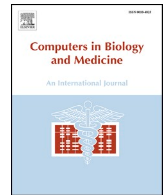
Computers in Biology and Medicine 135 (2021) 104603

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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



Application of miRNA-seq in neuropsychiatry: A methodological perspective

Daniel Pérez-Rodríguez [a,b], Hugo López-Fernández [c,d,e,*,1], Roberto C. Agís-Balboa [a,b,**,1]

[a] Translational Neuroscience Group-CIBERSAM, Galicia Sur Health Research Institute (IIS Galicia Sur), Área Sanitaria de Vigo-Hospital Álvaro Cunqueiro, SERGAS-UVIGO, 36213, Vigo, Spain
[b] NeuroEpigenetics Lab. University Hospital Complex of Vigo, SERGAS-UVIGO, 36213, Vigo, Spain
[c] Instituto de Investigação e Inovação Em Saúde (I3S), Universidade Do Porto, Rua Alfredo Allen, 208, 4200-135, Porto, Portugal
[d] CINBIO, Universidade de Vigo, Department of Computer Science, ESEI – Escuela Superior de Ingeniería Informática, 32004, Ourense, Spain
[e] SING Research Group, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Spain

A R T I C L E I N F O

Keywords:
RNA-Seq
microRNAs
Neuropsychiatric diseases
NGS
Bioinformatics workflow
Guide

A B S T R A C T

MiRNAs are emerging as key molecules to study neuropsychiatric diseases. However, despite the large number of methodologies and software for miRNA-seq analyses, there is little supporting literature for researchers in this area. This review focuses on evaluating how miRNA-seq has been used to study neuropsychiatric diseases to date, analyzing both the main findings discovered and the bioinformatics workflows and tools used from a methodological perspective. The objective of this review is two-fold: first, to evaluate current miRNA-seq procedures used in neuropsychiatry; and second, to offer comprehensive information that can serve as a guide to new researchers in bioinformatics. After conducting a systematic search (from 2016 to June 30, 2020) of articles using miRNA-seq in neuropsychiatry, we have seen that it has already been used for different types of studies in three main categories: diagnosis, prognosis, and mechanism. We carefully analyzed the bioinformatics workflows of each study, observing a high degree of variability with respect to the tools and methods used and several methodological complexities that are identified and discussed in this review.

1. Introduction

Unlike other diseases, mental illnesses diagnosis is mainly made by the identification of a wide set of unspecific symptoms. Patients diagnosed with the same psychiatric disorder can exhibit very different clinical manifestations, often resulting in poor treatment efficacy and management. The artificial classification of the mental disorders also restricts the power of discovering the underlying biological mechanism, as much variability is lost when assigning a diagnosis.

Micro RNAs (miRNAs) are small non-coding RNAs (<200 nucleotides) involved in the post-transcriptional regulation of gene expression, or RNA silencing, and participate in the epigenetic regulation of protein synthesis [1–3]. Most miRNAs are 21–24 nt in length, play a role in the regulation of most of the biological processes [4], and their expression is known to be affected by everyday events such as sleep, eating, stress or medications [3]. In the last decade, the importance of miRNAs as etiological mechanisms of neuropsychiatric disorders has been recognized and a large number of studies had begun to discover their roles in neuropsychiatry diseases such as Alzheimer's, schizophrenia depression or Parkinson [5–10]. In this context, miRNAs have recently gained attention as potential biomarkers in brain diseases due to their ability to epigenetically influence almost all aspects of brain functioning in a reversible manner [3,5,10,11].

The development of RNA-seq technologies focused on miRNA expression (miRNA-seq) allowed new approaches to understand, diagnose and treat these diseases. In this review, we focus on evaluating how miRNA-seq has been used to date in this area, with the goal of analyzing the usefulness and state of maturity of this technology in neuropsychiatry. These approaches have the benefit of capturing a lot of the variability inside these disorders, and setting a starting point to trace

* Corresponding author. Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Rua Alfredo Allen, 208, 4200-135, Porto, Portugal.
** Corresponding author. Translational Neuroscience Group-CIBERSAM, Galicia Sur Health Research Institute (IIS Galicia Sur), Área Sanitaria de Vigo-Hospital Álvaro Cunqueiro, SERGAS-UVIGO, 36213, Vigo, Spain.
E-mail addresses: hugo.fernandez@i3s.up.pt (H. López-Fernández), roberto.carlos.agis.balboa@sergas.es (R.C. Agís-Balboa).
[1] These authors contributed equally to the work.

https://doi.org/10.1016/j.compbiomed.2021.104603
Received 4 May 2021; Received in revised form 21 June 2021; Accepted 21 June 2021
Available online 24 June 2021
0010-4825/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

their etiology [7,12,13]. In addition, it is important to note that very little information and standardization regarding miRNA-seq data analysis is available, resulting in a growing concern about the veracity and reproducibility of its results and conclusions [14,15]. For this reason, this review also aims to offer clear and comprehensive information that can serve as a guide to new researchers in this field.

Recent reviews focused on analyzing the trends in the development of miRNA bioinformatics tools [16] or specific steps of the typical workflows such as preprocessing [17] or normalization [18]. In this review, we analyze the main objectives and bioinformatics workflows of miRNA-seq studies of neuropsychiatric diseases from a methodological perspective. Our objective is to create a valuable resource for those who want to work in this field by providing an overview of how miRNA-seq data is currently processed to study neuropsychiatric conditions.

For this, we conducted four systematic searches on PubMed Central using the terms detailed in the Supplementary Table 1. Our inclusion criteria were the use of a miRNA-seq methodology to study a neuropsychiatric condition (on human or animal models), and a publication date ranging from January 2016 to June 2020. After the initial search, we manually examined the 440 results and excluded bibliographic reviews, studies focused on specific miRNAs or those addressing traumatic brain injuries. We also discarded studies based on data acquired from microarray or RT-qPCR technologies, including those performing validation of a previous miRNA-seq study. We selected two studies [19,20] from the 73 results of the first search; one [21] from the 70 results of the second; nine [22–30] from the 225 results of the third, and four studies [31–34] from the 72 results of the fourth search. These 20 studies were used in the development of this review.

Once the bibliographic material was selected, we grouped the twenty studies by their main objectives (diagnosis, prognosis, and mechanism; which are explained in section 2), creating one diagram per article to analyze each bioinformatics workflow in detail. After analyzing each workflow, we built a summary diagram for each objective and used them to build a general diagram that encompasses all the articles pipelines (section 3). In addition, we analyzed each workflow step in detail to find the different approaches and software used in the literature.

2. Objectives

The aim of this section is to categorize the literature reviewed and analyze their main objectives from a functional point of view, bringing an overview of the questions addressed to date in the field of neuropsychiatry using miRNA-seq data. To complement this analysis, the next section focuses on the steps of the bioinformatics analyses required to accomplish such objectives. The 20 studies used miRNA-seq in animal and/or human samples oriented to study several psychiatric disorders and neurodegenerative diseases and fall into three main categories: diagnosis, mechanism, and prognosis. While the latter category has only one study, the remaining are almost equally distributed between diagnosis and mechanism, with one that falls in both categories.

Eight of the 20 studies reviewed fall within the diagnosis category [11,19,20,22,28,30,35,36]. Table 1 summarizes the main features of these studies in chronological order. As shown, all of them were carried out in humans and the most studied condition is Alzheimer's disease. These studies rely on the definition and discovery of biomarkers, that is, specific disease-related miRNAs. Once a set of biomarkers have been identified, two main analyses are performed: (i) building a predictive model to evaluate the potential diagnostic utility, and (ii) applying target prediction followed by enrichment analysis. Hicks [22]

Table 1
Studies in the diagnosis category.

Study	Date	Organism	miRNA sequencing target	Disease	Samples	Objectives
Hicks et al., 2016 [22]	April 2016	Human	Saliva	Autist Spectrum Disorder (ASD)	24 ASD, 21 controls	- Discover biomarkers by comparing miRNAs from ASD saliva samples and controls. - Evaluate the functional importance of ASD-related miRNAs.
Hoss et al., 2016 [19]	March 2016	Human	Brain (post-mortem)	Parkinson's Disease (PD)	29 PD, 33 controls	- Discover biomarkers by comparing PD patients and controls. - Study the disease-related phenotypes of onset age and dementia by comparing PD patients with dementia and dementia patients without PD.
Hara et al., 2017 [30]	January 2017	Human	Serum	Alzheimer's disease (AD)	Discovery set: 27 AD, 18 controls Validation set: 36 AD, 22 controls	- Discover blood biomarkers by comparing serum samples from AD patients and controls.
Wang et al., 2018 [26]	May 2018	Human	Blood samples/ White blood cells	Attention-deficit/ hyperactivity disorder (ADHD)	White blood cells (discovery) set: 5 ADHD, 5 controls Training set (blood samples): 68 ADHD, 54 controls Testing set: 20 ADHD, 20 controls	- Discover biomarkers by comparing white blood cells from ADHD patients with controls.
Gámez-Valero et al., 2019 [20]	October 2019	Human	Plasma extracellular vesicles (EVs)	Alzheimer's disease (AD) and dementia with Lewy bodies (DLB)	18 DLB, 10 AD, and 15 controls	- Discover biomarkers to improve misdiagnosis between AD and Dementia with Lewy bodies (DLB) by comparing miRNAs from plasma extracellular vesicles of AD patients with LBD patients.
Lee et al., 2020 [28]	January 2020	Human	Serum (peripheral blood)	Bipolar II disorder (BP-II)	Discovery set: 3 BP-II, 3 controls Training set: 79 BP-II, 95 controls Testing set: 20 BP-II, 20 controls	- Discover biomarkers by comparing miRNAs from BD-II blood samples and controls.
Wu et al., 2020 [36]	February 2020	Human	Blood samples	Alzheimer's disease (AD)	40 AD, 31 controls	- Discover biomarkers by comparing miRNAS from AD blood samples and controls.
Nie et al., 2020 [35]	May 2020	Human	Exosome (exo-miRNA)/peripheral blood	Parkinson's Disease (PD)	5 AD, 7 PD, 34 controls	- Discover biomarkers for AD and PD by comparing AD/PD plasma exosomal miRNAs with controls.

highlighted 14 miRNAs in saliva with potential for a diagnosis of Autism Spectrum Disorder (ASD). Wang [26] described a set of 13 miRNAs in peripheral blood that may assist in identifying Attention-deficit/hyperactivity disorder (ADHD). Lee [28] identified serum miRNAs (miR-7-5p, miR-142–3p, miR-221–5p, and miR-370–3p) that may be potential peripheral biomarkers in the diagnosis of bipolar II disorder (BP-II). Hoss [19] described 29 miRNAs altered in Parkinson Disease (PD) prefrontal cortex, 21 of which overlapped with Huntington Disease (HD). Nie [35] found eight miRNAs present in human plasma exosomes to be significantly altered in AD and PD samples. Hara [30] identified serum miR-501–3p as a potential biomarker related to the progression of Alzheimer's disease. Another study detected different miRNA expression levels comparing the miRNA profile of plasma-derived extracellular vesicles of dementia with Lewy bodies (DLB) and AD patients, which could help to improve the differential diagnosis between both diseases [20]. Wu [36] detected, in peripheral blood, 71 significantly differentially expressed miRNAs between the Alzheimer's disease (AD) and control groups.

Nine of the 20 studies reviewed fall within the mechanism category [21,22,24,25,27,29,31–34]. Table 2 summarizes the main features of these studies in chronological order. The aim of these articles is to deepen the understanding of neuropsychiatric pathology and expand the essential knowledge about the disease. Unlike the studies in the diagnosis category, these studies have a fundamental approach and usually do not seek direct clinical application, and cover topics such as discovering underlying biological mechanisms [22,25,27,32,34,37], investigating the relationship between miRNAs and an interesting feature [21,25,29,31,32,38], and studying the fluctuations of the miRNAome during an event of interest [27,34].

Hicks [22] is the only study falling into two categories that highlights

Table 2
Studies in the mechanism category.

Study	Date	Organism	miRNA sequencing target	Disease	Samples	Objectives
Hicks et al., 2016 [22]	April 2016	Human	Saliva	Autist Spectrum Disorder (ASD)	21 controls, 24 ASD	- Discover biomarkers by comparing miRNAs from ASD saliva samples and controls. - Evaluate the functional importance of ASD-related miRNAs.
Pfau et al., 2016 [24]	December 2016	Mouse	Nucleus Accumbens (NAc)	Stress-related disorders	NA	- Study the transcriptional and post-transcriptional profiles of NAc regarding to sexual differences in the behavioral response to subchronic variable stress (SCVS). For this, a comparison of the expression of miRNA and mRNA in NAc after exposure to such stress was conducted.
Martin et al., 2017 [25]	January 2017	Human	Peripheral blood	Post Traumatic Stress Disease (PTSD)	9 controls, 15 PTSD	- Study the role of miRNAs in PTSD symptoms by comparing PTSD patient blood samples with controls.
Faraji et al., 2017 [31]	May 2017	Long-Evans rats	Brain	Prenatal stress (PS)	34	- Study the possible relationships between multigenerational PS and focal cerebral ischemia outcomes by comparing miRNA expression levels between stressed rats, rats with a stroke, rats with both conditions and controls.
Rani et al., 2017 [21]	October 2017	Human	Extracellular microvesicle enriched plasma samples	Age-Related Cognitive Decline	97 healthy individuals	- Study the relationships of the structural and cognitive function of the brain in older adults by comparing the expression levels of miRNA between different ages and cognitive functions.
Si et al., 2018 [32]	August 2018	Mouse	Nucleus Accumbens	Depression	NA	- Improve understanding of the molecular mechanisms underlying stress-induced depression versus resilience. For this, a comparison of miRNA and mRNA expression levels from depressed and resistant mice after applying chronic unpredictable mild stress was conducted.
Sillivan et al., 2019 [29]	May 2019	Mouse	Basolateral amygdala complex (BLC)	Post Traumatic Stress Disease (PTSD)	NA	- Study the role of the miRNAs in the susceptibility to remote, stress-enhanced memories by comparing miRNA expression of stress-susceptible mice to stress resilient mice.
Hu et al., 2019 [27]	August 2019	Human	Human dorsolateral prefrontal cortex (DLPFC)	Schizophrenia	109 controls, 34 patients	- Study the temporal dynamics of miRNA expression and its dysregulation in the DLPFC of patients with schizophrenia by determining its expression at different ages and comparing it with controls.
Mavrikaki et al., 2019 [33]	October 2019	Sprague-Dawley rats	Bed nucleus of the stria terminalis (BNST)	Anxiety/Stress response	NA	- Study the differences in the expression of BNST miRNA between male and female rats exposed to social isolation stress on the adolescence by comparing their miRNA expression to controls.
Song et al., 2019 [34]	December 2019	Sprague-Dawley rats	Nucleus Accumbens	Chronic Unpredictable Mild Stress (CUMS)	8 controls, 8 CUMS, 8 CUMS with treatment	- Study the alterations on NAc miRNA expression during chronic stress and the effect of the medication on this alteration. For this, comparisons between control, CUMS and CUMS with treatment were made through a co-expression analysis.
Ma et al., 2020 [37]	February 2020	Mouse	Brain	Alzheimer's disease (AD)	NA	- Study the interactions between miRNAs, mRNAs and lncRNAs on AD to seek differentially expressed networks.
Weisz et al., 2020 [38]	Month 2020	Human/Rat	Hipocampus (rat)/ Serum (human)	Traumatic Brain Injuries (TBI)	Human: 51 (6 acute and 6 aged controls, 33 acute TBI, 6 chronic TBI) Rats: 56 (half controls, half TBI)	- Study the possible relationships between long-term dysregulation of hippocampal miRNAs and chronic neurodegeneration after TBI. For this, a comparison of miRNA expression levels between patient with a diffuse injury pattern and patients with focal contusion was conducted.

14 miRNAs as potential biomarkers for ASD. Pfau [24] and Si [32] suggested that miRNA profiles are associated with depression resilience. Song [34] showed that antidepressant treatment (Escitalopram) can reverse Nucleus Accumbens (NAc) miRNA abnormality induced by chronic stress. Hu [27] showed that miR-936 and miR-3162 are important in dorsolateral prefrontal cortex development, suggesting a relationship between schizophrenia and dysregulation of miRNAs augmented in infancy and prepuberty. Posttraumatic stress disorder (PTSD) was found to be related to eight differentially expressed miRNAs in peripheral blood which can target axon guidance and Wnt signaling pathways [25]. Another study showed that mir-135b-5p contributes to the storage of stress-enhanced fear memory within the amygdala and may be an important therapeutic target [29]. In addition, multigenerational prenatal stress downregulates miR-708 in the prefrontal cortex and upregulates the MAPK pathway involved in regulation of neuron development and cell proliferation [31]. Mavrikaki [33] raises the opportunity to develop sex-specific pharmacotherapies. Ma [37] used the APP/PS1 mouse brain to provide insights that facilitate AD diagnosis and future treatment strategies. Rani [21] identified 13 miRNAs in circulating microvesicles as potential biomarkers for age-related cognitive decline, and that could be associated with a pre-symptomatic stage of disease. To conclude, Weisz [38] proposed that miRNA-sequencing in biofluids might be used to distinguish traumatic brain injuries (TBI) as acute, chronic, focal or diffuse, and potentially, the existence of neuro-degenerative sequelae.

Finally, only one study belongs to the prognosis category [23]. In this study, authors identified biomarkers for disease risk estimation, looking for indications of a prognosis in the characteristics of the patient (sex, disease subtype) and attempted to find these variation patterns at the miRNA expression level. Like all the diagnosis studies, this study was also carried out in human samples of immune cells (leukocytes). The disease under study was psychosis (schizophrenia or a related disorder) and included 27 controls, 30 high-risk progressors, and 37 high-risk non-progressors. Authors were able to identify five regulatory miRNAs in leukocytes that could differentiate persons who develop psychosis from those who do not.

3. Bioinformatics analysis

This section is a compilation of the analyses, methods and software tools that were found in the literature. Table 3 is intended to serve as a guide to the subsequent content by providing an overview of the software used in each step of the miRNA-seq workflow; additionally, Supplementary Table 2 offers a summary of each analysis along with a list of the software used for each process, which are discussed in the following subsections.

3.1. miRNA/mRNA NGS data processing

The first process of the bioinformatics pipeline is the preparation of the miRNA sequenced data to run all the subsequent analysis. This step comprises data preprocessing, quality control, alignment to the reference genome and quantification and normalization of the results. Furthermore, the studies that analyze the miRNA-mRNA targets [24,32,37] also preprocess the mRNA data, adding an extra step: filtering the non-coding transcripts. These are all well established procedures across the studies and there is not much conceptual variation.

3.1.1. Preprocessing

Raw sequenced data need to be prepared before any type of interpretation. This preprocessing mainly implies trimming the 3′ or 5′ adapters needed by the HT-Seq technology and filtering the sequences by size to keep only those attributable to miRNAs (approximately 18–25 nt).

Cutadapt [39] was the most common software used for adaptor trimming, employed on six of the 20 studies [19,29–31,33,36]. The main features of cutadapt are trimming the 5′ and 3' adapter and the poli-A tails of the transcripts, and offering custom thresholds for mismatches, deletions and insertions.

Six of the remaining studies used different tools, namely FASTX Trimmer [24], Ion Torrent Suite [25], miRDeep2 [38], miRSeq [26], Partek Flow [21] and Trimmomatic [20]. The remaining studies did not specify the process.

Table 3
Software used per analysis on the reviewed studies. They are grouped chronologically on each objective. Processes not specified on the studies are denoted by "-".

Objective	Study	Software												
		Preprocessing	Quality control	Alignment	Quantification	Normalization	Filter nc-transcripts	Filter reads by abundance	DEA	FDR correction	Predictive model	Target prediction	Enrichment analysis	Network analysis
DIAGNOSIS	Hicks et al., 2016 [22]	-	-	Bowtie	BaseSpace	BaseSpace	-	-	-	-	Caret	miRDB	DAVID	-
	Hoss et al., 2016 [19]	Cutadapt	FastQC	Bowtie	FASTA/Q collapser	DESeq	-	DESeq2	Limma	Limma	GenePatterns	-		-
	Hara et al., 2017 [30]	Cutadapt	-	Bowtie	HT-Seq	DESeq	-	DESeq2	DESeq	DESeq	-	-	-	-
	Wang et al., 2018 [26]	miRSeq	miRSeq	miRSeq	miRSeq	-	-	-	SPSS	-	-	TargetScan	Partek-Genomic-Suite	-
	Gámez-Valero et al., 2019 [20]	Trimmomatic	-	Bowtie	-	-	-	-	-	-	-	-	-	-
	Lee et al., 2020 [28]	-	-	Bowtie	-	-	-	-	SPSS	-	SPSS	TargetScan	DAVID	-
	Wu et al., 2020 [36]	Cutadapt	FastQC	Bowtie	sRNAnalyzer	edgeR	-	-	edgeR	edgeR	-	IPA	DAVID	-
	Nie et al., 2020 [35]	-	-	-	-	DESeq	-	DESeq2	DESeq, edgeR, Limma	DESeq	-	TargetScan, miRanda, RNAhybrid	DIANA-miRPath	-
MECHANISM	Hicks et al., 2016 [22]	-	-	Bowtie	BaseSpace	BaseSpace	-	-	-	-	Caret	miRDB	DAVID	-
	Pfau et al., 2016 [24]	FASTX Trimmer	FastQC	Bowtie	HT-Seq	-	-	-	Limma	Limma	-	miRWalk	IPA, DAVID	Cytoscape
	Martin et al., 2017 [25]	Ion Torrent Suite	Ion Torrent Suite	Burrows Wheeler Aligner	miRDeep2	DESeq	-	DESeq2	DESeq	DESeq	-	miRWalk	miRWalk	-
	Faraji et al., 2017 [31]	Cutadapt	FastQC	MicroRazerS	MicroRazerS	DESeq	-	-	DESeq	DESeq	-	-	-	-
	Rani et al., 2017 [21]	Partek Flow	Ion Torrent Suite	Bowtie	-	-	-	-	-	-	-	DIANA-TarBase	DIANA-miRPath	-
	Si et al., 2018 [32]	-	NOIseq	Bowtie, Blast	-	DESeq	-	-	DESeq, NOIseq	DESeq, NOIseq	-	RNAhybrid, TargetScan, miRanda	-	Cytoscape
	Sillivan et al., 2019 [29]	Cutadapt	-	miRDeep2	miRDeep2	DESeq	-	DESeq2	DESeq	DESeq	-	TargetScan, DIANA-TarBase, DIANA-miRPath	DIANA-miRPath	-
	Hu et al., 2019 [27]	-	-	Bowtie	-	edgeR	-	-	edgeR	edgeR	-	TargetScan	JNLP-GSEA	-
	Mavrikaki et al., 2019 [33]	Cutadapt	FastQC	seqbuster, STAR	miRDeep2	DESeq	-	DESeq2	DESeq, Limma	DESeq, Limma	-		DIANA-miRPath	-
	Song et al., 2019 [34]	-	-	CLC genomics_workbench	-	-	-	-	DEGseq	DEGseq	-	TargetScan	DAVID	Cytoscape
	Ma et al., 2020 [37]	Custom script	Custom script	Bowtie	miRDeep2	DESeq	CNCI, CPC, PfamScan, PhyloCSF	-	DESeq, Cuffdiff	DESeq, Cuffdiff	-	miRanda	DAVID	Cytoscape
	Weisz et al., 2020 [38]	miRDeep2	-	miRDeep2	miRDeep2	-	-	DESeq2	edgeR, DESeq	edgeR, DESeq	-	miRDB	IPA	-
PROGNOSIS	Jeffries et al., 2016 [23]	-	-	-	-	-	-	-	-	-	-	-	-	-

3.1.2. Quality control

Before data interpretation, reads need to be analyzed for foreign DNA contamination, unwanted biases or artificial duplications that might have been generated by the sequencer itself. For this purpose, the reviewed studies used four different tools, namely FastQC [40], Ion Torrent Suite [41], miRSeq [42] and NOIseq [43].

FastQC was the most common software used for quality control and it was employed in five of the 20 studies to estimate overall sequence quality [19,24,29,31,33]. The main feature of FastQC is the generation of an html report from a BAM, SAM or FastQ file. This file contains basic statistics results such as total numbers of reads, reads length or GC content along with more specific analyses intended to test for several types of bias like overrepresented sequences or Kmer content.

Two of the remaining studies [21,25] used Ion Torrent Suite for this process and the other two miRSeq and NOIseq, respectively [26,32]. The remaining studies did not specify this process.

3.1.3. Alignment

Alignment is the process of mapping reads against a reference genome or transcriptome in order to identify the corresponding genomic positions and find the reads identity. It is the key step where the data obtained by a sequencer is linked to current knowledge, and all downstream analyses depend on its precision.

Aligning against a genome allows for the discovery of potentially new miRNAs, but it is a complex process and usually requires indexing the reference genome in order to increase the speed of the query process. Transcriptome alignment on its side is faster and more reliable in the reads identification, but is restricted to the current knowledge and prevents new, unannotated transcript discovery [44].

Eight of the 20 studies (40%) align to the transcriptome [11,23,28,32,34,36,38], twelve (60%) align to a reference genome [19–22,24,25,29–31,33,35,37] and only one (5%) performed both alignments [33]. Is important to note that nine studies used a deprecated version of a genome for alignment (Supplementary Table 3).

As all the studies used miRNAs data, they employed specialized aligners to match short reads to large genomes; these are known as short read aligners [45]. Bowtie [46] was by far the most widely used software being employed on 9 of the 20 studies (45%) [19–22,24,27,28,30,32,36,37]. Its two main features are great efficiency on short reads alignment, as a result of indexing the genome with a Burrows-Wheeler transformation, and high interoperability due to its standard output on SAM format. It is also integrated by tools like TopHat [47] and Crossbow [48]. Bowtie 2 [49], meanwhile, is recommended for reading lengths> 50 nt; however it was used in three studies [20,28,37]. As shown in Table 3, the remaining eight studies that did specify this step used CLC genomics workbench [50], miRDeep2 [51], miRSeq [42], MicroRazerS [52], Blast [53], Burrows-Wheeler Aligner [54], seqbuster [55] and STAR [56]. Lastly, in the study conducted by Mavrikaki [33] two different tools are used, namely seqbuster and STAR.

3.1.4. Quantification

The aligned sequences are the reads "tagged" with an identifier (usually their genome coordinate) along with some metadata and information about each alignment. Some common formats to store this type of data are the BAM, SAM and CRAM. In the quantification process, these sequences are grouped by their identifier and quantified, so that the output of this process is a list of non-repeated identifiers together with their number of occurrences in the sample. The coordinates of the sequence are usually intersected before the actual quantification process with an annotation file from some database (commonly on GFF/GTF format) to add current knowledge about each read. In addition, unannotated reads can be tested for a potential miRNA function.

In line with the above, miRbase [57] was the most employed database for miRNA annotation, having been used in 14 of the studies (70%) [19,21,26–30,32–38]. Other databases used were MirGeneDB [58], GenBank [59], Rfam [60] and ncRNA database [61]. Three studies [21,33,37] tested for potential miRNA function using miRDeep2, and only Ma [37] also used miREvo [62] for this purpose.

Regarding the quantification process itself, seven different tools were used for quantification, namely BaseSpace [63], FASTA/Q collapser [64], HTSeq [65], miRDeep2 [51], miRSeq [42], sRNAnalyzer [66] and MicroRazerS [52]. MiRDeep2 was the most prevalent tool, being used in four of the 20 studies [25,29,33,38]. It is a software package composed of three scripts that covers preprocessing (mapper.pl), quantification (quantifier.pl) and identification of known and novel miRNAs (miRDeep2.pl). The latter uses an algorithm based on Bayesian statistics to score and collect potential miRNA sequences between the aligned reads [51].

Sillivan [29], Weisz [38] and Martin [25] used quantifier.pl for quantification whereas Mavrikaki [33] did not specify the quantification method, although they probably used quantifier.pl as well. As shown in Table 3, Hara [30] and Pfau [24] used HTSeq Python package, and the five remaining studies used Illumina BaseSpace platform [22], FASTA/Q collapser [19], miRSeq [26], sRNAnalyzer [36], MicroRazerS [31] and a custom script [37]. The remaining articles did not specify this process.

3.1.5. Normalization

The number of reads mapped to a gene is affected by different factors aside from its abundance; factors such as reads length [67], GC-content [68] and sequencing depth [69] can cause significant alterations on the quantification results and produce false positives in differential expression analysis [70]. Therefore, normalization is essential to make reliable comparisons between samples. Five different normalization methods were used on the bibliography reviewed, namely Reads Per Kilobase Million (RPKM), Transcripts Per Kilobase Million (TPM), Trimmed Mean of M-values (TMM), Reads per million mapped reads (RPM) and median-of-ratios method (DESeq). RPKM and TPM are used to normalize for library size whereas TMM and DESeq normalize for sequence depth, while RPM is similar to RPKM but without taking into account the transcripts length. It has been suggested that TMM and DESeq are the most suitable methods for miRNA count data normalization, whereas the Total Count and RPKM normalization methods are discouraged [17,71].

As can be seen on Supplementary Table 4, five of the 20 studies (25%) normalized for library size using TPM (3), RPKM (1) and custom methods (1). 12 of the 20 studies (60%) normalized for sequence depth using DESeq (9) and TMM (3). Hicks [22] used RPM for normalization, Song [34] did not normalize the data due to DEGseq requirements [72] and Pfau [24] and Rani [21] did not specify the normalization process.

To carry out the counts normalization the reviewed studies used three different software tools, namely DESeq/DESeq2 [73,74], edgeR [75] and Illumina BaseSpace platform [63]. As shown in Table 3, the most used software was DESeq/DESeq2, having been employed in nine of the 20 studies [19,25,29–33,35,38]. DESeq is an R package for normalization, visualization and differential expression analysis of high dimensional count data. The DESeq normalization assumes that DE and non-DE genes behave the same and that the expression across conditions is balanced [70]. The use of edgeR, which is an R package specialized on differential expression analysis of RNA-seq data with a built-in normalization method (TMM) for sequence depth, is notable in three of the most recent studies [27,36,38]. It makes the same assumptions as DESeq. Finally, Hicks [22] used the Illumina Basespace platform for the normalization process. The remaining studies did not specify the software used.

3.1.6. Filtering non-coding transcripts

Several studies apply a "Network analysis" to search for relationships between miRNAs and mRNAs. This process involves the integration of miRNA and mRNA data, which requires an mRNA preprocessing. To obtain the mRNAs from the RNA-seq data, a filter is applied to non-coding reads to retain only those with potential mRNA function. Only Ma [37] specified the software tools used for this purpose, which were

CNCI [76], CPC [77], PfamScan [78] and PhyloCSF [79]. The remaining articles [24,32] applied this step but did not specify any particular software.

3.2. Expression analysis

After preprocessing comes expression analysis, which is the first stage of the workflow where the data begins to be analyzed: it comprises filtering reads by abundance, the differential expression analysis itself and a false discovery rate correction procedure. The expression analysis can be regarded as the set of processes aimed at detecting statistically significant variations in miRNA expression between different conditions.

3.2.1. Filter low-expression genes

RNA-seq provides a high dynamic range on gene expression profile quantification; however, quantification methods lack precision on low-expressed reads due to noise produced by the random sampling process typical of this technology [80]. The filtering of low-expressed genes can increase the number of differentially expressed reads and improve the robustness of its results [81].

There are several filtering procedures, usually based on an empirical threshold suggested by a differential expression analysis software [80]. For instance, in DESeq2 this filter can be applied automatically.

As shown in Supplementary Tables 5 and 15 of 20 articles (75%) explicitly stated that they filtered low-expression miRNAs/mRNA or used DESeq2, whereas five studies did not report any software. Supplementary Table 5 shows the filtering criteria in each study: usually the occurrence of a read between samples, a minimum threshold for reads on each sample or a combination of both criteria are used to discard low-expression reads. There was no agreement about the "occurrence across samples" threshold, which was 10% in the Hu study [27], 50% in the Hicks study [22], 80% in the Pfau study [24], 100% in the Rani and Jeffries studies [21,23] and a mixed criteria in the Gámez-Valero study [20]. Regarding the "minimum reads per sample threshold", four studies [20,24,32,36] established five reads as the minimum required to count a miRNA as present, whereas the remaining ones widely differ on their criteria. Finally, five studies [20,21,23,24,36] used both the threshold of "occurrence across samples" and "minimum reads per sample" to find low-expressed miRNAs.

3.2.2. Differential expression analysis

Once the miRNA-seq data for each condition studied is quantified, normalized, and low-expression reads have been removed, the miRNAs are compared to find transcriptional variations between cases and controls with the goal of finding over-expressed, under-expressed, or absent miRNAs among the conditions under study. This process is known as differential expression analysis (DEA) and its results are often critical to the study's conclusions. There are several complexities in this process that can lead to an inaccurate outcome. On one hand, there are many covariates besides the one of interest that may hide the association of expression levels with the primary factor of interest [82]; these include variables such as age, gender or medications, but also the batch effect. On the other hand, the available DEA software tools implement different methods that can lead to different results from the same data [83–85], making comparisons across studies difficult. In line with these problems, two of the reviewed studies [35,38] addressed the variability of the DEA tools by using more than one and keeping only the matching results.

Regarding the confounding factors effects, four studies [19,24,30,33] repeated the DEA with and without adjustment for covariates, and then compared the results to overview the magnitude of the non-relevant variability. Wang [26] applied a Multivariate Analysis of Covariance (MANCOVA) to test whether age, sex, or intelligence quotient functioned as confounding factors, whereas Rani used an ANOVA for sex differences and Pearson's regression analysis for age. Hu [27] previously applied a Principal Component Analysis (PCA) to test if diagnostic, age, sex or race cause a significant variation in expression levels, and then adjusted a general linear model (GLM) for the most relevant variables. This is important because accounting for the effect of irrelevant variables also reduces the power of detecting DE genes [86]. Finally, due to the small sample sizes only Hoss [19] used ComBat [87] to correct for batch effect before applying the DEA.

Regarding the DEA tools, seven different software packages were used for this process, namely DESeq/DESeq2 [73,74], edgeR [75], Limma [88], SPSS, Cuffdiff [89], DEGseq [72] and NOIseq [43]. DESeq and its latest version DESeq2 were the most frequent software, used in nine of the 20 studies (45%). These two programs use a parametric approach (the negative binomial distribution) to build a model to find differentially expressed reads between conditions. With the parametric approach it is possible to rely on the model to predict the value of unknown data, but the conclusions are limited by the fit of the model [86].

Also notable is the use of other two parametric DEA tools: edgeR and Limma, each one used in four different studies (Table 3). EdgeR uses negative binomial distribution whereas Limma relies on a linear approach. Lastly, Cuffdiff, DEGseq and NOIseq were each used in just one study. DEGseq has a parametric approach based on a Poisson model while Cuffdiff and NOIseq are non-parametric tools.

It is worth mentioning that 11 of the studies reviewed [20,23,24,27–30,32,33,37,84] used an experimental validation through RT-qPCR to confirm the in-silico DE miRNA expression. Wu [36] addressed the lack of experimental validation with the replication of the study in an independent cohort.

3.2.3. Multiple testing correction

On an RNA-seq DEA, each gene is tested for differential expression analysis with a certain threshold of significance. This means that tens of thousands of comparisons are made, one for each gene, and that the confidence level is applied to each test considered individually. This fact, known as multiple comparisons problem, may lead to high rates of false positives and produce an overestimation of the associations detected. Thus, a multiple testing correction procedure is applied.

The most common correction in the reviewed literature was the False Discovery Rate (FDR), which was proposed by Benjamini and Hochberg in 1995 and is defined as "an estimation of the proportion of errors committed by falsely rejecting null hypotheses'' [90]. The FDR correction was employed on 17 of the 20 articles (85%), and it is included on all the DEA software used (Cuffdiff [89], DEGseq [72], DESeq [74], edgeR [75], Limma [88], NOIseq [43]) on the studies [19,24,25,27,29–38]. Hicks [22], Gámez-Valero [20] and Rani [21] applied this correction but did not specify any software.

Only Jeffries [23] used a family-wise error rate approach using a Bonferroni correction. The Bonferroni correction is a very conservative correction when applied to high-dimensional data, since the significance level decreases as the number of comparisons increases. This results in a lack of power to detect true positives and has been the main reason for the disuse of this method in the RNA-seq analysis in favor of the FDR approaches [91]. Finally, Wang [26] and Lee [28] did not mention any multiple testing correction.

3.3. Evaluation of diagnostic utility

After DEA and multiple testing correction, a list of potential biomarkers is obtained. Then, a set of tests can be performed to assess the predictive power of these molecules and thus estimate their applicability in a clinical setting. The main approach is to build a predictive model using machine learning techniques with the data and then evaluate its performance to discriminate between the conditions under study. Predictive models are excellent tools for testing potential biomarkers, but their performance is tied to the data on which they are built. If the sample is small, as in the case of RNA-seq studies, there is a high risk of overfitting, as the model will not be able to generalize to new, unseen data.

Six of the 20 articles (30%) built predictive models to assess the diagnostic utility of the differentially expressed miRNAs, using different classification models and validation schemes [19,20,22,23,28,36]. Table 4 summarizes these methods along with the software packages used when available. These classification models were: supported vector machine (SVM) [28], Wilcoxon-rank sum test [20], partial least squares discriminant analysis (PLS-DA) [22], weighted voting classification [19], logistic regression analysis [36], and a combination of a greedy algorithm (GA) with a *t*-test [23].

Regarding the validation schemes (Table 4), two of the studies used Leave One Out Cross-Validation (LOOCV) [19,20], whereas the remaining articles applied different approaches. Lee et al. [28], used a train dataset for model building and an independent test set for model evaluation. Interestingly, Wu et al. [36] replicated the study on an independent dataset and compared the overlapping miRNAs found in both datasets. Then, they applied a ROC analysis to assess the discriminative power of the two replicated DE miRNAs. Other approaches include Monte Carlo cross validation (MCCV) [22] and the application of a permutation test [23].

3.4. Functional analysis

At this stage of the bioinformatics workflow, the miRNAs presumably related with the conditions under study have been discovered due to the DEA. Thus, the next logical step is to analyze these molecular measurements at the functional level. The group of methods intended to characterize the functions affected by these DE miRNAs constitute the functional analysis.

3.4.1. Target prediction

As post-transcriptional repressors, miRNAs block the protein-coding gene expression by binding their seed region to the mRNA 3′ UTR. An alteration in the miRNA biogenesis or their mRNA targets has been linked to the development and the evolution of diseases [92], including psychiatric and neurodevelopmental disorders [93]. Furthermore, since one miRNA can target several mRNAs, the number of genes affected by a miRNA deregulation is expected to be greater than the number of miRNAs altered in such dysregulation. The identification of the mRNAs targeted by the DE miRNAs is known as target prediction (TP) and is a key step on the interpretation of an RNA-seq analysis.

The process of TP is usually based on a comparison between the DE miRNA and an mRNA database, resulting in a list of candidate mRNA targets. Commonly, these comparisons are focused on a miRNA-mRNA seed match, the conservation of miRNA regions or the ease of the miRNA-mRNA hybridization [94].

Twelve of the reviewed studies (60%) performed a TP analysis using a total of eight different software tools [25–29,32,34–36], namely IPA [95], DIANA-miRPath [96], miRanda [97], miRDB [98], miRWalk [99], RNAhybrid [100], DIANA-TarBase [101] and TargetScan [102]. The most used tool was TargetScan, which was employed in seven of the 20 studies [26–29,32,34,35]. TargetScan is a web server focused on miRNA TP that uses the miRNA-mRNA seed matching and conserved sites to establish the relationship between the miRNA and its mRNA target. The output of TargetScan is a rank of the predicted targets. To obtain more robust results, three studies applied several TP software to their data and kept only those targets predicted by all the databases [32,35] or more than one [29]. The combinations used were TargetScan + miRanda + RNAhybrid [32,35] and TargetScan + TarBase + microCts [29]. The remaining two articles used Ingenuity Pathway Analysis (IPA) [36] and miRWalk [25].

Table 4
Classification and validation methods used for the evaluation of the diagnostic utility.

Study	Classifiation Model	Validation	Software
Lee et al., 2020 [28]	Support Vector Machines	Train/test	SPSS
Wu et al., 2020 [36]	Logistic regression analysis	Independent sample	–
Gámez-Valero et al., 2019 [20]	Wilconson-rank sum test	Leave One Out Cross-Validation	–
Hicks et al., 2016 [22]	Partial least squares discriminant analysis	Monte Carlo cross-validation	Caret
Hoss et al., 2016 [19]	Weighted Voting classification	Leave One Out cross-validation	GenePatterns
Jeffries et al., 2016 [23]	Greedy Algorithm and *t*-test	Permutation test	–

3.4.2. Pathway enrichment analysis

A pathway is a group of genes and their interactions that are related to a specific biological function [103]. The TP analysis brings information about the genes altered by the disease, so the process of finding the interactions and functional groups in these genes and then searching for differential expression is known as pathway enrichment analysis (PEA). PEA bridges the gap between gene alterations and biological functions, making RNA-seq information less abstract and offering clearer insights of the underlying biological processes.

There are two key concepts to understanding a PEA: (i) the databases with the information about the gene functions (e.g., Gene Ontology [104], GeneCards [105]) or pathways in which they are present (e.g., KEGG [106], Reactome [107], WikiPathways [108]); and (ii) the methodology used to find enriched pathways from the DE genes of the study.

There are three types of PEA according to the methodology used to find altered pathways: Over-Representation Analysis (ORA), Functional Class Scoring (FCS) and Pathway Topology (PT). They are also known as first, second and third generation approaches respectively [103]. In an ORA approach, a pathway is considered DE when more DE genes belong to it than expected compared to all the genes in the study. The main benefit of this method is its simplicity, but it treats each gene equally regardless of their expression, uses only a portion of the data (DE genes), and does not account for gene interactions [109]. On the other hand, FCS methods use a gene-level statistic to compute a pathway weight and thus find DE pathways. With FCS methods all data available is considered, and the expression of each gene is used to identify DE pathways. As a drawback, FCS also does not consider genetic interactions. Finally, PT methods address the ORA and FCS limitations, being similar to FCS but using the topology of the gene interactions to compute the gene-level statistics, and usually having a great performance when pathways do not overlap [110,111].

As Table 5 shows, 14 of the reviewed studies (70%) applied a PEA [21,22,24–29,33–35,37,38] using a total of six different software tools, namely DAVID [112], DIANA miRPath [96], IPA [95], JNLP-GSEA [113], miRWalk [99] and Partek Genomic Suite [114]. ORA was the most common approach for PEA, having been used in 13 of the 20 articles [21,22,24,25,27–29,33–38]. Only one article used an FCS method [27] and neither used a PT approach.

DAVID was the most frequent PEA tool (Table 5), having been employed in five of the 20 studies (30%) [22,28,34,36,37]. DAVID is an ORA tool for determining functionally essential genes in a gene list through the identification of the enriched functions and biological processes in which DE genes are involved. The remaining articles used DIANA miRPath [21,29,33,35], IPA [38], JNLP-GSEA [27], miRWalk [25] and Partek Genomic Suite [26].

Regarding the databases used for PEA (Table 5), five were used in the literature, namely KEGG, Gene Ontology (GO), Simons Foundation Autism Database (AutDB) [115], Genomatix [116] and the reference sets of IPA. As Table 5 shows, KEGG and Gene Ontology (GO) were the most used databases. It is worth mentioning that Hicks et al. [22] also compared the miRNA targets with the AutDB in order to find genes presumably involved in autism, and that Weisz et al. [38] only selected the pathways related to central nervous signaling or immune system

Table 5
Methods, software and databases used for pathway enrichment analysis.

Study	Method	Software	Database
Hicks et al., 2016 [22]	ORA	DAVID –	- AutDB
Pfau et al., 2016 [24]	ORA	IPA, DAVID	GO terms IPA reference set
Martin et al., 2017 [25]	ORA	miRWalk	–
Rani et al., 2017 [21]	ORA	DIANA miRPath	GO terms
Wang et al., 2018 [26]	ORA	Partek-Genomic-Suite	–
Sillivan et al., 2019 [29]	ORA	DIANA miRPath	–
Hu et al., 2019 [27]	FCS ORA	JNLP-GSEA	Genomatix database
Mavrikaki et al., 2019 [33]	ORA	DIANA miRPath	–
Song et al., 2019 [34]	ORA	DAVID	GO terms KEGG
Lee et al., 2020 [28]	ORA	DAVID	KEGG
Wu et al., 2020 [36]	ORA	DAVID	KEGG
Nie et al., 2020 [35]	ORA	DIANA-miRPath	KEGG
Ma et al., 2020 [37]	ORA	DAVID	GO terms KEGG
Weisz et al., 2020 [38]	ORA	IPA	IPA reference set

*ORA: Over-Representation methods, FCS: Functional Class Scoring methods.

after performing the analysis.

3.4.3. Network analysis

A large amount of information is collected in miRNA-seq studies. Along with the samples, not only is miRNA expression gathered, but also information about other transcripts such as mRNA or long non coding RNA (lncRNA), medication, sex and other secondary conditions. The process of integrating this information and constructing a network of clusters and interactions is known as network or integrative analysis.

Four of the reviewed studies (20%) performed a network analysis [24,27,32,34,37], all of them using Cytoscape [117]. Cytoscape is an open-source platform for building and analyzing complex networks that allows linking the resulting network with databases of functional annotation.

Two studies sequenced miRNAs and mRNAs and used the network analysis to integrate the differential expression of both [24,32]. As repressors, miRNAs are usually negatively correlated with mRNA expression; by analyzing the simultaneous changes between both, they explore the potential interactions between the DE miRNAs and their mRNA targets.

On the other hand, Ma et al. [37] applied the network analysis to integrate information of the DE miRNA, DE mRNA and DE lncRNA expression to construct a competing endogenous RNA (ceRNA) network. Finally, Song et al. [34] built a network focused on the co-expression patterns of the DE miRNAs to identify those with common biological functions.

4. Discussion

MiRNA-seq is a valuable approach to address the study of neuropsychiatric diseases and gain insights into their development and course. However, it is still a new, complex field that requires the integration of many disciplines, relying heavily on the use of bioinformatics tools for data processing, interpretation and analysis. A large number of different bioinformatics programs and protocols for miRNA-seq (and other RNA-Seq technologies) data analysis have been developed in the last years [16,118,119]. For this reason, different online repositories of miRNA-seq tools such as miRandb [120], Tools4miRs [121], and miR-ToolsGallery [122], have been also developed to help users discover and select the right tools. In this scenario, it can be cumbersome for new researchers to design an appropriate workflow to conduct a miRNA-seq study and, therefore, its results and conclusions may have limited reproducibility.

After reviewing 20 miRNA-seq studies focused on neuropsychiatric diseases, we achieved a general depiction of main interests and current methodologies applied in this field. We categorized them into one of the three main objectives (diagnosis, mechanism, and prognosis) and analyzed each study in detail to capture the bioinformatics workflows used. As noted by Simoneau et al. [15], we found that the description of the bioinformatics methods is often incompletely reported in some articles, with some steps much better described than others and important omissions made. Nevertheless, we were able to identify the main workflows, creating one figure with the workflow of each study that can be found in the Supplementary Material 1. In addition, the Supplementary Material 1 also includes the typical workflows for the diagnosis and mechanism objectives. Fig. 1 shows the general miRNA-seq data analysis in neuropsychiatric studies inferred from the reviewed literature, which has similarities with analyses applied to a conventional NGS assessment of miRNA-seq data in other areas. All this information may help researchers in neuropsychiatric diseases in designing a workflow for the analysis of miRNA-seq data and selecting the tools to use in each specific step.

As Fig. 1 shows, the preprocessing and expression analysis steps are common to all the studies. In addition, a significant number of studies include the target prediction and enrichment analysis steps. Despite the high degree of similarity between the workflows used for each objective, there are notable differences in post-DEA analysis. Studies with diagnosis objectives usually build predictive models to test the specificity and sensitivity of the DE miRNAs (Fig. 1, number 3), and usually explore the influence of several variables on the miRNA expression of human samples. On the other hand, those classified in the mechanism category have a tendency to integrate miRNA and mRNA data on a network analysis (Fig. 1, number 3'), and often compare the results from animal models with previous data from human studies. Lastly, the only study classified on the category prognosis used the DE miRNAs between different stages of psychosis risk to build a classification model and then used a miRNA-miRNA correlation network to study the degree of co-regulation within groups.

During the elaboration of this review, we also identified several complexities that are worth comment. As shown here, workflows for miRNA-seq data analysis are, in general, diverse and complex, with many sequential steps and many choices for each of them. In this sense, a recent study by Botvinik-Nezer et al. conducted in 2020 indicated that the analytical flexibility produced by complex workflows (many steps, many choices and each step) across studies was related with substantial effects of scientific conclusions [123]. This study also emphasizes the need to validate and share analysis workflows and the need to perform and report multiple analyses of the same data to address the results variability. As shown in the previous section, the number of different tools used for the same analysis steps is very high. For instance, up to seven different tools were used for the differential expression analysis step, with DESeq/DESeq2 being the most popular choice (used in nine of 20 studies). Alignment, filtering low-expression genes, and target prediction were the steps most subject to methodological variations as can be seen in Table 4 and Supplementary Tables 3 and 5. Four of the studies included in this review [29,32,35,38] used more than one software to perform the same analysis and try to diminish the results variability. In addition, and regarding the preprocessing steps, a paper from Tam et al. [17] made practical recommendations on the most suitable preprocessing methods for the extraction and interpretation of miRNA count data that should be taken into account when developing new pipelines.

Once the analysis workflow is established and the corresponding tools are selected, another technical complexity is managing its execution. To overcome this issue, workflow management systems (WMS) are typically used, although they require advanced bioinformatics skills and may require a significant amount of development time and effort.

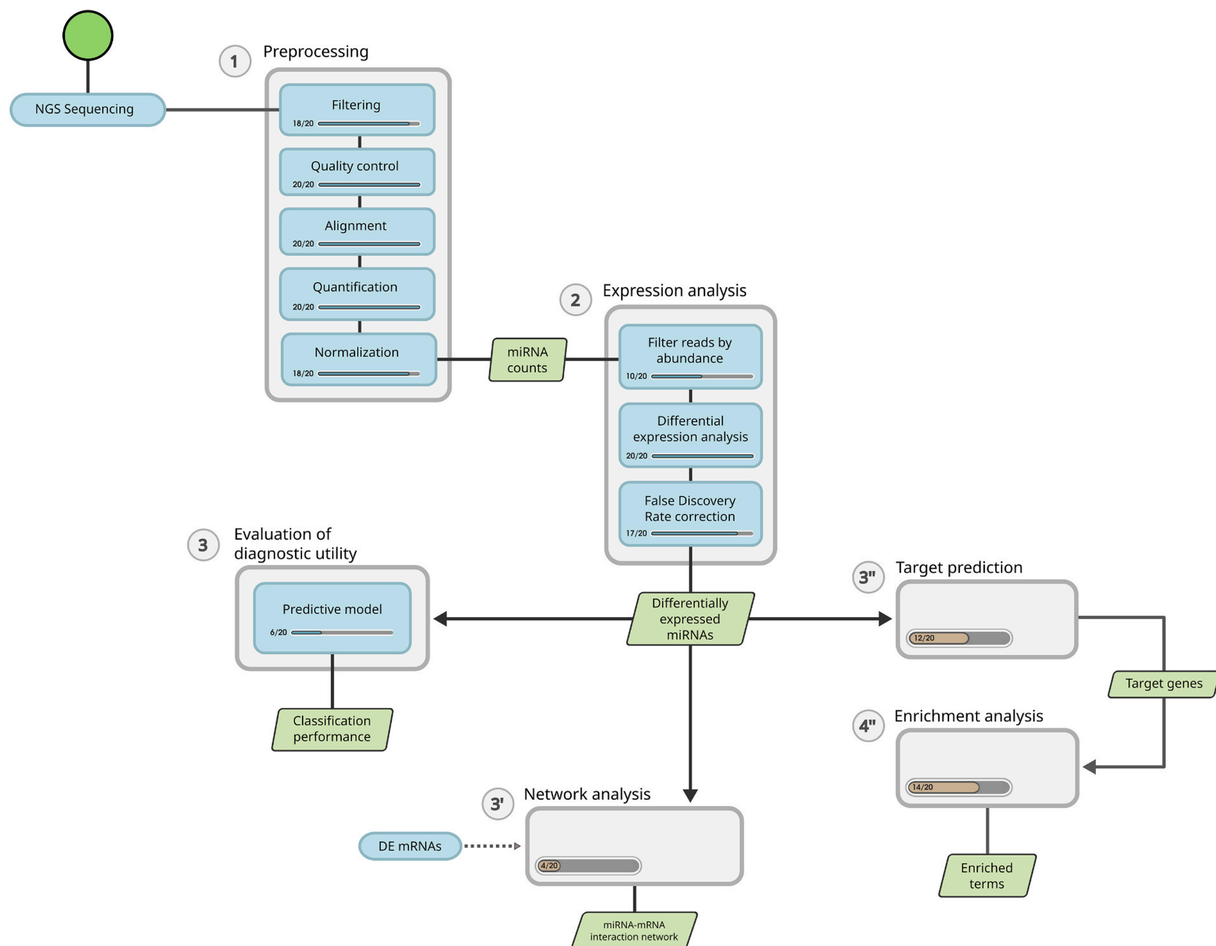


Fig. 1. Outline of the general bioinformatics workflow for a miRNA-seq study of a neuropsychiatric disease. The frequency of each process is indicated with a progress bar and the alternative steps are denoted with quotation marks.

Consequently, this is not always a feasible choice for many research groups. Given these facts, integrated tools have been developed to support typical analysis workflows, including miARma-Seq [124], a tool that includes almost all steps identified in the reviewed literature. None of the articles reviewed used such integrated software implementing a miRNA-seq workflow and do not even mention the use of a WMS to set up and execute their own analyses. The use of a WMS is recommended and a good practice providing several benefits, making the computational methods more maintainable, reproducible, and shareable [125]. Additionally, the use of a WMS would allow changing the software used in some steps more easily (e.g., changing the alignment tool) or using different tools and integrating their results (e.g., perform target prediction with different databases).

Finally, another sensitive issue found in the reviewed studies is a general use of outdated tools and databases, most notably in the alignment and enrichment analysis steps. Regarding the former, nine studies used a deprecated version of a genome for alignment Supplementary Table 3. This is probably due to the exclusive use of annotations from versions of miRBase prior to 21, where the annotations of the hg38 genome were included. Furthermore, it is easier to make comparisons with data from previous studies based on hg19 by using a deprecated version of the genome rather than reassigning the old coordinates. Regarding the latter, four studies conducted between 2019 and 2020 used DAVID, which had been outdated since October 2016. This may be the result of the variety and complexity of the analyses and tools for pathway analysis; there are many methodologies and software with ambiguous names and a lack of literature and consensus regarding procedures. In our opinion, this translates into a reduced understanding of the processes for pathway analysis, and a search for clear and simple solutions like those offered by DAVID or the ORA analyses.

Aside from the technical and methodological issues discussed so far, another important limitation of most studies reviewed is the small sample sizes used. Having an appropriate sample size is critical to conducting experiments with enough power to detect true effects; using small sample sizes can increase false positive results and produce augmented, unreliable effect sizes [126]. This issue is present in most of the articles reviewed and, in fact, most of them acknowledge it. As shown in Tables 1 and 2, most of the studies only collected a few tens of samples for their experiments. Such low sample sizes can hinder both the identification of true differentially expressed miRNAs and the creation of accurate predictive models.

In summary, many studies conducted over the last few years have combined microRNA-seq technologies with a large variety of experimental and bioinformatics approaches to identify and characterize microRNAs (key regulators of transcriptome plasticity) involved in several neuropsychiatric disorders. In such a context, microRNAs have emerged as potential candidates to be used as biomarkers (e.g.: disease prediction, diagnosis, prognosis and therapeutic response) and/or therapeutic targets to treat a large number of diseases, including brain diseases. Nevertheless, the methodological miRNA-seq perspective on the neuropsychiatric field remains somewhat generic. In the near future, the combination of methodologies, such as those shown in this review, together with predictive analysis (i.e. predictive modelling, machine learning) [127] may contribute to develop specific analytical approaches in the field of neuropsychiatry, thus improving medical care refining precision medicine (or personalized medicine). Future research

should be oriented in that direction.

Funding

This work was supported by Instituto de Salud Carlos III through the project PI18/01311 (co-funded by European Regional Development Fund, "A way to make Europe") and by a Ramon & Cajal grant [RYC2014-15246] to RCA-B. This work was partially supported by the Consellería de Educación, Universidades e Formación Profesional (Xunta de Galicia) under the scope of the strategic funding ED431C2018/55-GRC Competitive Reference Group. National funding by FCT, Foundation for Science and Technology, through the individual scientific employment program-contract with Hugo López-Fernández.

Authors' contributions

DPR, HLF and RCAB conceived the review. DPR performed the bibliographic searches, screened the literature, selected the final studies for the review, and designed and prepared the figures. DPR analyzed each article with the help of HLF. DPR and HLF drafted the manuscript with the help of RCAB. HLF and RCAB revised the manuscript. All authors read and approved the final manuscript.

Declaration of competing interest

None declared.

Acknowledgements

The authors would like to thank Galicia Sur Health Research Institute, Galicia Sur Biomedical Foundation, and the Area Sanitaria de Vigo for their support. We specially thank the Psychiatric Nursing Service and Psychiatrists at the Álvaro Cunqueiro Hospital and Nicolás Peña Hospital, and the Intensive Care Unit at the Álvaro Cunqueiro Hospital.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2021.104603.

References

[1] A.M.M.T. Reza, Y.-G. Yuan, microRNAs mediated regulation of the ribosomal proteins and its consequences on the global translation of proteins, Cells 10 (2021).
[2] E. Salta, B. De Strooper, Non-coding RNAs with essential roles in neurodegenerative disorders, Lancet Neurol. 11 (2012) 189–200.
[3] M. Esteller, Non-coding RNAs in human disease, Nat. Rev. Genet. 12 (2011) 861–874.
[4] J. Krol, I. Loedige, W. Filipowicz, The widespread regulation of microRNA biogenesis, function and decay, Nat. Rev. Genet. 11 (2010) 597–610.
[5] A. Zovoilis, H.Y. Agbemenyah, R.C. Agis-Balboa, et al., microRNA-34c is a novel target to treat dementias, EMBO J. 30 (2011) 4299–4308.
[6] X. You, Y. Zhang, Q. Long, et al., Investigating aberrantly expressed microRNAs in peripheral blood mononuclear cells from patients with treatment-resistant schizophrenia using miRNA sequencing and integrated bioinformatics, Mol. Med. Rep. 22 (2020) 4340–4350.
[7] A. Gibbons, S. Sundram, B. Dean, Changes in non-coding RNA in depression and bipolar disorder: can they Be used as diagnostic or theranostic biomarkers? Non-Coding RNA 6 (2020) 33.
[8] G.R. Fries, W. Zhang, D. Benevenuto, et al., MicroRNAs in major depressive disorder, Reviews on Biomarker Studies in Psychiatric and Neurodegenerative Disorders (2019) 175–190.
[9] Y. Yoshino, Y. Dwivedi, Non-coding RNAs in psychiatric disorders and suicidal behavior, Front. Psychiatr. 11 (2020).
[10] B. Alural, S. Genc, S.J. Haggarty, Diagnostic and therapeutic potential of microRNAs in neuropsychiatric disorders: past, present, and future, Prog. Neuro-Psychopharmacol. Biol. Psychiatry 73 (2017) 87–103.
[11] L. Wang, L. Zhang, Circulating exosomal miRNA as diagnostic biomarkers of neurodegenerative diseases, Front. Mol. Neurosci. 13 (2020).
[12] B. Roy, Y. Yoshino, L. Allen, et al., Exploiting circulating MicroRNAs as biomarkers in psychiatric disorders, Mol. Diagn. Ther. 24 (2020) 279–298.
[13] M.M.J. van den Berg, J. Krauskopf, J.G. Ramaekers, et al., Circulating microRNAs as potential biomarkers for psychiatric and neurodegenerative disorders, Prog. Neurobiol. 185 (2020), 101732.
[14] L. Peixoto, D. Risso, S.G. Poplawski, et al., How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets, Nucleic Acids Res. 43 (2015) 7664–7674.
[15] J. Simoneau, S. Dumontier, R. Gosselin, et al., Current RNA-seq methodology reporting limits reproducibility, Briefings Bioinf. 22 (2021) 140–145.
[16] L. Chen, L. Heikkinen, C. Wang, et al., Trends in the development of miRNA bioinformatics tools, Briefings Bioinf. 20 (2019) 1836–1852.
[17] S. Tam, M.-S. Tsao, J.D. McPherson, Optimization of miRNA-seq data preprocessing, Briefings Bioinf. 16 (2015) 950–963.
[18] X. Zhou, A. Oshlack, M.D. Robinson, miRNA-Seq normalization comparisons need improvement, RNA 19 (2013) 733–734.
[19] A.G. Hoss, A. Labadorf, T.G. Beach, et al., microRNA profiles in Parkinson's disease prefrontal cortex, Front. Aging Neurosci. 8 (2016).
[20] A. Gámez-Valero, J. Campdelacreu, D. Vilas, et al., Exploratory study on microRNA profiles from plasma-derived extracellular vesicles in Alzheimer's disease and dementia with Lewy bodies, Transl. Neurodegener. 8 (2019) 1–17.
[21] A. Rani, A. O'Shea, L. Ianov, et al., miRNA in circulating microvesicles as biomarkers for age-related cognitive decline, Front. Aging Neurosci. 9 (2017).
[22] S.D. Hicks, C. Ignacio, K. Gentile, et al., Salivary miRNA profiles identify children with autism spectrum disorder, correlate with adaptive behavior, and implicate ASD candidate genes involved in neurodevelopment, BMC Pediatr. 16 (2016).
[23] C.D. Jeffries, D.O. Perkins, S.D. Chandler, et al., Insights into psychosis risk from leukocyte microRNA expression, Transl. Psychiatry 6 (2016).
[24] M.L. Pfau, I. Purushothaman, J. Feng, et al., Integrative analysis of sex-specific microRNA networks following stress in mouse nucleus accumbens, Front. Mol. Neurosci. 9 (2016).
[25] C.G. Martin, H. Kim, S. Yun, et al., Circulating miRNA associated with posttraumatic stress disorder in a cohort of military combat veterans, Psychiatr. Res. 251 (2017) 261–265.
[26] L.J. Wang, S.C. Li, M.J. Lee, et al., Blood-bourne microRNA biomarker evaluation in attention-deficit/hyperactivity disorder of han Chinese individuals: an exploratory study, Front. Psychiatr. 9 (2018).
[27] Z. Hu, S.S.S. Gao, D. Lindberg, et al., Temporal dynamics of miRNAs in human DLPFC and its association with miRNA dysregulation in schizophrenia, Transl. Psychiatry 9 (2019) 237.
[28] S.Y. Lee, R.B. Lu, L.J. Wang, et al., Serum miRNA as a possible biomarker in the diagnosis of bipolar II disorder, Sci. Rep. 10 (2020) 1131.
[29] S.E. Sillivan, S. Jamieson, L. de Nijs, et al., MicroRNA regulation of persistent stress-enhanced memory, Mol. Psychiatr. (2019).
[30] N. Hara, M. Kikuchi, A. Miyashita, et al., Serum microRNA miR-501-3p as a potential biomarker related to the progression of Alzheimer's disease, Acta neuropathologica communications 5 (2017) 10.
[31] J. Faraji, N. Soltanpour, M. Ambeskovic, et al., Evidence for ancestral programming of resilience in a two-hit stress model, Front. Behav. Neurosci. 11 (2017).
[32] Y. Si, Z. Song, X. Sun, et al., microRNA and mRNA profiles in nucleus accumbens underlying depression versus resilience in response to chronic stress, Am. J. Med. Genet., Part B: Neuropsychiatric Genetics 177 (2018) 563–579.
[33] M. Mavrikaki, L. Pantano, D. Potter, et al., Sex-dependent changes in miRNA expression in the bed nucleus of the stria terminalis following stress, Front. Mol. Neurosci. 12 (2019).
[34] W. Song, Y. Shen, Y. Zhang, et al., Expression alteration of microRNAs in Nucleus Accumbens is associated with chronic stress and antidepressant treatment in rats, BMC Med. Inf. Decis. Making 19 (2019).
[35] C. Nie, Y. Sun, H. Zhen, et al., Differential expression of plasma exo-miRNA in neurodegenerative diseases by next-generation sequencing, Front. Neurosci. 14 (2020).
[36] H.Z.Y. Wu, A. Thalamuthu, L. Cheng, et al., Differential blood miRNA expression in brain amyloid imaging-defined Alzheimer's disease and controls, Alzheimer's Res. Ther. 12 (2020).
[37] N. Ma, C. Tie, B. Yu, et al., Identifying lncRNA–miRNA–mRNA networks to investigate Alzheimer's disease pathogenesis and therapy strategy, Aging (Albany NY) 12 (2020) 2897–2920.
[38] H.A. Weisz, D. Kennedy, S. Widen, et al., MicroRNA sequencing of rat hippocampus and human biofluids identifies acute, chronic, focal and diffuse traumatic brain injuries, Sci. Rep. 10 (2020).
[39] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.journal 17 (2011) 10–12.
[40] S. Andrews, FASTQC. A Quality Control Tool for High Throughput Sequence Data, 2010.
[41] Torrent Suite Software for Sequencing Data Analysis.
[42] C.-T. Pan, K.-W. Tsai, T.-M. Hung, et al., miRSeq: a user-friendly standalone toolkit for sequencing quality evaluation and miRNA profiling, BioMed Res. Int. 2014 (2014) 1–8.
[43] S. Tarazona, P. Furió-Tarí, D. Turrà, et al., Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package, Nucleic Acids Res. 43 (2015) e140.
[44] L. Yi, L. Liu, P. Melsted, et al., A direct comparison of genome alignment and transcriptome pseudoalignment, bioRxiv (2018), 444620.
[45] M. Ziemann, A. Kaspi, A. El-Osta, Evaluation of microRNA alignment techniques, RNA 22 (2016) 1120–1138.

[46] B. Langmead, C. Trapnell, M. Pop, et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.
[47] C. Trapnell, L. Pachter, S.L. Salzberg, TopHat: discovering splice junctions with RNA-Seq, Bioinformatics 25 (2009) 1105–1111.
[48] B. Langmead, M.C. Schatz, J. Lin, et al., Searching for SNPs with cloud computing, Genome Biol. 10 (2009) R134.
[49] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359.
[50] CLC Genomics Workbench. Bioinformatics Software and Services | QIAGEN Digital Insights.
[51] M.R. Friedländer, S.D. Mackowiak, N. Li, et al., miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, Nucleic Acids Res. 40 (2012) 37–52.
[52] A.-K. Emde, M. Grunert, D. Weese, et al., MicroRazerS: rapid alignment of small RNA reads, Bioinformatics 26 (2010) 123–124.
[53] T. Madden, The BLAST Sequence Analysis Tool, 2003. The NCBI Handbook [Internet].
[54] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26 (2010) 589–595.
[55] L. Pantano, X. Estivill, E. Martí, SeqBuster, a bioinformatic tool for the processing and analysis of small RNAs datasets, reveals ubiquitous miRNA modifications in human embryonic cells, Nucleic Acids Res. 38 (2010) e34.
[56] A. Dobin, C.A. Davis, F. Schlesinger, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21.
[57] A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function, Nucleic Acids Res. 47 (2019) D155–D162.
[58] B. Fromm, D. Domanska, E. Høye, et al., MirGeneDB 2.0: the metazoan microRNA complement, Nucleic Acids Res. 48 (2020) D132–D141.
[59] K. Clark, I. Karsch-Mizrachi, D.J. Lipman, et al., GenBank. Nucleic Acids Res 44 (2016) D67–D72.
[60] I. Kalvari, E.P. Nawrocki, J. Argasinska, et al., Non-coding RNA analysis using the Rfam database, Curr Protoc Bioinformatics 62 (2018) e51.
[61] M. Szymanski, V.A. Erdmann, J. Barciszewski, Noncoding RNAs database (ncRNAdb), Nucleic Acids Res. 35 (2007) D162–D164.
[62] M. Wen, Y. Shen, S. Shi, et al., miREvo: an integrative microRNA evolutionary analysis platform for next-generation sequencing experiments, BMC Bioinf. 13 (2012) 140.
[63] Illumina BaseSpace Apps..
[64] A. Gordon, agordon/fastx_toolkit (2020).
[65] S. Anders, P.T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169.
[66] X. Wu, T.-K. Kim, D. Baxter, et al., sRNAnalyzer-a flexible and customizable small RNA sequencing data analysis pipeline, Nucleic Acids Res. 45 (2017) 12140–12151.
[67] A. Oshlack, M.J. Wakefield, Transcript length bias in RNA-seq data confounds systems biology, Biol. Direct 4 (2009) 14.
[68] D. Risso, K. Schwartz, G. Sherlock, et al., GC-content normalization for RNA-seq data, BMC Bioinf. 12 (2011) 480.
[69] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, Genome Biol. 11 (2010) R25.
[70] C. Evans, J. Hardin, D.M. Stoebel, Selecting between-sample RNA-Seq normalization methods from the perspective of their assumptions, Briefings Bioinf. 19 (2018) 776–792.
[71] M.-A. Dillies, A. Rau, J. Aubert, et al., A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis, Briefings Bioinf. 14 (2013) 671–683.
[72] L. Wang, Z. Feng, X. Wang, et al., DEGseq: an R package for identifying differentially expressed genes from RNA-seq data, Bioinformatics 26 (2010) 136–138.
[73] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550.
[74] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.
[75] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data, Bioinformatics 26 (2010) 139–140.
[76] L. Sun, H. Luo, D. Bu, et al., Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts, Nucleic Acids Res. 41 (2013) e166, e166.
[77] Mengyuqi. Mengyuqi/cpc, 2020.
[78] J. Mistry, A. Bateman, R.D. Finn, Predicting active site residue annotations in the Pfam database, BMC Bioinf. 8 (2007) 298.
[79] M.F. Lin, I. Jungreis, M. Kellis, PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions, Bioinformatics 27 (2011) i275–i282.
[80] Ying Sha, J.H. Phan, M.D. Wang, Effect of low-expression gene filtering on detection of differentially expressed genes in RNA-seq data, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2015, pp. 6461–6464.
[81] R. Bourgon, R. Gentleman, W. Huber, Independent filtering increases detection power for high-throughput experiments, Proc. Natl. Acad. Sci. Unit. States Am. 107 (2010) 9546–9551.
[82] Y. Nguyen, D. Nettleton, H. Liu, et al., Detecting differentially expressed genes with RNA-seq data using backward selection to account for the effects of relevant covariates, J. Agric. Biol. Environ. Stat. 20 (2015) 577–597.
[83] C. Soneson, M. Delorenzi, A comparison of methods for differential expression analysis of RNA-seq data, BMC Bioinf. 14 (2013) 91.
[84] T. Wang, B. Li, C.E. Nelson, et al., Comparative analysis of differential gene expression analysis tools for single-cell RNA sequencing data, BMC Bioinf. 20 (2019) 40.
[85] B. Lin, Z. Pang, Stability of methods for differential expression analysis of RNA-seq data, BMC Genom. 20 (2019) 35.
[86] J. Costa-Silva, D. Domingues, F.M. Lopes, RNA-Seq differential expression analysis: an extended review and a software tool, PloS One 12 (2017).
[87] J.T. Leek, W.E. Johnson, H.S. Parker, et al., The sva package for removing batch effects and other unwanted variation in high-throughput experiments, Bioinformatics 28 (2012) 882–883.
[88] M.E. Ritchie, B. Phipson, D. Wu, et al., Limma powers differential expression analyses for RNA-sequencing and microarray studies, Nucleic Acids Res. 43 (2015) e47, e47.
[89] C. Trapnell, D.G. Hendrickson, M. Sauvageau, et al., Differential analysis of gene regulation at transcript resolution with RNA-seq, Nat. Biotechnol. 31 (2013) 46–53.
[90] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. Roy. Stat. Soc. B 57 (1995) 289–300.
[91] K. Korthauer, P.K. Kimes, C. Duvallet, et al., A practical guide to methods controlling false discoveries in computational biology, Genome Biol. 20 (2019) 118.
[92] Y. Li, K.V. Kowdley, MicroRNAs in common human diseases, Dev. Reprod. Biol. 10 (2012) 246–253.
[93] B. Xu, P.-K. Hsu, M. Karayiorgou, et al., MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction, Neurobiol. Dis. 46 (2012) 291–301.
[94] S.M. Peterson, J.A. Thompson, M.L. Ufkin, et al., Common features of microRNA target prediction tools, Front. Genet. 5 (2014).
[95] Ingenuity Pathway Analysis (IPA)..
[96] I.S. Vlachos, K. Zagganas, M.D. Paraskevopoulou, et al., DIANA-miRPath v3.0: deciphering microRNA function with experimental support, Nucleic Acids Res. 43 (2015) W460–W466.
[97] D. Betel, A. Koppal, P. Agius, et al., Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites, Genome Biol. 11 (2010) R90.
[98] Y. Chen, X. Wang, miRDB: an online database for prediction of functional microRNA targets, Nucleic Acids Res. 48 (2020) D127–D131.
[99] H. Dweep, C. Sticht, P. Pandey, et al., miRWalk–database: prediction of possible miRNA binding sites by 'walking' the genes of three genomes, J. Biomed. Inf. 44 (2011) 839–847.
[100] M. Rehmsmeier, P. Steffen, M. Höchsmann, et al., Fast and effective prediction of microRNA/target duplexes, RNA 10 (2004) 1507–1517.
[101] D. Karagkouni, M.D. Paraskevopoulou, S. Chatzopoulos, et al., DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA–gene interactions, Nucleic Acids Res. 46 (2018) D239–D245.
[102] V. Agarwal, G.W. Bell, J.-W. Nam, et al., Predicting effective microRNA target sites in mammalian mRNAs, eLife 4 (2015) e05005.
[103] P. Khatri, M. Sirota, A.J. Butte, Ten years of pathway analysis: current approaches and outstanding challenges, PLoS Comput. Biol. 8 (2012) e1002375–e1002375.
[104] M. Ashburner, C.A. Ball, J.A. Blake, et al., Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[105] M. Rebhan, V. Chalifa-Caspi, J. Prilusky, et al., GeneCards: integrating information about genes, proteins and diseases, Trends Genet. 13 (1997) 163.
[106] M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes, Nucleic Acids Res. 28 (2000) 27–30.
[107] B. Jassal, L. Matthews, G. Viteri, et al., The reactome pathway knowledgebase, Nucleic Acids Res. 48 (2020) D498–D503.
[108] M. Martens, A. Ammar, A. Riutta, et al., WikiPathways: connecting communities, Nucleic Acids Res. 49 (2021) D613–D621.
[109] J.-H. Hung, Gene set/pathway enrichment analysis, Data Mining for Systems Biology: Methods and Protocols (2013) 201–213.
[110] J. Ma, A. Shojaie, G. Michailidis, A comparative study of topology-based pathway enrichment analysis methods, BMC Bioinf. 20 (2019) 546.
[111] M. Bayerlová, K. Jung, F. Kramer, et al., Comparative study on gene set and pathway topology-based enrichment methods, BMC Bioinf. 16 (2015) 334.
[112] G. Dennis, B.T. Sherman, D.A. Hosack, et al., DAVID: database for annotation, visualization, and integrated discovery, Genome Biol. 4 (2003) R60.
[113] A. Subramanian, P. Tamayo, V.K. Mootha, et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 15545–15550.
[114] Gosche K. Partek Genomics Suite Analysis Software - NGS, (Microarray, & qPCR. Partek Inc).
[115] S.N. Basu, R. Kollu, S. Banerjee-Basu, AutDB: a gene reference resource for autism research, Nucleic Acids Res. 37 (2009) D832–D836.
[116] Genomatix Genome Analyzer - Our Integrated Solution for the Interpretation of Biological Data..
[117] P. Shannon, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504.
[118] F. Shaker, A. Nikravesh, R. Arezumand, et al., Web-based tools for miRNA studies analysis, Comput. Biol. Med. 127 (2020), 104060.
[119] V. Shukla, V.K. Varghese, S.P. Kabekkodu, et al., A compilation of Web-based research tools for miRNA analysis, Briefings in Functional Genomics 16 (2017) 249–273.

[120] S.H. Aghaee-Bakhtiari, E. Arefian, Lau P. miRandb, A resource of online services for miRNA research, Briefings Bioinf. (2017) bbw109.
[121] A. Lukasik, M. Wójcikowski, P. Zielenkiewicz, Tools4miRs – one place to gather all the tools for miRNA analysis, Bioinformatics 32 (2016) 2722–2724.
[122] L. Chen, L. Heikkinen, C. Wang, et al., miRToolsGallery: a tag-based and rankable microRNA bioinformatics resources database portal, Database 2018 (2018).
[123] R. Botvinik-Nezer, F. Holzmeister, C.F. Camerer, et al., Variability in the analysis of a single neuroimaging dataset by many teams, Nature 582 (2020) 84–88.
[124] E. Andrés-León, R. Núñez-Torres, A.M. Rojas, miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis, Sci. Rep. 6 (2016), 25749.
[125] J.M. Perkel, Workflow systems turn raw data into scientific knowledge, Nature 573 (2019) 149–150.
[126] M.G.M. Kok, de Ronde MWJ, P.D. Moerland, et al., Small sample sizes in high-throughput miRNA screens: a common pitfall for the identification of miRNA biomarkers, Biomolecular Detection and Quantification 15 (2018) 1–5.
[127] R.B. Parikh, Z. Obermeyer, A.S. Navathe, Regulation of predictive analytics in medicine, Science 363 (2019) 810–812.

Daniel Pérez-Rodríguez is a PhD Student of the Instituto de Investigación Sanitaria Galicia Sur (IISGS), Universidade de Vigo. His research interests include the development of bioinformatics workflows to study neuropsychiatric diseases.

Hugo López-Fernández is a Junior Researcher at Instituto de Investigação e Inovação em Saúde (i3S; Universidade do Porto). His research interests include the development of complex bioinformatics workflows.

Roberto C. Agís-Balboa is a Ramon y Cajal fellow at the Servicio Gallego de Salud (SERGAS, IISGS). He studies the epigenetic mechanisms and searches for biomarkers associated with neuropsychiatric diseases.