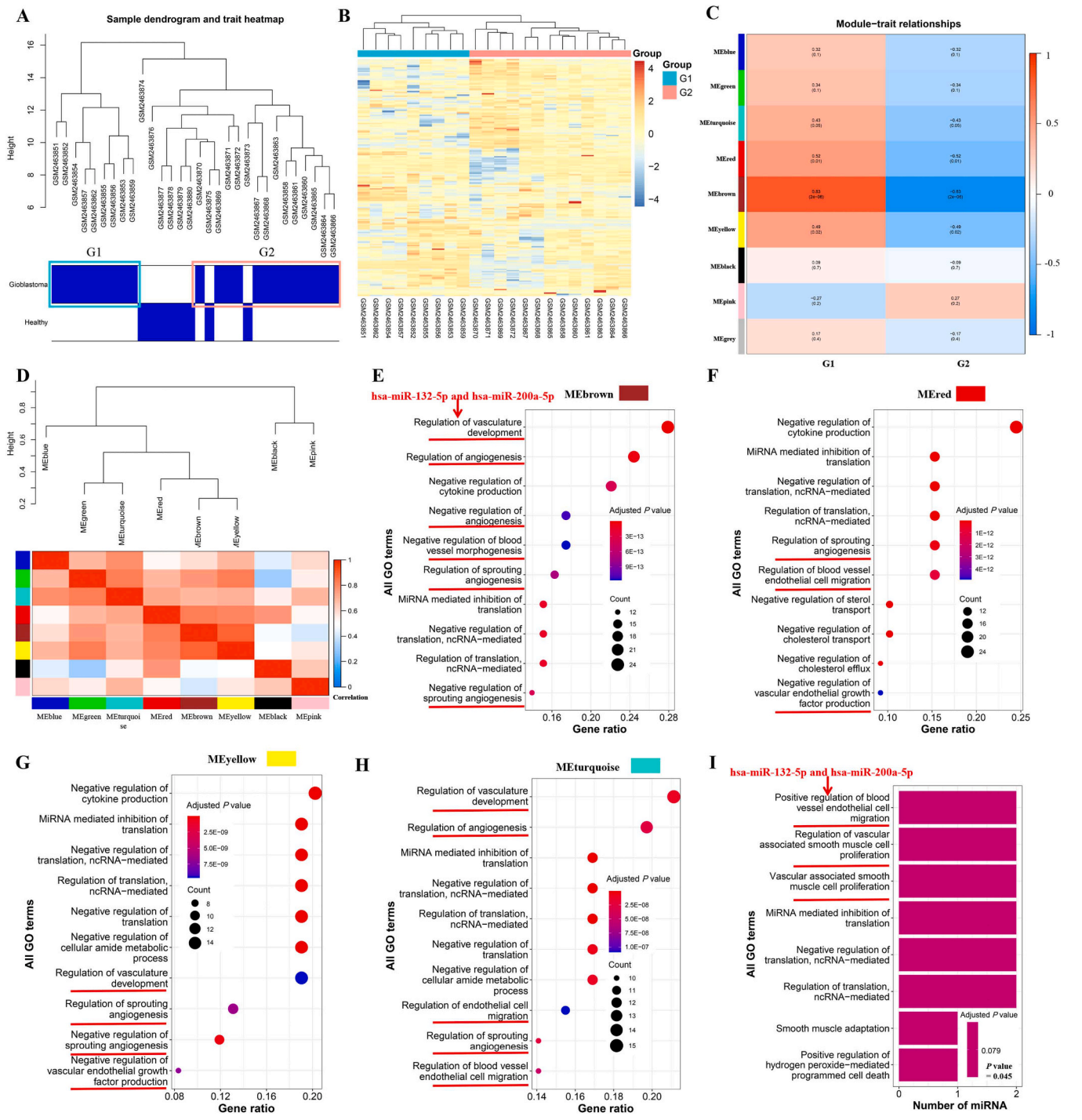

Fig. 4. Extracellular miRNA regulates angiogenesis related functional pathways in glioblastoma. A. Sample dendrogram and trait heatmap results of WGCNA analysis suggested that patients with glioblastoma have different exosomal miRNA expression profiles from healthy controls, and these patients can be divided to two subgroups named G1 and G2. B. The two 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 functional enrichment analysis of the MEbrown module. F. The GO functional enrichment analysis of the MEred module. G. The GO functional enrichment analysis of the MEyellow module. H. The GO functional enrichment analysis of the MEturquoise module. **I.** The GO functional enrichment of the differential exosomal miRNA genes between the two subgroups. (The method and min module gene size for the WGCNA analysis was set to average and 30, respectively).

≤ 0.05), including 21 downregulated and 27 upregulated exosomal miRNAs, between the two subgroups (Fig. 5A and Table S3). In line with the GO functional enrichment analysis of the module eigengenes, we confirmed that these differentially expressed exosomal miRNAs were enriched in angiogenesis-related pathways (Fig. 4I). For example, dysregulation of hsa-miR-132–5p and hsa-miR-200a–5p was associated with the positive regulation of blood vessel endothelial cell migration (Figs. 4I and 5A–C), and these two miRNAs were significantly correlated (Fig. 5D). Furthermore, the functions of hsa-miR-132–5p and hsa-miR-200a-5p were consistently enriched in angiogenesis-related pathways based on the GO annotations of their co-expressed exosomal miRNAs (Fig. S4).

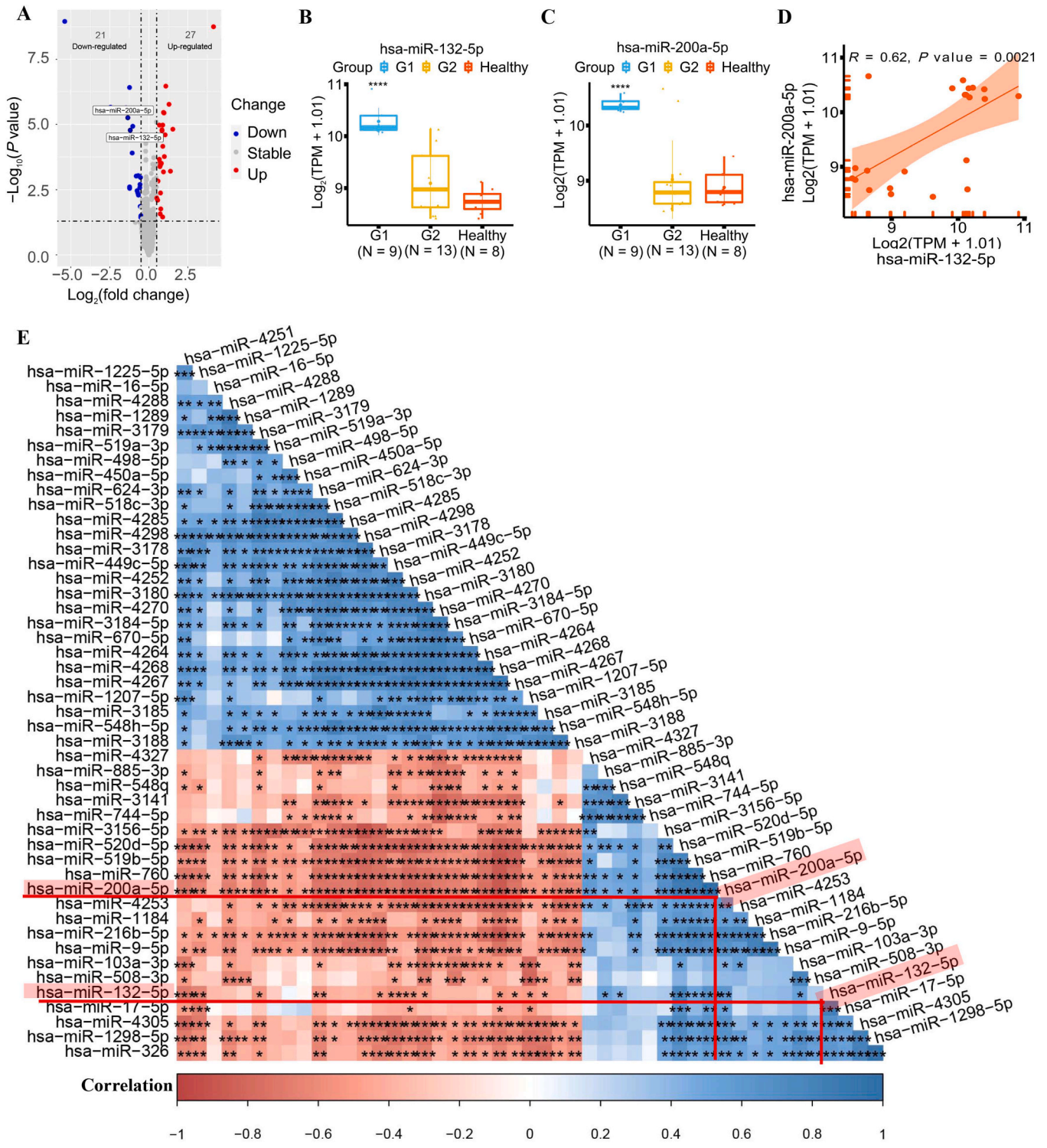

Fig. 5. Angiogenic exosomal miRNAs, hsa-miR-132–5p and hsa-miR-200a-5p, were upregulated in the G1 glioblastoma subgroup as opposed to that in the G2 glioblastoma subgroup. Moreover, a significant correlation between these miRNAs was observed. A. Volcano plot.for hsa-miR-132–5p and hsa-miR-200a–5p. B. Boxplot of hsa-miR-132–5p. C. Boxplot of hsa-miR-200a–5p. D. Expression correlation between hsa-miR-132–5p and hsa-miR-200a-5p. E. Corrplot shows hsa-miR-132–5p and hsa-miR-200a–5p associated with other differentially expressed exosomal miRNAs.

To further elucidate the clinical significance of these two exosomal miRNAs in gliomas, we conducted COX regression and survival analyses on glioma datasets integrated from TCGA within ExomiRHub. The results indicated that the downregulation of the angiogenic miRNAs hsa-mir-132 and hsa-mir-200a significantly decreased mortality rates (Fig. 6A) and prolonged the overall survival probability of patients with glioma (Fig. 6B). Consistent with the aforementioned results, the two miRNAs were upregulated in high-grade gliomas (HGG, WHO III, and WHO IV) compared with that in low-grade gliomas (LGG, WHO I, and WHO II) (Fig. 6C and D) and showed a significant correlation (Fig. 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 for

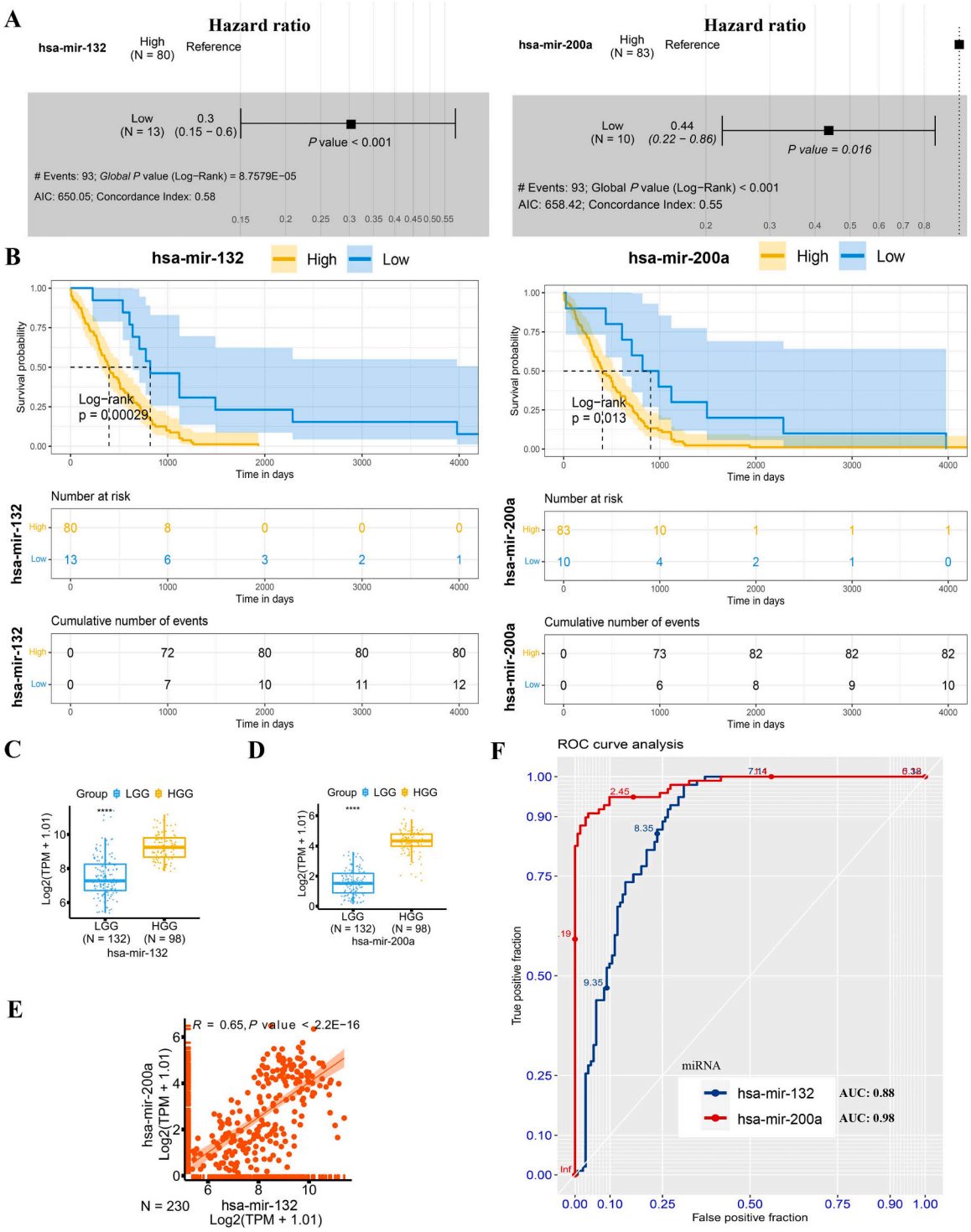

Fig. 6. Angiogenic exosomal miRNAs, hsa-miR-132–5p and hsa-miR-200a–5p, are associated with glioblastoma progression. A. COX regression analysis suggested that down-regulation of hsa-mir-132 and hsa-mir-200a significantly decreased the mortality rates and prolonged the overall survival rate of patients with glioma. B. Survival analysis suggested that down-regulation of hsa-mir-132 and hsa-mir-200a significantly decreased the mortality rates and prolonged the overall survival rate of patients with glioma. C. hsa-mir-132 was up-regulated in HGG. D. hsa-mir-200a was up-regulated in HGG. E. hsa-mir-132 and hsa-mir-200a were significantly correlated in glioblastoma tissues. F. ROC analysis indicated that hsa-miR-132 and hsa-miR-200a can serve as independent factors for distinguishing between HGG and LGG. HGG, high-grade glioma; LGG, low-grade glioma; * *** *, *P* value ≤ 0.0001.

distinguishing HGG from LGG (Fig. 6F). Additionally, analysis of an independent dataset from the Chinese Glioma Genome Atlas [32] indicated that hsa-mir-132 and hsa-mir-200a were upregulated in HGG and that their upregulation was related to the poor prognosis of glioma (Fig. S5). 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 by modulating angiogenesis, rendering them potential non-invasive biomarkers for the diagnosis and monitoring of gliomas.

Furthermore, differential expression and WGCNA analyses of the

miRNA expression datasets from the tissues of patients with LGG and HGG indicated significantly distinct miRNA expression patterns between the two grades of patients (Fig. 7). Differentially expressed miRNAs were consistently enriched in the functional pathways involved in angiogenesis regulation (Fig. 7). Clinically, glioblastomas (WHO Grade IV) can be differentiated from anaplastic gliomas (WHO Grade III) by the presence of neoplastic vasculature. Therefore, our non-invasive markers are aligned with known pathological features. These findings suggest that miRNAs play an important role in the progression of LGG to HGG by regulating angiogenesis-related functional pathways. However, these observations must be further validated through in vitro and in vivo, as well as through additional large-scale clinical analyses.

4. Discussion and perspectives

In summary, ExomiRHub, a biologist-oriented database, is designed to comprehensively host and explore human extracellular and intracellular miRNA transcriptomics, providing customizable features to enhance analysis capabilities. ExomiRHub surpasses similar databases (Table 1), such as Vesiclepedia [9], ExoCarta [11], CMEP [12], miREV [14], EVAtlas [15], and EVpedia [35], in terms of data volume, coverage, and functionality. Unlike these databases that compile and

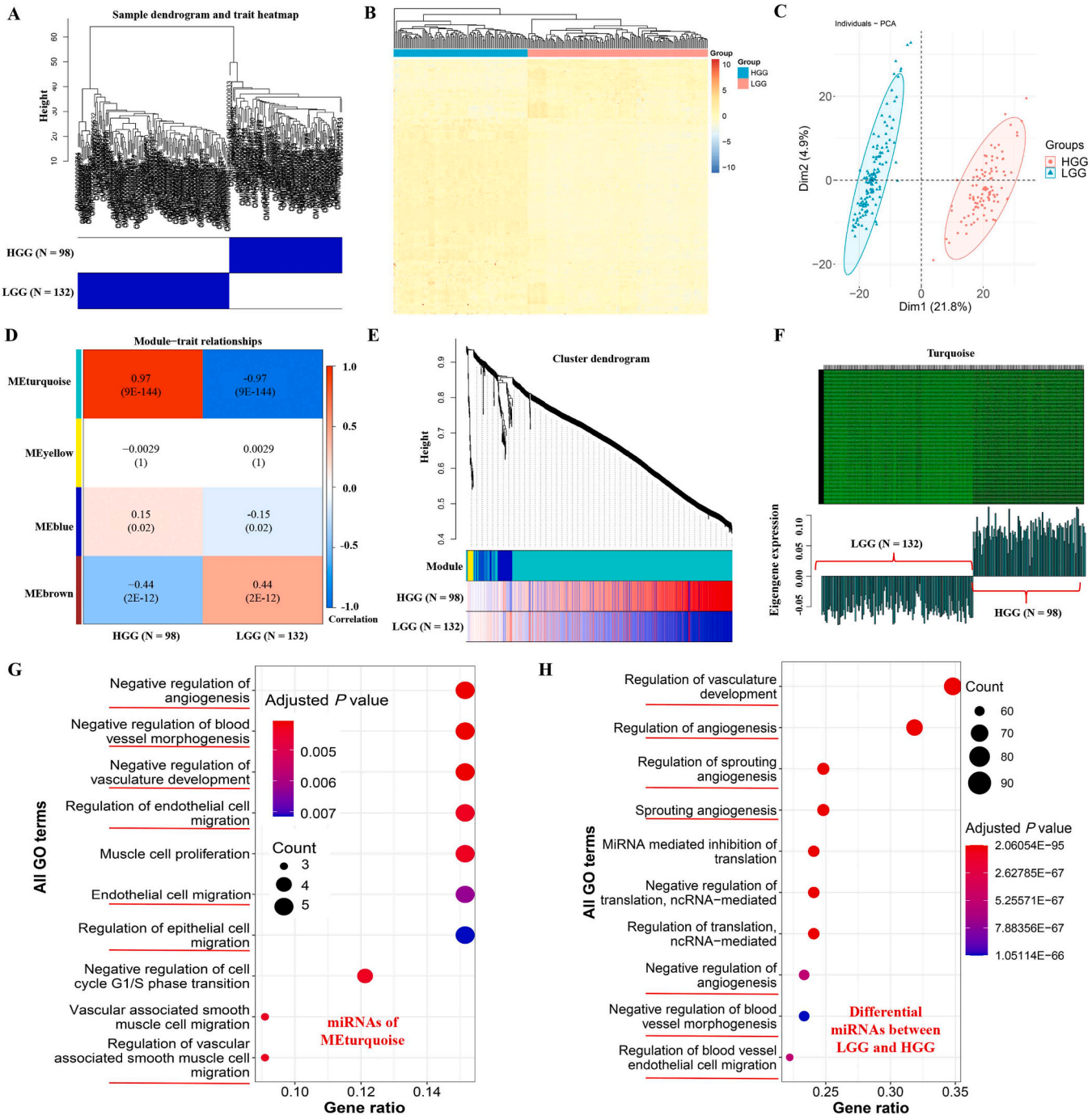

Fig. 7. Intracellular miRNA is associated with the progression of glioblastoma through regulating angiogenesis-related functional pathways. A. Sample dendrogram and trait. B. Heatmap. C. Principal component analysis. D. Module-trait relationship. E. Module eigengene dendrogram and triat heatmap. F. miRNA eigengenes expression of the MEturquoise module. G. GO functional enrichment analysis of the miRNA eigengenes in the MEturquoise module. H. GO functional enrichment analysis of the differential expression miRNA between LGG and HGG.

present experimentally supported disease-related extracellular miRNAs and other biological molecular association data through manual curation of literature and analysis of public expression datasets, ExomiRHub focuses on collecting and standardizing disease-related extracellular and intracellular miRNA expression datasets. It also offers extensive and customizable analytical and visualization functions for exploring these datasets, surpassing the capabilities of existing databases (Table 1).

For example, the datasets of extracellular miRNA in ExomiRHub are 2.2 times larger than that of CMEP and 2.4 times larger than that of EVAtlas. Moreover, ExomiRHub provides a comprehensive suite of functionalities, including differential expression, co-expression, co-expression network, and feature selection analyses, which CMEP and EVAtlas do not offer. Additionally, ExomiRHub provides thousands of miRNA transcriptomic datasets integrated from TCGA, a feature not offered by CMEP and EVAtlas. ExomiRHub also integrates advanced functionalities for miRNA function prediction, mutation evaluation, and target prediction and validation, encouraging in-depth studies on miRNA regulatory mechanisms. We believe these improvements further solidify the position of ExomiRHub as a valuable resource in the landscape of similar databases.

Compared with CancerMIRNome [17], ExomiRHub offers significant advantages(Table 1), including extracellular miRNA expression data for 112 human diseases from 23 sample types, whereas CancerMIRNome offers data only from plasma and whole blood samples of patients with human cancer (Table 1). Consequently, 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, of which 21 are not supported by CancerMIRNome(Table 1). Additionally, ExomiRHub provides more detailed biomedical annotations for each sample than that of CancerMIRNome, including comprehensive demographic information, drug/agent treatment, genotype, immune infiltration, and lifestyle information (Table 1), significantly broadening the application scenarios for ExomiRHub. For example, users can 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 groups, and set parameters for comparative analyses. It also offers a web service application enabling users to perform all analytical and visualization functions with their uploaded data, a feature not provided by existing databases (Table 1). Therefore, as a novel and powerful resource, ExomiRHub is a crucial resource for users delving into the study of extracellular miRNA in human diseases.

We anticipate an increase in the availability of human miRNA data for public resources in the foreseeable future. We are committed to continuously maintaining and updating the analytical and visualization features of ExomiRHub to incorporate new data. 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 lncRNA, circRNA, mRNA, and protein/peptide data. Additionally, we plan to integrate multi-omics data from NCBI GEO [19] and TCGA [22], which include miRNAs, mRNAs, lncRNAs, and circRNAs, to introduce new functions for exploring the 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 increase the utility of ExomiRHub for the miRNA research community.

CRediT authorship contribution statement

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

Declaration of Competing Interest

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 other 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); Basic and Applied Basic Research Foundation of Guangdong Province, China (Grant Nos. 2020A1515110528 and 2021B0101310002); Basic and Applied Basic Research Foundation of Guangzhou, China (Grant Nos. SL2023A04J00291); 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); the Science and Technology Program of Guangzhou, China (Grant No. 2024A04J3341).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.csbj.2024.07.024.

References

[1] Mori MA, Ludwig RG, Garcia-Martin R, Brandao BB, Kahn CR. Extracellular miRNAs: from biomarkers to mediators of physiology and disease. Cell Metab 2019;30:656–73.
[2] Diener C, Keller A, Meese E. Emerging concepts of miRNA therapeutics: from cells to clinic. Trends Genet 2022;38:613–26.
[3] Ho P, Clark IM, Le LTT. MicroRNA-based diagnosis and therapy. Int J Mol Sci 2022;23.
[4] Isaac R, Reis F, Ying W, Olefsky JM. Exosomes as mediators of intercellular crosstalk in metabolism. Cell Metab 2021;33:1744–62.
[5] Makarova J, Turchinovich A, Shkurnikov M, Tonevitsky A. Extracellular miRNAs and Cell-Cell Communication: Problems and Prospects. Trends Biochem Sci 2021;46:640–51.
[6] Xu R, Rai A, Chen M, Suwakulsiri W, Greening DW, Simpson RJ. Extracellular vesicles in cancer - implications for future improvements in cancer care. Nat Rev Clin Oncol 2018;15:617–38.
[7] Li C, Ni YQ, Xu H, Xiang QY, Zhao Y, Zhan JK, et al. Roles and mechanisms of exosomal non-coding RNAs in human health and diseases. Signal Transduct Tar 2021;6:383.
[8] Russo F, Di Bella S, Vannini F, Berti G, Scoyni F, Cook HV, et al. miRandola 2017: a curated knowledge base of non-invasive biomarkers. Nucleic Acids Res 2018;46:D354–9.
[9] Pathan M, Fonseka P, Chitti SV, Kang T, Sanwlani R, Van Deun J, et al. Vesiclepedia 2019: a compendium of RNA, proteins, lipids and metabolites in extracellular vesicles. Nucleic Acids Res 2019;47:D516–9.
[10] Kim DK, Lee J, Kim SR, Choi DS, Yoon YJ, Kim JH, et al. EVpedia: a community web portal for extracellular vesicles research. Bioinformatics 2015;31:933–9.
[11] Keerthikumar S, Chisanga D, Ariyaratne D, Al SH, Anand S, Zhao K, et al. ExoCarta: a web-based compendium of exosomal cargo. J Mol Biol 2016;428:688–92.
[12] Li JR, Tong CY, Sung TJ, Kang TY, Zhou XJ, Liu CC. CMEP: a database for circulating microRNA expression profiling. Bioinformatics 2019;35:3127–32.
[13] Wang X, Chai Z, Pan G, Hao Y, Li B, Ye T, et al. ExoBCD: a comprehensive database for exosomal biomarker discovery in breast cancer. Brief Bioinform 2021;22.

[14] Hildebrandt A, Kirchner B, Nolte-'T HE, Pfaffl MW. miREV: An Online Database and Tool to Uncover Potential Reference RNAs and Biomarkers in Small-RNA Sequencing Data Sets from Extracellular Vesicles Enriched Samples. J Mol Biol 2021;433:167070.
[15] Liu CJ, Xie GY, Miao YR, Xia M, Wang Y, Lei Q, et al. EVAtlas: a comprehensive database for ncRNA expression in human extracellular vesicles. Nucleic Acids Res 2022;50:D111–7.
[16] Huang HY, Lin YC, Cui S, Huang Y, Tang Y, Xu J, et al. miRTarBase update 2022: an informative resource for experimentally validated miRNA-target interactions. Nucleic Acids Res 2022;50:D222–30.
[17] Li R, Qu H, Wang S, Chater JM, Wang X, Cui Y, et al. CancerMIRNome: an interactive analysis and visualization database for miRNome profiles of human cancer. Nucleic Acids Res 2022;50:D1139–46.
[18] Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, et al. Database resources of the national center for biotechnology information. Nucleic Acids Res 2022;50:D20–6.
[19] Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, et al. NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013;41:D991–5.
[20] Davis S, Meltzer PS. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 2007;23:1846–7.
[21] Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res 2019;47:D155–62.
[22] Jensen MA, Ferretti V, Grossman RL, Staudt LM. The NCI Genomic Data Commons as an engine for precision medicine. Blood 2017;130:453–9.
[23] Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47.
[24] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. Bmc Bioinforma 2008;9:559.
[25] Sachs MC. plotROC: A Tool for Plotting ROC Curves. J Stat Softw 2017;79.
[26] Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R. Bioinformatics 2005;21:3940–1.
[27] Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. Bioinformatics 2013;29:1830–1.
[28] Zhang W, Zeng B, Yang M, Yang H, Wang J, Deng Y, et al. ncRNAVar: A Manually Curated Database for Identification of Noncoding RNA Variants Associated with Human Diseases. J Mol Biol 2021;433:166727.
[29] Frankish A, Diekhans M, Jungreis I, Lagarde J, Loveland JE, Mudge JM, et al. GENCODE 2021. Nucleic Acids Res 2021;49:D916–23.
[30] Glazar P, Papavasileiou P, Rajewsky N. circBase: a database for circular RNAs. Rna 2014;20:1666–70.
[31] Miranda KC, Huynh T, Tay Y, Ang YS, Tam WL, Thomson AM, et al. A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. Cell 2006;126:1203–17.
[32] Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov (Camb) 2021;2:100141.
[33] Zhang W, Zhang Y, Min Z, Mo J, Ju Z, Guan W, et al. COVID19db: a comprehensive database platform to discover potential drugs and targets of COVID-19 at whole transcriptomic scale. Nucleic Acids Res 2022;50:D747–57.
[34] Zhang W, Liu Y, Min Z, Liang G, Mo J, Ju Z, et al. circMine: a comprehensive database to integrate, analyze and visualize human disease-related circRNA transcriptome. Nucleic Acids Res 2022;50:D83–92.
[35] Kim DK, Lee J, Kim SR, Choi DS, Yoon YJ, Kim JH, et al. EVpedia: a community web portal for extracellular vesicles research. Bioinformatics 2015;31:933–9.