REVIEW

Challenges in microRNAs' targetome prediction and validation

Jesus Eduardo Rojo Arias, Volker Busskamp*
Center for Regenerative Therapies (CRTD), Technische Universität Dresden, Dresden, Germany

Funding: *VB was supported by a Volkswagen Foundation Freigeist fellowship (A110720) and by an European Research Council starting grant (678071-ProNeurons). JERA was supported by the Dresden International Graduate School for Biomedicine and Bioengineering (DIGS-BB) program.*

Abstract

MicroRNAs (miRNAs) are small RNA molecules with important roles in post-transcriptional regulation of gene expression. In recent years, the predicted number of miRNAs has skyrocketed, largely as a consequence of high-throughput sequencing technologies becoming ubiquitous. This dramatic increase in miRNA candidates poses multiple challenges in terms of data deposition, curation, and validation. Although multiple databases containing miRNA annotations and targets have been developed, ensuring data quality by validating miRNA-target interactions requires the efforts of the research community. In order to generate databases containing biologically active miRNAs, it is imperative to overcome a multitude of hurdles, including restricted miRNA expression patterns, distinct miRNA biogenesis machineries, and divergent miRNA-mRNA interaction dynamics. In the present review, we discuss recent advances and limitations in miRNA prediction, identification, and validation. Lastly, we focus on the most enriched neuronal miRNA, miR-124, and its gene regulatory network in human neurons, which has been revealed using a combined computational and experimental approach.

Key Words: *miRNAs; miRNA regulation; miR-124; wTO analysis; miRNA biogenesis; miRNA prediction; miRNA identification; miRNA validation*

***Correspondence to:**
Volker Busskamp,
volker.busskamp@tu-dresden.de.

orcid:
0000-0001-7517-8944
(Volker Busskamp)

doi: *10.4103/1673-5374.257514*

Received: *December 19, 2019*
Accepted: *January 14, 2019*

The Canonical MicroRNA Processing Machinery

In recent years, the central role of microRNAs (miRNAs) as post-transcriptional repressors has been extensively studied in development and disease. In animals, miRNAs are non-coding 22 nucleotide (nt) sequences that regulate gene expression by binding either partially- or fully-complementary RNA sequences - predominantly messenger RNA (mRNA), but also but also ribosomal RNAs, transfer RNAs, long intergenic non-coding RNAs, and other miRNAs (Helwak et al., 2013). The main determinants of target recognition reside within the miRNA 5′ end in positions 2 to 7, which together are commonly known as the "seed" sequence (Ha and Kim, 2014). Computational predictions suggest that individual miRNAs regulate large genetic networks by targeting hundreds of mRNA molecules (Agarwal et al., 2015), although experimental data on entire networks within specific cell types are rare. Most experimental studies only validate one or a few miRNA targets at once. In turn, an individual mRNA molecule can possess binding sites for several miRNA species (Friedman et al., 2009; Chou et al., 2018), though the latter might compete against each other for specific binding sites (Saetrom et al., 2007) and thereby lead to non-synergistic outcomes (Xu et al., 2011). Additionally, the existence of competing endogenous RNAs (Tay et al., 2014; Denzler et al., 2016), *i.e.* natural miRNA sponges such as long non-coding and circular RNAs, is further evidence of the intricacy of miRNA regulatory networks.

The biogenesis and maturation of miRNAs is highly conserved across species. miRNAs are generated from partially-complementary regions of primary RNA transcripts (pri-miRNA) produced mainly by RNA polymerase II, but also by RNA polymerase III. These transcripts anneal and give rise to short RNA hairpins or stem-loops (Ha and Kim, 2014). Initially contained within non-coding RNAs or within the introns of coding sequences (host genes), these hairpins are recognized and cleaved in the mammalian cell nucleus by the miRNA-processing complex, also known as microprocessor, formed by the Drosha ribonuclease and the DiGeorge critical region 8 (DGCR8) protein. By cutting 11 base pairs (bp) away from the base of the hairpin stem, the miRNA-processing complex produces a 70-nt-long sequence known as a precursor miRNA (pre-miRNA) and characterized by a 5′ phosphate and a 2-nt 3′ overhang (Han et al., 2004). Pre-miRNA hairpins are subsequently transported by exportin 5 to the cytoplasm, where they are cleaved by the Dicer endoribonuclease 22 nt away from the Microprocessor cutting site. This second cut generates a new 5′ monophosphate and one more 2-nt 3′ overhang on the opposite side of the double-stranded RNA sequence (Chiang et al., 2010). The resulting ~22-nt RNA duplexes are incorporated into the Argonaute protein and unwound to form, together with other proteins, the RNA-induced silencing complex (RISC). RISC then mediates the miRNA-guided scanning of target transcripts and the subsequent miRNA-target pairing (Bartel, 2009; Ha and Kim, 2014). Ultimately, miRNA-targeted mRNA molecules are degraded or exhibit reduced translational efficiency (Krol et al., 2010).

Challenges in the Identification of Novel MicroRNAs

Although miRNA identification was initially achieved by Sanger sequencing of complementary DNA clones produced from small-RNA libraries, the rapid advance in sequencing technologies towards higher-throughput methods has led to an exponential increase in the number of predicted miRNAs (Ko-

zomara and Griffiths-Jones, 2011). RNA sequencing (RNA-Seq), in particular, has revolutionized the field of transcriptomics and spearheaded the discovery of all types of novel transcripts (Kukurba and Montgomery, 2015). Offering higher coverage and resolution than traditional Sanger- and microarray-based sequencing methods, RNA-Seq experiments have revealed many novel miRNAs (Kozomara and Griffiths-Jones, 2011): this has resulted in an enormous challenge in miRNA annotation and validation. The latter is further compounded by the usage of heterogeneous RNA-Seq methods and deep-sequencing parameters. Similarly, low RNA quality has a notoriously detrimental effect on sequencing results. Sequencing low-quality RNA has the potential to bias miRNA quantifications and might lead to assuming that RNA fragments resulting from degradation are real miRNAs (Kukurba and Montgomery, 2015; Ludwig et al., 2017). Indeed, hundreds of miRNAs can be erroneously annotated from a single, poorly-analyzed RNA-Seq dataset and decimate the fraction of legitimate miRNAs annotated in databases (Kozomara and Griffiths-Jones, 2014). In order to deal with these difficulties, guidelines for miRNA annotation have been devised and are continuously being refined.

The first set of guidelines for annotating miRNAs was published in 2003 (Ambros, 2003). These guidelines stated that cloning, sequencing, or northern blotting should be used to provide evidence for the existence of a 22-nt-long sequence, and that the predicted origin of such sequence should be a region with the potential to generate a stem-loop structure. These principles, however, are subject to certain limitations. For instance, millions of hairpin structures are predicted to exist in the genome, yet not all of them give rise to miRNAs (Kozomara and Griffiths-Jones, 2011). Hence, stem loops are necessary but insufficient to predict the existence of a "true" miRNA. Furthermore, even "true" miRNA candidates might escape detection in sequencing experiments if tissues or cells are examined at time-points in which their expression levels are low. For this reason, complementing criteria for the validation of animal miRNAs were established. These criteria are often incorporated in modern miRNA identification tools such as miRDeep (An et al., 2013) and include: a) the presence of matching 3′ overhangs up to 4 nt-long; b) 5′ ends compatible with the miRNA processing machinery; c) a maximum number of permitted mismatches between pre-miRNA strands; d) a minimum number of required mapped reads; and e) a specific pattern for such reads (Friedlander et al., 2008; Hendrix et al., 2010). With regard to criterion e), a high number of reads is expected to map to the mature miRNA, whereas much lower read numbers should map to its loop region and to the second strand of the miRNA duplex (miR*; also known as "passenger" strand). Additionally, reads mapping to both loop and miRNA (or to both loop and miR*) are expected to be almost undetectable as they contain the cutting sites for the miRNA processing machinery. Although these principles are certainly useful in aiming for high-confidence annotations, defining a minimum number of required mapped reads as a guideline could cause problems, particularly because some miRNA species are expressed only at low levels and in a temporally- and tissue-restricted manner (Karali et al., 2010). Numerous candidate miRNAs have been annotated in miRNA repositories even with a single mapped read from a deep-sequencing experiment as evidence (Berezikov et al., 2006). Since many of these miRNAs have not yet been validated, it is still possible that they are artifacts.

"True" MicroRNAs: Needles or Haystack?

Several of the hurdles inherent in the identification of novel miRNAs are evident in multiple miRNA databases. miRBase, for instance, represents the primary public repository and online resource for miRNA sequences and annotations, and describes 48,860 mature miRNAs from 271 organisms in its 22nd version (Kozomara et al., 2019). Nonetheless, many miRNA annotations in this and other databases correspond to predicted miRNAs and might thus be inaccurate. As an example, over 150 annotated miRNAs have failed to yield RNAs with miRNA features in response to the overexpression of their hairpin sequences flanked on both sides by the ~100-nt genomic sequence required for their correct processing (Chiang et al., 2010). Moreover, although the mapping patterns of over 5.5 billion reads gathered from almost 1500 small RNA-Seq datasets support the validity of the miRBase annotations to varying degrees (up to 65% depending on the species), over 200 miRNA sequences have been removed from this database in the past 5 years (Kozomara et al., 2019). In contrast, more than 10,000 miRNAs were added within the same timeframe.

For *homo sapiens*, in particular, the miRBase presently contains 2654 mature miRNA sequences predicted to originate from 1917 hairpin precursors. Yet, only approximately one quarter of these miRNAs are annotated with high confidence, and no more than 20 reads are associated with each strand of the RNA duplex for about half of them (Kozomara et al., 2019). However, validating all candidate miRNAs *via* overexpression assays (Chiang et al., 2010) is a titanic task likely to fail in identifying miRNAs produced *via* non-canonical pathways and hence not complying with the features that arise from conventional miRNA processing (Yang and Lai, 2011; Ha and Kim, 2014; Kozomara et al., 2019). Thus, improved annotation approaches, and the development and implementation of stringent validation pipelines are imperative for defining whether a miRNA is "true" or falsely-identified.

Experimental Approaches to Identify and Validate MicroRNAs

As most of the RNA species obtained from biological tissues or cells in culture correspond to ribosomal RNA, a variety of depletion and enrichment strategies have emerged over time to preferentially sequence RNAs of interest. A prominent strategy to enrich miRNAs from total RNA is to isolate small RNAs, *i.e.* species of 15–30 nt. This is commonly achieved either by excising gel fragments after electrophoresis in the region corresponding to the expected RNA fragment size, or by using silica columns that bind small RNAs (Kukurba and Montgomery, 2015). These approaches, however, are prone to miRNA misidentification in the case of low-quality or degraded RNA. Recently, a number of methods based on the crosslinking of miRNAs and their targets, together with immunoprecipitation of RISC proteins, have emerged for the validation of miRNAs and for the identification of miRNA-target interactions. High-throughput sequencing of RNAs isolated by crosslinking and immunoprecipitation of Argonaute, for instance, allows miRNAs contained within AGO and the binding sites of Argonaute-miRNA complexes within distinct mRNA molecules to be identified (Chi et al., 2009; Boudreau et al., 2014). Building on this method, photoactivatable ribonucleoside-enhanced crosslinking and immu-

noprecipitation uses randomly-incorporated photoactivatable nucleosides, commonly 4-thiouridine, prior to crosslinking RNA molecules with RNA binding proteins and ribonucleoprotein complexes (Hafner et al., 2010). This enables enhanced crosslinking and subsequent RNA recovery with reduced ultraviolet irradiation and, more importantly, leads to characteristic mutations, depending on the nucleoside utilized, which correlate with RNA-protein crosslinking sites. Thus, by coupling photoactivatable ribonucleoside-enhanced crosslinking and immunoprecipitation with AGO immunoprecipitation, miRNA binding sites can be identified by mutational analyses, and discriminated from contamination by other RNA species. Similarly, by crosslinking the RNA sequences bound to the AGO-miRNA complex after immunoprecipitation and sequencing the hybrids generated, CLASH (crosslinking, ligation, and sequencing of hybrids) allows miRNAs and their targets to be identified with high confidence (Helwak et al., 2013; Helwak and Tollervey, 2014). Importantly, CLIP-based sequencing approaches, including CLASH, are among those considered reliable for miRNA-target validation within the miRTarBase (version 7.0) (http://mirtarbase.mbc.nctu.edu.tw/php/index.php), the most comprehensively annotated database of experimentally-validated miRNA-target interactions (Chou et al., 2018). Yet, in spite of its benefits, crosslinking poses the risk of generating undesired mutations, introducing background, and failing to detect mRNA-miRNAs interactions occurring *via* non-canonical "seed" pairing. Further, CLIP-based methods often require large amounts of biological material and are biased towards the identification of miRNAs that are highly expressed and that contain specific sequences (Tan et al., 2014). To overcome these limitations, RNA immunoprecipitation (RIP) followed by next generation sequencing (RIP-Seq) enables specific RNA-binding proteins and their target-RNA complexes to be pulled down either in their native state or after cross-linking (Malmevik et al., 2015; Gagliardi and Matarazzo, 2016). In this context, immunoprecipitation of AGO-bound RNAs in the absence of crosslinking, followed by sequencing (AGO-RIPseq), allows miRNAs and bound mRNAs to be identified (Petri and Jakobsson, 2018). RIP-Seq datasets are considered as evidence of miRNA-mRNA binding in the DIANA-TarBase (version 8.0) (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index), a manually-curated database containing over 790,000 experimentally-validated miRNA-gene interactions (Karagkouni et al., 2018). This repository also incorporates the results of the IMPACT-Seq approach (identification of miRNA-responsive elements by pull-down and alignment of captive transcripts-sequencing), in which RNA targets are sequenced after co-purification with transfected miRNA mimics biotinylated at their 3' end in the absence of crosslinking (Tan et al., 2014). Although these methods have enabled an enormous number of novel miRNA–mRNA interactions to be identified, such interactions remain limited by the expression levels of the miRNAs of interest within particular cells or tissues at the examined time-point.

Predicting the Targetome and Genetic Regulatory Network of a MicroRNA

The targetome of a miRNA refers to the set of RNA molecules that it targets. Several computational approaches have been developed to predict the targetome of specific miRNAs. To date, most of these prediction algorithms use the canonical rule of miRNAs binding to their target mRNAs mainly by base pairing between the seed region of the miRNA and fully-complementary sequences within the 3′ untranslated region of the mRNA (Tan et al., 2014). However, evidence indicates that bulges, G:U wobbles, "seedless" interactions, and binding regions within a gene's coding sequence are admissible and not uncommon, with canonical miRNA-mRNA interactions accounting for only 25% (Helwak et al., 2013). Hence, computational prediction of binding regions within mRNA molecules, and thus of miRNA target genes, remains a major challenge. Although non-canonical miRNA-mRNA interactions seem to have no effect on mRNA or protein levels in mammals (Agarwal et al., 2015), incomplete miRNA-mRNA base-pairing is common and enhances Argonaute-mediated target slicing and the rate of product release (or target turnover) in other classes of living organisms (Wee et al., 2012; Helwak et al., 2013; Chen et al., 2017). Thus, although the biological relevance of accurately predicting partially-complementary miRNA-mRNA interactions remains controversial, there is no doubt such interactions are important to ensure undesired mRNA regulation is avoided upon administration of RNA interference-based therapeutics.

In general, the prediction accuracy for individual miRNA targetomes is relatively high if distinct parameters are considered. For example, to predict the probability of a specific miRNA-mRNA, TargetScan has developed the context++ model to go beyond simply searching for conserved multimers (more specifically, sequences of 6, 7, or 8 nt) that match the miRNA seed region. The TargetScan context++ model factors in site accessibility (Robins et al., 2005; Tafer et al., 2008; Hölscher, 2014), and allows mismatches in the miRNA seed region (Agarwal et al., 2015) by incorporating the concepts of conserved 3′ pairing (Friedman et al., 2009) and centered sites (Shin et al., 2010). Moreover, many online platforms now even incorporate machine-learning algorithms and advanced text-mining approaches using natural language processing methods to explore the literature and devise distinct miRNA-targeting mechanisms, the probability that they occur, and their effects in terms of translational repression and mRNA degradation. Offering high-quality miRNA targetome predictions for distinct vertebrate species and model organisms, such platforms include the user DNA Intelligent Analysis (DIANA)-microT (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index) (Paraskevopoulou et al., 2013), miRDB (mirdb.org), and MIRZA-G (Gumienny and Zavolan, 2015). Although targetome and large-scale genetic network predictions for individual miRNAs are increasingly reliable, accounting for the combined action of multiple miRNAs and predicting their joint effects on mRNA or protein levels remain daunting endeavors. The distance between miRNA binding sites in an mRNA molecule, for example, defines whether different miRNAs will act cooperatively or competitively, with distant sites favoring the former, but short distances promoting the latter (Grimson et al., 2007). In this sense, the DIANA-miRPath online platform enables predictions on the combined effect of modulating multiple miRNAs to be run (Vlachos et al., 2012, 2015). Altogether, only the combination of prediction algorithms and experimental approaches will ultimately lead to a detailed understanding of miRNA regulatory networks.

The Role of MicroRNAs in the Central Nervous System

miRNAs are implicated in virtually all aspects of cell biology, and play essential roles in several developmental and pathological processes. Dysfunctional miRNA regulatory networks correlate with the emergence and/or progression of diabetes, heart failure, cancer, and neurodevelopmental disorders (Allen, 1940; Blenkiron and Miska, 2007; Rajman and Schratt, 2017). In the brain, for instance, miRNAs act not only as "fine-tuners" but also as master regulators of neuronal circuit development, maturation, and function, and influence processes such as cell-fate determination, cell migration, neuronal polarization, cognition, and synapse formation and plasticity (Ye et al., 2016; Rajman and Schratt, 2017). Early studies aimed at assessing the role of miRNAs in the central nervous system revealed that deletions of key components of the miRNA-processing pathway, including Dicer, DGCR8, and Argonaute, lead to defects in neurogenesis, impaired brain development, and neuronal atrophy (Schaefer et al., 2007; Davis et al., 2008, 2015; Im and Kenny, 2012). Moreover, Cre-loxP-mediated conditional deletion of Dicer in specific neuronal cell types leads to their premature death, often as a consequence of a depleted neuronal progenitor pool or of enhanced apoptosis (Barca-Mayo and De Pietri Tonelli, 2014). Thus, miRNAs are likely to regulate cell-type specification in the central nervous system. Remarkably, transcriptomic profiling studies have identified miR-124, in particular, as the most abundant small RNA in the central nervous system (Lagos-Quintana et al., 2002; Landgraf et al., 2007).

While ectopic expression of miRNA-124 in mice depletes the pool of neuronal progenitors within the subventricular zone by forcing their differentiation, its partial inhibition blocks neurogenesis and promotes progenitor self-renewal (Cheng et al., 2009; Åkerblom et al., 2012). Nonetheless, as the cellular and molecular composition of the human brain is different to that of other mammals, including primates (Somel et al., 2011), investigating the role of miRNAs in human model systems is indispensable. In the HeLa human cell line, miR-124 overexpression has previously been reported to drive the acquisition of a neuronal-like transcriptional profile (Conaco et al., 2006), strongly hinting at the crucial role of this miRNA species in also defining neuronal identity in humans. In order to gain a deeper insight into the role of miR-124 in human neurogenesis, we used a highly homogeneous cellular model system known as iNGN cells. In this system, overexpression of the transcription factors neurogenin-1 and neurogenin-2 is driven by a doxycycline-inducible promoter in induced pluripotent stem cells. These cells acquire a homogeneous bipolar neuronal morphology and become electrically active 7 days after doxycycline administration (Busskamp et al., 2014). More importantly, as in the murine central nervous system, miR-124 is also the most abundant miRNA species in iNGN cells. In our study, CRISPR/Cas9-mediated disruption of the six miR-124 paralogs revealed this miRNA species to not be required for neurogenesis (Kutsche et al., 2018). However, miR-124-depleted iNGN cells acquired more complex morphologies, reduced their dendritic extension speed, and, unexpectedly, shifted from a glutamatergic identity towards the preferential use of acetylcholine as a neurotransmitter, which are findings consistent with previous reports (Yu et al., 2008; Franke et al., 2012; Volvert et al., 2014). Of note, miR-124-deficient cells also exhibited increased apoptosis in long-term cultures, as was previously shown in mouse neurons *in vivo* upon knock-out of a single miR-124 locus (Sanuki et al., 2011). Since the molecular machinery mediating these phenotypic alterations is so far largely unexplored, we decided to combine experimental and computational approaches to deepen our understanding of the miR-124 regulatory network.

Taking its limitations into consideration, we used AGO-RIP-Seq in miR-124-depleted iNGN cells at distinct time-points over the course of differentiation to identify miR-124 high-confidence targets (Kutsche et al., 2018). Although 4024 human transcripts have been described as possessing binding sites for miR-124, our approach allowed us to reliably identify 127 transcripts with reduced expression in miR-124-deficient cells compared to their wildtype counterparts. Remarkably, 98 of these 127 transcripts (77%) had been predicted to be miR-124 targets, and 38 of them had also been annotated as previously validated in the miRTarBase (Chou et al., 2018). Using luciferase reporter assays, which are considered a reliable validation method in the miRTarBase, we confirmed 43 novel *bona fide* miR-124 targets. Despite this approach signifying important progress in the validation of mRNAs targeted by miR-124, its interaction capacity with the remaining ~4000 annotated transcripts needs to be confirmed, potentially by also considering changes in protein levels. However, targeting 98 mRNAs, of which 81 have now been validated, demonstrated that miR-124 exerts its regulatory function at the network level. Another striking observation in our study was that 24 of these 98 predicted miR-124 targets were transcription factors (TFs) (Kutsche et al., 2018), hinting that these TFs are potential effectors of miR-124. As TFs modulate entire genetic programs and thereby have a strong impact in cell fate decisions (Hobert, 2008; Lambert et al., 2018), miRNA-mediated perturbations in the expression levels of TFs are particularly relevant.

Knowledge of the existence of regulatory feedback mechanisms between miRNAs and TFs is not new. The *die-1* TF, required for controlling chemosensory laterality in *Caenorhabditis elegans*, was shown over a decade ago to perform its functions by activating the *lsy-6* miRNA in left-sided chemosensory neurons but having its expression levels modulated in left-sided neurons by *mir-273* (Chang et al., 2004). Similarly, multiple miRNA-TF feedback circuits have been described as essential for mammalian neurogenesis, including miR-133b/Pitx (Kim et al., 2007), miR-7/Pax6 (de Chevigny et al., 2012), miR-132/Nurr1 (Nr4a2) (Yang et al., 2012), and miR-17-3p/Olig2 (Chen et al., 2011). Moreover, specific miRNAs are known to regulate the expression of multiple TFs, as is the case for the neuron-enriched miR-9 in the murine brain (Shibata et al., 2011). As a consequence of their biological importance and widespread occurrence, multiple online platforms have been developed to predict miRNA-TF interactions, including TFmiR (Hamed et al., 2015) and DIANA-mirExtra (Vlachos et al., 2016). However, building an interaction network is not trivial. To generate gene regulatory networks, in particular, co-expression networks are widely used (Yang et al., 2014; van Dam et al., 2018). Such networks are characterized by connected pairs of nodes representing significantly correlated expression patterns which can also contain information regarding both the direction and the strength (or weight) of the interaction. Nonetheless, network analyses are often performed using defined subsets of entities, leading to biased results if relevant interactions *via* other intermediaries are not considered. To overcome this limitation, the

weighted topological (wTO) network analysis method implicitly includes correlations among nodes which are subsequently omitted in the analysis (Nowick et al., 2009; Gysi et al., 2018). Thereby, the wTO approach enables a more reliable evaluation of the interactions among sets of nodes of interest relative to other network analysis pipelines. By implementing the wTO network analysis method on the differentially regulated TFs in miR-124-deficient iNGN cells, we identified a number of factors likely to mediate the biological functions of miR-124. To validate the results, the function of the zinc finger 787 (*ZNF787*) was additionally characterized in iNGN cells, confirming its role as a repressor of neuronal features and revealing that it modulates a big fraction of the targets regulated by miR-124 (Kutsche et al., 2018).

Lastly, the experimental deletion of a highly abundant miRNA species did not result in a vacuum of miRNA regulation, as we detected other miRNAs being upregulated (Kutsche et al., 2018). Hence, our data suggest that this *de novo* miRNA regulation upon manipulation affects the primary phenotype. This likely contributes to the difficulties in interpreting experimental miRNA knockout and knockdown phenotypes, and must therefore be considered.

Outlook and Perspective

With less than 2% of the human genome encoding protein-coding genes, non-coding RNAs represent most of the human transcriptome (Tay et al., 2014). The central roles of such RNAs, including miRNAs, in health and disease is now widely recognized. Consequently, steady progress has been made over recent years in the annotation and validation of novel miRNA. However, there are still a number of important challenges in collecting highly heterogeneous data from multiple experimental sources and in developing highly accurate prediction tools. The latter are essential for enhancing the medical potential of miRNAs (Jimenez-Mateos et al., 2012). We foresee that machine learning will make an important contribution to the refinement of miRNA repositories and prediction tools acting at the systems level.

Acknowledgments: *The authors would also like to thank Sara Oakeley for critical feedback on the manuscript.*
Author contributions: *Manuscript conception, design and writing: JERA and VB.*
Conflicts of interest: *The authors declare no conflict of interest.*
Financial support: *VB was supported by a Volkswagen Foundation Freigeist fellowship (A110720) and by an European Research Council starting grant (678071-ProNeurons). JERA was supported by the Dresden International Graduate School for Biomedicine and Bioengineering (DIGS-BB) program.*
Copyright license agreement: *The Copyright License Agreement has been signed by both authors before publication.*
Plagiarism check: *Checked twice by iThenticate.*
Peer review: *Externally peer reviewed.*

References

Agarwal V, Bell GW, Nam JW, Bartel DP (2015) Predicting effective microRNA target sites in mammalian mRNAs. Elife 4.
Åkerblom M, Sachdeva R, Barde I, Verp S, Gentner B, Trono D, Jakobsson J (2012) MicroRNA-124 is a subventricular zone neuronal fate determinant. J Neurosci 32:8879-8889.
Allen IM (1940) The immediate and remote effects of minor lesions of the cervical portion of the spinal cord following head injury. Aust N Z J Surg 10:157-172.
Ambros V (2003) A uniform system for microRNA annotation. RNA 9:277-279.
An J, Lai J, Lehman ML, Nelson CC (2013) miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. Nucleic Acids Res 41:727-737.
Barca-Mayo O, De Pietri Tonelli D (2014) Convergent microRNA actions coordinate neocortical development. Cell Mol Life Sci 71:2975-2995.
Bartel DP (2009) MicroRNAs: target recognition and regulatory functions. Cell 136:215-233.
Berezikov E, van Tetering G, Verheul M, van de Belt J, van Laake L, Vos J, Verloop R, van de Wetering M, Guryev V, Takada S, van Zonneveld AJ, Mano H, Plasterk R, Cuppen E (2006) Many novel mammalian microRNA candidates identified by extensive cloning and RAKE analysis. Genome Res 16:1289-1298.
Blenkiron C, Miska EA (2007) miRNAs in cancer: approaches, aetiology, diagnostics and therapy. Hum Mol Genet 16 Spec No 1:R106-113.
Boudreau RL, Jiang P, Gilmore BL, Spengler RM, Tirabassi R, Nelson JA, Ross CA, Xing Y, Davidson BL (2014) Transcriptome-wide discovery of microRNA binding sites in human brain. Neuron 81:294-305.
Busskamp V, Lewis NE, Guye P, Ng AH, Shipman SL, Byrne SM, Sanjana NE, Murn J, Li Y, Li S, Stadler M, Weiss R, Church GM (2014) Rapid neurogenesis through transcriptional activation in human stem cells. Mol Syst Biol 10:760.
Chang S, Johnston Jr RJ, Frøkjær-Jensen C, Lockery S, Hobert O (2004) MicroRNAs act sequentially and asymmetrically to control chemosensory laterality in the nematode. Nature 430:785-789.
Chen GR, Sive H, Bartel DP (2017) A seed mismatch enhances argonaute2-catalyzed cleavage and partially rescues severely impaired cleavage found in fish. Mol Cell 68:1095-1107.e5.
Chen JA, Huang YP, Mazzoni EO, Tan GC, Zavadil J, Wichterle H (2011) Mir-17-3p controls spinal neural progenitor patterning by regulating Olig2/Irx3 cross-repressive loop. Neuron 69:721-735.
Cheng LC, Pastrana E, Tavazoie M, Doetsch F (2009) miR-124 regulates adult neurogenesis in the subventricular zone stem cell niche. Nat Neurosci 12:399-408.
Chi SW, Zang JB, Mele A, Darnell RB (2009) Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460:479-486.
Chiang HR, Schoenfeld LW, Ruby JG, Auyeung VC, Spies N, Baek D, Johnston WK, Russ C, Luo S, Babiarz JE, Blelloch R, Schroth GP, Nusbaum C, Bartel DP (2010) Mammalian microRNAs: experimental evaluation of novel and previously annotated genes. Genes Dev 24:992-1009.
Chou CH, Shrestha S, Yang CD, Chang NW, Lin YL, Liao KW, Huang WC, Sun TH, Tu SJ, Lee WH, Chiew MY, Tai CS, Wei TY, Tsai TR, Huang HT, Wang CY, Wu HY, Ho SY, Chen PR, Chuang CH, et al. (2018) miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Res 46:D296-302.
Conaco C, Otto S, Han JJ, Mandel G (2006) Reciprocal actions of REST and a microRNA promote neuronal identity. Proc Natl Acad Sci U S A 103:2422-2427.
Davis GM, Haas MA, Pocock R (2015) MicroRNAs: Not "Fine-Tuners" but key regulators of neuronal development and function. Front Neurol 6:245.
Davis TH, Cuellar TL, Koch SM, Barker AJ, Harfe BD, McManus MT, Ullian EM (2008) Conditional loss of Dicer disrupts cellular and tissue morphogenesis in the cortex and hippocampus. J Neurosci 28:4322-4330.
de Chevigny A, Core N, Follert P, Gaudin M, Barbry P, Beclin C, Cremer H (2012) miR-7a regulation of Pax6 controls spatial origin of forebrain dopaminergic neurons. Nat Neurosci 15:1120-1126.
Denzler R, McGeary SE, Title AC, Agarwal V, Bartel DP, Stoffel M (2016) Impact of microRNA levels, target-site complementarity, and cooperativity on competing endogenous RNA-regulated gene expression. Mol Cell 64:565-579.
Franke K, Otto W, Johannes S, Baumgart J, Nitsch R, Schumacher S (2012) miR-124-regulated RhoG reduces neuronal process complexity via ELMO/Dock180/Rac1 and Cdc42 signalling. EMBO J 31:2908-2921.
Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N (2008) Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 26:407-415.
Friedman RC, Farh KK, Burge CB, Bartel DP (2009) Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19:92-105.
Gagliardi M, Matarazzo M (2016) RIP: RNA Immunoprecipitation. New York, NY: Humana Press.
Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27:91-105.
Gumienny R, Zavolan M (2015) Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G. Nucleic Acids Res 43:1380-1391.
Gysi DM, Voigt A, Fragoso TM, Almaas E, Nowick K (2018) wTO: an R package for computing weighted topological overlap and a consensus network with integrated visualization tool. BMC Bioinformatics 19:392.

Hölscher C (2014) The incretin hormones glucagonlike peptide 1 and glucose-dependent insulinotropic polypeptide are neuroprotective in mouse models of Alzheimer's disease. Alzheimers Dement 10:S47-54.
Ha M, Kim VN (2014) Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol 15:509-524.
Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M, Jr., Jungkamp AC, Munschauer M, Ulrich A, Wardle GS, Dewell S, Zavolan M, Tuschl T (2010) Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. Cell 141:129-141.
Hamed M, Spaniol C, Nazarieh M, Helms V (2015) TFmiR: a web server for constructing and analyzing disease-specific transcription factor and miRNA co-regulatory networks. Nucleic Acids Res 43:W283-288.
Han J, Lee Y, Yeom KH, Kim YK, Jin H, Kim VN (2004) The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev 18:3016-3027.
Helwak A, Tollervey D (2014) Mapping the miRNA interactome by cross-linking ligation and sequencing of hybrids (CLASH). Nat Protoc 9:711-728.
Helwak A, Kudla G, Dudnakova T, Tollervey D (2013) Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. Cell 153:654-665.
Hendrix D, Levine M, Shi W (2010) miRTRAP, a computational method for the systematic identification of miRNAs from high throughput sequencing data. Genome Biol 11:12.
Hobert O (2008) Gene regulation by transcription factors and microRNAs. Science 319:1785-1786.
Im HI, Kenny PJ (2012) MicroRNAs in neuronal function and dysfunction. Trends Neurosci 35:325-334.
Jimenez-Mateos EM, Engel T, Merino-Serrais P, McKiernan RC, Tanaka K, Mouri G, Sano T, O'Tuathaigh C, Waddington JL, Prenter S, Delanty N, Farrell MA, O'Brien DF, Conroy RM, Stallings RL, DeFelipe J, Henshall DC (2012) Silencing microRNA-134 produces neuroprotective and prolonged seizure-suppressive effects. Nat Med 18:1087-1094.
Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, Papadimitriou D, Kavakiotis I, Maniou S, Skoufos G, Vergoulis T, Dalamagas T, Hatzigeorgiou AG (2018) DIANA-TarBase v8: a decade-long collection of experimentally supported miRNA-gene interactions. Nucleic Acids Res 46:D239-245.
Karali M, Peluso I, Gennarino VA, Bilio M, Verde R, Lago G, Dollé P, Banfi S (2010) miRNeye: a microRNA expression atlas of the mouse eye. BMC Genomics 11:715.
Kim J, Inoue K, Ishii J, Vanti WB, Voronov SV, Murchison E, Hannon G, Abeliovich A (2007) A microRNA feedback circuit in midbrain dopamine neurons. Science 317:1220-1224.
Kozomara A, Griffiths-Jones S (2011) miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res 39:D152-157.
Kozomara A, Griffiths-Jones S (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42:D68-73.
Kozomara A, Birgaoanu M, Griffiths-Jones S (2019) miRBase: from microRNA sequences to function. Nucleic Acids Res 47:D155-162.
Krol J, Loedige I, Filipowicz W (2010) The widespread regulation of microRNA biogenesis, function and decay. Nat Rev Genet 11:597-610.
Kukurba KR, Montgomery SB (2015) RNA sequencing and analysis. Cold Spring Harb Protoc 2015:951-969.
Kutsche LK, Gysi DM, Fallmann J, Lenk K, Petri R, Swiersy A, Klapper SD, Pircs K, Khattak S, Stadler PF, Jakobsson J, Nowick K, Busskamp V (2018) Combined experimental and system-level analyses reveal the complex regulatory network of miR-124 during human neurogenesis. Cell Syst 7:438-452.
Lagos-Quintana M, Rauhut R, Yalcin A, Meyer J, Lendeckel W, Tuschl T (2002) Identification of tissue-specific microRNAs from mouse. Curr Biol 12:5.
Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, Chen X, Taipale J, Hughes TR, Weirauch MT (2018) The Human Transcription Factors. Cell 172:650-665.
Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M, Lin C, Socci ND, Hermida L, Fulci V, Chiaretti S, Foa R, Schliwka J, Fuchs U, Novosel A, Muller RU, et al. (2007) A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129:1401-1414.
Ludwig N, Becker M, Schumann T, Speer T, Fehlmann T, Keller A, Meese E (2017) Bias in recent miRBase annotations potentially associated with RNA quality issues. Sci Rep 7:5162.
Malmevik J, Petri R, Klussendorf T, Knauff P, Akerblom M, Johansson J, Soneji S, Jakobsson J (2015) Identification of the miRNA targetome in hippocampal neurons using RIP-seq. Sci Rep 5:12609.
Nowick K, Gernat T, Almaas E, Stubbs L (2009) Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. Proc Natl Acad Sci U S A 106:22358-22363.
Paraskevopoulou MD, Georgakilas G, Kostoulas N, Vlachos IS, Vergoulis T, Reczko M, Filippidis C, Dalamagas T, Hatzigeorgiou AG (2013) DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. Nucleic Acids Res 41:W169-173.
Petri R, Jakobsson J (2018) Identifying miRNA Targets Using AGO-RIPseq. In: mRNA Decay: Methods and Protocols (Lamandé SR, ed), pp 131-140. New York, NY: Humana Press.
Rajman M, Schratt G (2017) MicroRNAs in neural development: from master regulators to fine-tuners. Development 144:2310-2322.
Robins H, Li Y, Padgett RW (2005) Incorporating structure to predict microRNA targets. Proc Natl Acad Sci U S A 102:4.
Saetrom P, Heale BS, Snove O, Jr., Aagaard L, Alluin J, Rossi JJ (2007) Distance constraints between microRNA target sites dictate efficacy and cooperativity. Nucleic Acids Res 35:2333-2342.
Sanuki R, Onishi A, Koike C, Muramatsu R, Watanabe S, Muranishi Y, Irie S, Uneo S, Koyasu T, Matsui R, Cherasse Y, Urade Y, Watanabe D, Kondo M, Yamashita T, Furukawa T (2011) miR-124a is required for hippocampal axogenesis and retinal cone survival through Lhx2 suppression. Nat Neurosci 14:1125-1134.
Schaefer A, O'Carroll D, Tan CL, Hillman D, Sugimori M, Llinas R, Greengard P (2007) Cerebellar neurodegeneration in the absence of microRNAs. J Exp Med 204:1553-1558.
Shibata M, Nakao H, Kiyonari H, Abe T, Aizawa S (2011) MicroRNA-9 regulates neurogenesis in mouse telencephalon by targeting multiple transcription factors. J Neurosci 31:3407-3422.
Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP (2010) Expanding the microRNA targeting code: functional sites with centered pairing. Mol Cell 38:789-802.
Somel M, Liu X, Tang L, Yan Z, Hu H, Guo S, Jiang X, Zhang X, Xu G, Xie G, Li N, Hu Y, Chen W, Paabo S, Khaitovich P (2011) MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. PLoS Biol 9:e1001214.
Tafer H, Ameres SL, Obernosterer G, Gebeshuber CA, Schroeder R, Martinez J, Hofacker IL (2008) The impact of target site accessibility on the design of effective siRNAs. Nat Biotechnol 26:578-583.
Tan SM, Kirchner R, Jin J, Hofmann O, McReynolds L, Hide W, Lieberman J (2014) Sequencing of captive target transcripts identifies the network of regulated genes and functions of primate-specific miR-522. Cell Rep 8:1225-1239.
Tay Y, Rinn J, Pandolfi PP (2014) The multilayered complexity of ceRNA crosstalk and competition. Nature 505:344-352.
van Dam S, Vosa U, van der Graaf A, Franke L, de Magalhaes JP (2018) Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform 19:575-592.
Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG (2015) DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic Acids Res 43:W460-466.
Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, Paraskevopoulou MD, Prionidis K, Dalamagas T, Hatzigeorgiou AG (2012) DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. Nucleic Acids Res 40:W498-504.
Vlachos IS, Vergoulis T, Paraskevopoulou MD, Lykokanellos F, Georgakilas G, Georgiou P, Chatzopoulos S, Karagkouni D, Christodoulou F, Dalamagas T, Hatzigeorgiou AG (2016) DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data. Nucleic Acids Res 44:W128-134.
Volvert ML, Prevot PP, Close P, Laguesse S, Pirotte S, Hemphill J, Rogister F, Kruzy N, Sacheli R, Moonen G, Deiters A, Merkenschlager M, Chariot A, Malgrange B, Godin JD, Nguyen L (2014) MicroRNA targeting of CoREST controls polarization of migrating cortical neurons. Cell Rep 7:1168-1183.
Wee LM, Flores-Jasso CF, Salomon WE, Zamore PD (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. Cell 151:1055-1067.
Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, Xu LD, Wang YY, Du L, Zhang YP, Jiang W, Li CQ, Xiao Y, Li X (2011) MiRNA-miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. Nucleic Acids Res 39:825-836.
Yang D, Li T, Wang Y, Tang Y, Cui H, Tang Y, Zhang X, Chen D, Shen N, Le W (2012) miR-132 regulates the differentiation of dopamine neurons by directly targeting Nurr1 expression. J Cell Sci 125:1673-1682.
Yang JS, Lai EC (2011) Alternative miRNA biogenesis pathways and the interpretation of core miRNA pathway mutants. Mol Cell 43:892-903.
Yang Y, Han L, Yuan Y, Li J, Hei N, Liang H (2014) Gene co-expression network analysis reveals common system-level properties of prognostic genes across cancer types. Nat Commun 5:3231.
Ye Y, Xu H, Su X, He X (2016) Role of microRNA in governing synaptic plasticity. Neural Plast 2016:4959523.
Yu JY, Chung KH, Deo M, Thompson RC, Turner DL (2008) MicroRNA miR-124 regulates neurite outgrowth during neuronal differentiation. Exp Cell Res 314:2618-2633.

C-Editors: Zhao M, Yu J; T-Editor: Liu XL