



## Article

# MyBrain-Seq: A Pipeline for MiRNA-Seq Data Analysis in Neuropsychiatric Disorders

Daniel Pérez-Rodríguez <sup>1,2,\*</sup> , Roberto Carlos Agís-Balboa <sup>1,2,3,4,\*</sup> and Hugo López-Fernández <sup>5,6,†</sup>

<sup>1</sup> Neuro Epigenetics Lab, Health Research Institute of Santiago de Compostela (IDIS), Santiago University Hospital Complex, 15706 Santiago de Compostela, Spain

<sup>2</sup> Translational Neuroscience Group, Galicia Sur Health Research Institute (IIS Galicia Sur), Área Sanitaria de Vigo-Hospital Álvaro Cunqueiro, SERGAS-UVIGO, CIBERSAM-ISCIII, 36213 Vigo, Spain

<sup>3</sup> Translational Research in Neurological Diseases Group, Health Research Institute of Santiago de Compostela (IDIS), Santiago University Hospital Complex, 15706 Santiago de Compostela, Spain

<sup>4</sup> Servicio de Neurología, Hospital Clínico Universitario de Santiago, 15706 Santiago de Compostela, Spain

<sup>5</sup> CINBIO, Department of Computer Science, ESEI-Escuela Superior de Ingeniería Informática, Universidade de Vigo, 32004 Ourense, Spain; hlfernandez@uvigo.es

<sup>6</sup> SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36213 Vigo, Spain

\* Correspondence: daniel.prz.rodriguez@gmail.com (D.P.-R.); roberto.carlos.agis.balboa@sergas.es (R.C.A.-B.)

† Co-senior authors: these authors contributed equally to this work.

**Abstract:** High-throughput sequencing of small RNA molecules such as microRNAs (miRNAs) has become a widely used approach for studying gene expression and regulation. However, analyzing miRNA-Seq data can be challenging because it requires multiple steps, from quality control and preprocessing to differential expression and pathway-enrichment analyses, with many tools and databases available for each step. Furthermore, reproducibility of the analysis pipeline is crucial to ensure that the results are accurate and reliable. Here, we present myBrain-Seq, a comprehensive and reproducible pipeline for analyzing miRNA-Seq data that incorporates miRNA-specific solutions at each step of the analysis. The pipeline was designed to be flexible and user-friendly, allowing researchers with different levels of expertise to perform the analysis in a standardized and reproducible manner, using the most common and widely used tools for each step. In this work, we describe the implementation of myBrain-Seq and demonstrate its capacity to consistently and reproducibly identify differentially expressed miRNAs and enriched pathways by applying it to a real case study in which we compared schizophrenia patients who responded to medication with treatment-resistant schizophrenia patients to obtain a 16-miRNA treatment-resistant schizophrenia profile.

**Keywords:** miRNA-Seq; miRNAs; differential expression; functional analysis; neuropsychiatry; reproducibility; myBrain-Seq; Compi; NGS; epigenetics



**Citation:** Pérez-Rodríguez, D.; Agís-Balboa, R.C.; López-Fernández, H. MyBrain-Seq: A Pipeline for MiRNA-Seq Data Analysis in Neuropsychiatric Disorders. *Biomedicines* **2023**, *11*, 1230. <https://doi.org/10.3390/biomedicines11041230>

Academic Editor: Thomas Mohr

Received: 16 March 2023

Revised: 14 April 2023

Accepted: 18 April 2023

Published: 21 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid development in the field of transcriptomics has allowed the existing knowledge about the molecular mechanisms of pathogenesis to be expanded. Over the last two decades, RNA-Seq technology has been used in translational medicine as a valuable tool for disease profiling and biomarker identification. This has led to important discoveries [1–3] and the foundation of precious resources such as ENCODE [4], the Cancer Genome Atlas [5] or the pathway database Reactome [6]. RNA-Seq technology has also allowed a new approach to study some of the most heterogeneous disorders, such as many neuropsychiatric conditions. The study of gene-expression regulators such as microRNAs (miRNAs) has offered a new angle from which neuropsychiatric conditions could be understood: the environmental influence on gene expression and the different responses to that environment. This application of the RNA-Seq technology is commonly known as miRNA-Seq.

MiRNAs are short non-coding RNA molecules involved in mRNA silencing. As epigenetic regulators, they are closely associated with adaptation processes and their expression levels are affected by events such as diet, sleep, stress or medications [7]. Nowadays, miRNAs are recognized as important etiological factors of neuropsychiatric diseases such as schizophrenia [8,9], depression [10,11] or Alzheimer's disease [12,13], and thus have potential to be biomarkers and therapeutic targets. As a result, the number of studies using miRNA-Seq has increased in the last decade [7], leading to the creation of miRNA databases such as the human microRNA disease database (HMDD) [14], the Central Nervous System microRNA Profiles (CNS microRNA Profiles) [15] and the Human miRNA Expression Database (miRmine) [16].

However, the extensive use of the miRNA-Seq technology, together with the absence of a standardized methodology for bioinformatics analysis, has raised concerns about its reproducibility [17]. The main concerns are the great variability in the analysis procedures [18,19], the large influence that biological references have on the results [19,20], the application of generic RNA methods to miRNA data [21] and the biological variability itself, that makes the comparison between studies complex [22,23].

Many bioinformatic tools were designed to perform specific tasks in a miRNA-Seq analysis. These tasks range from miRNA identification, such as in the case of Mirnova [24], UEA sRNA workbench [25] or miRDeep [26], to annotation tools as miRBase [27], Rfam [28] or miRIAD [29], to interaction databases such as Diana Tarbase [30], starBase [31] or PceRBase [32]. Importantly, few tools integrate these solutions into a single pipeline aiming for a complete miRNA-Seq analysis [33]. Some pipelines such as the ENCODE miRNA-Seq pipeline [34] or miRge3.0 [35] process the raw data until the transcript annotation and quantification step. Others, such as CAP-miRSeq [36], process the data and perform a differential expression analysis. Finally, very few pipelines perform a target annotation and a functional analysis, as in the case of miARma-Seq [37]. To the best of our knowledge, there are neither bioinformatic pipelines with a functional analysis corrected for the enrichment bias of the miRNA annotations [7,21] nor a pipeline that generates a network of miRNA–protein molecular interactions.

In this context, we present myBrain-Seq (<https://github.com/sing-group/my-brain-seq>), a highly modular pipeline for performing replicable miRNA-Seq analysis, from the preprocessing of the raw data to the generation of a network of miRNA–protein interactions. It covers most of the necessities of a miRNA-Seq study of neuropsychiatric data, which were identified by thoroughly reviewing current miRNA-Seq methodologies in the field [7] and translating that information into the pipeline design [17,38]. Finally, myBrain-Seq uses Docker technology, which ensures a stable environment regarding the dependencies and biological annotations used for data analysis, thus enhancing the replicability of the whole process [17,38].

In this study, we demonstrate the usefulness of myBrain-Seq by showing how it is applied to a real case study in which we compared schizophrenia patients who responded to medication with treatment-resistant schizophrenia patients to obtain a 16-miRNA treatment-resistant schizophrenia profile [8].

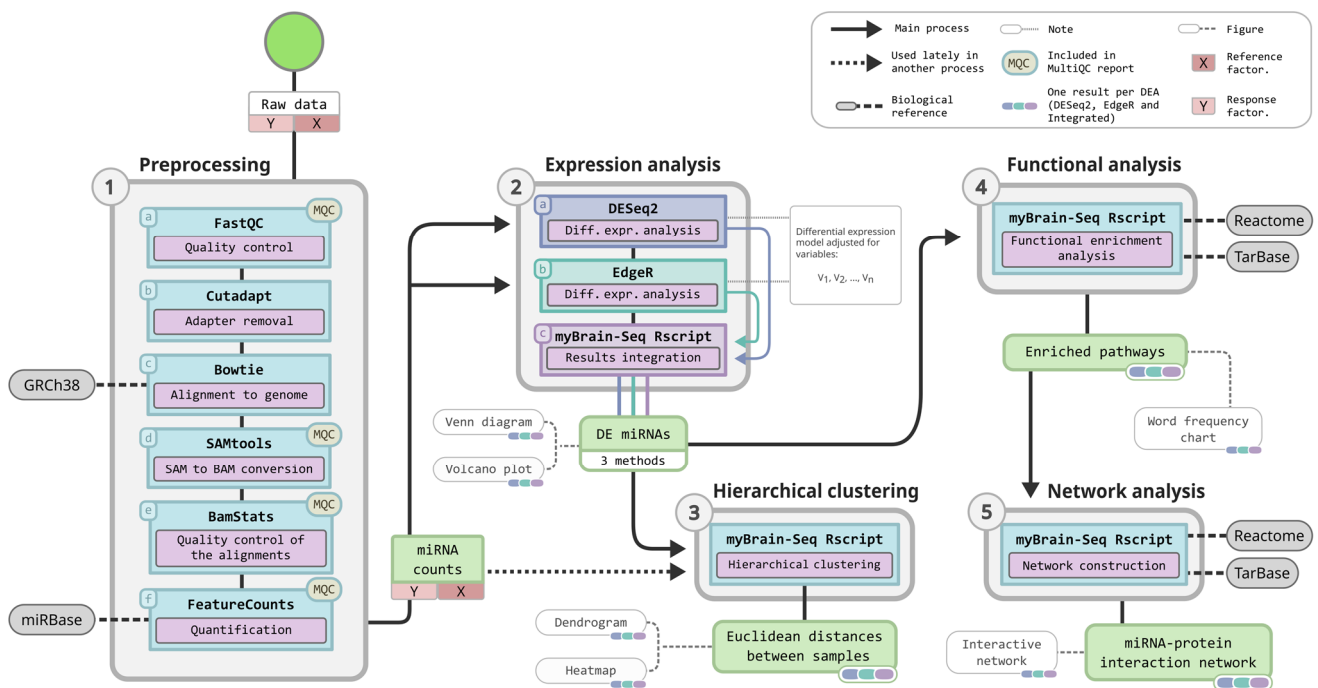
## 2. Materials and Methods

### 2.1. MyBrain-Seq, a Pipeline for miRNA-Seq Analysis

MyBrain-Seq is a Compi [39,40] pipeline to automatically analyze miRNA-Seq data in a highly reproducible way. It helps to find a profile of differentially expressed miRNAs between two conditions, assesses its potential classification power using hierarchical clustering analysis and aids in the discovery of biological pathways potentially affected by the conditions. It also helps to discover limitations in the quality of the data that may affect the conclusions of the study.

As depicted in Figure 1, the workflow entails the following main steps: (i) preprocessing, (ii) expression analysis, (iii) hierarchical clustering, (iv) functional analysis and (v) network analysis. Each of these steps comprises several tasks. The first step, preprocessing,

includes quality control of the raw data, removal of adapter sequences, alignment to a reference genome, quality control of the alignments and annotation and quantification of the aligned transcripts. The second step, expression analysis, includes two differential expression analyses (DEA) performed using two different software packages as well as the integration of both results using a custom myBrain-Seq script. Finally, the third, fourth and fifth steps are single-task steps specifically designed for myBrain-Seq. The following subsections provide more details about each one.



**Figure 1.** Summary of myBrain-Seq analysis with its five main steps: (1) preprocessing, (2) expression analysis, (3) hierarchical clustering, (4) functional analysis and (5) network analysis. Processes named “myBrain-Seq Rscript” are analysis scripts developed specifically for myBrain-Seq.

## 2.2. Quality Control and Adapter Removal

The first two preprocessing steps (Figure 1, step 1, a & b) are to discover samples with low-quality sequences as well as to detect the presence of foreign DNA contamination such as sequencing adapters. An evaluation of the results of this module is important for discarding samples from the downstream analysis and therefore it is the first step to be performed.

On one hand, myBrain-Seq includes the analysis of sequence quality scores, sequence length distribution, overrepresented sequences, adapter content and other parameters included in the FastQC tool [41]. On the other hand, the adapter trimming is performed using the Cutadapt tool [42]. Both quality control and adapter trimming analyze several samples at once, thus accelerating the analysis of large sample sizes.

## 2.3. Alignment to the Reference Genome

Alignment is the process of mapping reads against a biological reference (usually a genome or transcriptome) in order to assign genomic positions to each of the sequences of the raw data. The alignment in myBrain-Seq (Figure 1, step 1, c–e) encompasses building of a genome index (if needed), the alignment of the reads to a reference genome, a file format conversion of the results and a quality control of the alignments. First, myBrain-Seq performs the alignments with the Bowtie 1 tool [43], a short sequence aligner specializing in mapping short transcripts (such as miRNAs) to large genomes [7]. Bowtie uses a Burrows–Wheeler index of the genome to speed the mapping process and reduce the impact on

computer memory. This small memory footprint is used by myBrain-Seq to parallelize the alignments of several samples to accelerate the processing of large batches of files. Next, a file format conversion from Sequence Alignment Map (SAM) to Binary Alignment Map (BAM) is performed with SAMtools [44]. Finally, a quality control with SAMtools stats and SAMtools bcftools [45] brings information about the sequencing depth of the samples, an important parameter that allows an estimate of the reliability of the RNA-Seq data. Other parameters such as ACGT cycles, sequencing coverage and GC content are also reported.

#### 2.4. Transcripts Annotation and Quantification

Aligned sequences in BAM are then mapped to a reference transcriptome. This process, usually known as “annotation”, uses the genomic coordinates of an annotation file to convert the genomic locations of the aligned samples into transcript IDs. Those transcripts are then grouped by ID and quantified in a process known as read summarization or quantification. MyBrain-Seq performs both processes, annotation and quantification, using the software FeatureCounts [46] (Figure 1, step 1, f). It also requires a user-uploaded GTF/GFF file as the biological reference for annotations. Results of the quantification are presented as plain transcript counts.

#### 2.5. Differential Expression Analysis

A differential expression analysis (DEA) is the process of finding statistically significant differences in the transcript expression between two different conditions. Those differences can be potentially related to biological alterations and usually are the starting point of the functional analysis and the interpretation of results. MyBrain-Seq uses two well-known software packages for DEA, namely DESeq2 [47] and EdgeR [48] (Figure 1, step 2, a & b), and implements a volcano plot for the visualization of the differentially expressed miRNAs (DE miRNAs). Additionally, myBrain-Seq implements an option to integrate the results of both pieces of software (“integrated results” henceforth), thus offering a more conservative analysis of the data (Figure 1, step 2, c).

Regarding the software for differential expression analysis, DESeq2 normalizes the counts using the median of ratios method [49] prior to the DEA using a negative binomial distribution. On the other hand, EdgeR uses a weighted trimmed mean of the log expression ratios between the samples for the normalization (TMM method) [50] and an exact test for calculating the statistical differences in the miRNA expression. Both software packages apply the Benjamini–Hochberg FDR correction to the  $p$ -values [51]. MyBrain-Seq is also able to adjust the DEA model for covariates using the user input (more details in the Results section).

The integration of the DESeq2 and EdgeR results is performed by finding their common miRNAs; then, for each of those miRNAs, FDR and  $p$ -values are averaged. Finally, FDR and  $\log_2$  FC thresholds are set (default FDR < 0.05;  $|\log_2$  FC|  $\geq$  0.5) to obtain the list of DE miRNAs. The user can manually adjust both of these thresholds as well as the ones used for subsequent tasks (refer to myBrain-Seq documentation). Additionally, a Venn diagram is created to offer a visual representation of the coincidences: first DESeq2 and EdgeR results are filtered by FDR (default FDR < 0.05;  $|\log_2$  FC|  $\geq$  0.5), then coincidences and differences are counted and plotted using the R package “VennDiagram” [52]. Volcano plots of all the results are created using the R package “EnhancedVolcano” [53].

#### 2.6. Hierarchical Clustering

In the hierarchical clustering step, samples are assigned into clusters using the expression of the DE miRNAs. Samples with similar expression levels will group closely. This grouping can ease the identification of unknown relationships between the data as well as being an estimation of the classificatory power of the DE miRNAs. MyBrain-Seq automatically performs a hierarchical clustering analysis after the DEA using the R package “hclust” [54] (Figure 1, step 3). The whole process is divided into two steps: first, the data are prepared for the hierarchical clustering; second, the hierarchical clustering

and figures are generated. The first step, preparation of the data, comprises the optional normalization of all the samples with DESeq2 [47] (flag `-deseqNormalizationHclust`) and the filtering of the DESeq2 and/or EdgeR results by FDR and  $\log_2$  FC (default  $FDR \leq 0.05$ ;  $|\log_2 FC| \geq 0.5$ ) to obtain the DE miRNAs. After that, a table is built with one column per sample, each column being the DE miRNAs counts for a specific sample. In the second step, the counts are scaled using the R “scale” function, the matrix of Euclidean distances between samples is created and the clustering is performed using the “ward.D2” method of the R package “hclust” [54]. Finally, a dendrogram and a heatmap are generated using the R packages “dendextend” [55] and the function “heatmap.2” of the package gplots [56], respectively.

### 2.7. Functional Analysis

A functional analysis aims to put the differences spotted in the DEA into a biological context, suggesting biological pathways that might be useful for the investigator to analyze. The functional enrichment analysis of myBrain-Seq (Figure 1, step 4) uses two annotation sources which are included in the myBrain-Seq Docker image: the miRNA-gene annotations from the Diana TarBase [30] and the gene-pathway annotations of the Reactome databases [6]. We selected Diana TarBase as our target annotation database because it contains only miRNA–target interactions that have been experimentally validated. Additionally, both Diana TarBase and Reactome are publicly accessible and updated on a regular basis. Regarding the enrichment analysis, myBrain-Seq follows the strategy proposed by Godard and van Eyll [21] to be specific for miRNA data, briefly: (i) using TarBase annotations, protein coding genes in Reactome pathways are converted into lists of miRNAs that target at least one of these genes; (ii) enrichment analysis is performed by comparing DE miRNAs of DESeq2, EdgeR or integrated results to the lists of miRNAs previously associated with the different pathways. Enrichment scores are calculated using a Fisher hypergeometric test. Finally, a word analysis on the enriched terms is performed using the R package “tidytext” [57] and the results are summarized in a figure and presented along with the pathway-enriched table.

Conventional enrichment analysis is performed on genes rather than miRNAs. In a miRNA-Seq analysis, genes are indirectly selected by target prediction; therefore, genes with more targets have more chances to be selected. The consequence of this bias is a non-specific result, usually identifying as enriched the most studied pathways such as cancer-related or generic signaling pathways [21]. MyBrain-Seq enrichment analysis deals with this bias as each miRNA is only represented once in each pathway, thus ensuring the specificity of the results.

### 2.8. MiRNA–Protein Interaction Network

After the functional analysis, the most enriched pathway is used to build a network of miRNA–protein interactions, providing the researcher with a possible molecular context for the observed differences in miRNA expression. The miRNA–protein interaction network step of myBrain-Seq (Figure 1, step 5) uses the same annotation files as in the functional enrichment analysis step plus a protein–protein interaction file from the Reactome database [6]. The network is built by expanding the miRNA–protein interactions present in the most enriched pathway with the Reactome protein–protein interactions. The process is as follows: (i) miRNAs and genes that participate in the most enriched pathway are found using the functional analysis result; (ii) miRNA–protein interactions are found by using the Tarbase annotations file [30]; (iii) protein–protein interactions that participate in the most enrichment pathways are found using the protein–protein interaction file; and (iv) miRNA–protein interactions and protein–protein interactions are merged into a single table. Finally, a table with all the interactions is generated. This table can be easily imported into network analysis software such as Cytoscape [58] for further analysis and expansions. Additionally, an interactive network file in HTML is generated using the R packages “networkD3” [59] and “htmlwidgets” [60].

### 2.9. Summarization of the Quality Controls

Results of the quality control of the samples and the alignment are calculated on a per-sample basis, making overall interpretation of data quality difficult. To avoid this, the last step of myBrain-Seq analysis is the summarization of featureCounts [46], Samtools [45] and FastQC [41] results using MultiQC [61] (see Figure 1). The output of this step is a single HTML report from which different tables and graphs can be generated.

### 2.10. MyBrain-Seq Implementation

MyBrain-Seq is implemented as a Compi pipeline [39,40] and distributed as a Docker image that allows running it effortlessly. All external dependencies (Table 1) are satisfied using Docker images from the pegi3s Bioinformatics Docker images Project [62] (<https://pegi3s.github.io/dockerfiles/>). The source code of the pipeline is publicly available at GitHub (<https://github.com/sing-group/my-brain-seq>) under an MIT LICENSE, and the Docker image is available at Docker Hub (<https://hub.docker.com/r/singgroup/my-brain-seq>) and at Compi Hub (<https://www.sing-group.org/compihub/explore/625e719acc1507001943ab7f#overview>).

**Table 1.** MyBrain-Seq software dependencies and databases.

Dependencies	Version	Dependencies	Version
pegi3s/r_deseq2	1.32.0	pegi3s/samtools_bcftools	1.10
pegi3s/r_edger	3.36.0	pegi3s/r_data-analysis	4.1.1_v2
pegi3s/r_enhanced-volcano	1.12.0	pegi3s/r_venn-diagram	1.7.0
pegi3s/cutadapt	1.16	pegi3s/r_network	4.1.1_v2_v3
pegi3s/fastqc	0.11.9	pegi3s/multiqc	1.14.0
pegi3s/bowtie1	1.2.3	python3	3.8.5
pegi3s/feature-counts	2.0.0	DIANA Tarbase annotations	8
pegi3s/samtools_bcftools	1.9	Reactome annotations	83

### 2.11. Case Study Dataset: Treatment Resistant Schizophrenia

The application of myBrain-Seq to biomedical research is illustrated in a recent study [8] with the comparison of the miRNA profile of patients with schizophrenia (SZ) and treatment-resistant schizophrenia (TRS). The dataset comprises reads of circulating miRNA of 40 human patients with schizophrenia, of which 19 patients have a normal response to medication (MR; n = 19) and 21 have an insufficient response to medication (MNR; n = 21). The dataset can be downloaded from Gene Expression Omnibus (GEO) under the accession number GSE223043. We also provide a script in Supplementary Material Script S1 to automatically download and perform myBrain-Seq analysis on this dataset.

## 3. Results and Discussion

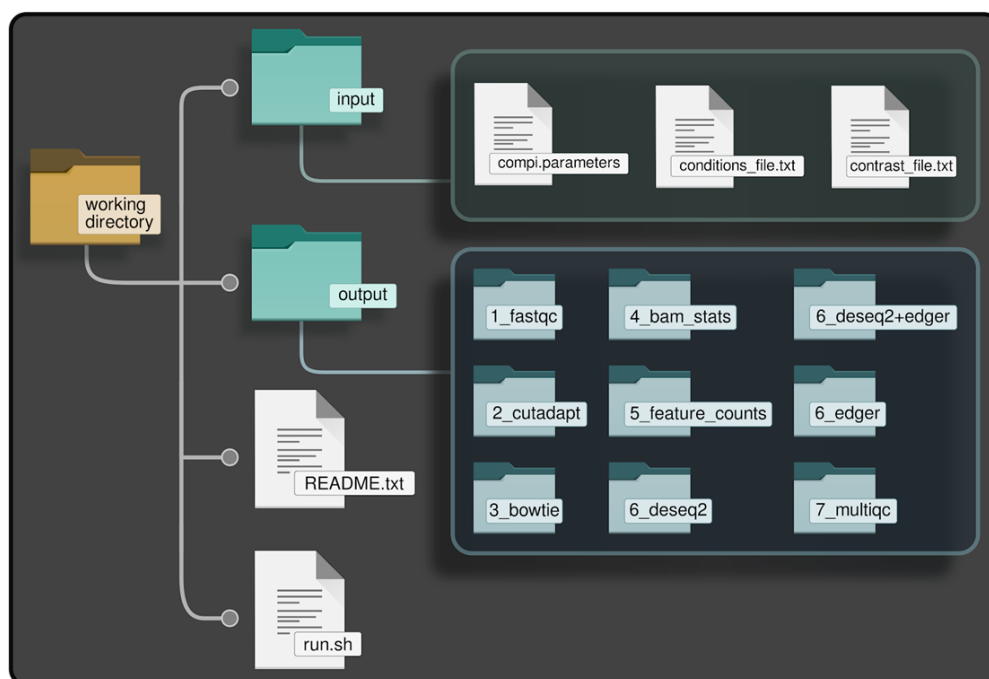
The goal of myBrain-Seq is to offer a modular and highly customizable tool for miRNA-Seq analysis that allows performing replicable studies. It offers a straightforward analysis process that brings together the most common tools in the field embedded in a portable and customizable pipeline. Among myBrain-Seq's main contributions are the options designed to solve typical problems in the analysis of transcriptomic data such as the high variability of the transcripts abundance (covariate adjustment), bias in the pathway-enrichment analysis resulting from the indirect selection of genes (miRNA-oriented pathway analysis strategy), low replicability of the results (containerized processes, DEA integration) or the need to explore the results in a biological context (miRNA-protein interaction network). The following subsections describe the contributions of myBrain-Seq with more details.

### 3.1. MyBrain-Seq Execution

The sequence of steps needed to start a myBrain-Seq can be described as follows:

1. Creation of the directory tree in the local file system, referred to as "working directory", shown in Figure 2. The working directory consists of a main directory with

two subdirectories: “/input” and “/output”. The input subdirectory is where the parameter files of myBrain-Seq should be placed; the output subdirectory will contain the results after myBrain-Seq execution. This working directory can be initialized using the utilities included in the myBrain-Seq Docker image. This initialization creates a “run.sh” file, used to run the pipeline and templates of the other files required by myBrain-Seq (those inside “/input”). A “README.txt” file is also created with the instructions to fill the template files and run the pipeline.



**Figure 2.** Working directory of myBrain-Seq.

2. The second step is the preparation of the data. In addition to the FASTQ files, myBrain-Seq needs a reference genome (or Bowtie index) and a GFF file with miRNA annotations as biological references to perform the analysis. It is recommended, but not mandatory, to put all these files inside subdirectories under “/input”. Nevertheless, if they are in other locations (e.g., a shared directory to save disk space), the provided “run.sh” will take care of this and create the appropriate Docker volume bindings in a transparent manner for the user.
3. The third step is the configuration of the analysis. This comprises the creation of three files, namely: “compi.parameters”, “conditions\_file.txt” and “contrast\_file.txt”. These files are usually placed into the “input” directory.
  - a. The “**compi.parameters**” file contains the paths and parameters needed for the analysis, i.e.: path of the working directory, paths to FASTQ files and biological references, paths to “conditions\_file.txt” and to “contrast\_file.txt” and the adapter sequence. For more information about the optional parameters that can be added, refer to the myBrain-Seq user manual (<https://github.com/sing-group/my-brain-seq>).
  - b. The “**conditions\_file.txt**” contains the metadata regarding names and conditions of each fastQ file. This file is used by myBrain-Seq to link each sample with a condition and its covariates. Each row of this file contains the name of the FASTQ file, its condition, a user label for that sample and zero or more columns describing the covariates for that sample (e.g., age, sex). All the covariates added in this file will be used in the DEA to adjust the statistical model.

- c. The “**contrast\_file.txt**” contains the conditions to compare during the analysis and a label for each contrast. Conditions included in this file must be the same as those stated in “contitions\_file.txt”. MyBrain-Seq can perform several contrasts in the same pipeline execution if several contrasts are specified in this file, one per line.
4. The final step is running myBrain-Seq analysis using the “run.sh” script created during the working directory initialization (step number 1). This script will use “compi.parameters” as reference, mount all the needed Docker volumes (by extracting the path from the Compi parameters file) and create a directory for the log files of the current execution. MyBrain-Seq users do not need to modify this file, as it is ready to use. Thus, users only need to run the script using the path to “compi.parameters” as the unique argument to start the myBrain-Seq analysis.
5. Both final and intermediate results are saved in the “/output” directory. Such output files are placed in directories corresponding to the different steps of the workflow, namely: “1\_fastqc”, “2\_cutadapt”, “3\_bowtie”, “4\_bam\_stats”, “5\_feature\_counts”, “6\_deseq2”, “6\_deseq2+edger”, “6\_edger” and “7\_multiqc”. Results from the hierarchical clustering, functional analysis and network analysis are placed in the directories prefixed with “6\_”, according to the data from which they were generated. Files from the same contrast are grouped in subdirectories named with the contrast label.

### 3.2. MyBrain-Seq Results

The results of myBrain-Seq are placed in sub-directories inside the “output” directory (see Figure 2). The main results stem from the DEA and are presented for each contrast specified in “contrast\_file.txt” and for each DEA method. Figure 3 illustrates the graphical results of a myBrain-Seq analysis:

- Volcano plot with the results of each DEA; Figure 3A.
- Venn diagram with the DE miRNA coincidences between DESeq2 and EdgeR; Figure 3B.
- Dendrogram with the result of the hierarchical clustering; Figure 3C.
- Heatmap with the result of the hierarchical clustering; Figure 3D.
- HTML file with a miRNA–protein interaction network of the most enriched pathway; Figure 3E.
- Lollipop chart with the word frequency of the enriched terms; Figure 3F.

Figure 4 illustrates the tabular results in TSV:

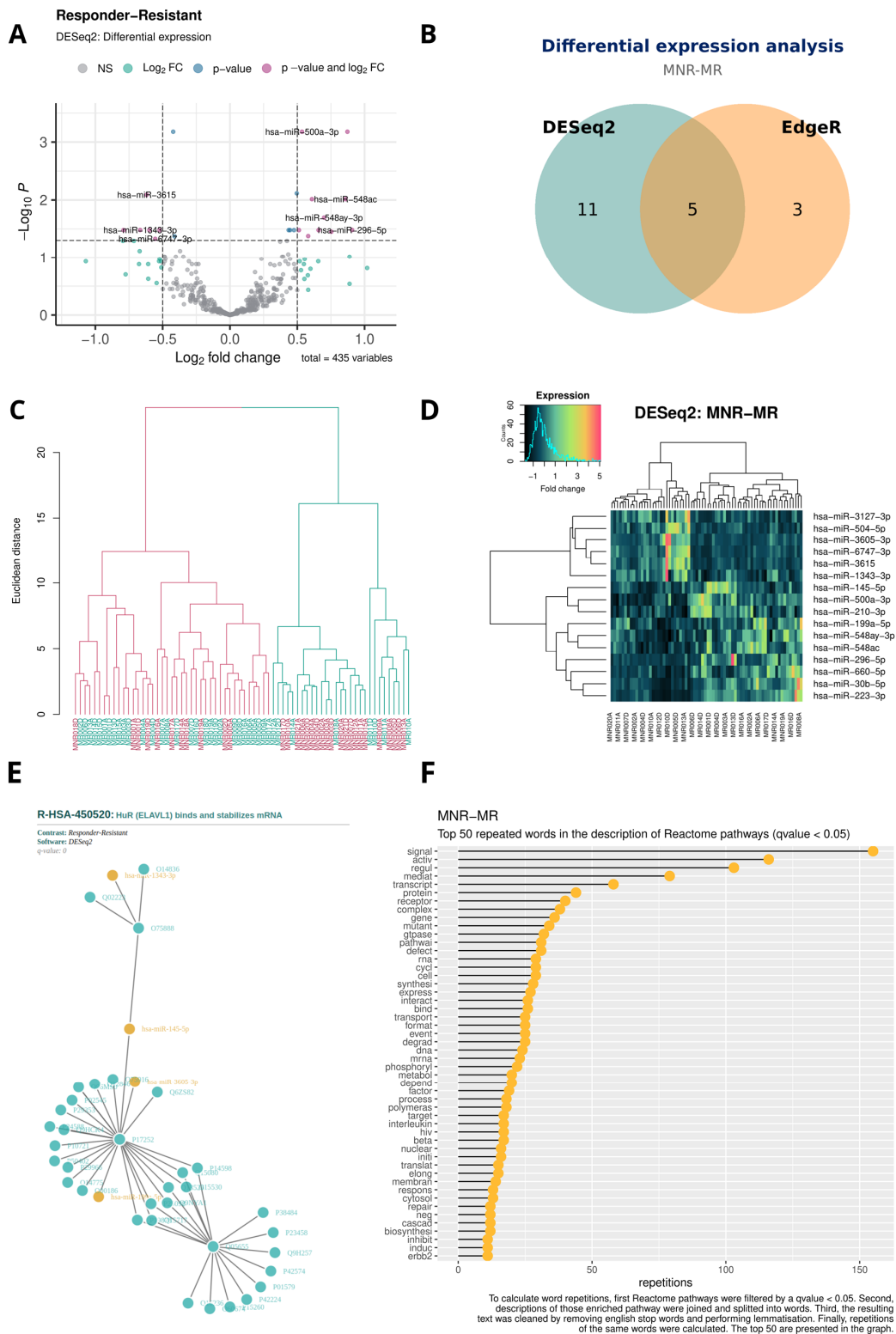
- Results of the DEA; Figure 4A. Full table in supplementary Table S1.
- List of DE miRNAs; Figure 4B.
- Enriched pathways; Figure 4C. Full table in supplementary Table S2.
- miRNA–protein interaction network; Figure 4D. Full table in supplementary Table S3.

Additionally, myBrain-Seq offers the intermediate files of the analysis to reuse or inspect, namely:

- Adapter-trimmed FASTQ files.
- BAM and SAM files resulting from the alignment.
- A TXT file with the counts of miRNA per sample.
- A summary of the quantification results.
- A file per contrast with a subset of counts for that contrast.
- A TSV file with the expression per sample of each DE miRNA, used for the hierarchical clustering.

Finally, myBrain-Seq generates an HTML file with a summary of the results of the quality control, alignments, assignments and with the quantification of all the samples.





**Figure 3.** Figures generated after myBrain-Seq analysis: (A) Volcano plot of the differential expression analysis; (B) Venn diagram comparing the differentially expressed miRNAs resulting from DESeq2 and EdgeR methods; (C) dendrogram result of a hierarchical clustering of the samples using the differentially expressed miRNAs; (D) heatmap with the levels of expression of the differentially expressed miRNAs per sample; (E) miRNA–protein interaction network (portion); (F) lollipop chart with the word frequency of the top 50 enriched pathways.

A	Feature	baseMean	log2FC	lfcSE	stat	pvalue	qvalue
	hsa-miR-520g-3p	0.135036368499093	15.2694825004486	3.1127941859134	4.90539418556772	9.32398959360394E-07	NA
	hsa-miR-500a-3p	177.882883142614	0.535144108199585	0.113970966460276	4.69544239923689	2.66030327552587E-06	0.000662669226033714
	hsa-miR-5698	0.802536936057229	8.03984812709738	1.75003521356067	4.59410648699992	4.34607417257374E-06	NA
	hsa-miR-145-5p	41.8142047756387	0.872361183656924	0.190206693800889	4.58638529604076	4.50985989603239E-06	0.000662669226033714
	hsa-miR-744-5p	490.643127363411	-0.42168442195939	0.0919982927433184	-4.58361138435383	4.57013259333596E-06	0.000662669226033714
	hsa-miR-502-3p	281.086299780003	0.496165091221384	0.124836266694889	3.97452682908288	7.05193130130609E-05	0.00766897529017037
	hsa-miR-3615	65338.2568962364	-0.619056414906553	0.158430785870534	-3.9074250090032	9.32849808175308E-05	0.00811579333112518

B	hsa-miR-3605-3p
	hsa-miR-548ac
	hsa-miR-199a-5p
	hsa-miR-210-3p
	hsa-miR-1343-3p
	hsa-miR-6747-3p
	hsa-miR-3615
	hsa-miR-3127-3p

C	reactome_id	description	url	pvalue	qvalue	DE_mirnas
	R-HSA-1989781	PPARA activates gene expression	<a href="https://reactome.org">https://reactome.org</a>	5.673E-08	1.122E-05	hsa-miR-210-3p hsa-miR-502-3p
	R-HSA-8856828	Clathrin-mediated endocytosis	<a href="https://reactome.org">https://reactome.org</a>	9.198E-08	1.122E-05	hsa-miR-145-5p hsa-miR-502-3p
	R-HSA-2500257	Resolution of Sister Chromatid Cohesion	<a href="https://reactome.org">https://reactome.org</a>	8.501E-08	1.122E-05	hsa-miR-1343-3p hsa-miR-502-3p
	R-HSA-9679191	Potential therapeutics for SARS	<a href="https://reactome.org">https://reactome.org</a>	5.993E-08	1.122E-05	hsa-miR-210-3p hsa-miR-502-3p
	R-HSA-68867	Assembly of the pre-replicative complex	<a href="https://reactome.org">https://reactome.org</a>	7.157E-08	1.122E-05	hsa-miR-30b-5p hsa-miR-502-3p
	R-HSA-5578749	Transcriptional regulation by small RNAs	<a href="https://reactome.org">https://reactome.org</a>	5.22E-08	1.122E-05	hsa-miR-210-3p hsa-miR-502-3p
	R-HSA-5617472	Activation of anterior HOX genes in hindbrain	<a href="https://reactome.org">https://reactome.org</a>	4.8E-08	1.122E-05	hsa-miR-744-5p hsa-miR-502-3p

D	interactor_1	interactor_2	Interaction.type
	hsa-miR-502-3p	O14980	DIRECT
	hsa-miR-532-3p	O14980	DIRECT
	hsa-miR-660-5p	O14980	DIRECT
	O14867	O14980	physical association
	O14980	Q9HCE7	physical association
	O14980	O15105	physical association
	O14980	O14867	physical association

**Figure 4.** Subset of the main results produced by myBrain-Seq analysis: (A) Results of a differential expression analysis for a specific contrast and software (DESeq2, EdgeR or integrated); (B) list of differentially expressed miRNAs for a specific contrast and software; (C) list of enriched pathways and DE miRNAs implicated in each pathway; (D) miRNA–protein interaction table, compatible with network software such as Cytoscape [58].

### 3.3. Case Study

An early version of myBrain-Seq (v0.1.0) was used in Pérez-Rodríguez et al. 2023 to perform all analysis steps up to quantification [8]. In this study, myBrain-Seq was used to identify 16 differentially expressed miRNAs (DE miRNAs) between the MR and MNR conditions using DESeq2. The analysis performed can be described using Figure 1, where the reference factor “X” is the schizophrenia condition (MR), the response factor “Y” is the treatment-resistant schizophrenia condition (MNR) and six variables ( $n = 6$ ) were used to adjust the differential expression model ( $V_1, V_2, \dots, V_6$ ). These six variables are: processing bath, sex, drug consumption (alcohol OR tobacco OR illegal), time (hospital arrival/discharge), treatment based on diazepam, oxazepam, thiazepam and oxepin and treatment based on other antipsychotics.

However, there are remarkable differences in the functional analysis and in the miRNA–protein network due to the application of different methodologies applied in both analyses. In Pérez-Rodríguez et al. 2023, we performed a bibliographic search followed by a target prediction to enrich Tarbase [30] annotations. After that, the network was built on Cytoscape [58], expanded and filtered using custom Cytoscape [58] filters and StringApp [63]. The pathway enrichment was later performed using the molecules present in the resulting network. On the other hand, as explained before, myBrain-Seq v1.0.0 takes a simpler but effective approach to the functional analysis: to avoid the artificial miRNA target overrepresentation and construct a reduced miRNA–protein network. With this approach, the network is produced after the functional enrichment, thus being smaller and with no

overrepresentation biases. This network can be further expanded and filtered externally as we did in Pérez-Rodríguez et al. 2023.

Regarding the enriched pathways, Table 2 offers a comparison between the top ten Reactome-enriched pathways [6] in Pérez-Rodríguez et al. 2023 and myBrain-Seq v1.0.0. In the original study, several pathway databases were used in order to perform the enrichment analysis. On the other hand, myBrain-Seq v1.0.0 only uses Reactome annotations. There are no coincidences between these top ten enriched pathways, probably because StringApp enrichment uses annotations of all levels of the Reactome pathway hierarchy whereas myBrain-Seq uses only the lowest level annotations. This has the advantage of providing more useful results by discarding big unspecific pathways such as “Metabolism of proteins”, “Developmental Biology” or “Disease”, which provide little help in getting to the molecular causes of a condition. Regarding this matter, 217 of the enriched pathways that were discovered using myBrain-Seq v1.0.0 were also detected in the enriched results of Pérez-Rodríguez et al. 2023 (refer to Supplementary Table S4). However, their significance in the context of the entire table differs significantly. For example, the second most enriched pathway in myBrain-Seq, “Activation of anterior HOX genes in hindbrain development during early embryogenesis” (see Table 2), which has a q-value of  $1.22 \times 10^{-5}$ , is ranked at position 86 and has a q-value of  $2.46 \times 10^{-7}$  in the results of Pérez-Rodríguez et al. 2023, which is two orders of magnitude lower (see Supplementary Table S4).

**Table 2.** Top 10 enriched pathways in Pérez-Rodríguez et al. 2023 [8] and in the myBrain-Seq reanalysis. Only Reactome pathways [6] were included in this table.

Pérez-Rodríguez et al. 2023 [8]			MyBrain-Seq		
Pathway	p-Value	q-Value	Pathway	p-Value	q-Value
Metabolism of proteins	$2.32 \times 10^{-55}$	$5.03 \times 10^{-52}$	HuR (ELAVL1) binds and stabilizes mRNA	$4.22 \times 10^{-8}$	$1.22 \times 10^{-5}$
Gene expression (Transcription)	$1.47 \times 10^{-54}$	$1.6 \times 10^{-51}$	Activation of anterior HOX genes in hindbrain development during early embryogenesis	$4.80 \times 10^{-8}$	$1.22 \times 10^{-5}$
Cellular responses to stress	$3.18 \times 10^{-47}$	$1.73 \times 10^{-44}$	Transcriptional regulation by small RNAs	$5.22 \times 10^{-8}$	$1.22 \times 10^{-5}$
Disease	$1.08 \times 10^{-44}$	$3.89 \times 10^{-42}$	Cyclin E associated events during G1/S transition	$5.31 \times 10^{-8}$	$1.22 \times 10^{-5}$
Metabolism of RNA	$7.31 \times 10^{-43}$	$1.99 \times 10^{-40}$	MAPK6/MAPK4 signaling	$5.52 \times 10^{-8}$	$1.22 \times 10^{-5}$
Cell Cycle	$1.33 \times 10^{-42}$	$3.22 \times 10^{-40}$	PPARA activates gene expression	$5.67 \times 10^{-8}$	$1.22 \times 10^{-5}$
Developmental Biology	$8.61 \times 10^{-34}$	$1.7 \times 10^{-31}$	Cyclin A:Cdk2 associated events at S phase entry	$5.68 \times 10^{-8}$	$1.22 \times 10^{-5}$
Transcriptional Regulation by TP53	$5.17 \times 10^{-33}$	$9.37 \times 10^{-31}$	Potential therapeutics for SARS	$5.99 \times 10^{-8}$	$1.22 \times 10^{-5}$
DNA Repair	$2.5 \times 10^{-32}$	$4.18 \times 10^{-30}$	Assembly of the pre-replicative complex	$7.16 \times 10^{-8}$	$1.22 \times 10^{-5}$
Innate Immune System	$1.94 \times 10^{-31}$	$3.01 \times 10^{-29}$	SUMOylation of ubiquitylation proteins	$8.15 \times 10^{-8}$	$1.22 \times 10^{-5}$

In relation to this, it is also worth noting the differences in the scale of the p and q values. These differences are likely due to both the lack of specificity of the pathways and the overrepresentation of miRNA targets, resulting in imbalances in the enrichment values (see functional analysis section). Both phenomena produce a high number of false positives, which in turn forces the use of additional filtering strategies for the identification of relevant

pathways. Thanks to myBrain-Seq correction, interpretations can be made straight from the enriched table.

#### 4. Conclusions

MyBrain-Seq is a highly modular bioinformatics pipeline that specializes in replicable miRNA-Seq data analyses. Created using Compi, it provides a complete set of analyses ranging from raw data preprocessing and DEA to hierarchical clustering, functional analysis and network creation. MyBrain-Seq adaptations to miRNA data include the use of a short-sequence aligner, correction of the DEA model for confounding factors, correction of the artificial target overrepresentation in the functional analysis and creation of a miRNA–protein interaction network.

MyBrain-Seq has already been used in a real case study in which we compared schizophrenia patients who responded to medication to treatment-resistant schizophrenia patients to obtain a 16-miRNA treatment-resistant schizophrenia profile [8]. We were able to reproduce all the findings from the case study up to the functional analysis stage, including the miRNA profile suggested in Pérez-Rodríguez et al. 2023, as well as the hierarchical clustering. By making adjustments to myBrain-Seq functional analysis, we were able to generate more succinct and insightful enriched pathways, while also preventing any biases in the results that could have been caused by an overrepresentation of miRNA targets.

Overall, myBrain-Seq offers a powerful and reliable tool for miRNA-Seq data analysis, which can help researchers identify meaningful biological insights with greater confidence and ease. By reducing the need for additional analysis and providing a complete pipeline for data processing, DEA, hierarchical clustering, functional analysis and network creation, myBrain-Seq can streamline the research process and promote greater reproducibility of the results.

The tool is open for further extension and new features will be included as it is used in new studies. MyBrain-Seq is freely distributed under an MIT license and a complete manual is available at <https://github.com/sing-group/my-brain-seq>.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biomedicines11041230/s1>, Table S1: Results of the DEA; Table S2: Enriched pathways; Table S3: miRNA–protein interaction network; Table S4: Comparison of enriched pathways; Script S1: Replication of the case study.

**Author Contributions:** Conceptualization, D.P.-R., R.C.A.-B. and H.L.-F.; methodology, D.P.-R., R.C.A.-B. and H.L.-F.; software, D.P.-R., R.C.A.-B. and H.L.-F.; validation, D.P.-R. and H.L.-F.; formal analysis, D.P.-R.; investigation, D.P.-R., R.C.A.-B. and H.L.-F.; resources, R.C.A.-B. and H.L.-F.; data curation, D.P.-R.; writing—original draft preparation, D.P.-R., R.C.A.-B. and H.L.-F.; writing—review and editing, D.P.-R., R.C.A.-B. and H.L.-F.; visualization, D.P.-R.; supervision, R.C.A.-B. and H.L.-F.; project administration, R.C.A.-B. and H.L.-F.; funding acquisition, R.C.A.-B. and H.L.-F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was partially supported by: (i) Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding ED431C2018/55-GRC and ED431C 2022/03-GRC Competitive Reference Group; (ii) Instituto de Salud Carlos III (ISCIII) through the project PI18/01311 (co-funded by European Regional Development Fund (FEDER), “A way to make Europe”, UE) to R.C. Agís-Balboa and (iii) Investigo Program, TR349V, predoctoral contract from Xunta de Galicia to D. Pérez-Rodríguez.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** MyBrain-Seq is freely distributed under an MIT license at <https://github.com/sing-group/my-brain-seq>. The dataset used for the case study is publicly available under Gene Expression Omnibus (GEO) accession number GSE223043.

**Acknowledgments:** SING group thanks the CITI (Centro de Investigación, Transferencia e Innovación) from the University of Vigo for hosting its IT infrastructure. The authors would like to thank Galicia Sur Health Research Institute and the Área Sanitaria de Vigo for their support, and the Health Research Institute of Santiago de Compostela (IDIS) and Área Sanitaria de Santiago de Compostela & Barbanza. We would also like to thank Patricia Pérez Crespo (Clinical Laboratory Technician) and Mateo Pérez-Rodríguez (Facultade de Matemáticas, Universidade de Santiago de Compostela) for their valuable help and support during the development of myBrain-Seq.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Saliba, A.-E.; Santos, S.C.; Vogel, J. New RNA-Seq Approaches for the Study of Bacterial Pathogens. *Curr. Opin. Microbiol.* **2017**, *35*, 78–87. [CrossRef] [PubMed]
2. Sudhagar, A.; Kumar, G.; El-Matbouli, M. Transcriptome Analysis Based on RNA-Seq in Understanding Pathogenic Mechanisms of Diseases and the Immune System of Fish: A Comprehensive Review. *Int. J. Mol. Sci.* **2018**, *19*, 245. [CrossRef] [PubMed]
3. Kaartokallio, T.; Cervera, A.; Kyllönen, A.; Laivuori, K.; Kere, J.; Laivuori, H. Gene Expression Profiling of Pre-Eclamptic Placentae by RNA Sequencing. *Sci. Rep.* **2015**, *5*, 14107. [CrossRef] [PubMed]
4. ENCODE Project Consortium. An Integrated Encyclopedia of DNA Elements in the Human Genome. *Nature* **2012**, *489*, 57–74. [CrossRef]
5. The Cancer Genome Atlas Program—NCI. Available online: <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga> (accessed on 13 March 2023).
6. Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; et al. The Reactome Pathway Knowledgebase 2022. *Nucleic Acids Res.* **2022**, *50*, D687–D692. [CrossRef]
7. Pérez-Rodríguez, D.; López-Fernández, H.; Agís-Balboa, R.C. Application of MiRNA-Seq in Neuropsychiatry: A Methodological Perspective. *Comput. Biol. Med.* **2021**, *135*, 31–42. [CrossRef]
8. Pérez-Rodríguez, D.; Penedo, M.A.; Rivera-Baltanás, T.; Peña-Centeno, T.; Burkhardt, S.; Fischer, A.; Prieto-González, J.M.; Olivares, J.M.; López-Fernández, H.; Agís-Balboa, R.C. MiRNA Differences Related to Treatment-Resistant Schizophrenia. *IJMS* **2023**, *24*, 1891. [CrossRef]
9. Chang, X.; Liu, Y.; Hahn, C.-G.; Gur, R.E.; Sleiman, P.M.A.; Hakonarson, H. RNA-Seq Analysis of Amygdala Tissue Reveals Characteristic Expression Profiles in Schizophrenia. *Transl. Psychiatry* **2017**, *7*, e1203. [CrossRef]
10. Pantazatos, S.P.; Huang, Y.; Rosoklija, G.B.; Dwork, A.J.; Arango, V.; Mann, J.J. Whole-Transcriptome Brain Expression and Exon-Usage Profiling in Major Depression and Suicide: Evidence for Altered Glial, Endothelial and ATPase Activity. *Mol. Psychiatry* **2017**, *22*, 760–773. [CrossRef]
11. Labonté, B.; Engmann, O.; Purushothaman, I.; Menard, C.; Wang, J.; Tan, C.; Scarpa, J.R.; Moy, G.; Loh, Y.-H.E.; Cahill, M.; et al. Sex-Specific Transcriptional Signatures in Human Depression. *Nat. Med.* **2017**, *23*, 1102–1111. [CrossRef]
12. Zovoilis, A.; Agbemenyah, H.Y.; Agis-Balboa, R.C.; Stilling, R.M.; Edbauer, D.; Rao, P.; Farinelli, L.; Delalle, I.; Schmitt, A.; Falkai, P.; et al. MicroRNA-34c Is a Novel Target to Treat Dementias. *EMBO J.* **2011**, *30*, 4299–4308. [CrossRef] [PubMed]
13. Neff, R.A.; Wang, M.; Vatansever, S.; Guo, L.; Ming, C.; Wang, Q.; Wang, E.; Horgusluoglu-Moloch, E.; Song, W.; Li, A.; et al. Molecular Subtyping of Alzheimer’s Disease Using RNA Sequencing Data Reveals Novel Mechanisms and Targets. *Sci. Adv.* **2021**, *7*, eabb5398. [CrossRef] [PubMed]
14. Huang, Z.; Shi, J.; Gao, Y.; Cui, C.; Zhang, S.; Li, J.; Zhou, Y.; Cui, Q. HMDD v3.0: A Database for Experimentally Supported Human MicroRNA–Disease Associations. *Nucleic Acids Res.* **2019**, *47*, D1013–D1017. [CrossRef] [PubMed]
15. Pomper, N.; Liu, Y.; Hoye, M.L.; Dougherty, J.D.; Miller, T.M. CNS MicroRNA Profiles: A Database for Cell Type Enriched MicroRNA Expression across the Mouse Central Nervous System. *Sci. Rep.* **2020**, *10*, 4921. [CrossRef] [PubMed]
16. Panwar, B.; Omenn, G.S.; Guan, Y. MiRmine: A Database of Human MiRNA Expression Profiles. *Bioinformatics* **2017**, *33*, 1554–1560. [CrossRef]
17. Pérez-Rodríguez, D.; López-Fernández, H.; Agís-Balboa, R.C. On the Reproducibility of MiRNA-Seq Differential Expression Analyses in Neuropsychiatric Diseases. In *Practical Applications of Computational Biology & Bioinformatics, 15th International Conference (PACBB 2021)*; Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Casado-Vara, R., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 41–51.
18. Peixoto, L.; Risso, D.; Poplawski, S.G.; Wimmer, M.E.; Speed, T.P.; Wood, M.A.; Abel, T. How Data Analysis Affects Power, Reproducibility and Biological Insight of RNA-Seq Studies in Complex Datasets. *Nucleic Acids Res.* **2015**, *43*, 7664–7674. [CrossRef]
19. Simoneau, J.; Dumontier, S.; Gosselin, R.; Scott, M.S. Current RNA-Seq Methodology Reporting Limits Reproducibility. *Brief. Bioinform.* **2021**, *22*, 140–145. [CrossRef]
20. Zhao, S.; Zhang, B. A Comprehensive Evaluation of Ensembl, RefSeq, and UCSC Annotations in the Context of RNA-Seq Read Mapping and Gene Quantification. *BMC Genom.* **2015**, *16*, 97. [CrossRef]
21. Godard, P.; van Eyll, J. Pathway Analysis from Lists of MicroRNAs: Common Pitfalls and Alternative Strategy. *Nucleic Acids Res.* **2015**, *43*, 3490–3497. [CrossRef]

22. Hansen, K.D.; Wu, Z.; Irizarry, R.A.; Leek, J.T. Sequencing Technology Does Not Eliminate Biological Variability. *Nat. Biotechnol.* **2011**, *29*, 572–573. [[CrossRef](#)]
23. McIntyre, L.M.; Lopiano, K.K.; Morse, A.M.; Amin, V.; Oberg, A.L.; Young, L.J.; Nuzhdin, S.V. RNA-Seq: Technical Variability and Sampling. *BMC Genom.* **2011**, *12*, 293. [[CrossRef](#)] [[PubMed](#)]
24. Vitsios, D.M.; Kentepozidou, E.; Quintais, L.; Benito-Gutiérrez, E.; van Dongen, S.; Davis, M.P.; Enright, A.J. MirNovo: Genome-Free Prediction of MicroRNAs from Small RNA Sequencing Data and Single-Cells Using Decision Forests. *Nucleic Acids Res.* **2017**, *45*, e177. [[CrossRef](#)] [[PubMed](#)]
25. Stocks, M.B.; Moxon, S.; Mapleson, D.; Woolfenden, H.C.; Mohorianu, I.; Folkes, L.; Schwach, F.; Dalmay, T.; Moulton, V. The UEA SRNA Workbench: A Suite of Tools for Analysing and Visualizing next Generation Sequencing MicroRNA and Small RNA Datasets. *Bioinformatics* **2012**, *28*, 2059–2061. [[CrossRef](#)] [[PubMed](#)]
26. An, J.; Lai, J.; Lehman, M.L.; Nelson, C.C. MiRDeep\*: An Integrated Application Tool for MiRNA Identification from RNA Sequencing Data. *Nucleic Acids Res.* **2013**, *41*, 727–737. [[CrossRef](#)]
27. Kozomara, A.; Birgaoanu, M.; Griffiths-Jones, S. MiRBase: From MicroRNA Sequences to Function. *Nucleic Acids Res.* **2019**, *47*, D155–D162. [[CrossRef](#)]
28. Kalvari, I.; Nawrocki, E.P.; Argasinska, J.; Quinones-Olvera, N.; Finn, R.D.; Bateman, A.; Petrov, A.I. Non-Coding RNA Analysis Using the Rfam Database. *Curr. Protoc. Bioinform.* **2018**, *62*, e51. [[CrossRef](#)]
29. Hinske, L.C.; França, G.S.; Torres, H.A.M.; Ohara, D.T.; Lopes-Ramos, C.M.; Heyn, J.; Reis, L.F.L.; Ohno-Machado, L.; Kreth, S.; Galante, P.A.F. MiRIAD-Integrating MicroRNA Inter- and Intragenic Data. *Database* **2014**, *2014*, bau099. [[CrossRef](#)]
30. Karagkouni, D.; Paraskevopoulou, M.D.; Chatzopoulos, S.; Vlachos, I.S.; Tastsoglou, S.; Kanellos, I.; Papadimitriou, D.; Kavakiotis, I.; Maniou, S.; Skoufos, G.; et al. DIANA-TarBase v8: A Decade-Long Collection of Experimentally Supported MiRNA–Gene Interactions. *Nucleic Acids Res.* **2018**, *46*, D239–D245. [[CrossRef](#)]
31. Li, J.-H.; Liu, S.; Zhou, H.; Qu, L.-H.; Yang, J.-H. StarBase v2.0: Decoding MiRNA–CeRNA, MiRNA–NcRNA and Protein–RNA Interaction Networks from Large-Scale CLIP-Seq Data. *Nucleic Acids Res.* **2014**, *42*, D92–D97. [[CrossRef](#)]
32. Yuan, C.; Meng, X.; Li, X.; Illing, N.; Ingle, R.A.; Wang, J.; Chen, M. PceRBase: A Database of Plant Competing Endogenous RNA. *Nucleic Acids Res.* **2017**, *45*, D1009–D1014. [[CrossRef](#)]
33. Chen, L.; Heikkinen, L.; Wang, C.; Yang, Y.; Sun, H.; Wong, G. Trends in the Development of MiRNA Bioinformatics Tools. *Brief. Bioinform.* **2019**, *20*, 1836–1852. [[CrossRef](#)] [[PubMed](#)]
34. MicroRNA-Seq Data Standards and Processing Pipeline—ENCODE. Available online: <https://www.encodeproject.org/microna/microna-seq/#references> (accessed on 13 March 2023).
35. Patil, A.H.; Halushka, M.K. MiRge3.0: A Comprehensive MicroRNA and TRF Sequencing Analysis Pipeline. *NAR Genom. Bioinform.* **2021**, *3*, lqab068. [[CrossRef](#)] [[PubMed](#)]
36. Sun, Z.; Evans, J.; Bhagwate, A.; Middha, S.; Bockol, M.; Yan, H.; Kocher, J.-P. CAP-MiRSeq: A Comprehensive Analysis Pipeline for MicroRNA Sequencing Data. *BMC Genom.* **2014**, *15*, 423. [[CrossRef](#)]
37. Andrés-León, E.; Núñez-Torres, R.; Rojas, A.M. MiARma-Seq: A Comprehensive Tool for MiRNA, MRNA and CircRNA Analysis. *Sci. Rep.* **2016**, *6*, 25749. [[CrossRef](#)] [[PubMed](#)]
38. Pérez-Rodríguez, D.; Pérez-Rodríguez, M.; Agís-Balboa, R.C.; López-Fernández, H. Towards a Flexible and Portable Workflow for Analyzing MiRNA-Seq Neuropsychiatric Data: An Initial Replicability Assessment. In *Practical Applications of Computational Biology and Bioinformatics, 16th International Conference (PACBB 2022)*; Fdez-Riverola, F., Rocha, M., Mohamad, M.S., Caraiman, S., Gil-González, A.B., Eds.; Springer International Publishing: Cham, Switzerland, 2023; pp. 31–42.
39. López-Fernández, H.; Graña-Castro, O.; Nogueira-Rodríguez, A.; Reboiro-Jato, M.; Glez-Peña, D. Compi: A Framework for Portable and Reproducible Pipelines. *PeerJ Comput. Sci.* **2021**, *7*, e593. [[CrossRef](#)]
40. Nogueira-Rodríguez, A.; López-Fernández, H.; Graña-Castro, O.; Reboiro-Jato, M.; Glez-Peña, D. Compi Hub: A Public Repository for Sharing and Discovering Compi Pipelines. In *Practical Applications of Computational Biology & Bioinformatics, 14th International Conference (PACBB 2020)*; Panuccio, G., Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Casado-Vara, R., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 51–59.
41. Andrews, S. FASTQC. A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: [https://scholar.google.com/scholar?hl=en&q=FASTQC.+A+quality+control+tool+for+high+throughput+sequence+data#d=gs\\_cit&t=1681893164969&u=%2Fscholar%3Fq%3Dinfo%3A7Au96aB8tVoj%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den](https://scholar.google.com/scholar?hl=en&q=FASTQC.+A+quality+control+tool+for+high+throughput+sequence+data#d=gs_cit&t=1681893164969&u=%2Fscholar%3Fq%3Dinfo%3A7Au96aB8tVoj%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D0%26hl%3Den) (accessed on 19 April 2023).
42. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet. J.* **2011**, *17*, 10–12. [[CrossRef](#)]
43. Langmead, B.; Trapnell, C.; Pop, M.; Salzberg, S.L. Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome. *Genome Biol.* **2009**, *10*, R25. [[CrossRef](#)]
44. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [[CrossRef](#)]
45. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve Years of SAMtools and BCFtools. *Gigascience* **2021**, *10*, giab008. [[CrossRef](#)]
46. Liao, Y.; Smyth, G.K.; Shi, W. FeatureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features. *Bioinformatics* **2014**, *30*, 923–930. [[CrossRef](#)]

47. Love, M.I.; Huber, W.; Anders, S. Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2. *Genome Biol.* **2014**, *15*, 550. [[CrossRef](#)]
48. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. EdgeR: A Bioconductor Package for Differential Expression Analysis of Digital Gene Expression Data. *Bioinformatics* **2010**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
49. Anders, S.; Huber, W. Differential Expression Analysis for Sequence Count Data. *Genome Biol.* **2010**, *11*, R106. [[CrossRef](#)] [[PubMed](#)]
50. Robinson, M.D.; Oshlack, A. A Scaling Normalization Method for Differential Expression Analysis of RNA-Seq Data. *Genome Biol.* **2010**, *11*, R25. [[CrossRef](#)] [[PubMed](#)]
51. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [[CrossRef](#)]
52. Chen, H.; Boutros, P.C. VennDiagram: A Package for the Generation of Highly-Customizable Venn and Euler Diagrams in R. *BMC Bioinform.* **2011**, *12*, 35. [[CrossRef](#)] [[PubMed](#)]
53. Blighe, K. EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling. 2022. Available online: <https://bioconductor.org/packages/devel/bioc/vignettes/EnhancedVolcano/inst/doc/EnhancedVolcano.html> (accessed on 19 April 2023).
54. Murtagh, F.; Legendre, P. Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion? *J. Classif.* **2014**, *31*, 274–295. [[CrossRef](#)]
55. Galili, T. Dendextend: An R Package for Visualizing, Adjusting and Comparing Trees of Hierarchical Clustering. *Bioinformatics* **2015**, *31*, 3718–3720. [[CrossRef](#)]
56. Warnes, G.R.; Bolker, B.; Bonebakker, L.; Gentleman, R.; Huber, W.; Liaw, A.; Lumley, T.; Maechler, M.; Magnusson, A.; Moeller, S.; et al. Gplots: Various R Programming Tools for Plotting Data. 2022. Available online: <https://cran.r-project.org/web/packages/gplots/gplots.pdf> (accessed on 19 April 2023).
57. Silge, J.; Robinson, D. Tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *JOSS* **2016**, *1*, 37. [[CrossRef](#)]
58. Shannon, P. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
59. Allaire, J.J.; Ellis, P.; Gandrud, C.; Kuo, K.; Lewis, B.W.; Owen, J.; Russell, K.; Rogers, J.; Sese, C.; Yetman, C.J. NetworkD3: D3 JavaScript Network Graphs from R. 2017. Available online: <https://cran.r-project.org/web/packages/networkD3/networkD3.pdf> (accessed on 19 April 2023).
60. Vaidyanathan, R.; Xie, Y.; Allaire, J.J.; Cheng, J.; Sievert, C.; Russell, K.; Hughes, E. RStudio Htmlwidgets: HTML Widgets for R. 2023. Available online: <https://www.htmlwidgets.org/> (accessed on 19 April 2023).
61. Ewels, P.; Magnusson, M.; Lundin, S.; Källér, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)] [[PubMed](#)]
62. López-Fernández, H.; Ferreira, P.; Reboiro-Jato, M.; Vieira, C.P.; Vieira, J. The Pegi3s Bioinformatics Docker Images Project. In *Practical Applications of Computational Biology & Bioinformatics, 15th International Conference (PACBB 2021)*; Rocha, M., Fdez-Riverola, F., Mohamad, M.S., Casado-Vara, R., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 31–40.
63. Doncheva, N.T.; Morris, J.H.; Gorodkin, J.; Jensen, L.J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **2019**, *18*, 623–632. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.