



# Exo-miRExplorer: A Comprehensive Resource for Exploring and Comparatively Analyzing Exogenous MicroRNAs

Ling-Ling Zheng, Kai-Wen Deng, An-Cheng Deng, Jie Wu, Jian-Hua Yang, Zhao-Rong Lun and Liang-Hu Qu\*

Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory of Biocontrol, RNA Information Center, School of Life Sciences, Sun Yat-sen University, Guangzhou, China

## OPEN ACCESS

### Edited by:

Andrea Masotti,  
Bambino Gesù Children's Hospital  
(IRCCS), Italy

### Reviewed by:

Zhumur Ghosh,  
Bose Institute, India  
Anindya Bhattacharya,  
University of California San Diego,  
USA

### \*Correspondence:

Liang-Hu Qu  
lssqth@mail.sysu.edu.cn

### Specialty section:

This article was submitted to  
Systems Microbiology,  
a section of the journal  
Frontiers in Microbiology

**Received:** 16 December 2016

**Accepted:** 18 January 2017

**Published:** 01 February 2017

### Citation:

Zheng L-L, Deng K-W, Deng A-C,  
Wu J, Yang J-H, Lun Z-R and Qu L-H  
(2017) Exo-miRExplorer: A  
Comprehensive Resource for  
Exploring and Comparatively  
Analyzing Exogenous MicroRNAs.  
*Front. Microbiol.* 8:126.  
doi: 10.3389/fmicb.2017.00126

MicroRNAs (miRNAs) are small regulatory RNAs that play important roles in animals, plants, and viruses. Deep-sequencing technology has been widely adopted in miRNA investigations. However, it is still a big mysterious why nearly all sequencing data contain miRNA sequences from exogenous species, called exo-miRNAs. In this study, we developed a novel platform, exo-miRExplorer, for mining and identifying exo-miRNAs from high-throughput small RNA sequencing experiments which originated from tissues and cell lines of multiple organisms. Thousands of exo-miRNAs are characterized with their expression abundance, the RNA families, original organisms and the sequencing platforms presented in exo-miRExplorer. Subsequently, we used exo-miRExplorer to perform further analysis. Comparative analysis of the exo-miRNAs between different sequencing datasets revealed significant correlation of exo-miRNAs between experiments in the same study. The plant-derived exo-miRNAs analysis provided robust evidence for non-diet source of exo-miRNAs. Virus-derived exo-miRNA analysis showed that pathogen RNAs could transfer to host cells and exist in deep-sequencing result at abundance level. In conclusion, exo-miRExplorer provides users with an integrative resource to facilitate detection and analysis of exo-miRNAs. exo-miRExplorer is available at the following URL: <http://rna.sysu.edu.cn/exomiRDB/>.

**Keywords:** exogenous, microRNA, deep-sequencing, organisms, contamination

## INTRODUCTION

MicroRNA (miRNA) is a class of small RNA with 18–24 nt length, and widespread distribution in animals, plants, and viruses (Lee et al., 1993; Bartel, 2004). microRNA genes are first transcribed into primary miRNA transcript (pri-miRNA), and then cleaved by RNase III enzyme to generate precursor miRNA (pre-miRNA) (Lee et al., 2003). The pre-miRNAs are about 70–100 nt length and formed hairpin structure. They are further processed into a pair of small RNAs—mature miRNA and miRNA star in the cytoplasm and loaded into the Argonaute complex (Lee and Doudna, 2012). miRNAs could regulate gene expression at the post-transcriptional level through RNA interference (RNAi) mechanism (Fire et al., 1998) by pairing to the 3' UTR of mRNA (He and Hannon, 2004). miRNAs are involved in almost all biological process in cell, especially in development and differentiation, while their abnormal expression can lead to cause cancer and other diseases (Esquela-Kerscher and Slack, 2006; Bartel, 2009; Ameres and Zamore, 2013).

Next-generation sequencing (NGS) or deep sequencing technology could rapidly retrieve almost all RNA sequences in cells, which becomes a powerful method to investigate microRNAs. Interestingly, most, if not all, sequenced datasets contain small RNA fragments from the genome of exogenous species, called exo-miRNAs (exo-miRNAs) (Jeang, 2012). These fragments were always being discarded from subsequent analysis. Recently, several studies found exogenous plant microRNAs in the blood plasma of humans and other animals by small RNA deep sequencing experiments (Wang et al., 2012; Zhang L. et al., 2012; Liang et al., 2014; Zhou et al., 2015). A representative study found 25 known plant miRNAs in Chinese healthy donors, among which miR168a and miR156a showed considerable levels of expression (Zhang L. et al., 2012). These authors further considered that these exo-miRNAs derived from food intake, and believed that these exo-miRNAs have cross-kingdom regulation functions (Zhang L. et al., 2012; Zhou et al., 2015). Later, they even found that these exo-miRNAs could cross the placental barrier to regulate the fetal gene expression (Li et al., 2015). Some researchers followed these idea to successfully detect dietary RNAs in ingesting mammals (Liang et al., 2015; Yang et al., 2015; Chin et al., 2016). Other researchers, however, disputed that these exo-miRNAs were most likely originated from environmental RNAs pollution in the process of experiments (Zhang Y. J. et al., 2012; Dickinson et al., 2013; Snow et al., 2013; Witwer et al., 2013; Lusk, 2014; Tosar et al., 2014; Witwer and Hirschi, 2014; Witwer, 2015; Bagci and Allmer, 2016), which could disturb the assessment of real gene expression and possibly affecting conclusions derived from the NGS data analysis (Tosar et al., 2014).

These studies sparked keen debate and drew much attention from both academic and commercial community (Jiang et al., 2012; Vaucheret and Chupeau, 2012; Wang et al., 2012; Witwer, 2012; Chen et al., 2013). Several controversial questions have been raised by researchers. What organisms are the origin of these exo-miRNAs? How can these exo-miRNAs get into the sequencing datasets? Could the same exo-miRNA also be re-detected in the sequencing result performed by another research group? To answer these questions, we need both a comprehensive view of exo-miRNA spectra in the whole cell, and comparative analysis of them between different studies using different sequencing methods and platforms.

To address these questions, we established a specific resource, exo-miRExplorer, to comprehensively explore the exo-miRNAs from high-throughput small RNA sequencing experiments. A total of 563 published small RNA sequencing datasets were used for exo-miRNAs classification and annotation, including the messages of expression abundance, the RNA families, original organisms, and sequencing platforms (Table S3). Subsequently, we used exo-miRExplorer to perform comparative analysis of the exo-miRNAs. In addition, we also developed a dynamic web interface to facilitate the integration, interactive evaluation, and visualization of these exo-miRNAs. Our exo-miRExplorer will facilitate scientists to effectively recognize the exo-miRNAs in the deep-sequencing experiments and reduce the potential risk from experimental contamination.

## MATERIALS AND METHODS

### Obtain Small RNA Sequencing Data

A total of 563 small RNA deep-sequencing datasets were compiled from multiple related studies (Table 1), which were downloaded from NCBI SRA and GEO databases (Acland et al., 2014). The raw data contain more than 192 million sequencing reads. These datasets were classified into different clades, species, tissues, and cell lines according to the description on the website or related literature.

### Obtain Genome and Annotation Files

All known miRNAs were downloaded from miRBase (release 21) (Kozomara and Griffiths-Jones, 2014). The genome sequences and transcript sequences were downloaded from NCBI Reference Sequences (RefSeq) (Pruitt et al., 2012), UCSC Bioinformatics websites (Meyer et al., 2013) and other special databases (Table 2). Human (*Homo sapiens*, UCSC hg19), mouse (*Mus musculus*, UCSC mm10), chicken (*Gallus gallus*, v3) and *Ciona intestinalis* (JGI v2.0) genome sequences were downloaded from the UCSC Bioinformatics website; *Caenorhabditis remanei*, *Caenorhabditis elegans*, and *Caenorhabditis briggsae* genome sequences were downloaded from WormBase (Harris et al., 2014); *Drosophila melanogaster* genome sequences were download from Flybase (Tweedie et al., 2009); *Bombyx mori* genome sequences were download from silkDB (Duan et al., 2010). The *Trypanosoma brucei* genome sequences were download from TriTrypDB (Aslett et al., 2010). Known non-coding RNAs were downloaded from UCSC (Karolchik et al., 2014), Ensembl (Flicek et al., 2014), and Rfam (Gardner et al., 2009).

### Filtering Endosome Small RNAs Derived from Host

So far as we know, many studies have used the “map and remove” approach for the identification of exogenous RNAs (Wang et al., 2012; Tang et al., 2013; Pandya et al., 2014; Tosar et al., 2014), and we applied the same strategy with more detailed

**TABLE 1 | List of species collected in exo-miRExplorer, and the number of exo-miRNAs in each species.**

Clade	Full	Samples	exo-miRNA Number
Mammal	<i>Homo sapiens</i>	382	7143
Mammal	<i>Mus musculus</i>	128	2574
Mammal	<i>Sus scrofa</i>	6	268
Vertebrate	<i>Gallus gallus</i>	4	342
Cephalochordata	<i>Branchiostoma floridae</i>	1	119
Deuterostome	<i>Ciona intestinalis</i>	4	62
Insect	<i>Bombyx mori</i>	3	226
Insect	<i>Drosophila melanogaster</i>	8	84
Nematode	<i>Caenorhabditis briggsae</i>	4	88
Nematode	<i>Caenorhabditis elegans</i>	16	549
Nematode	<i>Caenorhabditis remanei</i>	4	126
Protozoan parasite	<i>Trypanosoma brucei</i>	2	61
Prokaryote	<i>Escherichia coli</i>	1	1

**TABLE 2 | Names of species and databases where the genome sequences were downloaded.**

Specie	Database	Edition/Reference
Human	UCSC Bioinformatics website	UCSC hg19
Mouse		UCSC mm10
Chicken		<i>Gallus gallus</i> , v3
<i>Ciona intestinalis</i>		JGI v2.0
<i>Caenorhabditis remanei</i>	WormBase	Harris et al., 2014
<i>Caenorhabditis elegans</i>		
<i>Caenorhabditis briggsae</i>		
<i>Drosophila melanogaster</i>	Flybase	Tweedie et al., 2009
<i>Bombyx mori</i>	silkDB	Duan et al., 2010
<i>Trypanosoma brucei</i>	TriTrypDB	Aslett et al., 2010

annotation. We firstly build a reference dataset which contained a variety of host sequences, including genome sequences, mRNAs, tRNAs, rRNAs, snoRNAs, lncRNAs, and microRNA precursors. In this dataset, we collected all possible transcripts. We mapped the reads to genome and then to transcript, in order to solve the cases like small RNAs derived from exon-exon junction region of mRNAs and gene fusion. Small RNA sequencing reads were firstly mapped to these reference sequences with bowtie program (Langmead et al., 2009). If a read can map to any sequence in the reference dataset with one mismatch allowed, it will be filtered as the endosome small RNAs. The left reads are used for exo-miRNAs identification. We firstly classified all the mature microRNAs from miRBase according to their source species and source kingdom. Sequencing reads were exactly aligned to these microRNAs and classified as “exo-miRNAs from the same kingdom” and “exo-miRNAs from another kingdom.” To reduce the false positive as far as possible, exo-miRExplorer provides multiple features of the potential exo-miRNAs, such as the “Supporting number,” which represents the number of particular miRNA in multiple datasets. An RPM (Reads Per Million) value was used to evaluate the expression of exo-miRNAs (Motameny et al., 2010). These features may provide users to gain exo-miRNAs for more evidence. The identified exo-miRNAs were clearly classified, annotated, and loaded into a local database. The microRNA gene names in the exo-miRExplorer follow the naming guidelines described by MicroRNA Registry (Kozomara and Griffiths-Jones, 2014).

## The Expression of Exo-miRNAs

“Expression” tool was developed to demonstrate the abundance and frequency of “non-metazoa” microRNAs. It will provide a comprehensive view of exo-miRNAs in a particular tissue or cell line. Users can choose their source of interest by clicking the “select” box in the top middle, after which a heat map graph will be displayed on the web page. The horizontal list represents the miRNA family, and the vertical list indicates the exogenous species. Each color grid demonstrates the “frequency value” of the particular exo-miRNA. The “frequency value” of each exo-miRNA is evaluated by a method modified from reads per million

(RPM) values (Mortazavi et al., 2008) and calculated by the following formula:

$$F_{ij} = \sum_{k=1}^n \frac{S_{ijk}}{T_k} * 10^9$$

where *i* represents the species; *j* represents the miRNA family; *k* represents the source of the experiment, which contain *n* datasets; *S* represents the reads number of the exo-miRNA in the *k*th dataset; *T* represents all the sequencing reads in the *k*th dataset. There are 3 ways to show the order of the heatmap graph: (1) by miRNAs, sorting the miRNAs by name on the horizontal axis; (2). by species, ordering the species by name on the vertical axis; (3). by frequency, sorting the values by “frequency value” on both the horizontal and vertical axes. These three ordering can be chosen from the pull-down box on the top right. Changes of the order trigger dynamic effects on the website that provide the user with comprehensive information about exo-miRNAs in whole tissues or cell lines from multiple angles.

## RESULTS

### A Comprehensively Annotated Catalog of Exo-miRNAs

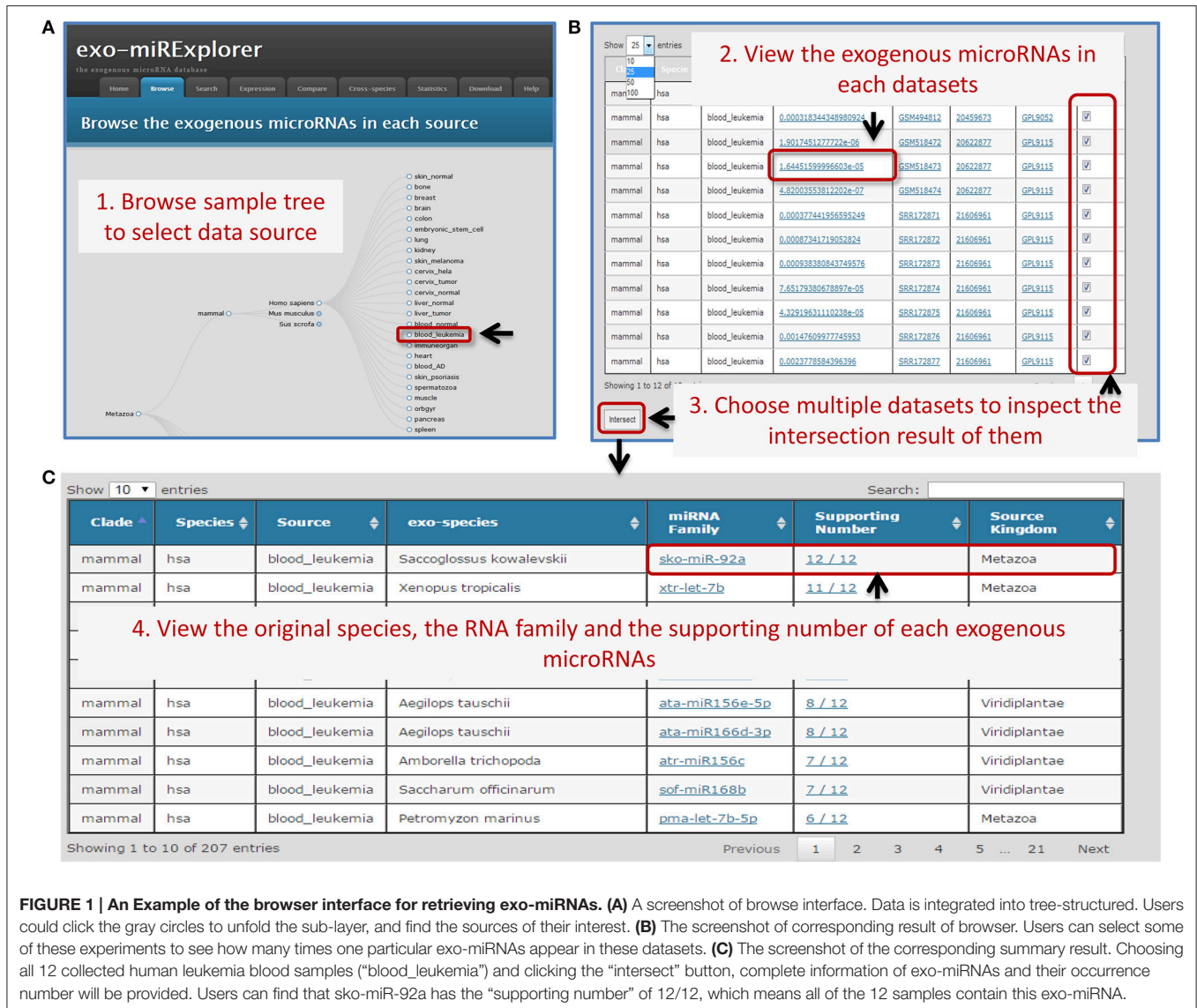
Exo-miRExplorer has been designed to focus primarily on the exo-microRNAs presented in the small RNA deep-sequencing datasets. A total of 563 small RNA deep-sequencing datasets were compiled from multiple studies. The experiments were performed by 56 distinct laboratories with multiple sequencing platforms. Thirteen species were selected in exo-miRExplorer (Table 1).

In exo-miRExplorer, 2754 exo-miRNAs have been clearly annotated and classified according to their original organisms, miRNA family and sequencing source dataset. The sequencing dataset can be viewed from the tree-structure browse page (Figure 1A). Each leaf node represents an experiment sample (tissue, cell line, or development stage). exo-miRNAs in multiple samples can be retrieved by integrated samples (Figure 1). Other features of exo-miRNAs are also listed, including the expression abundance, the source kingdom of species and their portion in the sequencing library. Detailed information about the microRNA, the sample description and the sequencing platform can link to their source annotation website (including miRBase, GEO, and PubMed etc.) by the dynamic hyperlink.

To fully demonstrate the expression profiles of exo-miRNAs, we provide a heat map view of each sample. To fully exclude the endogenous microRNAs (endo-miRNAs), a strict filtering standard was used to get rid of all the metazoa miRNAs (Materials and Methods). In the “Expression” tool, only “non-metazoa” exo-miRNAs are shown in the expression page (Figure 2). Next, we will use exo-miRExplorer to verify if these exo-miRNAs actually present in cells or result from experimental contamination.

### Comparative Analysis of Exo-miRNAs between “Intra-Study” and “Inter-Study”

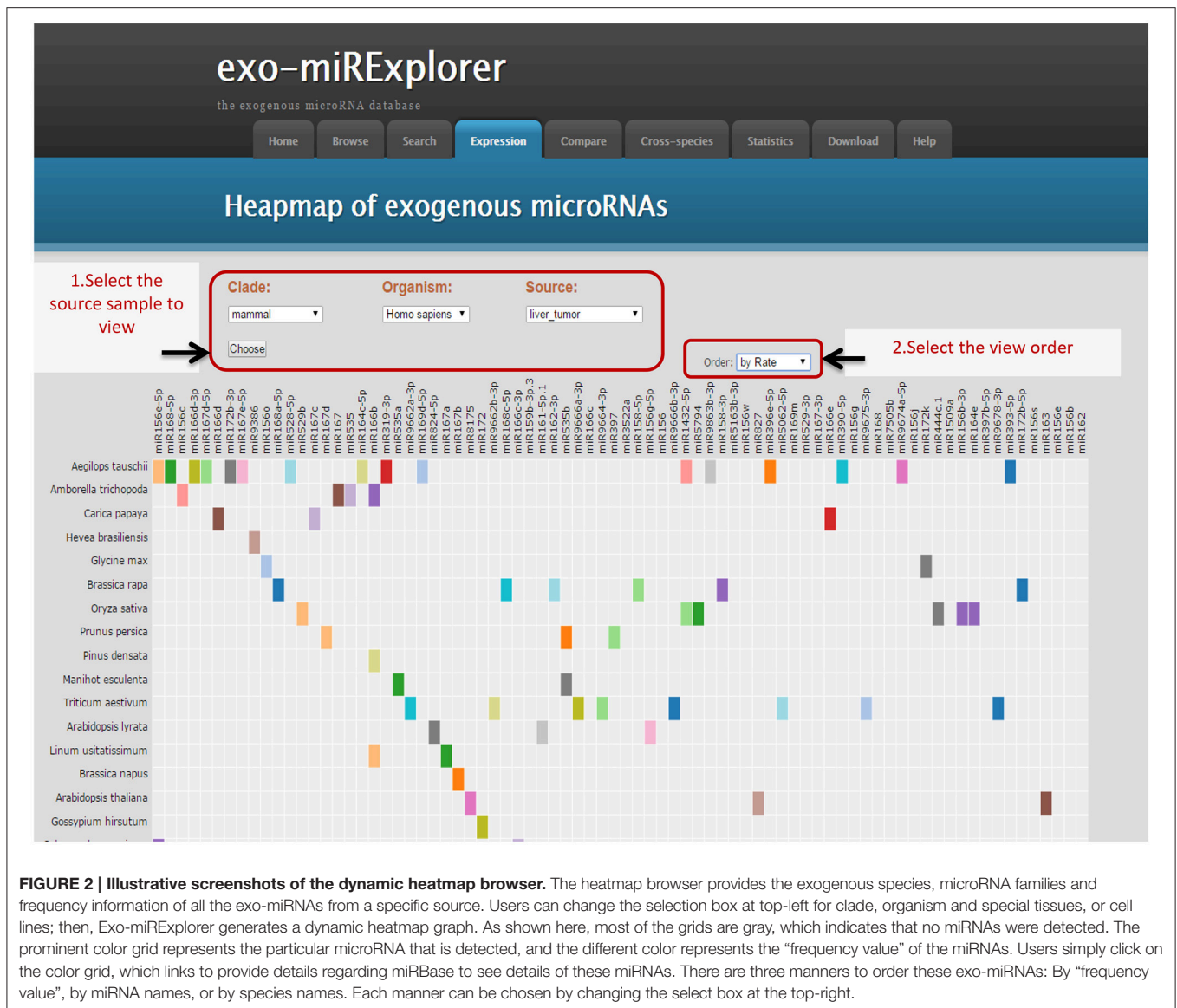
Exo-miRExplorer provides a “compare” tool for researchers to carry out a comparative analysis of exo-miRNAs in different



**FIGURE 1 | An Example of the browser interface for retrieving exo-miRNAs. (A)** A screenshot of browse interface. Data is integrated into tree-structured. Users could click the gray circles to unfold the sub-layer, and find the sources of their interest. **(B)** The screenshot of corresponding result of browser. Users can select some of these experiments to see how many times one particular exo-miRNAs appear in these datasets. **(C)** The screenshot of the corresponding summary result. Choosing all 12 collected human leukemia blood samples (“blood\_leukemia”) and clicking the “intersect” button, complete information of exo-miRNAs and their occurrence number will be provided. Users can find that sko-miR-92a has the “supporting number” of 12/12, which means all of the 12 samples contain this exo-miRNA.

sequencing datasets. Users would like to find whether there are common exo-miRNAs between different datasets. If one exo-miRNA frequently presents in multiple datasets produced by the same research group (the “intra-study”), it has a high probability to be the pollution during the process of sequencing. On the other hand, if one exo-miRNA frequently presents in the datasets performed by different research groups (“inter-study”), it may be genuine functional exo-miRNA in the cell. Following this idea, we used the “compare” tool to perform the analysis. For “intra-study” analysis, two groups of small RNA sequencing datasets which were performed by the same laboratory we chosen (Joyce et al., 2011). Dataset 1 was from the normal skin tissues, while dataset 2 was from the psoriatic skin tissues (Figure 3A and Figure S1A). The comparison result showed that, in total, there are 275 exo-miRNAs in either of the datasets, 44.36% (122 out of 275) of them were observed in both of the datasets. Pearson correlation analysis revealed

a strong positive association between these two datasets ( $p$ -value < 0.0001) (Figure 3B). We next performed “inter-study” comparison. These datasets were obtained from melanoma skin tissue, but were performed by another laboratory (Figure 3A) (Stark et al., 2010). When comparing “dataset 1” and “dataset 3” (Figure S1B), only 3.2% (6 in 186) of exo-miRNAs was found to be shared between the two groups, and no significant correlation was observed between the two datasets ( $p$ -value = 0.68); A similar result was observed between the comparison of “dataset 2” and “dataset 3” ( $p$ -value = 0.73) (Figure 3 and Figure S1C). The comparative analysis results demonstrated a significant correlation of exo-miRNAs between datasets within “intra-study,” but no correlation between “inter-study” which were performed using the same type of tissue (Figure 3A). Combined with previous investigations (Tosar et al., 2014), we can deduce that these “intra-study” shared exo-miRNAs are more likely to derive from contamination. These contaminations may



**FIGURE 2 | Illustrative screenshots of the dynamic heatmap browser.** The heatmap browser provides the exogenous species, microRNA families and frequency information of all the exo-miRNAs from a specific source. Users can change the selection box at top-left for clade, organism and special tissues, or cell lines; then, Exo-miRExplorer generates a dynamic heatmap graph. As shown here, most of the grids are gray, which indicates that no miRNAs were detected. The prominent color grid represents the particular microRNA that is detected, and the different color represents the “frequency value” of the miRNAs. Users simply click on the color grid, which links to provide details regarding miRBase to see details of these miRNAs. There are three manners to order these exo-miRNAs: By “frequency value”, by miRNA names, or by species names. Each manner can be chosen by changing the select box at the top-right.

be derived from the preparation of samples in the laboratory or from the sequencing process in the sequencing center.

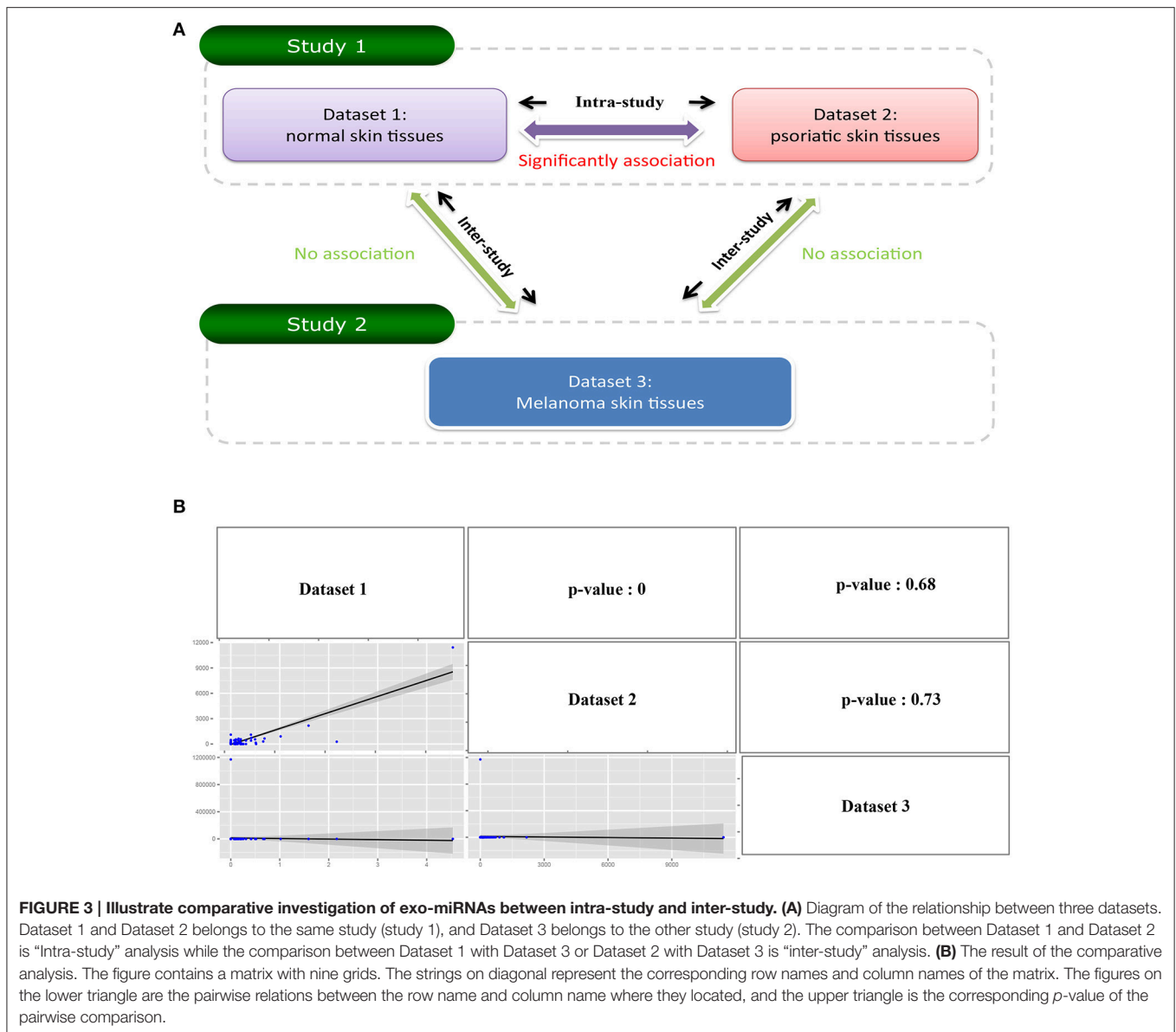
## Inspect Exogenous Plant-Derived microRNAs in Animal Samples

To determine whether the plant miRNAs exist in the samples of animal and other species, all samples collected in exo-miRExplorer were examined to find exo-miRNAs derived from plants. The results showed that plant-derived exo-miRNAs existed in all of the categories in exo-miRExplorer (Figure 4). In human samples, 237 plant-derived miRNAs were detected in 382 sequencing samples (Table S1). These miRNA families include miR168, miR156, miR166, miR167, and miR172 etc.

The most frequently presented miRNA is miR168a, which was detected in 95 samples from 11 distinct studies. The average abundance is 22114.9 RPM. In total, miR168a is presented in 155 samples and multiple species, including human, mouse,

chicken, worm, even in the lower eukaryote parasite *T. brucei* and prokaryote *Escherichia coli*. It should be noted that *T. brucei* is an early-branched single-cell animal. It has been believed that microRNA genes are absent from this ancient organism (Wen et al., 2011; Zheng et al., 2013). In addition, prokaryote *E. coli* even lacks the RNAi mechanism, which is replaced by a distinct system called CRISPR-Cas (clustered regularly interspaced short palindromic repeat). This system is used to against invading phages and plasmids (van der Oost et al., 2009). Moreover, both of these two organisms do not rely on a plant as their food and, therefore, there is no chance for plant-derived small RNAs transferred into these two organisms by food-intake.

miR156a is another frequently observed exo-miRNA from plants. It could be detected in 65 samples in exo-miRExplorer with the average abundance of 86.69 RPM. miR156 is a conserved family in plants, which has been shown to play important roles in controlling the agronomic traits of plants (Jiao et al.,

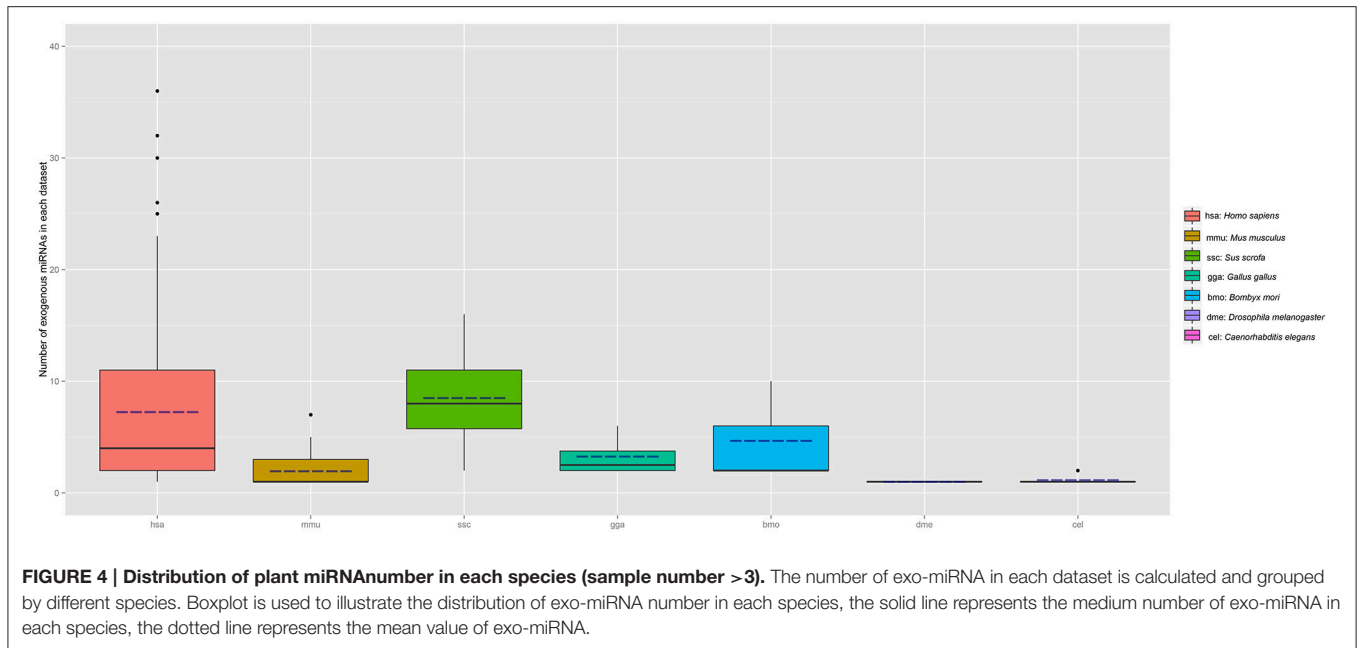


2010; Miura et al., 2010; Zheng and Qu, 2015). Moreover, a recent investigation found that plant miR156a has significant sequence identity to the microRNA in a representative of Cnidaria (sea anemone *Nematostella*) (Moran et al., 2014). We then inspected this sequence in other animal species. To our surprise, miR156a has highly sequence similarity to multiple animal sequences, including human, chimp, Rhesus monkey, mouse, pig, chicken, insects, and *Branchiostoma floridae* (Figure 5). In addition, the *Helianthus argophyllus* miR156c, which is in the same family with miR156a, could perfectly match to the human genome in two regions. These two regions are located in the intron region of two distinct genes, anaplastic lymphoma receptor tyrosine kinase (ALK), and autism susceptibility candidate 2 (AUTS2) (Figure S2). Therefore, we consider that it is highly possible that the miR156 sequence found in NGS data is actually derived from mammal itself.

Therefore, the high sequence similarity of plant-derived exo-miRNA with the animal reference genome sequence will remind us to take special caution of these exo-miRNAs in our own analysis.

## Investigation of Virus-Derived MiRNAs in B-Cell Samples

In addition to contamination, scientists are more concerned about genuine exo-miRNAs in the samples. Recently, several researchers have found viral transcripts in lymphoma cell lines by deep-sequencing technology (Cao et al., 2015). Considering that virus could be transferred into host cells, it is intriguing to assume that exo-miRNAs from the virus could be detected in host cells. We then chose immune organ as an example to investigate all the virus-derived exo-miRNAs.



Rice	TGACAGAAGAGAGTGAGCAC
Human	aGACAaAAGAGAGTGAGCAC
Chimp	aGACAaAAGAGAGTGAGCAC
Rhesus	aGACAaAAGAGAGTGAGCAC
Mouse	TGcCAGcAGAGAGTGAGCAC
Sus scrofa	TGACAGAAGtGAcTGAGCAC
Chicken	aGACAaAAGAGAGTGAGCAC
Drosophila	aGACAaAAGAGAGTGAGCAC
Silkworm	aGACAaAAGAGAGTGAGCAC
C.intestinalis	TGggAGAAGAGAGTGAGCAC
C.elegans	aGACAaAAGAGAGTGAGCAC
Branchiostoma	TGACAaaaTAGAGTGAGCAC
Nematostella	TGACAGAAGAtAG-AAGcCTG

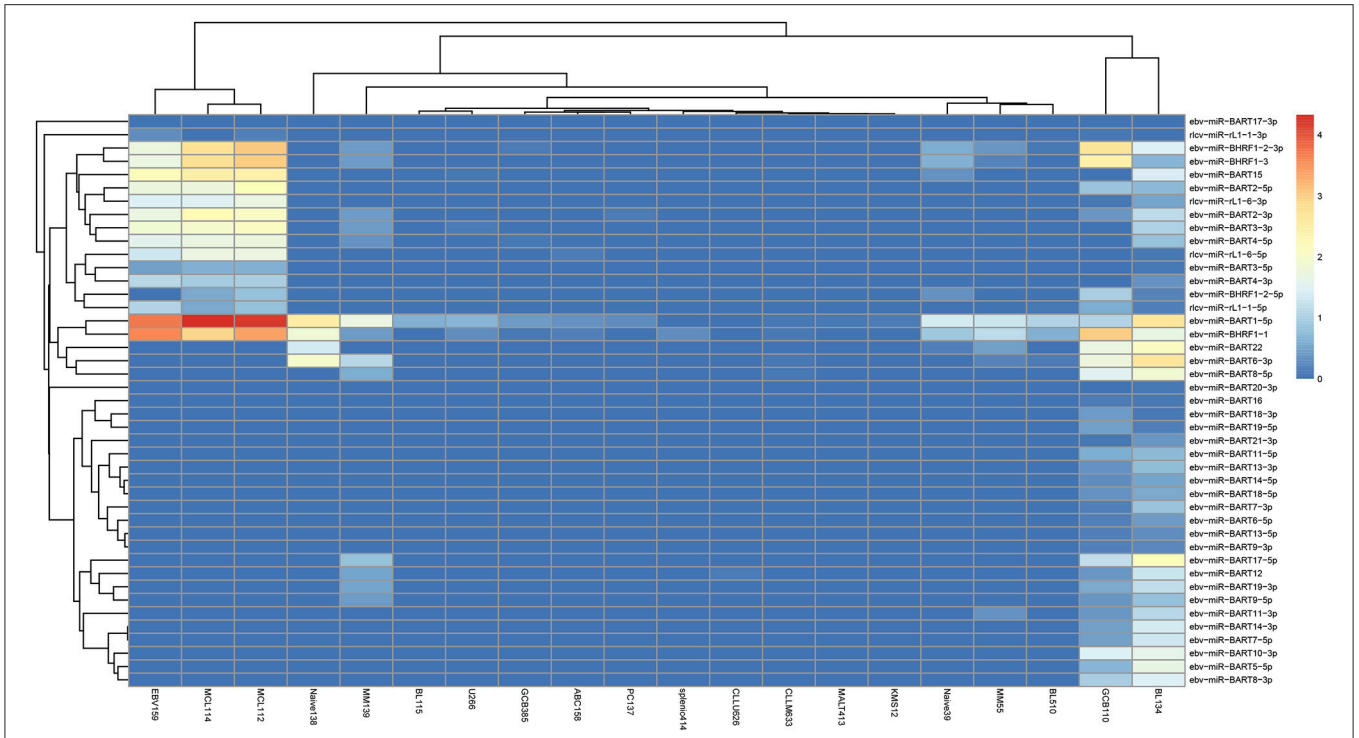
**FIGURE 5 | The similarity sequence of osa-miR156a in other organisms.**

In exo-miRExplorer, there are 43 samples related to lymphoid cells and bone marrow samples, which were performed by two different groups (Jima et al., 2010; Schotte et al., 2011). In total, 43 virus-derived exo-miRNAs were found (Figure 6). Most of them are derived from Epstein-Barr virus (EBV) (38 in 43 = 88.37%), a human herpesvirus. Among them, ebv-miR-BART1-5p and ebv-miR-BHRF1-1 are two most frequently observed miRNAs (Table S2). These miRNAs are significantly high expressed in the EBV activated B-cell line (ebv-159) and two mantle cell lymphoma samples (MCL112 and MCL114) (GSM497062 and GSM497061) (Figure 6). However, these EBV-derived miRNAs are rarely observed in other tissues and cell lines. EBV has been demonstrated to be associated with a number

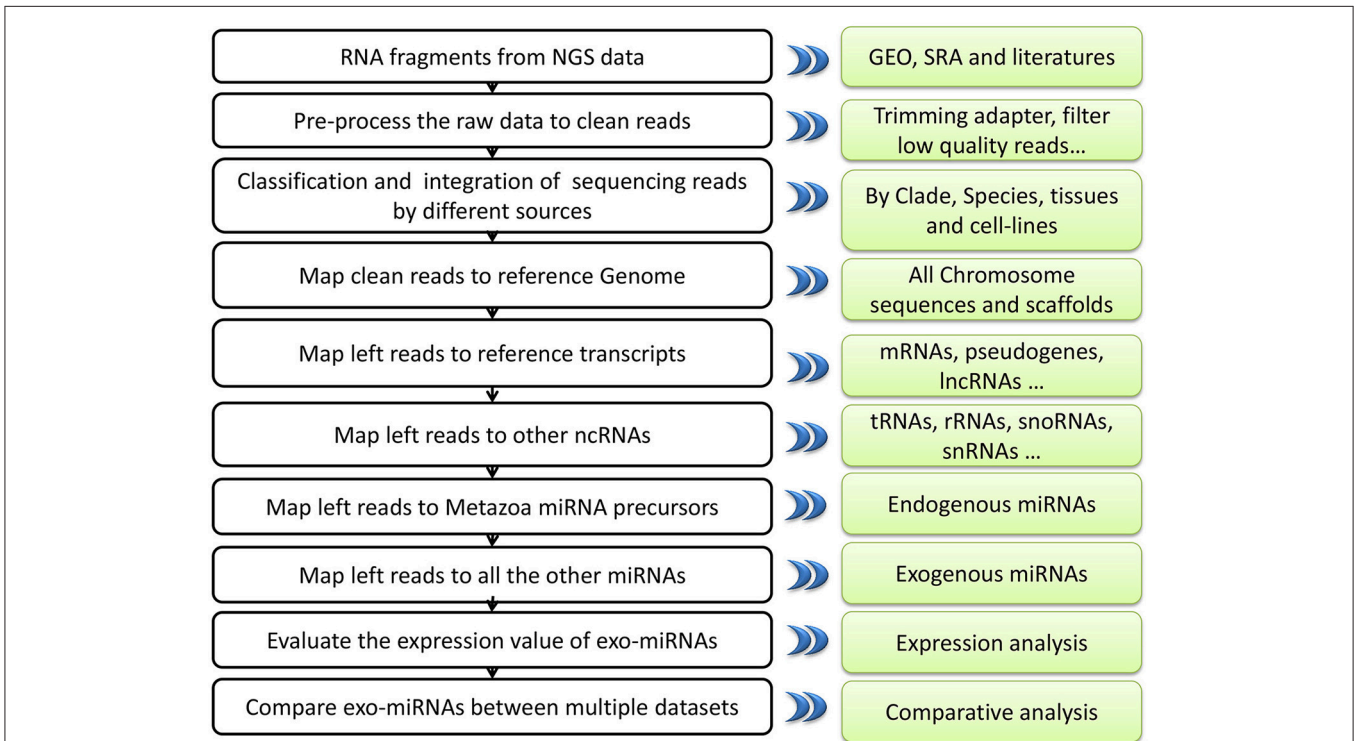
of B-cell cancers and lymphoproliferative disorders (Saha and Robertson, 2011). The finding that exo-miRNAs presented in the deep-sequencing results of human samples and cell lines will help to reveal the interaction relationship of viruses and hosts. More experiments should be designed to reveal the potential functions and application value of these virus-derived exo-miRNAs.

## DISCUSSION

Next-generation sequencing technology can deeply detect nearly all of the RNAs in samples. Nevertheless, it can also bring in some pollution fragments to the result (Olarerin-George and Hogenesch, 2015). The major goal of exo-miRExplorer is to help the researchers to identify the exo-miRNAs from small RNA deep sequencing data (Figure 7). These exo-miRNAs might either be common contaminations or transferred from other species. In one hand, if they are frequently observed in exo-miRExplorer, they have a high probability of a common contaminations. exo-miRExplorer has demonstrated 30 most frequently presented exo-miRNAs in statistics page (Figure S3), researchers must be cautious about these exo-miRNAs if they present in their data. These contaminations are quite similar with the endogenous RNAs, it will make researchers overestimate the expression level of endogenous genes, which will possibly lead to bias or even completely wrong conclusions for the experiments. In addition, if most of the sequencing reads are derived from exo-miRNAs, which point out the data cannot be used for further analysis. In another hand, if these exo-miRNAs are transferred from other species, we should find the source of these exo-miRNAs in exo-miRExplorer, and try to reveal the mechanisms of their cross-species regulation. Therefore, it's of great importance to our database that we want to remind the researchers to be careful



**FIGURE 6 | The expression of virus-derived miRNAs in each B-cell samples.** The horizontal axis illustrates the abbreviation of the cell-line according to the original names in GEO. The vertical axis shows the name of exogenous RNA. The color represents the normalized RPM value of the exo-miRNA.



**FIGURE 7 | A systematic overview of EXO-miRExplorer core framework.** All results generated by EXO-miRExplorer are deposited in a MySQL relational database and displayed in the modern web browsers.



about the frequent potential exo-miRNAs that might either be common contaminant or new mechanism yet to discover.

When the sequencing reads are mapped to the reference genome, analysis methods always allow 1–2 nt error tolerance. However, our finding revealed that exogenous plant miR156a could match multiple animal genomes with 2 mismatches. miR156a had a relatively high abundance in the sequencing datasets and could be found in many public datasets. If the mapping criterion is restricted to 2 mismatches, these contamination exo-miRNAs will interfere with the measurement of actual gene expression number. Therefore, based on our experience, we strongly suggest that the parameter of mapping sequencing reads to genome should be only one mismatch allowed to exclude the impact of exo-miRNAs. Moreover, these highly similar exo-miRNAs need deeply classification and clearly excluded by special parameters when processing small RNAs sequencing data.

The latest investigations revealed that exogenous RNA from virus or parasites could be detected in the sequencing datasets, and could be used as novel biomarkers for diagnosis of pathogen infection (Li et al., 2010; Hoy et al., 2014; Kawano et al., 2014; Tritten et al., 2014). Interestingly, results revealed that EBV viral microRNAs could induce tumor metastasis in nasopharyngeal carcinoma (Cai et al., 2015). In fact, viral miRNAs are also found in exo-miRExplorer at an abundance level. In addition, these viral miRNAs show tissue-specific expression, especially in immune associated organs. In most viral infections, the host innate immune system can block the viral replication cycle. However, whether these viral miRNAs observed in immune organs are the cause or the consequence of the host immune response is unclear. It is necessary therefore to investigate the relationship between viral miRNAs and cancers systematically. Obviously, further works are required to understand how these virus-derived RNAs contribute to the capacity of pathogens in communicated with the host cells. We hope that our results could inspire the creative scientists who are interesting to investigate this fascinating research field.

In our experience, we consider that the most important and difficult thing is to distinguish the contamination exo-miRNAs from the actual infected exo-miRNAs. Here we provide three major features for users to recognize the actual exo-miRNAs from contamination: (1) The observed frequency. If one miRNA could be observed in multiple samples performed by independent groups, it should highly possible be the real exo-miRNA. exo-miRExplorer provide “supporting number” for users to inspect this feature. (2) The evolutionary distance between the species which contain the same type of exo-miRNAs. In our analysis, human and *E. coli* have the same type of plant miRNAs, it is hard to believe that these exo-miRNAs are derived from food. When users find the expression of one particular exo-miRNA, they can retrieve all the species which contain the same exo-miRNA in exo-miRExplorer. (3) Whether other miRNAs from the same species could also be observed in the sample? The ratio of miRNAs observed in other species to the total miRNAs in their original specie. If one species could release miRNA to another species, it should not release only one or two types of microRNAs. Although rice miRNAs are frequently observed in other species,

most of them are miR-168a and miR-156a. Of 713 miRNAs recorded in rice, only 6.17% of them could be observed in other species. In comparison, 86.36% of recorded 44 EBV miRNAs could be observed in human samples. We consider that these three features can be used to distinguish the real exo-miRNAs from the potential contamination RNAs.

Considering the false positive in miRNA identification and high homology of miRNAs between different organisms, we cannot completely exclude the potential miRNAs in which some fragments are actually derived from endogenous RNAs, though strict filter standard has been used in our analysis. Other possibilities also exist of course. For instance, if genome-sequencing is incomplete, it will inevitably contain sequencing error, and if the RNAs go through some RNA process events like novel splicing, RNA editing, or gene fusion events, they can result in endogenous fragments exactly the same to exo-miRNAs. exo-miRExplorer will keep updating by improving the algorithms and integrating more sequencing datasets to filter false positive results. We also try to answer the questions regarding the mechanism of exo-miRNAs transfer or even communicate with the host molecules.

## AUTHOR CONTRIBUTIONS

LQ, JY, ZL, and LZ conceived and designed the study. LZ, KD performed the data analysis pipeline. AD, JW collected, classified, and handled the source data. LQ, JY, ZL, JW, AD, KD, and LZ wrote the paper. All authors read and approved the manuscript.

## FUNDING

This research is supported by the Ministry of Science and Technology of China; National Basic Research Program (No. 2011CB811300) from the National Basic Research program (“973” program) to LQ; and the National Natural Science Foundation of China (No. 31230042, 31472058, 31401975, 31471223, 31370791); Guangzhou science and technology plan projects (201504010022). This research is supported in part by the Guangdong Province Key Laboratory of Computational Science and the Guangdong Province Computational Science Innovative Research Team.

## ACKNOWLEDGMENTS

The authors greatly appreciated Dr. Jun-Hao Li and Shun Liu of Sun Yat-sen University for their valuable suggestions on the manuscript. We are also extremely grateful to the reviewers for their important advice and guidance.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2017.00126/full#supplementary-material>

**Table S1 | List of plant-derived miRNAs in exo-miRExplorer.**

**Table S2 | List of virus-derived miRNAs in exo-miRExplorer.**

**Table S3 | List of datasets used for exo-miRExplorer.**

**Figure S1 | The manipulating steps and results of comparative analysis between “intra-study” (A) and “inter-study” (B,C).**

**Figure S2 | Illustrate the two locations of plant microRNA sequence in the human genome.** The sequence of har-miR156c is “TGACAGAAGAGAGGG AGCA,” which can perfectly match two locations in the human genome at the

antisense strand. They are chr2:29,682,257–29,682,275 (A) and chr7:69,461, 270–69,461,288 (B). This figure illustrates the screen shots from the UCSC Genome Browser of these two regions. The black block represents the location of har-miR156c, and the thin lines represent the intron region of gene according to the annotation from multiple databases, including UCSC genes, RefSeq Genes, Ensembl genes, GECODE, and GenBank. The arrows show the transcribe direction of genes.

**Figure S3 | Top 30 exo-miRNAs types and number.**

## REFERENCES

- Acland, A., Agarwala, R., Barrett, T., Beck, J., Benson, D. A., Bolln, C., et al. (2014). Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 42, D7–D17. doi: 10.1093/nar/gkv1290
- Ameres, S. L., and Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* 14, 475–488. doi: 10.1038/nrm3611
- Aslett, M., Aurrecochea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., et al. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res.* 38, D457–D462. doi: 10.1093/nar/gkp851
- Bagci, C., and Allmer, J. (2016). One step forward, two steps back; xenomicroRNAs reported in breast milk are artifacts. *PLoS ONE* 11:e0145065. doi: 10.1371/journal.pone.0145065
- Bartel, D. P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116, 281–297. doi: 10.1016/S0092-8674(04)00045-5
- Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233. doi: 10.1016/j.cell.2009.01.002
- Cai, L., Ye, Y., Jiang, Q., Chen, Y., Lyu, X., Li, J., et al. (2015). Epstein-Barr virus-encoded microRNA BART1 induces tumour metastasis by regulating PTEN-dependent pathways in *Nasopharyngeal carcinoma*. *Nat. Commun.* 6, 7353. doi: 10.1038/ncomms8353
- Cao, S., Strong, M. J., Wang, X., Moss, W. N., Concha, M., Lin, Z., et al. (2015). High-throughput RNA sequencing-based virome analysis of 50 lymphoma cell lines from the cancer cell line encyclopedia project. *J. Virol.* 89, 713–729. doi: 10.1128/JVI.02570-14
- Chen, X., Zen, K., and Zhang, C. Y. (2013). Lack of detectable oral bioavailability of plant microRNAs after feeding in mice reply. *Nat. Biotechnol.* 31, 967–969. doi: 10.1038/nbt.2741
- Chin, A. R., Fong, M. Y., Somlo, G., Wu, J., Swiderski, P., Wu, X., et al. (2016). Cross-kingdom inhibition of breast cancer growth by plant miR159. *Cell Res.* 26, 217–228. doi: 10.1038/cr.2016.13
- Dickinson, B., Zhang, Y. J., Petrick, J. S., Heck, G., Ivashuta, S., and Marshall, W. S. (2013). Lack of detectable oral bioavailability of plant microRNAs after feeding in mice. *Nat. Biotechnol.* 31, 965–967. doi: 10.1038/nbt.2737
- Duan, J., Li, R. Q., Cheng, D. J., Fan, W., Zha, X. F., Cheng, T. C., et al. (2010). SilkDB v2.0: a platform for silkworm (*Bombyx mori*) genome biology. *Nucleic Acids Res.* 38, D453–D456. doi: 10.1093/nar/gkp801
- Esquela-Kerscher, A., and Slack, F. J. (2006). Oncomirs - microRNAs with a role in cancer. *Nat. Rev. Cancer* 6, 259–269. doi: 10.1038/nrc1840
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806–811. doi: 10.1038/35888
- Flicek, P., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2014). Ensembl 2014. *Nucleic Acids Res.* 42, D749–D755. doi: 10.1093/nar/gkt1196
- Gardner, P. P., Daub, J., Tate, J. G., Nawrocki, E. P., Kolbe, D. L., Lindgreen, S., et al. (2009). Rfam: updates to the RNA families database. *Nucleic Acids Res.* 37, D136–D140. doi: 10.1093/nar/gkn766
- Harris, T. W., Baran, J., Bieri, T., Cabunoc, A., Chan, J., Chen, W. J., et al. (2014). WormBase 2014: new views of curated biology. *Nucleic Acids Res.* 42, D789–D793. doi: 10.1093/nar/gkt1063
- He, L., and Hannon, G. J. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5, 522–531. doi: 10.1038/nrg1379
- Hoy, A. M., Lundie, R. J., Ivens, A., Quintana, J. F., Nausch, N., Forster, T., et al. (2014). Parasite-derived MicroRNAs in host serum as novel biomarkers of Helminth infection. *PLoS Negl. Trop. Dis.* 8:e2701. doi: 10.1371/journal.pntd.0002701
- Jiang, K. T. (2012). RNAi in the regulation of mammalian viral infections. *BMC Biol.* 10:58. doi: 10.1186/1741-7007-10-58
- Jiang, M. X., Sang, X. L., and Hong, Z. (2012). Beyond nutrients: food-derived microRNAs provide cross-kingdom regulation. *Bioessays* 34, 280–284. doi: 10.1002/bies.201100181
- Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591
- Jima, D. D., Zhang, J., Jacobs, C., Richards, K. L., Dunphy, C. H., Choi, W. W., et al. (2010). Deep sequencing of the small RNA transcriptome of normal and malignant human B cells identifies hundreds of novel microRNAs. *Blood* 116, e118–e127. doi: 10.1182/blood-2010-05-285403
- Joyce, C. E., Zhou, X., Xia, J., Ryan, C., Thrash, B., Menter, A., et al. (2011). Deep sequencing of small RNAs from human skin reveals major alterations in the psoriasis miRNAome. *Hum. Mol. Genet.* 20, 4025–4040. doi: 10.1093/hmg/ddr331
- Karolchik, D., Barber, G. P., Casper, J., Clawson, H., Cline, M. S., Diekhans, M., et al. (2014). The UCSC genome browser database: 2014 update. *Nucleic Acids Res.* 42, D764–D770. doi: 10.1093/nar/gkt1168
- Kawano, Y., Kawada, J., and Ito, Y. (2014). Epstein-Barr Virus MicroRNAs in plasma as potential biomarkers for chronic infections. *J. Infect. Dis.* 209, 1298–1300. doi: 10.1093/infdis/jit676
- Kozomara, A., and Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* 42, D68–D73. doi: 10.1093/nar/gkt1181
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Lee, H. Y., and Doudna, J. A. (2012). TRBP alters human precursor microRNA processing *in vitro*. *RNA* 18, 2012–2019. doi: 10.1261/rna.035501.112
- Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854.
- Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., et al. (2003). The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425, 415–419. doi: 10.1038/nature01957
- Li, J., Zhang, Y., Li, D., Liu, Y., Chu, D., Jiang, X., et al. (2015). Small non-coding RNAs transfer through mammalian placenta and directly regulate fetal gene expression. *Protein Cell.* 6, 391–396. doi: 10.1007/s13238-015-0156-2
- Li, L. M., Hu, Z. B., Zhou, Z. X., Chen, X., Liu, F. Y., Zhang, J. F., et al. (2010). Serum microRNA profiles serve as novel biomarkers for HBV infection and diagnosis of HBV-positive Hepatocarcinoma. *Cancer Res.* 70, 9798–9807. doi: 10.1158/0008-5472.CAN-10-1001
- Liang, G., Zhu, Y., Sun, B., Shao, Y., Jing, A., Wang, J., et al. (2014). Assessing the survival of exogenous plant microRNA in mice. *Food Sci. Nutr.* 2, 380–388. doi: 10.1002/fsn3.113
- Liang, H., Zhang, S., Fu, Z., Wang, Y., Wang, N., Liu, Y., et al. (2015). Effective detection and quantification of dietetically absorbed plant microRNAs in human plasma. *J. Nutr. Biochem.* 26, 505–512. doi: 10.1016/j.jnutbio.2014.12.002
- Lusk, R. W. (2014). Diverse and widespread contamination evident in the unmapped depths of high throughput sequencing data. *PLoS ONE* 9:e110808. doi: 10.1371/journal.pone.0110808

- Meyer, L. R., Zweig, A. S., Hinrichs, A. S., Karolchik, D., Kuhn, R. M., Wong, M., et al. (2013). The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.* 41, D64–D69. doi: 10.1093/nar/gks1048
- Miura, K., Ikeda, M., Matsubara, A., Song, X. J., Ito, M., Asano, K., et al. (2010). OsSPL14 promotes panicle branching and higher grain productivity in rice. *Nat. Genet.* 42, 545–549. doi: 10.1038/ng.592
- Moran, Y., Fredman, D., Praher, D., Li, X. Z., Wee, L. M., Rentzsch, F., et al. (2014). Cnidarian microRNAs frequently regulate targets by cleavage. *Genome Res.* 24, 651–663. doi: 10.1101/gr.162503.113
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Motameny, S., Wolters, S., Nurnberg, P., and Schumacher, B. (2010). Next generation sequencing of miRNAs - strategies, resources and methods. *Genes* 1, 70–84. doi: 10.3390/genes1010070
- Olarerin-George, A. O., and Hogenesch, J. B. (2015). Assessing the prevalence of mycoplasma contamination in cell culture via a survey of NCBI's RNA-seq archive. *Nucleic Acids Res.* 43, 2535–2542. doi: 10.1093/nar/gkv136
- Pandya, D., Mariani, M., McHugh, M., Andreoli, M., Sieber, S., He, S., et al. (2014). Herpes virus microRNA expression and significance in serous ovarian cancer. *PLoS ONE* 9:e114750. doi: 10.1371/journal.pone.0114750
- Pruitt, K. D., Tatusova, T., Brown, G. R., and Maglott, D. R. (2012). NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. *Nucleic Acids Res.* 40, D130–D135. doi: 10.1093/nar/gkr1079
- Saha, A., and Robertson, E. S. (2011). Epstein-Barr virus-associated B-cell lymphomas: pathogenesis and clinical outcomes. *Clin. Cancer Res.* 17, 3056–3063. doi: 10.1158/1078-0432.CCR-10-2578
- Schotte, D., Akbari Moqadam, F., Lange-Turenhout, E. A., Chen, C., van Ijcken, W. F., Pieters, R., et al. (2011). Discovery of new microRNAs by small RNAome deep sequencing in childhood acute lymphoblastic leukemia. *Leukemia* 25, 1389–1399. doi: 10.1038/leu.2011.105
- Snow, J. W., Hale, A. E., Isaacs, S. K., Baggish, A. L., and Chan, S. Y. (2013). Ineffective delivery of diet-derived microRNAs to recipient animal organisms. *RNA Biol.* 10, 1107–1116. doi: 10.4161/rna.24909
- Stark, M. S., Tyagi, S., Nancarrow, D. J., Boyle, G. M., Cook, A. L., Whiteman, D. C., et al. (2010). Characterization of the Melanoma miRNAome by Deep Sequencing. *PLoS ONE* 5:e9685. doi: 10.1371/journal.pone.0009685
- Tang, K. W., Alaei-Mahabadi, B., Samuelsson, T., Lindh, M., and Larsson, E. (2013). The landscape of viral expression and host gene fusion and adaptation in human cancer. *Nat. Commun.* 4, 2513. doi: 10.1038/ncomms3513
- Tosar, J. P., Rovira, C., Naya, H., and Cayota, A. (2014). Mining of public sequencing databases supports a non-dietary origin for putative foreign miRNAs: underestimated effects of contamination in NGS. *RNA* 20, 754–757. doi: 10.1261/rna.044263.114
- Tritten, L., Burkman, E., Moorhead, A., Satti, M., Geary, J., Mackenzie, C., et al. (2014). Detection of circulating parasite-derived microRNAs in filarial infections. *PLoS Negl. Trop. Dis.* 8:e2971. doi: 10.1371/journal.pntd.0002971
- Tweedie, S., Ashburner, M., Falls, K., Leyland, P., McQuilton, P., Marygold, S., et al. (2009). FlyBase: enhancing Drosophila gene Ontology annotations. *Nucleic Acids Res.* 37, D555–D559. doi: 10.1093/nar/gkn788
- van der Oost, J., Jore, M. M., Westra, E. R., Lundgren, M., and Brouns, S. J. J. (2009). CRISPR-based adaptive and heritable immunity in prokaryotes. *Trends Biochem. Sci.* 34, 401–407. doi: 10.1016/j.tibs.2009.05.002
- Vaucheret, H., and Chupeau, Y. (2012). Ingested plant miRNAs regulate gene expression in animals. *Cell Res.* 22, 3–5. doi: 10.1038/cr.2011.164
- Wang, K., Li, H., Yuan, Y., Etheridge, A., Zhou, Y., Huang, D., et al. (2012). The Complex exogenous RNA spectra in human plasma: an interface with human Gut Biota? *PLoS ONE* 7:e51009. doi: 10.1371/journal.pone.0051009
- Wen, Y. Z., Zheng, L. L., Liao, J. Y., Wang, M. H., Wei, Y., Guo, X. M., et al. (2011). Pseudogene-derived small interference RNAs regulate gene expression in African *Trypanosoma brucei*. *Proc. Natl. Acad. Sci. U.S.A.* 108, 8345–8350. doi: 10.1073/pnas.1103894108
- Witwer, K. W. (2012). XenomiRs and miRNA homeostasis in health and disease Evidence that diet and dietary miRNAs directly and indirectly influence circulating miRNA profiles. *RNA Biol.* 9, 1147–1154. doi: 10.4161/rna.21619
- Witwer, K. W. (2015). Contamination or artifacts may explain reports of plant miRNAs in humans. *J. Nutr. Biochem.* 26, 1685. doi: 10.1016/j.jnutbio.2015.09.004
- Witwer, K. W., and Hirschi, K. D. (2014). Transfer and functional consequences of dietary microRNAs in vertebrates: concepts in search of corroboration: negative results challenge the hypothesis that dietary xenomiRs cross the gut and regulate genes in ingesting vertebrates, but important questions persist. *Bioessays* 36, 394–406. doi: 10.1002/bies.201300150
- Witwer, K. W., McAlexander, M. A., Queen, S. E., and Adams, R. J. (2013). Real-time quantitative PCR and droplet digital PCR for plant miRNAs in mammalian blood provide little evidence for general uptake of dietary miRNAs limited evidence for general uptake of dietary plant xenomiRs. *RNA Biol.* 10, 1080–1086. doi: 10.4161/rna.25246
- Yang, J., Farmer, L. M., Agyekum, A. A., Elbaz-Younes, I., and Hirschi, K. D. (2015). Detection of an abundant plant-based small RNA in healthy consumers. *PLoS ONE* 10:e0137516. doi: 10.1371/journal.pone.0137516
- Zhang, L., Hou, D., Chen, X., Li, D., Zhu, L., Zhang, Y., et al. (2012). Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA. *Cell Res.* 22, 107–126. doi: 10.1038/cr.2011.158
- Zhang, Y. J., Wiggins, B. E., Lawrence, C., Petrick, J., Ivashuta, S., and Heck, G. (2012). Analysis of plant-derived miRNAs in animal small RNA datasets. *BMC Genomics* 13:381. doi: 10.1186/1471-2164-13-381
- Zheng, L. L., and Qu, L. H. (2015). Application of microRNA gene resources in the improvement of agronomic traits in rice. *Plant Biotechnol. J.* 13, 329–336. doi: 10.1111/pbi.12321
- Zheng, L. L., Wen, Y. Z., Yang, J. H., Liao, J. Y., Shao, P., Xu, H., et al. (2013). Comparative transcriptome analysis of small noncoding RNAs in different stages of *Trypanosoma brucei*. *RNA* 19, 863–875. doi: 10.1261/rna.035683.112
- Zhou, Z., Li, X., Liu, J., Dong, L., Chen, Q., Liu, J., et al. (2015). Honeysuckle-encoded atypical microRNA2911 directly targets influenza A viruses. *Cell Res.* 25, 39–49. doi: 10.1038/cr.2014.130

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Zheng, Deng, Wu, Yang, Lun and Qu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.