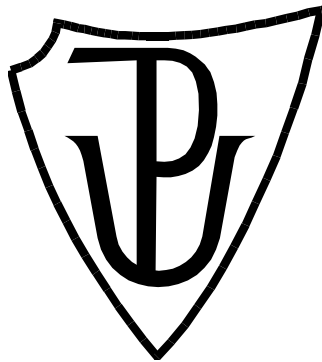


PALACKÝ UNIVERSITY OLOMOUC
Faculty of Science
Laboratory of Growth Regulators



**Methodology and application of quantitative
microRNA analysis**

Ph.D. thesis

Author: Ing. Peter Androvič
Study programme: P1527 Biology
Field of Study: 1501V019 Experimental Biology
Form of Study: Daily
Supervisor: prof. Mikael Kubista, Ph.D.
Consultant: Ing. Lukáš Valihrač, Ph.D.

Bibliographical identification

Author's first name and surname	Ing. Peter Androvič
Title of thesis	Methodology and application of quantitative microRNA analysis
Type of thesis	Ph.D. Thesis
Department	Laboratory of Growth Regulators, Faculty of Science UPOL & Laboratory of Gene Expression, Institute of Biotechnology, Czech Academy of Sciences
Supervisor	prof. Mikael Kubista, Ph.D.
The year of presentation	2020
Abstract	<p>MicroRNAs (miRNAs) are a class of small non-coding RNAs that serve as important regulators of gene expression at the posttranscriptional level. Their dysregulation is implicated in many physiological and pathological conditions including development, cancer or neurological disease. In addition, miRNAs are stable in body fluids and pose a great potential to serve as minimally invasive clinical biomarkers. All these applications require precise and sensitive methods to quantify miRNA levels. However, reliable miRNA quantification is challenging due to their inherent properties as well as various pre-analytical variables that are sources of bias. This thesis focuses on improvement of miRNA quantification workflows. It presents novel specific and cost-effective RT-qPCR system, novel tool for quality control of circulating miRNA studies, and it comprehensively evaluates the performance of commercial library preparation methods for miRNA sequencing. Next, these and complementary techniques are applied in biological models to unravel mRNA and miRNA expression changes after injuries to the central nervous system (CNS). Taken together, this thesis provides novel tools for reliable miRNA analysis and provides new insights to the mechanisms underlying CNS injuries.</p>
Keywords	microRNA, miRNA, RT-qPCR, RNA-Seq, quality control, liquid biopsies, ischemic stroke, spinal cord injury
Number of pages	64
Number of appendices	5
Language	English

Bibliografická identifikace

Jméno a příjmení autora	Ing. Peter Androvič
Název práce	Metodologie a aplikace kvantitativní analýzy mikroRNA
Typ práce	Disertační
Pracoviště	Laboratoř růstových regulátorů, Přírodovědecká fakulta UPOL & Laboratoř genové exprese, Biotechnologický ústav Akademie věd České Republiky
Vedoucí práce	prof. Mikael Kubista, Ph.D.
Rok obhajoby práce	2020
Abstrakt	<p>MikroRNA (miRNA) jsou malé nekódující RNA, které slouží jako důležité regulátory genové exprese na posttranskripční úrovni. Jejich dysregulace provází mnohé fyziologické a patologické stavy, včetně poruch vývoje, rakoviny nebo neurologických onemocnění. Kromě toho jsou miRNA stabilní v tělních tekutinách, což jim dává velký potenciál jako minimálně invazivním klinický biomarkerům. Všechny tyto způsoby uplatnění vyžadují přesné a citlivé metody pro kvantifikaci hladin miRNA. Spolehlivá kvantifikace miRNA je však náročná kvůli jejich přirozeným vlastnostem a preanalytickým krokům, které jsou zdrojem chyb. Tato práce se zaměřuje na zlepšení metod kvantifikace miRNA. Představuje novou, vysoce specifickou a cenově efektivní RT-qPCR metodu, nový nástroj pro kontrolu kvality studií zaměřených na analýzu cirkulujících miRNA a komplexně vyhodnocuje vlastnosti komerčních metod pro přípravu knihoven pro sekvenování miRNA. Dále jsou tyto techniky spolu s komplementárními metodami aplikovány v biologických modelech za účelem porozumění změn exprese mRNA a miRNA při poškození centrálního nervového systému (CNS). Souhrnem jsou v rámci této práce představeny nové nástroje pro spolehlivou analýzu miRNA a prezentovány nové poznatky o mechanismech poškození CNS.</p>
Klíčová slova	mikroRNA, miRNA, RT-qPCR, RNA-Seq, kontrola kvality, biopsie tělních tekutin, mozková ischemie, poranění míchy
Počet stran	64
Počet příloh	5
Jazyk	Anglický

Declaration

Hereby, I declare that this thesis summarizes my original results or results obtained with my significant contribution and I have composed the thesis by myself using the cited literature.

Prohlašuji, že předložená práce shrnuje moje originální výsledky nebo výsledky, které byly získány s mým významným příspěvkem a práci jsem vypracoval samostatně za použití citované literatury.

In Olomouc,

.....

Acknowledgement

First, I would like to thank my supervisor prof. Mikael Kubista for his support, and opportunities provided to me throughout the duration of my Ph.D. studies.

I would like to say a massive thank you to Dr. Lukáš Valihrač for his guidance, advice on all matters, tremendous patience, incredible attitude and optimism (which I often needed), and for the effort he put into my supervision. Without him, the research presented in this thesis, and most likely my passion for science, would not exist.

A big thank you to all my friends from the Laboratory of Gene Expression for help in and outside the work, and for creating a truly unmatched working atmosphere. Thank you for the countless memorable moments in and outside the lab and thank you for making my Monday mornings feel like most people's Friday evenings.

Thank you to our collaborators from a team of Dr. Miroslava Anděrová at the Institute of Experimental Medicine, and to prof. Michael Pfaffl and his team at Technical University of Munich for their help and supervision during my visit in Germany.

I would like to thank my family for their continuous support in doing what I like and pursuing my Ph.D. degree.

I would also like to acknowledge all the funding bodies, which made this research possible.

Contents

List of abbreviations	8
List of publications	9
Contribution report.....	10
1 Introduction.....	11
2 Aims and scope	12
3 Literature review	13
3.1 MicroRNAs.....	13
3.1.1 MiRNA genomic organization.....	13
3.1.2 MiRNA biogenesis	13
3.1.3 MiRNA regulatory mechanisms.....	15
3.1.4 Plant miRNAs.....	15
3.1.5 MiRNA annotation and nomenclature	16
3.1.6 IsomiRs – miRNA variants.....	17
3.1.7 Extracellular miRNAs	18
3.1.8 Circulating miRNAs as clinical biomarkers	18
3.2 Quantitative miRNA profiling	19
3.2.1 Experimental design.....	19
3.2.2 Sampling and storage	20
3.2.3 MiRNA extraction	20
3.2.4 Quality control.....	21
3.2.5 MiRNA quantification	23
3.2.6 Data analysis.....	28
3.3 MiRNAs in central nervous system	30
3.3.1 MiRNAs in acute CNS injury	31
4 Materials and methods	33
4.1 Samples	33
4.2 Primers and synthetic oligonucleotides	33
4.3 RNA isolation	33
4.4 RT-qPCR of miRNAs	34
4.5 Library preparation and sequencing	34
4.6 High-throughput RT-qPCR	34
4.7 Bioinformatic and statistical data analysis.....	35
5 Survey of results	37
5.1 Development of novel RT-qPCR-based method for highly accurate miRNA quantification	37

5.2 Development of quality control tool for circulating miRNA studies	38
5.3 Comprehensive performance comparison of small RNA-Seq library preparation methods from biofluids	39
5.4 Decoding the transcriptional response to ischemic stroke in young and aged mice.....	41
5.5 mRNA-miRNA regulatory networks in CNS injury (unpublished results)	42
6 Conclusion	47
7 References	48
8 Supplements.....	64

List of abbreviations

AGO	Argonaute protein
circRNA	circular RNA
CNS	central nervous system
EVs	extracellular vesicles
FFPE	formalin-fixation and paraffin embedding
GO	gene ontology
GSEA	gene set enrichment analysis
IFN	interferon
IFN-I	type-I interferon
ISGs	interferon-stimulated genes
LNA	locked nucleic acid
lncRNA	long non-coding RNA
MCAo	middle cerebral artery occlusion
miRNA	microRNA
NGS	next generation sequencing
PAGE	polyacrylamide gel electrophoresis
PAP	poly(A) polymerase
PCA	principal component analysis
PCR	polymerase chain reaction
piRNA	piwi-interacting RNA
Pre-miRNA	precursor microRNA
Pri-miRNA	primary microRNA
qPCR	quantitative polymerase chain reaction
RBP	RNA-binding protein
RIN	RNA integrity number
RISC	RNA-induced silencing complex
RNAi	RNA interference
RNA-Seq	RNA sequencing
RQN	RNA quality number
rRNA	ribosomal RNA
RT	reverse transcription
SCI	spinal cord injury
siRNA	small interfering RNA
snRNA	small nuclear RNA
tRNA	transfer RNA
UMIs	unique molecular identifiers
WGCNA	weighted gene co-expression network analysis

List of publications

This thesis summarizes the following publications that are attached at the end of the thesis in the supplementary section:

- I. **Androvič, P.**, Valihrač, L., Elling, J., Sjöback, R., and Kubista, M. (2017). Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Research* 45, e144–e144
- II. **Androvič, P.**, Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., and Valihrač, L. (2019). Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Scientific Reports* 9, 4255.
- III. **Androvič, P.***, Benesova, S.*, Kubista, M., Valihrač, L. (2020). Performance comparison of small RNA-Seq library preparation methods for biofluid samples. *Manuscript in preparation*. *These authors contributed equally
- IV. Valihrač, L., **Androvič, P.**, Kubista, M. (2019). Circulating miRNA Analysis for Cancer Diagnostics and Therapy. *Molecular Aspects of Medicine*, <https://doi.org/10.1016/j.mam.2019.10.002>.
- V. **Androvič, P.**, Kirdajova, D., Tureckova, J., Zucha, D., Rohlova, E., Abaffy, P., Kriska, J., Anderova, M., Kubista, M., and Valihrač, L. (2019). Decoding the transcriptional response to ischemic stroke in young and aged mouse brain. BioRxiv 769331. *Online preprint. Manuscript is under review in an international peer-reviewed journal*.

In addition, Peter Androvič contributed to the following publications:

- VI. Zucha, D., **Androvič, P.**, Kubista M., Valihrač, L. (2019). Performance comparison of reverse transcriptases for single-cell studies. *Clinical Chemistry* 66, 217-228
- VII. Kolenicova, D., Tureckova, J., Pukajova, B., Harantova, L., Kriska, J., Kirdajova, D., Vorisek, I., Kamenicka, M., Valihrač, L., **Androvič, P.**, Kubista, M., Vargova L., and Anderova, M. (2019). High potassium exposure reveals the altered ability of astrocytes to regulate their volume in the aged hippocampus of GFAP/EGFP mice. *Neurobiology of Aging*. 86, 162–181
- VIII. Smieszek, A., Kornicka, K., Szłapka-Kosarzewska, J., **Androvič, P.**, Valihrač, L., Langerova, L., Rohlova, E., Kubista, M., and Marycz, K. (2019). Metformin Increases Proliferative Activity and Viability of Multipotent Stromal Stem Cells Isolated from Adipose Tissue Derived from Horses with Equine Metabolic Syndrome. *Cells* 8, 80.
- IX. Valihrač, L., **Androvič, P.**, and Kubista, M. (2018). Platforms for Single-Cell Collection and Analysis. *International Journal of Molecular Sciences* 19, 807.
- X. Valny, M., Honsa, P., Waloschkova, E., Matuskova, H., Kriska, J., Kirdajova, D., **Androvič, P.**, Valihrač, L., Kubista, M., and Anderova, M. (2018). A single-cell analysis reveals multiple roles of oligodendroglial lineage cells during post-ischemic regeneration. *Glia* 66, 1068–1081.
- XI. Honsa, P., Valny, M., Kriska, J., Matuskova, H., Harantova, L., Kirdajova, D., Valihrač, L., **Androvič, P.**, Kubista, M., and Anderova, M. (2016). Generation of reactive astrocytes from NG2 cells is regulated by sonic hedgehog. *Glia* 64, 1518–1531.

Contribution report

- I. As a first author, Peter Androvič contributed to the design of the experiments, performed all experiments, analyzed the data, prepared figures 2-7, and wrote the manuscript with contributions from Lukáš Valihrach and Mikael Kubista.
- II. As a first author, Peter Androvič conceived and designed the study together with Lukáš Valihrach. Peter Androvič performed all experiments, analyzed the data, prepared all figures and wrote the manuscript with contributions from Lukáš Valihrach and Mikael Kubista.
- III. As a co-first author, Peter Androvič conceived and designed the study together with Lukáš Valihrach. Peter Androvič performed wet-lab experiments and contributed to the analysis of the data and writing of the manuscript.
- IV. As a contributing author, Peter Androvič reviewed the literature with focus on miRNA quantification and quality control and contributed to the writing of the manuscript.
- V. As a first author, Peter Androvič prepared sequencing libraries, analyzed and interpreted the data, prepared all figures and wrote the manuscript with contributions from Lukáš Valihrach, Miroslava Anděrová, Denisa Kirdajová and Mikael Kubista.

1 Introduction

According to the latest reports, the human genome encodes for approximately 19 000 protein-coding genes, which represents only 1.0 – 1.5% of 3 billion-plus base pairs it contains [1]. The rest was originally considered “junk” DNA. However, closer exploration of the genome revealed that in fact about 80% has biochemical role and is integral to the function of cells, particularly for the control of the gene activity [2]. In addition to various regulatory elements, the genome harbors at least 18 000 non-coding genes that produce over 47 000 non-coding RNAs including tRNAs, rRNAs, lncRNAs, snRNAs, piRNAs, circRNAs, siRNAs, miRNAs and many others, often with unknown function. One of the relatively well-characterized classes are microRNAs (miRNAs) – tiny, but all the more important regulators of gene expression. First miRNA was discovered by Victor Ambros and his co-workers Rosalind Lee and Rhonda Feinbaum in 1993 [3]. They found that *lin-4*, a gene known to control timing of *C. elegans* larval development by repression of the *lin-14* gene, does not code for a protein, but instead produces a pair of small RNAs. This discovery represented dramatic breakthrough in our understanding of the transcriptome. Since then, miRNAs have been shown to play important roles in modulation of an array of physiological and pathological processes ranging from embryonic development to neoplastic progression. This brought them significant attention as potential therapeutic targets. Another wave of excitement came with the discovery that miRNAs are released into extracellular environment and are stably present in the circulatory system and various body fluids. The realization that circulating miRNA levels change in response to pathophysiological processes meant that they might serve as promising and non-invasive clinical biomarkers to aid diagnosis, prognosis and monitoring of the response to treatment.

Whether the goal is to elucidate regulatory roles of the miRNAs, or find novel biomarkers, miRNA expression patterns provide essential information for these endeavors. Therefore, since the very beginning quantitative profiling of miRNA expression has played a pivotal role in the miRNA research. Several techniques for miRNA profiling are nowadays available for both targeted and global measurements, including microarrays, RT-qPCR and next generation sequencing methods. However, to obtain precise and reliable readouts of miRNA profiles is not straightforward because of many technical challenges associated with the workflow.

This thesis focuses on the development of novel methods and tools for improved miRNA quantitative workflows and on the assessment of small-RNA sequencing methods for better understanding of their impact on the resulting data. After establishment of the technical base, these methods are applied together with global mRNA expression profiling to dissect molecular and cellular response to acute central nervous system injuries including stroke and spinal cord injury in rodent models.

2 Aims and scope

The broader goal of this thesis was to develop and establish robust and reliable workflows for both, targeted and global miRNA expression profiling, including profiling of challenging samples such as biofluids. The ultimate goal was to utilize these workflows together with techniques for gene expression profiling to improve our understanding of molecular mechanisms of central nervous system (CNS) injury.

The specific aims of the work described in this thesis were:

- To develop, optimize and validate new method for miRNA quantification based on RT-qPCR that would allow precise and cost-effective quantification from various samples including animal and plant tissues, cells and biofluids.
- To develop easy-to-use tool that would allow convenient optimization and troubleshooting of the wet-lab workflow of quantitative miRNA studies as well as routine control of sample quality.
- To comprehensively evaluate all currently available methods for small RNA-Seq library preparation with focus on their performance with challenging samples such as biofluids to reveal their strengths, drawbacks, biases and impact on resulting data quality.
- To dissect global gene and miRNA expression changes after ischemic stroke and spinal cord injury (SCI) to reveal underlying molecular and cellular mechanisms.

3 Literature review

3.1 MicroRNAs

MicroRNAs (miRNAs) are short non-coding regulatory RNA molecules with length of ~ 22 nt [4]. They are found in animals, plants and some viruses where they regulate expression of protein-coding genes at the post-transcriptional level. The miRNA pathway is derived from a closely related RNA-silencing pathway called RNA interference (RNAi). While RNAi pathway starts with longer dsRNA producing various small interfering RNAs (siRNAs), the hallmark of evolutionally younger miRNA pathway is the production of short hairpin RNAs giving rise to well-defined class of miRNAs [5]. In mammals, hundreds of miRNAs have been identified, including 1978 mature miRNAs in mouse and 2654 miRNAs in human according to latest release of the miRBase database [6]. Many of these miRNAs are conserved across species and it is estimated that over 60% of the mammalian protein-coding genes are subject to miRNA-mediated regulation [7]. It is perhaps not surprising that miRNAs play important regulatory roles in nearly every physiological and pathological aspect of biology including developmental processes and various diseases [8].

3.1.1 MiRNA genomic organization

MiRNA genes are located in all chromosomes in humans, except for Y chromosome [9]. Around half of all currently identified miRNAs are intragenic and processed mostly from introns and few exons of protein coding genes, while the remaining are intergenic, transcribed independently of a host gene and regulated by their own promoters. About 50% of intergenic miRNAs are found in close proximity to other miRNAs, forming extended clusters transcribed as a single polycistronic unit [10]. MiRNAs originating from a single pri-miRNA are often related, and share their mRNA targets [11]. MiRNAs are grouped into families based on their targeting properties, which depend primarily on the identity of their extended seed region (see below) and members of the same seed family are usually evolutionarily related [5]. A spectrum of miRNAs in most metazoan animals usually consists of species-specific miRNAs that are evolutionary novelties and evolutionary conserved miRNA families that are found also in other animal groups [12]. For example, of the 500-plus canonical miRNA genes confidently identified in the human genome, 296 fall within 177 seed families conserved among placental mammals [5].

3.1.2 MiRNA biogenesis

In animals, miRNA genes are transcribed by RNA polymerase II into primary miRNA transcripts (pri-miRNAs) several hundreds to thousands of nucleotides long. Each pri-miRNA contains at least one 60- to 80- nucleotide stem-loop structure [13]. In the first miRNA biogenesis step, pri-miRNAs are processed in the nucleus by a protein complex called microprocessor, composed of RNase III Drosha and its binding partner DGCR8 (known as Pasha in flies and nematodes). Drosha cuts each strand of the stem of the pri-miRNA hairpin with a 2 bp overhang, which releases a ~ 60 nt stem-loop precursor called pre-miRNA [5]. The pre-miRNA is then exported into the cytoplasm by Exportin-5, where it is further processed by another RNase III enzyme Dicer. Dicer cuts both RNA strands close to the loop region of the hairpin, which liberates the ~ 22-25 nt miRNA duplex. The miRNA duplex has a ~ 2 nt 3' overhang on each end and is composed of the miRNA guide strand hybridized to its passenger strand, often called the miRNA* („miRNA star“). The miRNA duplex is subsequently loaded by one of the members of the Argonaute (AGO) protein family, which selects

one strand to become the mature miRNA (guide strand), while the other strand is degraded (passenger strand, miRNA*). AGO proteins loaded with mature miRNAs form RNA-induced silencing complex (RISC), which is driven by guide miRNA to the target mRNAs and other transcripts, promoting their repression (Figure 1A) [14].

In addition to the canonical biogenesis described above, several non-canonical pathways, which do not require all of the proteins mentioned above, have been characterized [14]. Mirtrons are a major class of non-canonical miRNAs produced by Drosha-independent pathway. They originate from introns that, once spliced, form pre-miRNA hairpins that are exported into the cytoplasm and immediately processed by Dicer, bypassing the microprocessor cleavage [15]. Mirtrons are present in *A. thaliana*, *C. elegans*, *D. melanogaster* [16, 17] and mammals, where hundreds of mirtron loci have been identified [18]. Drosha-mediated processing is also bypassed in miRNAs derived from other non-coding RNAs, such as tRNAs or snoRNAs [19–21]. Although Dicer-independent miRNA processing is rare, there is an example of miRNA generated via this pathway which is all the more interesting (Figure 1A). Mir-451 is one of the most highly expressed miRNAs in erythrocytes and it is required for proper erythroblast maturation [22]. In contrast to other non-canonical miRNAs, miR-451 does require Drosha for its processing. However, the resulting pre-miRNA is too short to be processed by Dicer and instead is cleaved directly by Ago2 and further trimmed by a poly(A)-specific ribonuclease [23–25]. The existence of alternative biogenesis pathways demonstrates evolutionary flexibility of miRNA biogenesis. However, it is notable that vast majority of vertebrate miRNAs follow canonical biogenesis pathway, and functional relevance of most non-canonical miRNAs, with notable exceptions such as miR-320 or miR-451, remains elusive [13].

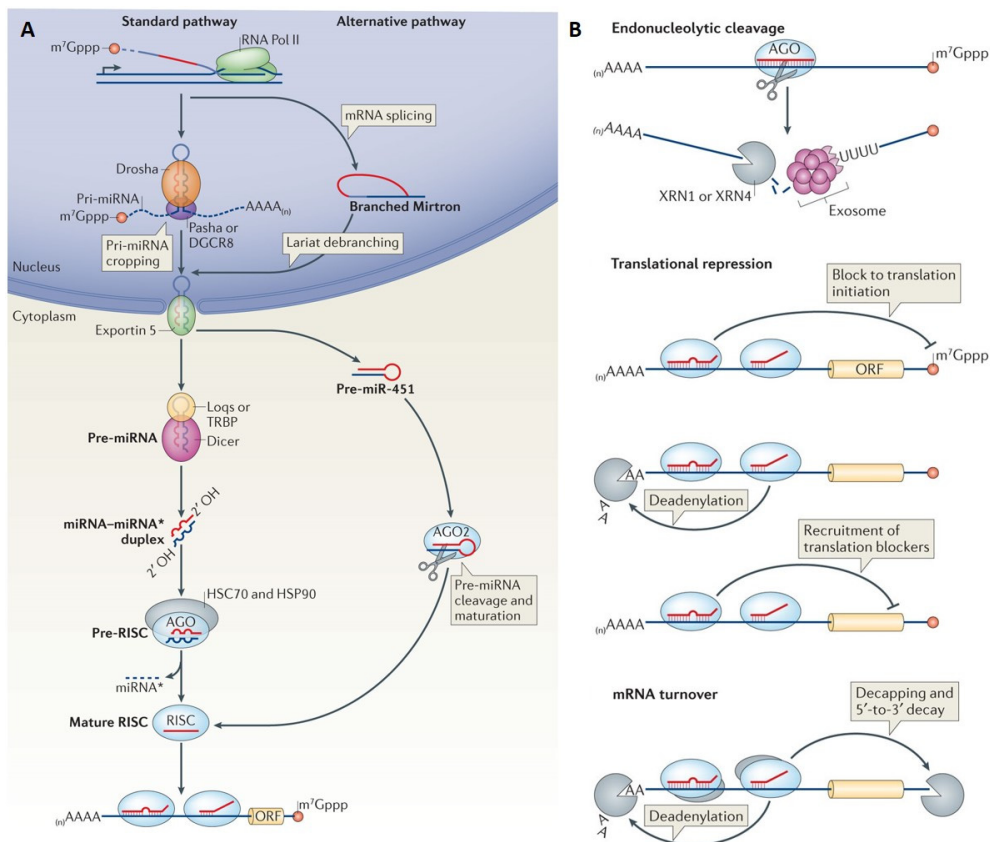


Figure 1. A) Scheme describing canonical and alternative miRNA biogenesis in animals. B) Scheme depicting various modes of miRNA target repression. Figure adapted from [26].

3.1.3 MiRNA regulatory mechanisms

The mature miRNA guides RISC to complementary sequences located mainly in the 3' untranslated region of its target mRNAs. Several mechanisms for subsequent post-transcriptional repression have been proposed, including target slicing, repression of translation, or acceleration of mRNA decay (Figure 1B) [4]. MiRNAs with highly complementary binding can trigger mRNA cleavage, a mechanism derived from the ancestral RNAi pathway, in which target slicing provides an anti-viral and anti-transposon defense. While this mode of miRNA action is common in plants [27], in animals it is very rare, with miRNA-directed slicing reported for only 20 cellular transcripts [5]. Moreover, three of the four human AGO clade proteins are catalytically inactive. Therefore, most animal miRNAs mediate target gene regulation through a mechanism that is independent of RNA endonucleolytic cleavage.

The vast majority of miRNAs in animals recognize their targets through much less extensive pairing of their seed sequence. The seed represents nucleotides at positions 2 to 7 from the 5' end of the miRNA, which can be often supplemented by nucleotide at the 8th position (so called "mate") or an A across from miRNA nucleotide 1, or both, to make 7 or 8 nt long binding sites [4]. These 7–8 nt sites mediate the bulk of the repression for each miRNA [28]. Upon the formation of a miRNA–mRNA interaction, AGO recruits a member of the GW protein family TNRC6. TNRC6 interacts with the poly(A)-binding protein (PABPC) associated with the mRNA poly(A) tail and also recruits deadenylase complexes. The deadenylases shorten the poly(A) tail, which in most systems causes mRNA destabilization through decapping and 5'-to-3' exonucleolytic decay [29]. Recruitment of TNRC6 also causes translational repression [26]. However, an overwhelming evidence from miRNA knock-in or knock-out experiments with subsequent quantitative readouts of mRNA and protein levels revealed that destruction of target mRNAs is a dominant mode of miRNA action [30–33]. mRNA destruction explains as much as 84% of miRNA-directed repression in mammals, with only modest (11-16%) effect attributable to translational repression [30, 31]. From an experimental perspective, this is good news for researchers aiming to study miRNA-mediated regulation. It means that changes of mRNA levels (which are measured much easier than protein or translational efficiency effects) provide a good quantitative readout of miRNA action, e.g. after its experimental perturbation. Notable exception to this phenomenon is the early zebrafish embryo, where miR-430 reduces translational occupancy of target mRNAs, thus repressing their translation, and only later triggers the mRNA decay [34]. This effect appears to be dependent on the cellular context, rather than specific miRNA properties. In cells with robust surveillance mechanisms, such as post-embryonic cells, shortening the tail of an mRNA reduces its stability but does not change its translational efficiency, whereas in early embryos, shortening the tail of an mRNA does not change its stability but dramatically decreases its translational efficiency [5, 35, 36].

3.1.4 Plant miRNAs

While plant miRNA pathway bears many similarities to the animal miRNAs, there are also considerable differences in plant miRNA genomic organization, biogenesis as well as the way miRNAs mediate post-transcriptional repression of their targets. Plant genomes typically encode a hundred to several hundreds of miRNA genes, which represent independent transcription units [37]. As opposed to animals, intronic miRNAs in plants are very rare, with only few cases identified to date [37–39]. While in animals roughly half of miRNA genes are located in clusters often composed of different mature miRNAs, plant genomes contain fewer cases of miRNA clusters and

they mostly encode for homologous miRNAs [40]. Plant pri-miRNA transcripts are much more variable in length (ranging from ~ 60 nt to several hundred nt) and bear more complex structures than animal pri-miRNAs [41]. In plants, DICER-Like 1 (DCL1), a dicer homolog, is responsible for both cropping the pri-miRNA and dicing the pre-miRNA, and these steps of miRNA biogenesis both occur in nucleus in specialized subnuclear regions called D-bodies, while in animals second cleavage occurs in cytoplasm [37, 42].

A unique feature of plant miRNAs is the 2'-O-methylation of the 3' most nucleotides of the miRNA/miRNA* duplex by the action of HEN1 methyltransferase. This modification prevents untemplated uridylation of the 3' ends, which is crucial for miRNA stability in plants [43, 44]. Importantly, 2'-O-methylation can also significantly impair detection and quantification of miRNAs due to severe inhibition of polyadenylation and ligation reactions employed by numerous quantification assays [45]. Therefore researchers should keep this in mind as appropriate methods and/or modifications to the protocols have to be chosen for quantification of miRNAs bearing 2'-O-Me modification [46, 47]. Another difference between plant and animal miRNAs is in their mode of action. Unlike in animals, target binding of plant miRNAs requires a nearly full complementarity, leading to the endonucleolytic cleavage of the target by the AGO protein, although translational inhibition also occurs [48, 49]. Due to requirement for extensive complementarity, each plant miRNA has at least two orders of magnitude fewer targets compared to animal miRNAs [41]. This also makes it much easier to identify targets *in silico*.

3.1.5 MiRNA annotation and nomenclature

Canonical miRNAs can be characterized by distinctive set of features and striking sequence conservation, not seen in other types of small RNAs [50, 51]. These structural features are typically used in conjunction with expression criteria to identify and annotate novel miRNAs (reviewed in [52]). Undoubtedly, the most popular miRNA sequence database is miRBase (www.mirbase.org; Kozomara et al., 2019). miRBase accepts new entries from published miRNA papers, and in its latest release (version 22.1) contains 38589 entries across 271 organisms. However, identifying what is and what is not a miRNA has not been a straightforward task, which in combination with ever-increasing amount of next generation sequencing (NGS) data and relatively lenient criteria for new entries to miRBase led to significant amount of false-positive entries. It was recently found that almost half of all animal entries in miRBase are not derived from *bona fide* miRNA genes [53]. In accordance, numerous other studies called large part of miRBase into question, showing that many entries actually originate from fragments of longer RNAs as a result of degradation [54–57]. At the same time, novel miRNAs without miRBase annotation have been identified and validated by several studies [57–59]. As the nomenclature for miRNA products evolved over the years, changes to the naming and/or sequence of miRNAs between miRBase releases were also introduced. Such misannotations can lead to erroneous conclusions on miRNA biology (e.g. Engkvist et al., 2017). In response, efforts have been made to identify and track changes to the miRBase [61, 62] as well as to build high-confident miRNA databases with uniform annotation system [53, 63, 64]. The aforementioned observations warrant caution for researchers studying miRNAs that have been introduced in later miRBase releases, species-specific miRNAs without evolutionary conservation, or when working with degraded samples, which is often unavoidable in clinical setting.

3.1.6 IsomiRs – miRNA variants

Often, miRNAs are thought of as a single defined sequence, perhaps owing to the way they are annotated in popular miRNA databases, such as miRBase [6]. However, with the advent of NGS technologies, it has become clear that mature miRNAs exist in several variants differing in their length and/or sequence composition. These variants are called isomiRs and can be produced by several mechanisms (Figure 2) [65]. IsomiRs heterogeneous in length at either or both ends can arise due to imprecise cleavage of pre-miRNA by Drosha and/or Dicer [26], or through action of exonucleases “nibbling” of the 3' end [66, 67], which leads to production of isomiRs with matching sequence to the pre-miRNA (templated modifications). In addition, post-transcriptional addition of one or more bases or RNA editing can lead to isomiRs whose sequence may not exactly match the parent gene (non-templated modifications) [5, 68].

Although many questions regarding the regulation of isomiR biogenesis and their functional relevance remain to be elucidated, isomiRs have been demonstrated to associate with RISC and regulate target mRNAs, demonstrating they are real physiological miRNA variants [69]. The most frequently observed type of isomiRs both in plants and animals are 3' variants, while 5' variants are less common [65]. Any variation at the 5' end of the miRNAs has a potential to alter targeting repertoire due to alterations to the seed sequence. The relative rarity of the 5' isomiRs thus suggests that evolutionary selection has prioritized processing precision to maximize targeting specificity [5]. Nevertheless, there are cases where 5' isomiRs expand regulatory repertoire of conserved miRNA genes, such as miR-10-5p in *Drosophila* [70] or miR-223 in mouse neutrophils [71], and at least 12 out of 90 conserved miRNA families produce one or more 5' isomiRs at considerable levels [72]. It has been hypothesized that 5' isomiRs may potentially function to increase signal-to-noise ratio of miRNA targeting by distributing off-target effects while still cooperatively targeting the core network of targets [73]. Unlike 5' isomiRs, variations at the 3' ends are not expected to directly alter miRNA targeting specificity. Instead, they may be associated with miRNA processing, stability or trafficking to various cellular compartments [74, 75].

Importantly, it appears to be common that particular isomiR variant is highly abundant, or even surpasses the abundance of the canonical variant listed in miRBase [65, 76]. In addition, expression of isomiRs can be cell-type specific and can change in response to biological stimuli [76–79]. For example, a recent reanalysis of small RNA-Seq data from 126 primary cell and 82 cancer cell samples have revealed that canonical miRNA was always the most abundant variant only for 33% of the detected miRNAs [76]. In addition, on average only 45% of sequencing reads for each isomiR group (i.e. isomiRs originating from the same gene) were assigned to the most abundant isomiR variant, suggesting that each miRNA is present as a relatively heterogeneous isomiR pool, rather than a single defined sequence [76]. Similar observations have been made in other biological models [80, 81]. This heterogeneity of sequences is increased in cancer samples, suggesting that miRNA processing in cancer is skewed towards randomness [76]. These observations have potentially several important implications for researchers studying miRNAs. First, isomiRs can have profound effect on miRNA quantification by commercial hybridization-based assays, or experimental manipulation by miRNA mimics/inhibitors, which are typically designed for the canonical miRNA variants listed in databases. Indeed, Blondal and colleagues found that their qPCR-based assay failed to quantify 60 out of 517 miRNAs in serum, due to their presence as non-canonical isomiRs [82]. Second, methods or analytical pipelines that fail to include

isomiRs may underestimate the presence of miRNAs or miss on relevant biological signal. For example, the expression profile of full set of isomiRs better separated healthy from tumor samples compared to canonical miRNAs only [83, 84]. Third, computational prediction of miRNA targeting may benefit from increased efficiency by taking into account all isomiR variants, as has been demonstrated with *A. thaliana* and human sequencing data [85].

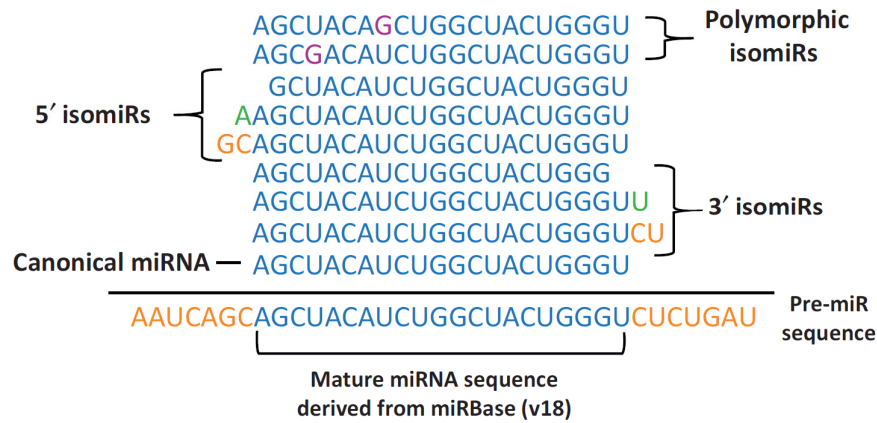


Figure 2. Schematic representation of isomiR variants. Figure reprinted from [65].

3.1.7 Extracellular miRNAs

MiRNAs were originally characterized only within cells. However, first evidence that miRNA may be released into extracellular environment came in 2008 when Chim and colleagues identified miRNAs of placental origin in the plasma of pregnant women [86]. Several reports on circulating miRNAs in serum of patients followed the same year [87, 88] and since then the presence of miRNAs has been reported in virtually every biofluid including serum, plasma, saliva, urine, breast milk, cerebrospinal fluid and others [89–94]. Two populations of extracellular miRNAs exist in biofluids. One can be found in extracellular vesicles (EVs), which is an umbrella term used to describe all secreted vesicles independent of their size and origin [95, 96]. The other population is associated with proteins or lipoproteins, especially AGO2 [97]. These carriers confer high stability to circulating miRNAs, as storage at room temperature, boiling, repeated freeze-thaw cycles, pH changes or even chemical agents and enzymes do not lead to significant degradation [98–100]. MiRNAs are released from cells either through active transport via EVs or as part of protein-miRNA complexes. In addition, passive release from broken or damaged cells, e.g. after injury, or during inflammation, apoptotic and necrotic processes, can occur [101]. Actively secreted miRNAs may act as signaling molecules and exert biological functions in recipient cells, thereby serving for hormone-like communication [101, 102]. EVs that reach recipient cells may activate intracellular signaling pathways, release their content into the cell by membrane fusion or enter the cell via phagocytosis or receptor-mediated endocytosis. Afterwards, various molecules carried in EVs may be released into the cytoplasm, where they may elicit functional response. However, what determines the miRNA content of particular EVs and how they are taken up by potential target cells is not well understood [102].

3.1.8 Circulating miRNAs as clinical biomarkers

The discoveries that miRNAs are present in quantifiable amounts in biofluids, are highly stable and reflect physiological and pathological status of the tissue of origin have brought them to attention

as a new promising group of biomarkers and sparked a whole new avenue of research. Contrary to standard tissue biopsies, sampling of biofluids is quick, minimally invasive and painless and individual miRNAs may be detected with high sensitivity and specificity, which makes them attractive biomarker candidates for diagnosis and monitoring of patients' responses to therapy [103]. In recent years, myriad of studies aimed at identification of miRNA biomarkers for diagnosis, prognosis and monitoring of multitude of human diseases has been published [103, 104]. This led to initiation of several clinical trials as well as international efforts aimed at deeper understanding of circulating miRNA function and standardization of the miRNA analysis field, e.g. Extracellular RNA Communication Consortium (ERCC; www.exrna.org) or CANCER-ID (www.cancer-id.eu). Despite the great promise of circulating miRNAs, there have been several issues that hamper the progress in the field [105, 106]. These are partly associated with technical difficulties accompanying workflow of miRNA quantitative experiments [107, 108]. Therefore, attention must be paid to the technological aspects of the measurements and they are discussed in the next chapters.

3.2 Quantitative miRNA profiling

Since the early days of miRNA research, quantitative profiling of miRNA expression has been crucial for understanding biological roles of miRNAs in health and disease and has become even more important for utilization of miRNAs for potential diagnostics. The process of obtaining quantitative miRNA profile from the sample involves multiple steps (Figure 3), which are discussed below.

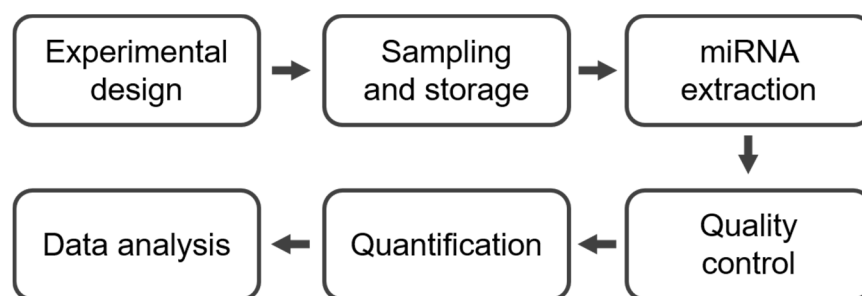


Figure 3. Schematic of the main steps of miRNA analysis workflow.

3.2.1 Experimental design

Proper experimental design, including a reasonable number of biological and technical replicates are crucial to every RNA expression study. Biological replicates are used to evaluate the variability between individuals to allow generalized conclusions about particular condition, whereas technical replicates help to tackle inevitable variability arising from technical workflow. The exact number of biological replicates required for each study is dependent of various factors, including inter-individual genetic variation (e.g. heterogeneous human population vs inbred mice), sampling bias and expected amplitude of differences between studied conditions [109]. Special attention is required for circulating miRNA analysis from biofluids, as it is known that the miRNA spectrum may vary not only due to pathophysiological processes, but also with the time of the day, diet, gender, age, alcohol consumption, medications, etc. [110–112] and well documented is also inter-individual variability of the basal miRNA levels [113, 114].

The character and number of technical replicates is also dependent on the nature of the sample, quantification platform and the amount of variability introduced by particular steps in the workflow.

Tichopad and colleagues investigated the effect of replication on different levels of PCR-based workflow and found that sampling introduced largest variability with solid tissues, cell cultures and single cells. The authors also found that bias introduced by reverse transcription (RT) is substantially higher compared to the bias from qPCR and suggested to invest technical replicates into the RT step, rather than qPCR as is typically done by many researchers [115]. With RNA-Seq-based readouts, it has been a question whether a higher number of biological replicates or higher sequencing depth (i.e. number of reads obtained from each sample) is better. Today, this has been addressed by several studies, which unequivocally arrived to the same conclusion that higher number of biological replicates is preferable over deeper sequencing [116, 117]. At least three biological replicates per group have been recommended [118], but recent study reported that such experiments are typically underpowered, and while results improve gradually with increasing replication, at least 12 biological replicates are needed to correctly detect 90% of differentially expressed genes [119]. RNA-Seq is considered highly reproducible and thus technical replicates are not typically performed [116].

3.2.2 Sampling and storage

Sampling and storage of material is a crucial step as it can have critical impact on the quality of RNA and subsequently on the resulting data and conclusions drawn. Widely used methods for tissue preservation include snap freezing, storage in commercial preservation buffers such as RNAlater (Thermo Fisher) or formalin fixation and paraffin-embedding (FFPE), which is common in clinical setting. However, FFPE leads to RNA degradation and therefore other methods are preferable. It has been initially argued that miRNAs are relatively stable in degraded RNA samples [120, 121] and it is true that there are differences in the rate and susceptibility of various RNA classes to degradation [122]. However, it has become clear that, at least with NGS and microarray-based readouts [56, 123], but also some qPCR assays [124], degraded RNA samples introduce bias into measured miRNA profiles, as short fragments of high molecular RNAs interfere with the measurement. Thus, assessment of RNA integrity should not be omitted for miRNA profiling experiments from cells and tissues (see chapter 3.2.4).

In biofluids, several studies have documented extraordinary miRNA stability in samples stored at various temperatures and after repeated freeze-thaw cycles [88, 100, 125], although the impact of storage conditions is still debated [126]. A major complication in miRNA analysis of blood-based samples, such as serum or plasma is contamination with miRNAs derived from lysed blood cells, and in particular hemolysed erythrocytes [127, 128]. Even minimal contamination with blood cells can increase certain miRNAs by up to 50-fold, which seriously confounds the analysis [128]. Therefore, care should be taken during pre-analytical phase to minimize cellular contamination, e.g. by proper blood collection and centrifugation steps, and plasma and serum samples should be screened for hemolysis prior to analysis (see chapter 3.2.4) [127, 128].

3.2.3 MiRNA extraction

General principles for the isolation of miRNAs are similar to that of longer RNAs, except that some protocols are modified to retain, alternatively enrich the small RNA fraction. The extraction methods may broadly be grouped into three categories: i) organic extraction (guanidine-phenol-chloroform based method), ii) filter-based methods (derivatized silica), and iii) magnetic particles-based methods [129]. While standard organic extraction using TRIzol reagent is typically sufficient for

cells and tissues, there are some caveats that one should keep in mind. It has been reported that TRIzol extraction leads to selective loss of short structured RNAs with low GC content when starting with low amount of cells [130]. This problem can be minimized by extracting the aqueous phase after TRIzol lysis with commercial spin column [131] or increase the concentration of ethanol in the purification step to 80%, which leads to decreased solubility of RNA pellet and minimization of small RNA losses [132]. Commercial isolation kits typically offer two regimes for RNA extraction: i) isolation of total RNA including small RNAs and ii) specific enrichment for small RNA fraction < 200 nt. While enrichment for small RNA can lead to lower background signals with some of the older quantification technologies such as northern blot and microarray, it usually results in lower total yield of small RNAs and may introduce bias to relative miRNA profiles. Isolation of total RNA is therefore preferred [133].

MiRNA extraction becomes particularly important when working with biofluids as the total level of RNA may be very low. Many efforts have been made to compare different isolation protocols from biofluids and although there is no clear consensus, commercial kits from Qiagen and Exiqon (now part of Qiagen) tend to rank higher than others [129]. Common observation from these studies is unexpectedly large variability of replicated extractions, stressing the need for rigorous quality control (see chapter 3.2.4). In addition, all methods tend to introduce some bias due to preference to certain miRNAs relative to others. A recent RNA-Seq study compared five isolation kits for serum and found that sample clustering was clearly driven by kit-specific differences [134]. Therefore, results obtained using different isolation protocols are essentially incomparable.

Regardless of the isolation method used, it is advisable to first perform a small pilot experiment to explore the performance of the protocol, before a more complex study is initiated [115, 135]. Such empirical optimization can identify optimal input and elution volumes for the highest yield, while maintaining low level of carryover inhibition [136–138]. In addition, usage of carriers during RNA isolation from low-concentrated samples may improve yield as well as reproducibility [136, 139]. However, care should be taken to choose non-RNA-based carriers such as glycogen, BSA or linear acrylamide if RNA-Seq is used for readout, due to risk of interference [109].

It has been argued that most circulating miRNAs in plasma and serum are derived from blood or endothelial cells and the contribution of miRNAs derived from diseased cells is rather low [106]. In such cases, isolation of EVs may provide fraction that is enriched in disease or tissue-specific miRNAs. However, EVs released from different cell types (and even from a single cell type) are heterogeneous in size and in protein, nucleic acid and lipid content [96]. Different separation methods enrich for single or multiple EV subtypes with diverse composition and variable purity, thus identifying method-dependent EV content and function [140, 141]. A more than 1000 unique isolation protocols have been identified from a survey of 1742 experiments, demonstrating enormous heterogeneity of EV isolation procedures [140]. As a response, international efforts to establish experimental guidelines, reporting and standardization have been launched [95, 140, 142]. Detailed discussion of EV extraction and miRNA extraction from EVs is out of the scope of this thesis and readers are referred to the recent papers addressing this issue [143, 144].

3.2.4 Quality control

Controls that test for confounding technical variation are essential for any gene expression study [145]. When analyzing tissues and/or cells, sufficient RNA quality is critical. RNA quality is typically

assessed by automatic electrophoresis systems, such as Bioanalyzer or Fragment Analyzer (both from Agilent), which calculate a so called RIN (RNA integrity number) or RQN (RNA quality number) values from 1 to 10, usually by assessing the ratios of 18S and 28S ribosomal peaks as well as other parameters [146]. However, as this control is based on RNAs other than miRNAs, it essentially represents an indirect measure for the quality of miRNAs. Similarly, RNA quantities are typically normalized based on quantity of total RNA. Nevertheless, the seeming increase of small RNA fraction, measured as percentage of total RNA, is observed in samples with low RIN values, resulting from degradation products, which may negatively influence the results, as discussed previously. Therefore, assessment of RNA quality should be performed also for profiling of miRNAs from tissues and cells [124].

When analyzing circulating miRNAs from biofluids, standard quantification and quality control methods are not reliably applicable. Instead, a fixed volume of the biofluid is typically used and efficiency of the extraction procedure can be tested by spike-in controls. A spike-in control is a foreign small RNA molecule of same length as native miRNAs without sequence homology to the endogenous miRNAs [147, 148]. The spike-ins are added to the sample prior to the extraction and their recovery is then measured (e.g. by RT-qPCR) and compared to the standard sample to identify problematic extractions (Figure 4). Several different spike-ins can be added at various stages of the workflow to identify steps introducing technical variation, e.g. addition before RT can identify potential inhibition in downstream RT-qPCR (Figure 4). As the methods for reliable quantification of isolated miRNA from liquid biopsies are not currently available, sufficient quantity can be judged by measurement of few selected miRNAs that are expected to be present in the sample and serve as endogenous control [136, 149].

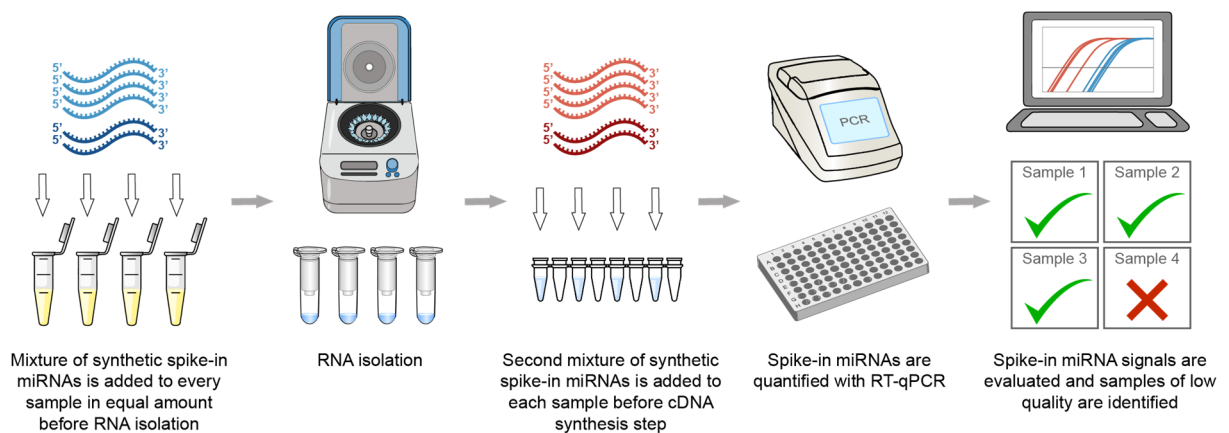


Figure 4. Concept of miRNA spike-in quality control. Samples are spiked with known amounts of exogenous spike-in miRNAs before extraction and RT. Quantification of the spike-in levels via RT-qPCR identifies material losses, inhibition, outlier samples and steps introducing severe variability.

A major complication in miRNA analysis of serum/plasma samples is contamination with miRNAs derived from lysed blood cells, and in particular hemolysed erythrocytes [128, 150]. Hemolysis can be assessed prior to extraction visually or spectroscopically by measuring the absorbance of hemoglobin at 414 nm [127]. An alternative approach is to measure the ratio of miR-23a, which is insensitive to hemolysis, and miR-451, which is highly enriched in erythrocytes [149]. This ratio increases with increased level of hemolysis and can be used to assess hemolysis levels even when the original sample is no longer available. The same strategy can be used to assess contamination

with other cell-types if needed. For example, miR-425 level may reflect contamination with platelets [151]. To ease the complete quality control of circulating miRNAs from biofluids, commercial [149] as well as freely-available [136] protocols combining exogenous, endogenous and hemolysis controls have been developed.

3.2.5 MiRNA quantification

3.2.5.1 Challenges of miRNA quantification

Quantification of miRNA profiles bears several unique challenges that are not present with long RNA transcripts and that are mainly associated with inherent properties of these small non-coding RNAs [108, 152–154]. These are:

- MiRNAs are very short (~22nt), which impedes efficient hybridization of detection probes or PCR primers and therefore miRNA elongation is typically one of the first steps of quantitative assays.
- Variable GC content of miRNAs results in relatively large interval of melting temperatures, representing a problem for efficient hybridization of nucleic acid duplexes.
- Unlike mRNAs, mature miRNAs lack a common sequence, such as poly(A) tail that would facilitate their selective capture or universal RT.
- Some mature miRNA sequences can be nearly identical and differ in only few nucleotides, which imposes a great challenge for specificity of the assays. Particularly variations in 5' region where the seed match is located can have profound impact on miRNA targeting repertoire, and high specificity is also required for detection of miRNA biomarkers in clinical setting.
- Due to the nature of miRNA biogenesis, the sequence of mature product is also contained in its precursors and it may be desirable to quantify both, or distinguish between these classes, depending on the research question.
- MiRNAs exist as pool of isomiRs heterogeneous in their length and/or sequence, which may interfere with the quantitative assays usually designed for the canonical variant. Without prior sequencing data it is essentially impossible to tell in advance which of the isomiRs is dominant and thus relying on the canonical sequence may lead to biased results.
- Biochemical modifications of terminal nucleotides, such as 3' methylation in plant miRNAs or animal piRNAs inhibit certain enzymatic steps such as polyadenylation or ligation of the adapters to the miRNAs and certain methods are therefore not suitable for their quantification.

Despite these challenges, we and others developed approaches for miRNA quantification, and today three most commonly used platforms are microarrays, RT-qPCR and small RNA-Seq. In the following chapters, RT-qPCR and RNA-Seq-based quantification workflow is described, as these were two major technologies utilized in this thesis.

3.2.5.2 RT-qPCR

RT-qPCR is a gold standard technique for miRNA quantification. It offers highest sensitivity (down to tens of molecules per reaction) and high specificity (in some cases down to single nucleotide difference) [46, 155, 156] with large dynamic range over 7 orders of magnitude [154]. The wet-lab workflow is fast, with sample-to-data time in order of few hours, the methodology is easily adaptable

for any laboratory familiar with qPCR and the data analysis workflows are well established [157]. In addition, RT-qPCR data can be calibrated for absolute quantification. On the other hand, similarly to hybridization-based techniques, only previously known miRNAs and limited spectrum of isomiRs can be detected [152]. Although RT-qPCR is considered low- to mid-throughput method, microfluidic and array formats capable of profiling hundreds of miRNAs in parallel are also available. However, use of RT-qPCR for high-throughput screening may be financially more demanding compared to other global techniques.

The first step in RT-qPCR is to convert the miRNAs into the cDNA by reverse transcription. As discussed above, the short length of miRNAs does not allow usage of standard PCR primers (~18-25 nt) and miRNAs have to be elongated. There are two main strategies to do this: i) extension of miRNAs during RT via specific RT-primers (Figure 5A, 5B) and ii) extension of all miRNAs with a universal sequence prior to cDNA synthesis so that subsequent RT can be performed with universal primer (Figure 5C, 5D). Each approach has its advantages and limitations. Specific RT primers have 3' ends complementary to the miRNA sequence, whereas their 5' ends serve as extensions. The specific RT primers may be linear (Figure 5B) [158], or bear secondary structures (Figure 5A) [46, 159, 160]. While linear primers are easier to design, they suffer from higher background signal as they easily bind to precursor miRNAs and other sequences [154]. The stem-loop structured primers [159], commercially sold as TaqMan miRNA assays (Thermo Fisher), are better at targeting the mature miRNAs and lowering the background signal, as their unspecific binding is reduced, but the disadvantage is hindered RT of isomiRs [46, 161]. Regardless of the primer structure, the limitation of specific RT approach is that only miRNAs that are targeted with RT primers can be quantified. Hence, if researchers want to add new targets later, they have to revert to the RNA (rather than starting from the cDNA) and include additional primers in the RT reaction. The complex design of structured RT primers and higher cost of some commercial assays also represent a limiting factor, although cost-efficient alternatives were recently introduced [46].

The possibility to reverse transcribe all miRNAs in a single reaction is the greatest advantage of the second, universal RT-priming approach (Figure 5C, 5D). It relies either on addition of artificial poly(A) tail by poly(A) polymerase (PAP) and subsequent priming with oligo(dT) primer (Figure 5C) [162], or on ligation of an adapter to the 3' end of miRNAs by T4 RNA ligase and universal priming with oligo complementary to the adapter sequence (Figure 5D) [163]. In addition, combination of 3' polyadenylation and 5' adapter ligation is commercially available (TaqMan Advanced miRNA Assays, Thermo Fisher). Downside to the universal RT approaches is the extension of other RNA classes, which increases the background, lower specificity for mature miRNA sequences and variable efficiency of the enzymatic steps due to sequence preferences or terminal modifications of miRNAs [108, 152, 154]. Since the universal RT-primers reverse transcribe all miRNAs, they do not contribute to the specificity of the analysis. Specificity is then conferred only by single qPCR primer, as the second qPCR primer is typically universal and binds to the extension sequence. To improve specificity and optimize melting temperature, primers containing LNA bases with improved hybridization properties were introduced (commercially sold as miRCURY LNA assays by Qiagen) [158, 164]. However, their higher price and complicated design hinder their wide utilization [152].

Whichever the RT strategy, cDNA is subsequently amplified using standard qPCR with either dye- or probe-based detection. Using dyes lowers the cost and allows assessment of the reaction specificity by melting curve analysis, although certain risk of false-positive detection prevails [165].

As an alternative, hydrolysis probes binding to a specific site in the target amplicon and giving rise to specific fluorescence can be used. They were introduced for miRNA applications by Chen and colleagues (2005), commercially provided as TaqMan miRNA assays (Thermo Fisher). However, in this system the hydrolysis probes bind to a universal sequence present in all RT primers and thus do not contribute to the specificity of the reaction. The probe design is improved in the newer generation of TaqMan advanced miRNA assays (Thermo Fisher).

In summary, RT-qPCR-based miRNA analysis is suitable for smaller studies analyzing a predefined set of miRNAs, for validation of high-throughput techniques, for applications that require absolute quantification, and for routine diagnostics when high sensitivity and/or specificity is required [129].

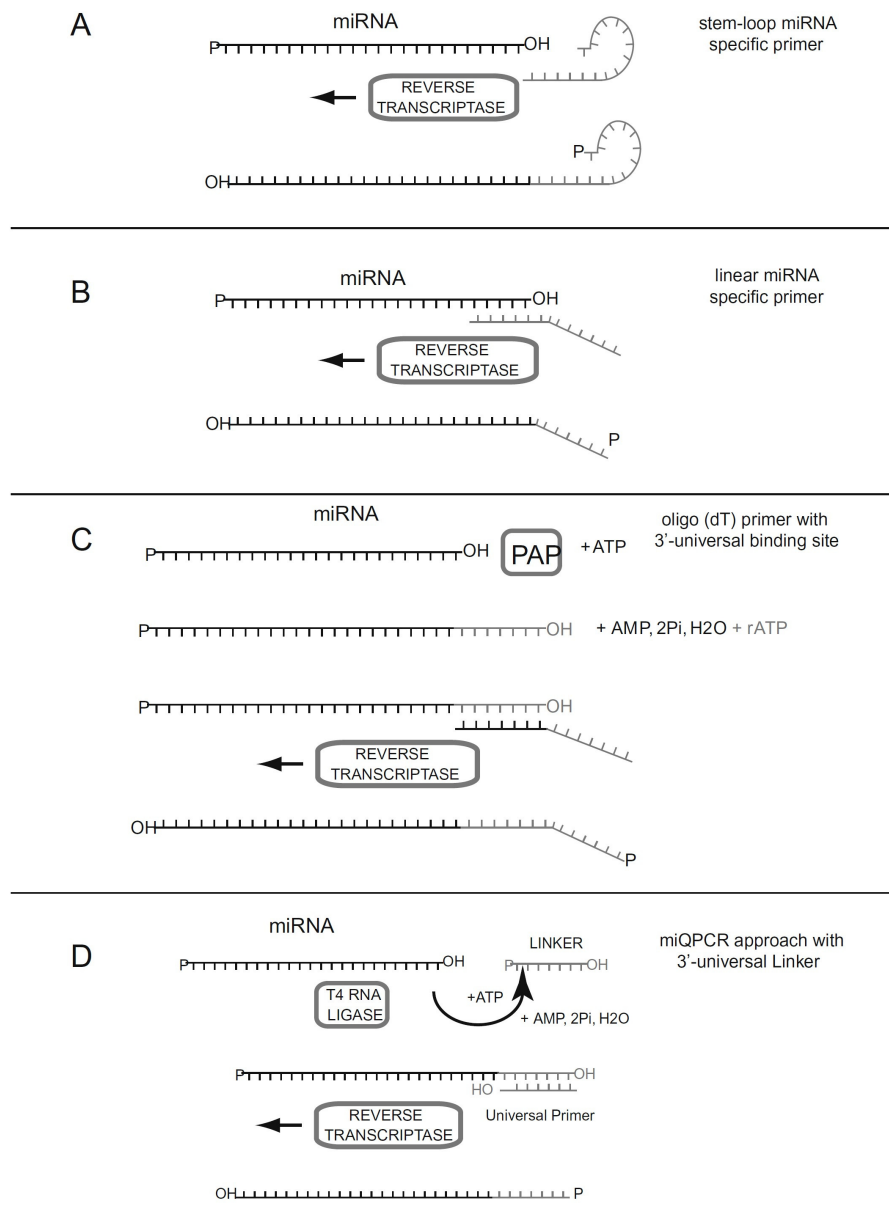


Figure 5. Schematic representation of strategies for cDNA generation from miRNAs. A) Targeted RT with specific stem-loop primers. B) Targeted RT with specific linear primers. C) Universal RT of all miRNAs by poly(A) tailing and oligo(dT) priming. D) Universal RT of all miRNAs by linker ligation and universal priming. Figure reprinted from [152].

3.2.5.3 RNA sequencing

Small RNA-Seq is becoming a leading technology for miRNA research. RNA-Seq does not suffer from saturation effects and offers higher sensitivity and broader dynamic range than microarrays, although it may not match optimized RT-qPCR [155, 166]. It is a high-throughput technology, currently allowing convenient multiplexing of up to 96 samples with commercially available kits and even more with in-house technologies [167]. Unlike previous technologies, RNA-Seq can detect both novel and known miRNAs and allows precise identification of closely related miRNAs, isomiRs, edited miRNAs as well as other classes of small RNAs [166]. However, not every small RNA obtained is a functional miRNA; degradation fragments from longer RNAs can introduce artefacts into the data [56] and significant computational effort is needed to properly analyze and interpret the results, which can be considered as one of the main drawbacks of the technology. Another downsides include biases introduced during library preparation (see below), lack of absolute quantification, challenging data analysis or lack of standardization [129].

For small RNA-Seq, RNA transcripts need to be converted into sequencing libraries. Several protocols for miRNA library preparation are available (Figure 6). The most standard protocol employs two sequential ligations of adapters to the 3' and 5' ends of the miRNA (Figure 6A) [168]. It takes advantage of the 5' phosphate and the 3' hydroxyl groups on the miRNA termini to selectively target and enrich miRNAs using ligases that require these terminal groups. After ligation of the adapters, a universal RT-primer complementary to the 3' adapter is used for cDNA synthesis. The cDNA library is then PCR-amplified using primers complementary to the adapters. These primers also introduce the flow-cell binding sequences and sample-specific barcodes. The final amplified library typically consists of approximately 120-150 bp of adapter sequences plus an insert of the original miRNA sequence of 20–30 bp, which makes a total of 140–180 bp. Longer products are generated from adapter ligation to non-miRNA species, including tRNAs, snoRNAs, piRNAs and other RNAs having 3' hydroxyl and 5' phosphate termini. These can either be retained as part of the library or removed by bead-based size selection or polyacrylamide gel electrophoresis (PAGE) purification [129].

One of the issues with the ligation-based library preparation methods is the accumulation of unwanted adapter dimers formed by the direct ligation of the 3' and 5' adapters without an RNA insert. These artefacts amplify during the PCR and may consume substantial portion of sequencing reads. One way to exclude them from the library is to purify only desired products via PAGE after PCR amplification. Alternatively, several modifications to the standard protocol have been proposed in order to minimize the formation of adapter dimers [169–173].

Another problem is that ligations of the adapters introduce substantial bias to the representation of miRNA sequences in the final data [174]. It arises due to variable efficiency of the adapter ligation to different miRNAs, in dependence of their sequence composition and secondary structure [175, 176]. Consequently, many miRNAs are over or under represented in the data, and some may even completely drop out. The measured values provide therefore a distorted picture of the real miRNA levels and the quantification must be performed relative to a reference or standard sample (i.e. case vs control). To reduce the ligation bias, three approaches have been developed so far. First method uses adapters with a stretch of random nucleotides at their ends adjacent to the ligation site, which provide a preferred pair of sequences for each miRNA and equalizes the ligation efficiency (Figure 6B) [177]. Second method omits ligation altogether and employs poly(A) tailing

and subsequent template-switching during RT to introduce adapter sequences (Figure 6C) (commercial kits by Takara and Diagenode). Third method relies on the ligation of a single combo adapter to the 3' end and subsequent circularization of the molecule prior to RT (Figure 6D) [178]. The intramolecular circularization is more efficient than intermolecular ligation and should introduce less bias. In addition, strategies to computationally correct ligation bias *post hoc* are also being developed [179, 180].

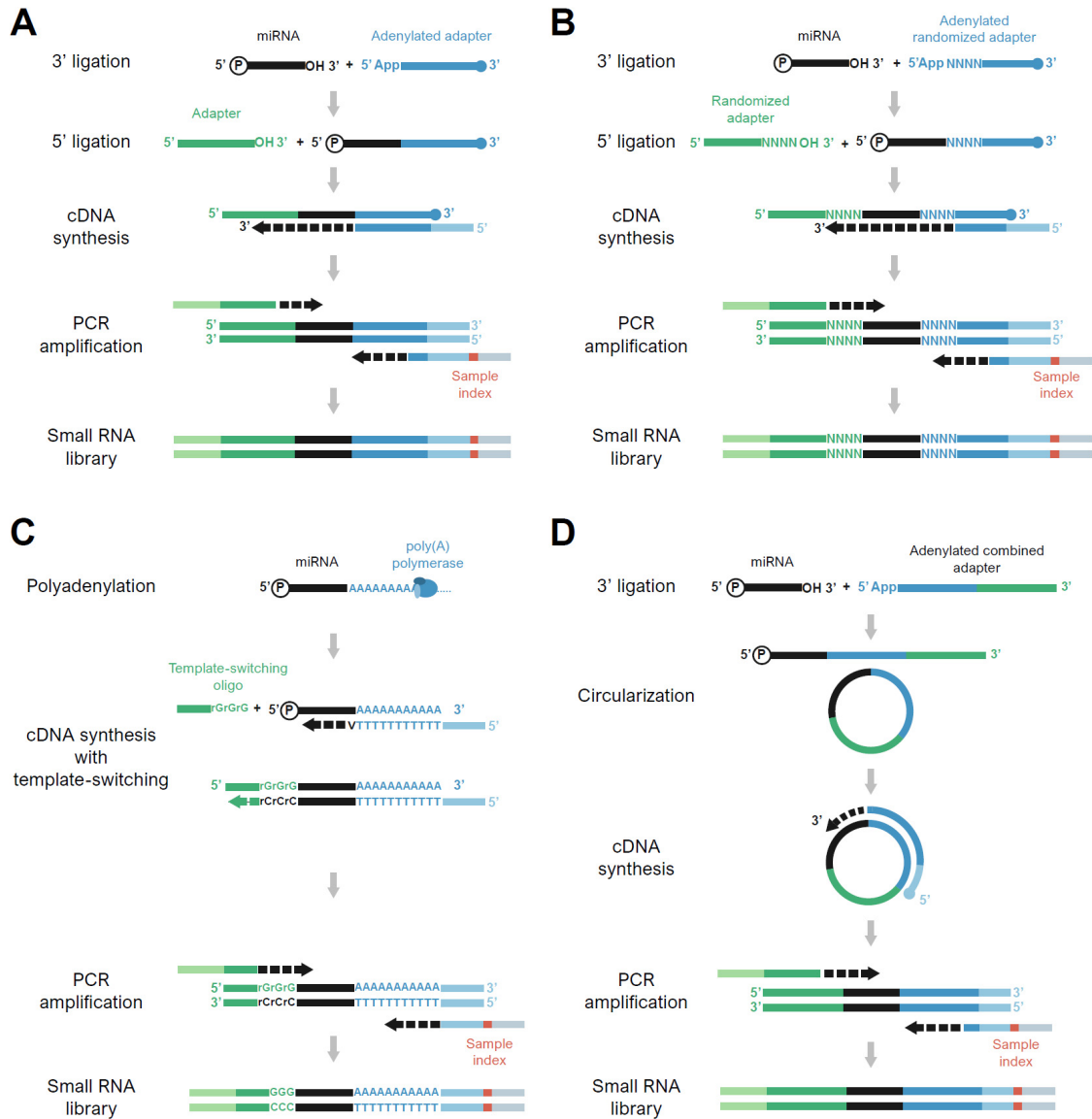


Figure 6. Schematic representation of methods for small RNA-Seq library preparation. A) Sequential ligation of two defined adapters. B) Sequential ligation of two adapters with randomized nucleotides to reduce ligation bias. C) Poly(A) tailing and subsequent oligo(dT)-primed RT with template switching. D) Single adapter ligation and circularization.

Another source of bias, although less prominent than adapter ligation, is the PCR amplification [109, 174]. PCR is known to suffer from unequal efficiencies in most RNA-Seq workflows. It has been argued that for small RNA-Seq, PCR bias is negligible, owing to short and even length of cDNA templates [176, 181, 182]. However, recent studies that assessed the impact of PCR bias via unique molecular identifiers (UMIs) call this into question [183, 184]. UMIs are stretches of

random sequences that tag each RNA molecule with a unique barcode [185]. Therefore, after sequencing it is possible to computationally identify PCR duplicates and correct for PCR bias. The usage of UMIs has become standard in the single cell RNA-Seq field, and recently it has been introduced also for small RNA-Seq in a commercial kit from Qiagen.

Several recent studies compared the performance of small RNA library preparation protocols [47, 184, 186–188]. In general, all protocols showed similar performance in terms of sensitivity, reproducibility, accuracy or diversity of detected miRNAs, but strongly differed in the ability to capture true representation of miRNA levels, proportion of miRNA reads in the libraries, or detection of isomiRs. Polyadenylation-based protocols minimized sequence bias, but produced considerably more unwanted side-products, low mapping rates, higher proportion of false isomiR detection and as a result did not perform well for biological samples [47, 184, 186]. Among ligation-based protocols, methods utilizing adapters with randomized nucleotides had less bias and better miRNA coverage at low sequencing depths [47, 184, 186, 188]. The studies also concluded that relative quantification is consistently accurate despite biases and that the results are reproducible between various laboratories, as long as the same protocol is used [47, 187].

Small RNA-Seq is still undergoing novel developments and currently will find its applications mainly for screening and discovery studies that will benefit from its immense power to profile global miRNA profiles with single-nucleotide resolution in large amount of samples in parallel [129].

3.2.5.4 Other techniques

In addition to previous approaches, other techniques for miRNA quantification, often aimed at standardized routine applications are emerging. Examples include the HTG EdgeSeq miRNA Whole Transcriptome Assay (abbr. EdgeSeq; HTG Molecular Diagnostics) and nCounter system (NanoString Technologies) [189, 190]. Both are based on hybridization for recognition of targets, but the hybridization is performed in solution, which increases its efficiency. EdgeSeq utilizes a large pool of capture probes to bind target miRNAs (>2000) and single strand specific nuclease is subsequently used to remove all unbound probes. The remaining probes are then amplified and sequencing is used for readout. Compared to conventional small RNA-Seq, the EdgeSeq workflow completely avoids ligation bias and allows the analysis of crude biofluids (down to 15 μ l), which also eliminates bias in the extraction procedure. nCounter relies on multi-step hybridization of miRNAs to probes followed by splinted ligation. After ligation of targets to miRNA-specific extension sequences, streptavidin capture probes co-hybridize with reporter probes carrying fluorescent barcodes to create target-probe complexes. These are immobilized and, after several washing steps, fluorescent barcodes, which correspond to the amount of specific miRNAs, are digitally counted. This leads to higher precision and dynamic range compared to microarrays [153]. In addition to these technologies, various other methods based on probe-ligations [191, 192], isothermal amplifications [193, 194] or biosensors [195, 196] have been described.

3.2.6 Data analysis

Although each platform includes technology-specific data processing steps, these can be broadly categorized into i) data quality control and pre-processing, ii) normalization, and iii) secondary analysis.

The first assessment of data quality is performed during data pre-processing. It includes analysis of positive and negative controls, replicates, melting curves in RT-qPCR, or examination of various parameters reflecting the quality of a sequencing run such as the number of identified reads, quality score distribution, flow cell loading, presence of over-represented sequences, etc. Data are then pre-processed to remove technical artefacts, identify unreliable measurements, and produce a matrix of measured target quantities. Instruments for qPCR are usually provided with software that performs most pre-processing steps with minimal user intervention, such as amplification curve modelling, fluorescence normalization, baseline subtraction, thresholds setting and C_q readout [129]. In contrast, RNA-Seq data pre-processing is time consuming, computationally demanding and not fully standardized. It includes the trimming of sequencing reads from adapters, removal of sequences with inadequate length or low quality, and finally alignment to a reference sequence [197]. miRNA reads are typically mapped against a reference genome or miRBase. Mapping against the genome sequence provides the most comprehensive view on the data and allows for the discovery of novel miRNAs [109]. Mapping against miRBase or similar databases is significantly faster and avoids issues with reads mapped to multiple genomic locations, which can introduce severe bias if handled inappropriately [109].

After data pre-processing, another round of quality control follows, focusing on the identification of samples of compromised quality. Typically, low quality samples show reduced number of detected miRNAs, reduced overall signal or total number of sequencing reads, deviations from typical expression profiles, or high proportion of reads mapped to longer, degraded RNAs [197, 198]. To avoid subjective calling of failed samples, pass/fail criteria may be set or statistical tests for outlier detection applied [199]. The sample quality may also be evaluated based on the quantification of typically expressed miRNAs, indicators of hemolysis in case of blood-derived biofluids, or other cell type contamination markers, and the examination of exogenous spike-in controls [136].

Next step in the data analysis workflow is normalization. This is arguably one of the most critical steps for comparison of samples and may severely compromise the quality of the data if done inappropriately, in worst case leading to erroneous conclusions [200]. The goal of the normalization is to reduce between-sample variation arising during the experimental procedure. RT-qPCR data of mRNAs are typically normalized to the expression of reference genes with stable expression. Application of this concept to miRNA expression is more problematic, as there are no universal reference miRNAs and it is critical to identify and validate them for every study [152, 200]. A widely used alternative is normalization to other small RNA molecules such as RNU6, RNU6A, RNU44, and 5S rRNA. Although these may be suitable for cell cultures and some tissues, they are inappropriate for normalization of biofluids due to their intracellular character. In addition, they have different biogenesis and processing steps, which may not mirror properties of miRNAs [200]. Normalization to mean expression may also be applied to RT-qPCR, if a large number of miRNAs is quantified [201]. Other normalization strategies include the volume of biofluid, total amount of miRNAs, and spike-in molecules. Main problems of these approaches is that they do not account for variation in RNA quality, input quantity, and individual variation. Recommended strategy is to perform a pilot experiment with a small representative set of samples that is screened for all miRNAs using a global platform and from those select the most promising candidates for validation by RT-qPCR [106, 200].

Normalization of RNA-Seq data aims to minimize differences in library size and composition caused by varying sequencing depth between samples. The general strategy is based on calculating a scaling factor for each sample, which is used to adjust for library size. Numerous strategies have been developed for the sequencing of long RNAs and many are applicable to miRNAs as well (reviewed in [202, 203]).

Secondary analysis depends on the particular objectives of the experiment and biological questions. First step is usually descriptive statistics to identify the number of positive and negative miRNAs. Further, miRNAs with prohibitively low readouts for reliable analysis are discarded. If complete data matrices are needed for analysis, imputation methods may be applied to replace missing values [204]. As the goal of majority of experiments is to identify differences between groups of samples, statistical tests are used to identify likely differentially expressed miRNAs. Standard statistical methods can be usually applied directly on RT-qPCR data analysis. However, RNA-Seq data, where the number of miRNAs typically is much larger than the number of samples, may require more sophisticated models to identify differential expression. The most popular tools are DESeq2 [205], edgeR [206], and limma [207], although plethora of others has been developed [208, 209] and the selection of proper tool depends on the properties of individual dataset, such as number of replicates [119]. A major area of interest is *in silico* identification of miRNA targets. As animal miRNAs only require partial match to a potential target and targeting regions are not clearly defined, prediction algorithms mostly rely on 3' UTR matches are generally known to suffer from relatively high false positivity, although improve strategies employing machine learning are emerging [28]. For in-depth description of prediction algorithms, several reviews are available [210, 211]. Another attractive application of miRNA profiling analysis is integration of miRNA expression with mRNA profiles from same samples to uncover mRNA-miRNA regulatory networks. Several *in silico* tools are available and have been reviewed elsewhere [212].

3.3 MiRNAs in central nervous system

Non-coding RNAs, particularly miRNAs are of prime importance in the CNS, because neural cells are highly transcriptionally active and exhibit robust expression of miRNAs [213, 214]. The proportion of non-coding DNA sequence in the genome correlates with organismal complexity (in contrast to number protein-coding genes) [215], and it has been argued that non-coding RNAs mediated the complex CNS evolution and underlie the unique functional repertoire of the brain [216–218]. Indeed, among all tissues, brain contains the largest number of unique miRNAs [59, 219]. Among miRNAs highly enriched in the brain are miR-124, miR-9, miR-219, miR-29, miR-128 and others, which play a role in a variety of specialized processes required for CNS function including cell proliferation [220, 221], cell specification [222], synaptic plasticity [223] or neurogenesis [224, 225]. Although neurons harbor majority of brain-specific miRNAs, the various cell types in the brain, (that is neurons, astrocytes, oligodendrocytes and microglia) have distinct miRNA profiles [226, 227]. Various neuronal subpopulations, for example, glutamatergic and GABAergic neurons also show distinct miRNA composition [228, 229]. Moreover, miRNAs are localized in many different subcellular compartments such as axons and synapses, where they play roles in axon extension, pathfinding and local protein synthesis [230, 231]. Because of their potential roles in regulating individual genes as well as large gene networks, miRNAs confer neural cells the capacity to exert very precise control over the spatiotemporal deployment of genes, which is crucial for executing complex neurobiological processes [213]. For example, a network of four

non-coding RNAs impacting neuronal excitability, formed by circRNA *Cdr1as*, lncRNA *Cyrano* and two miRNAs (miR-7 and miR-671) has been recently identified [232]. Another examples include transcriptional network mediating neural cell-fate decisions composed of miR-124 and miR-9 and transcription factors CREB, REST and coREST [233, 234], or network of four miRNAs (miR-124, miR-29, miR-9 and miR-128) and two RBPs acting as primary determinant of mRNA stability profile in the brain [235]. Since miRNAs participate in such a variety of cellular processes, changes in their levels can have profound effects on the CNS function. Indeed, numerous studies have shown involvement of miRNAs in the neurodevelopmental [236], neuropsychiatric [237, 238], neurodegenerative [239] as well as acute CNS disorders [240, 241].

3.3.1 MiRNAs in acute CNS injury

Acute injuries to the CNS including stroke and SCI affect significant proportion of the population and cause long-term functional disability and huge social and economic burden [242]. The progress of CNS injury can be divided into two phases: a primary trauma that affects the neural tissue directly and secondary injury induced by multiple biological processes including temporal changes in gene expression [243]. MiRNAs serve as important regulators in many pathological aspects of CNS injury. Global expression studies revealed significant reorganization of miRNA expression after SCI, pointing to their neuroprotective as well as detrimental roles [243–246]. Several miRNAs undergo expression changes that can be related to immune response, associated either with the invasion of immune cells or modulation of inflammatory pathways [247]. For example, neutrophil infiltration explains the upregulation of miR-223 [248], whereas increased expression of the lymphocyte specific miR-142 correlates with the access of these immune cells to the injury site [249]. Several studies demonstrate that increased levels of pro-inflammatory factor TNF- α may result from downregulation of miR-181 and miR-125b [250, 251] and increased levels of cytokines IL-6 and IL-1 β correlate with reduced expression of its regulators let-7a, miR-181a, miR-30b and miR-30c [243, 244, 250, 252]. MiRNAs are also implicated in regulation of cell death after SCI, as administration of miR-20a decreased apoptosis and functional deficits in mouse model [253]. Similarly, silencing miR-486 resulted in decreased neural death and led to an improvement in motor recovery after SCI [254]. Other studies implicate miRNA function in astrogliosis, myelination and intrinsic and extrinsic control of axonal regeneration, important hallmarks of response to SCI [240, 255, 256].

Several global profiling studies also demonstrated altered miRNA expression following experimental stroke, both in the brain and blood [257–260]. Up to 19-25% of assayed miRNAs were dysregulated within 3 days of injury. Ontological analyses predicted that the targets of the dysregulated miRNAs were involved in angiogenesis, hypoxia, endothelial cell regulation, and the immune response – pivotal pathophysiological processes of ischemic stroke. Interestingly, the miRNA expression pattern correlates with the extent of the infarct area, and allowed distinguishing between different etiologies and predict the clinical outcome [259]. Recent study employed AGO CLIP and RNA-Seq approaches to identify miRNA binding sites and expression changes following stroke [261]. The authors identified a key protective role for miR-29 family, which acts in astrocytes and regulates glutamate homeostasis. In addition to global profiling approaches, targeted studies implicated roles of miR-497, miR-29b, miR-134 or miR-21 in stroke pathology [262].

Emerging studies also highlight roles for miRNAs in EVs, where they contribute to cell-cell communication in the brain and throughout the nervous system [263]. For example, increased levels of miR-124 were found in microglial EVs after brain injury and it was observed that this miRNA inhibits neuronal inflammation and promotes neurite outgrowth via transfer to neurons [264]. As such, EVs are being evaluated for their potential as biomarkers or for targeted delivery of therapeutic agents for the treatment of CNS injuries [265]. Overall, these studies show various and multifaceted roles of miRNAs in progression of SCI and stroke. They also demonstrate viability of examining global expression changes and subsequent targeted manipulation of dysregulated miRNAs to influence functional outcomes of the injury. Nevertheless, some miRNAs are reported with contradictory data and information about roles and targets of many other miRNAs is limited and needs to be clarified.

4 Materials and methods

This section describes the methods with focus on the procedures performed by the author. Detailed description of all methods can be found in the publications provided in the supplementary section.

4.1 Samples

Animal samples were obtained from C57Black/6 or FVB mice, or from Wistar rats (Supp. I, V). Experimental stroke was induced by permanent MCAO on female mice (Supp. V), experimental SCI was induced by balloon compression lesion on male rats. For the preparation of rat serum, 1 ml of blood was collected from orbital plexus into 2 ml tubes (Eppendorf) using glass capillary. Blood was allowed to clot for 1 hour at room temperature and then centrifuged at 1000 g for 10 min. The clot was mechanically retracted from the tube wall before the centrifugation. Serum was transferred to another 2 ml tube and centrifuged a second time at 3000 g for 10 min. The supernatant was then transferred to cryovials (Biologix) and stored at -80°C until analysis. All procedures involving the use of laboratory animals were performed in accordance with appropriate regulations and efforts were made to minimize both the suffering and the number of animals used.

Human blood samples were obtained from healthy volunteers (Supp. II, III). Informed consent was obtained from all volunteers participating in the study. For the preparation of human serum, blood was collected into 8.5 ml BD Vacutainer SST II Advance tubes (Beckman Dickinson) and allowed to clot for at least 30 min before centrifugation at 1500 g for 10 min at room temperature. The serum was then transferred to 2 ml tubes (Eppendorf) and stored at -80°C . For the preparation of human plasma, blood was collected from four healthy volunteers into K₂EDTA BD Vacutainer tubes (Beckman Dickinson) and centrifuged within 30 min at 1500 g for 15 min at room temperature. The plasma fraction was aspirated and transferred to 2 ml tubes (Eppendorf) and centrifuged again for 15 min at 3000 g. The supernatant was transferred to new 2 ml tubes and stored at -80°C until analysis. Standardized human plasma sample was prepared by pooling high-quality RNA eluates from four healthy individuals (Supp. III).

4.2 Primers and synthetic oligonucleotides

Sequences of the miRNA oligonucleotides were obtained from the miRBase (www.mirbase.org). RNA oligonucleotides were synthesized and quantified by Integrated DNA Technologies. DNA primers were synthesized and quantified by Invitrogen (Supp. I, II, III, V). Precursor miRNAs were synthesized by *in vitro* transcription from corresponding PCR products using T7 RNA polymerase (New England Biolabs) according to the manufacturer's protocol (Supp. I). Secondary structure of the Two-tailed RT primers were predicted using the UNAFold web server (<http://unafold.rna.albany.edu/>) (Supp. I, II). Spike-in miRNA sequences were screened *in silico* for homology against human, mouse and rat miRBase records (Supp. II).

4.3 RNA isolation

For RNA isolation from tissue, samples were homogenized using the Tissue-Lyser (Qiagen). Total RNA was extracted with TRI Reagent (Sigma-Aldrich) according to the manufacturer's protocol and treated with TURBO DNA-free kit (Thermo Fisher). RNA quantity and purity was assessed using NanoDrop 2000 spectrophotometer (Thermo Fisher) and RNA integrity was assessed using Fragment Analyzer (Agilent) (Supp. I, III, V). For total RNA isolation from human plasma and human

and rat serum miRNeasy Serum/Plasma Advanced Kit (Qiagen) was used according to the manufacturer's instructions (Supp. II, III). 1 μ l of isolation spike-in mix containing synthetic cel-miR-54 (1e+7 copies/ μ l), spike-A (2e+5 copies/ μ l), spike-B (4e+3 copies/ μ l) and, when appropriate 1 μ l of GlycoBlue Coprecipitant (15 mg/mL) (Invitrogen), per sample was added at the lysis step. RNA was eluted into 20 μ l nuclease-free water and stored at -80°C (Supp. II, III).

4.4 RT-qPCR of miRNAs

RT reactions were performed with the qScript flex cDNA kit (Quantabio) in a total reaction volume of 10 μ l. The reaction mixture contained either 10 ng of total RNA or synthetic miRNA template, 1 \times RT buffer, 0.05 μ M RT primer, 1 μ l GSP enhancer and 0.5 μ l RT enzyme. RT reactions were incubated in a 96-well plate in a Bio-Rad CFX 1000 thermocycler for 45 min at 25°C , 5 min at 85°C and then held at 4°C (Supp. I, II, III). Reactions using TaqMan miRNA assays (Thermo Fisher) and Quantabio qScript miRNA system (Quantabio) were performed according to the manufacturer's protocol except that the total reaction volume was scaled down to 10 μ l. Reactions using miQPCR method were performed as described in [163] (Supp. I). Quantitative PCR (qPCR) was performed in a total volume of 10 μ l. One reaction contained 1 \times SYBR Grandmaster Mix (Tataa Biocenter), forward and reverse primer (final concentration 0.4 μ M), and 2 μ l of diluted cDNA template. qPCR was performed in a total reaction volume of 10 μ l containing 1 \times SYBR Grandmaster Mix (TATAA Biocenter), 0.4 μ M forward and reverse primer and the cDNA product diluted at least 10 \times . Reactions were performed in duplicates and incubated in a 96- or 384-well plate in a CFX 96 or CFX 384 Real Time Detection System (Bio-Rad) at 95°C for 30 s, followed by 45 cycles of 95°C for 5 s and 60°C for 15 s. Reaction specificity was assessed by melting curve analysis immediately after the qPCR (Supp. I, II, III). qPCR with TaqMan miRNA assays and Quantabio qScript miRNA system were performed according to manufacturers' protocols in a total reaction volume of 10 μ l (Supp. I).

4.5 Library preparation and sequencing

Small RNA libraries were prepared in technical duplicates starting from 5 μ l of plasma RNA pool and 5 μ l of miRXplore Universal Reference (2e+6 copies/ μ l) with six commercial kits (from Lexogen, Norgen Biotek, Bioo Scientific, Takara, Qiagen and Somagenics) according to each manufacturer's protocol (Supp. III). Libraries were quantified on the Qubit 3 fluorometer (Thermo Fisher) and Fragment Analyzer (Agilent). Libraries generated by the same kit were pooled and run on 5% TBE-PAGE on Mini-PROTEAN tetra cell (BioRad). A region representing fragments with RNA inserts of length 22 nt \pm \sim 10 nt (i.e. fragments corresponding to the size of miRNAs) was excised from the gel and purified. All libraries were sequenced in one sequencing run on NextSeq 500 high-output (Illumina) with 85bp single-end reads. Small RNA libraries for tissue profiling were prepared from 100 ng total brain RNA with RealSeq kit (Somagenics) according to manufacturer's protocol. mRNA libraries were prepared from 400 ng total brain RNA with QuantSeq 3' Library Prep Kit FWD (Lexogen) according to manufacturer's protocol (Supp. V). 1 μ l of ERCC spike-in (c = 0.01 \times ; Thermo Fisher) per library was included. This library preparation method generates stranded libraries predominantly covering the 3' end of the transcript, thus producing gene-centric expression values.

4.6 High-throughput RT-qPCR

Samples were reverse transcribed in a reaction volume of 10 μ l containing: 5 μ l template (either 125 ng total tissue RNA or 100 sorted cells after direct lysis), 0.5 μ l spike-in RNA (Tataa Biocenter;

c = 0.1x for tissues or 0.01x for sorted cells), 0.5 µl equimolar mixture of random hexamers with oligo(dT) (c = 50 µM), 0.5 µl dNTPs (c = 10 mM), 2 µl 5× RT buffer, 0.5 µl RNaseOUT, 0.5 µl Maxima H-Reverse Transcriptase (all Thermo Fisher) and 0.5 µl nuclease-free water. After the pre-incubation step at 65°C (t = 5 min), followed by the immediate cooling on ice, the main incubation was performed at 25°C (t = 10 min), 50°C (t = 30 min), 85°C (t = 5 min), after which the samples were immediately cooled on ice. cDNA from tissue samples was diluted 4x in nuclease-free water; sorted cell-cDNA was left undiluted. All cDNA samples were pre-amplified immediately after RT in 40 µl total reaction volume containing 4 µl cDNA, 20 µl IQ Supermix buffer (Bio-Rad), 4 µl primer mix of 96 assays (c = 250 nM each), and 12 µl of nuclease-free water. Reactions were incubated at 95°C (t = 3 min) following by 18 cycles of 95°C (t = 20 s), 57°C (t = 4 min) and 72°C (t = 20 s). After thermal cycling, reactions were immediately cooled on ice and diluted in nuclease-free water (sorted cells 4x, tissue 50x). High-throughput qPCR was then performed on a 96.96 microfluidic platform BioMark (Fluidigm) as previously described [266]. Cycling program consisted of activation at 95°C (t = 3 min), followed by 40 cycles of 95°C (t = 5 s), 60°C (t = 15 s) and 72°C (t = 20 s) and melting curve analysis (Supp. V).

4.7 Bioinformatic and statistical data analysis

4.7.1 Differential gene expression analysis

Differential gene expressions from RNA-Seq data between desired groups of samples were analysed in R project using DESeq2 package (Supp. V).

4.7.2 Gene set enrichment analysis (GSEA)

GSEA was performed for pairwise differential expression comparisons. First, a gene score was calculated for every gene using DESeq2 output as $-\log_{10}(p_{adj})$ and assigned a positive or negative sign based on direction of regulation. Genes were ranked by their gene-scores and GSEA was run in a weighted pre-ranked mode with 1000 permutations. Gene sets were downloaded from http://download.baderlab.org/EM_Genesets/. Significantly overrepresented gene sets were visualized as a network using Enrichment Map. In the network, each node represents gene set and highly overlapping gene sets are connected with edges, resulting in a tight clustering of highly redundant gene sets. For functional annotation of discrete sets of genes, we used Cytoscape plugin ClueGO with all expressed genes (16048 genes) as the background set (Supp. V).

4.7.3 Cell-specific gene sets and cell type proportion estimation

Marker genes for major cell types specifically in the mouse cortex region were taken as an initial reference. In order to acquire marker genes with stable expression regardless of activation states, we have further removed the genes previously found to be differentially expressed under similar conditions in studies on purified cell types (Supp. V). DESeq2-normalized gene expression data and the cell-specific gene lists were used as an input into the marker-GeneProfile R package or the estimation of marker gene profiles (MGPs), which serve as a proxy for relative cell type proportion changes. Differences in expression of final marker gene sets were analyzed by linear mixed model in R project v3.6.0 using lmerTest package. To validate the first estimations, we employed CIBERSORT, a transcriptome deconvolution algorithm that uses gene expression matrix of individual cell types as a reference, and deconvolutes the cellular composition of mixed sample by linear support vector regression. We used published single-cell RNA-Seq dataset of adult mouse cortex as a reference gene expression signature (Supp. V).

4.7.4 Weighted gene co-expression network analysis (WGCNA)

Standard WGCNA procedure was followed to create gene co-expression networks using `blockwiseModules` function from the WGCNA R package. This analysis groups genes with highly correlated expression pattern across samples into modules (Supp. V).

4.7.5 Motif and transcription factor enrichment analysis

Cytoscape plugin iRegulon with default parameters was used to search for over-represented motifs and their associated transcription factors 500 bp upstream of the transcription start site. All genes from a particular WGCNA module were used as an input (Supp. V).

4.7.6 Protein-protein interaction network

Known interactions between genes in the desired gene sets were downloaded from STRING database v10.5. The resulting interaction network was then visualized and analyzed in Cytoscape. Spectral partition-based network clustering algorithm via Cytoscape ReactomeFI plugin was used for network clustering (Supp. V).

4.7.7 Custom gene set enrichment

Gene sets of interest were collected directly from relevant publications. R package `GeneOverlap` v1.20 was used to calculate the odds ratio (OR) and the significance of the overlap of the gene sets of interest with the Fisher's exact test (Supp. V).

4.7.8 RT-qPCR data analysis

Raw data were pre-processed with the Real-Time PCR analysis software v4.1.3 (Fluidigm); unspecific values were deleted based on melting-curve analysis. Further processing was done in GenEx v6.0.1 (MultiD Analyses AB): Cq value cutoff was applied; gDNA background was subtracted using ValidPrime (Tataa Biocenter); data were normalized to the mean expression of 5 reference genes (*Actb*, *Gapdh*, *Ppia*, *Ywhaz*, *Tubb5*); outliers were deleted (within group Grubbs test, $p < 0.05$) and a gene was considered undetected for given group if more than 75% values per group were missing; technical replicates (RT and FACS) were averaged; if appropriate, missing data were imputed on a within-group basis and remaining missing data were replaced with $Cq_{\max} + 2$ for tissue samples or $Cq_{\max} + 0.5$ for sorted cells (Supp. V).

Temporal expression of individual genes was first analyzed with two-way ANOVA in R project v3.6., then differences between time-points were tested separately for each age by one-way ANOVA and post-hoc t-tests using `emmeans` package v1.3.5. P-values were adjusted with Benjamini-Hochberg method. Temporal expression of groups of cellular marker genes was first analyzed using linear mixed model in R project v3.6.0 (`lmerTest` package), then differences between time-points were tested separately for each age. Significance was tested by Satterthwaite's method. Post-hoc pairwise t-tests were performed using `emmeans` package v1.3.5 and p-values were adjusted by Holm method. Temporal expression of IFN-I pathway was analyzed in similar steps using random slope and random intercept mixed model with 23 interferon-stimulated genes as response variables. IFN-I pathway expression in sorted cells was analyzed in the same way with two-factor design (age, injury) with interaction, using only detected ISGs per each cell type. Differential expression of individual genes (relative to age-matched control) in sorted cells was tested in GenEx v6.0.1 (MultiD Analyses AB) using ANOVA with Bonferroni's post-hoc test for selected pairwise comparisons and p-values were corrected using Benjamini-Hochberg method (Supp. V).

5 Survey of results

5.1 Development of novel RT-qPCR-based method for highly accurate miRNA quantification

As discussed in previous chapters, sensitive and specific quantification of miRNAs is challenging. Although several methods for RT-qPCR analysis of miRNAs have been developed, they suffer from one or more drawbacks and are typically available only commercially, and costly. We have therefore aimed to develop novel specific and cost-effective approach to quantify miRNA expression that would improve on the available methods.

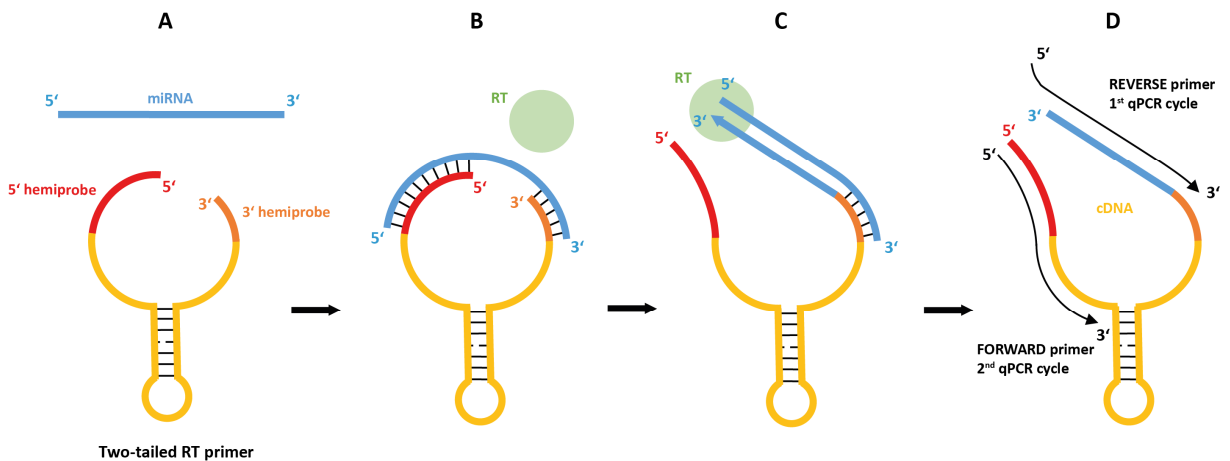


Figure 7. Schematic of Two-tailed RT-qPCR. A) Two-tailed RT primer with two hemiprobos connected by a hairpin folding sequence. B) The hemiprobos bind cooperatively, one at each end of the target miRNA, forming a stable complex. C) Reverse transcriptase binds the 3'-end of the hybridized Two-tailed RT primer and elongates it to form tailed cDNA. D) The cDNA is amplified by qPCR using two target-specific primers.

We have designed a RT-qPCR system that utilizes specific structured primers for reverse transcription and SYBR-Green-based qPCR, which we named “Two-tailed RT-qPCR” (Figure 7 and Supp. I). Briefly, RT primers containing two binding probes (“hemiprobos”) complementary to the target miRNA joined by sequence forming hairpin structure were designed (Figure 7A). We hypothesized that the introduction of 5' binding sequence would improve sensitivity and specificity as larger part of the miRNA sequence will be interrogated compared to previously available methods, which target only 3' region of the miRNA. Importance of the 5' hemiprobe was assessed on synthetic miRNA oligonucleotides with positive results (Supp. I, Fig. 2). Next, parameters of the method, including repeatability, sensitivity and dynamic range, ability to discriminate between highly similar miRNAs, between mature and precursor miRNAs, and ability to capture isomiRs were assessed and compared against three other previously available methods, each of which employs different technical approach (Supp. I., Table 1, Fig. 3, 4 and 7). These experiments revealed that Two-tailed RT-qPCR matches or outperforms other methods, while simultaneously being less costly (Supp. I, Table 2). In addition, Two-tailed RT-qPCR was used to profile seven miRNA targets in various mouse tissues, both in singleplexed and multiplexed setting and compared against then-industry-standard TaqMan miRNA assays (Thermo Fisher), which revealed good agreement of relative quantification both between single and multiplexed setting as well as between the two methods (Supp. I, Fig. 5 and 6).

In summary, we have developed a highly sensitive and exceedingly specific method called Two-tailed RT-qPCR, suitable for rapid and cost-effective miRNA profiling. At the same time, Two-tailed RT-qPCR reflects on the current state of miRNA field and confers several advantages over current RT-qPCR methods, including increased specificity and ability to capture the full isomiR profile.

5.2 Development of quality control tool for circulating miRNA studies

As discussed in chapter 3.1.9, application of circulating miRNAs as clinical biomarkers is an exciting avenue of miRNA research. In our laboratory, we were interested to study circulating miRNA profiles from serum and plasma samples to identify potential biomarker candidates for acute spinal cord injury and its severity. However, the miRNA profiling workflow from liquid biopsy samples is even more challenging than from typical tissue or cellular samples, and proper quality control becomes even more critical. Because standard quantification and quality control tools are inappropriate for biofluid profiling (see chapter 3.2.4), we have developed an RT-qPCR-based quality control tool for circulating miRNA studies (Supp. II). It is freely available and allows users to monitor quality of miRNA isolation, degree of inhibition, and erythrocyte contamination to ensure technical soundness of the obtained results.

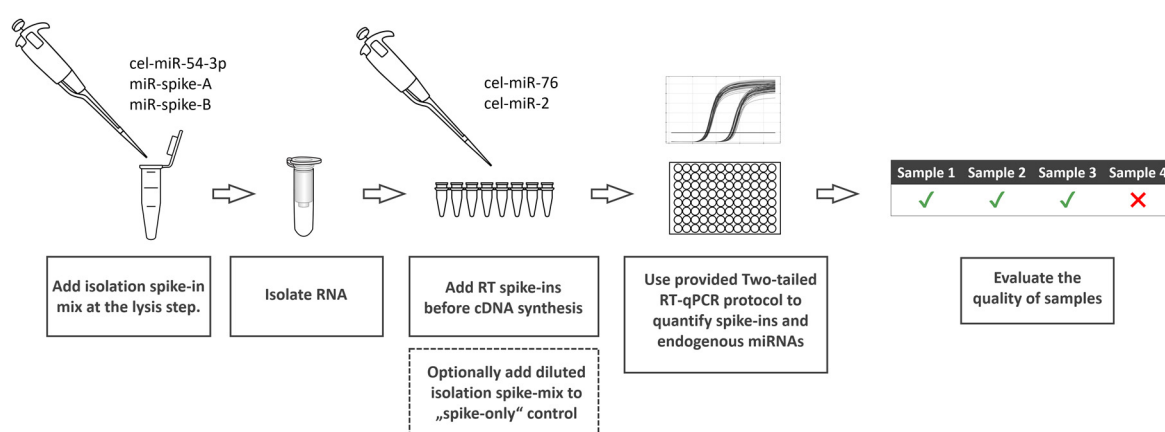


Figure 8. Quality control workflow with the Two-tailed QC panel. A mix of three synthetic RNA spike-ins is added prior to RNA isolation from the biofluid sample. A second mix of two spike-ins is added before cDNA synthesis step. Optionally, a diluted isolation spike-in mix is used as a template in a “spike-only” control reaction to determine spike-in baseline signal. Two-tailed RT-qPCR is used to quantify the spike-ins along with three endogenous microRNAs to evaluate the technical quality of RNA isolation, effect of inhibition and the level of hemolysis.

Briefly, we have designed five synthetic spike-in miRNAs and eight two-tailed RT-qPCR assays targeting these synthetic spike-ins, and three endogenous miRNAs serving as controls for miRNA yield and indicators of hemolysis (Figure 8 and Supp. II, Fig. 1). We then demonstrated how the protocol can be utilized to optimize input volume of the sample to obtain the best yield and purity using human plasma, human serum and rat serum and identified optimal volumes for isolation of these biofluids with miRNeasy Serum/Plasma Advanced Kit (Qiagen) (Supp. II, Fig. 2). Next, we assessed the effect of carriers (specifically glycogen) on the isolation procedure and demonstrated that it improves yield and reproducibility of the repeated isolations (Supp. II, Fig. 3). Next, we prepared hemolysis dilution series and constructed calibration curve to correlate hemolysis indicator based on ΔCq values (miR-23a – miR-451a) to the absorbance at 414 nm (wavelength

indicative of hemolysis). This allowed us to establish reference ΔCq values that can be used to identify level of hemolysis in any new sample as long as the described workflow is used (Supp. II, Fig. 4).

Taken together, we have developed Two-tailed RT-qPCR panel for quality control, monitoring of technical performance, and optimization of miRNA profiling experiments from biofluid samples. The detailed experimental protocol including guide to data interpretation, and sequences of the RNA oligonucleotides and RT-qPCR assays is provided to the community and will hopefully contribute to the increased quality and reliability of the results from circulating miRNA studies.

5.3 Comprehensive performance comparison of small RNA-Seq library preparation methods from biofluids

During the exploratory phase of miRNA studies, it is desirable to interrogate expression of the whole miRNome in several samples at once. For this purpose, small RNA-Seq is becoming a leading technology. At the same time, library preparation methods are known to suffer from several biases (see chapter 3.2.5.3), and new technical solutions including usage of randomized adapters, usage of poly(A)-tailing and template switching or usage of single-adaptor ligation and circularization were recently developed. In the past years, several studies compared different protocols for small RNA-seq. Although they provided valuable insights, they focused on the profiling of tissues and cells and typically covered only some of the available protocols. How various library preparation protocols perform with biofluid samples is not established. Therefore, we carried out comprehensive comparison of the six commercially available protocols for small RNA-Seq library preparation (from following vendors: Lexogen, Norgen Biotek, Bioo Scientific, Qiagen, Takara, Somagenics) with focus on their technical biases and performance in biofluids (Figure 9 and Supp. III). To our knowledge, this is the first study that covers all currently available approaches for small RNA library preparation.

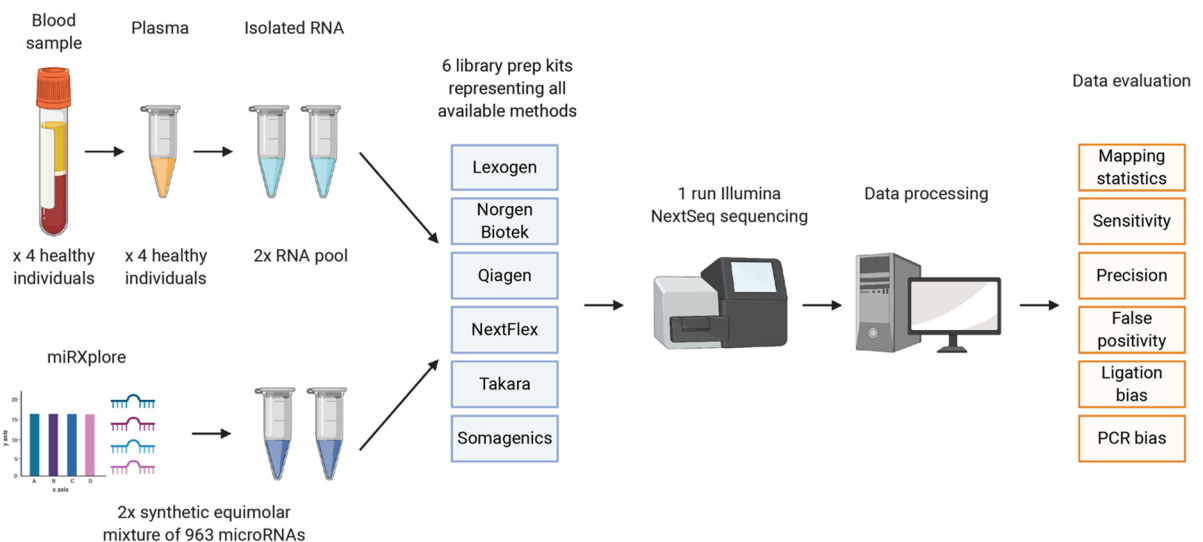


Figure 9. Design of the benchmarking study. Two types of samples were used for comparison of library preparation approaches – human plasma and miRXplore universal reference. After RNA isolation, six different protocols were used for library preparation and resulting libraries were sequenced on Illumina NextSeq platform. Data were trimmed, filtered according to length and further mapped to the respective reference sequences. After normalization, various metrics were evaluated.

Briefly, plasma samples were obtained from healthy volunteers and RNA was isolated using commercial kit. Quality of the isolation procedure was assessed using previously developed Two-tailed QC panel and RNA samples passing QC criteria were used to prepare standardized human plasma sample. Small RNA-Seq libraries were prepared with each kit in duplicates from human plasma as well as from synthetic mixture of 962 miRNAs in equimolar amount (miRXplore), which allowed us to assess technical biases of each method (Figure 9 and Supp. III, Fig. 1).

Mapping statistics for both plasma and miRXplore samples were examined, which revealed varying performance, particularly low mapping rate to miRNAs for the polyadenylation-based kit from Takara (SMARTer) and surprisingly low mapping rate for all kits in human plasma (Supp. III, Fig. 2). Next, ligation bias was assessed for each kit on the miRXplore samples as the fold-deviation from the expected number of reads for each miRNA (Supp. III, Fig. 3A). This analysis revealed that all kits suffer from substantial ligation bias, as percentage of miRNAs that can be considered unbiased ranged from 13.1% for Norgen kit to 50.7% for SMARTer. Notably, single-molecule ligation and circularization approach from Somagenics (RealSeq), that was recently claimed to be bias-free [178] showed only 21% unbiased miRNAs and was outperformed in this metric by three other kits (Supp. III, Fig. 3A). We also assessed the contribution of ligation bias vs PCR bias through UMIs that are incorporated in the Qiagen kit (QIAseq) (Supp. III, Fig. 3B). We found that ligation is a dominant source of bias (ligation bias explained more than 75% variability for 518 out of 962 miRNAs), although for approximately quarter (227) of miRNAs, PCR is significant contributor to bias too and explains more than half of their variation. We then assessed how the sequencing reads are allocated to the detected spectrum of miRNAs for each kit (Supp. III, Fig. 4 and 5) and how sensitive for miRNA detection is each kit at various sequencing depths (Supp. III, Fig. 6). We found that few most abundant miRNAs consumed majority of the reads in all kits (Supp. III, Fig. 2 and 6) and that the kits that detected most miRNAs at all sequencing depths in human plasma were SMARTer and NextFlex (Supp. III, Fig. 6). Arguably, this is due to their lower rates of bias, which causes that reads are progressively allocated to the larger spectrum of miRNAs. To examine which kit quantified abundances in plasma samples most truthfully, we have performed absolute quantification of 19 miRNAs in our plasma samples by Two-tailed RT-qPCR and compared absolute quantities to abundances reported by each kit (Supp. III, Fig. 7). SMARTer RNA-seq expression was closest to the true expression (Pearson $r = 0.94$), followed by RealSeq, QIAseq and NextFlex. Lexogen and Norgen data showed lower correlation (Pearson $r = 0.81$ and 0.73 ; Supp. III, Fig. 7). In order to compare kits at the level of individual miRNAs and identify which miRNAs are most affected by kit-specific technical performance, we clustered kits based on miRNAs with highest differential expression between all kits (Supp. III, Fig. 8). Such clustering of miRXplore samples showed that measured expression was similar in kits with similar technical procedures.

In summary, this study provides comprehensive overview of the performance of all currently available technical approaches for library preparation from biofluids and can serve as a guide for selection of optimal kit for each experiment. In addition, it contributes to our understanding of various technical biases, which is important for proper interpretation of the data and potential development of novel wet-lab protocols as well as *in silico* correction algorithms and data analysis pipelines.

5.4 Decoding the transcriptional response to ischemic stroke in young and aged mice

The previously described work covers the laboratory methodology for quantitative analysis of miRNA ranging from pre-analytical aspects to the establishment of methods for low and high-throughput measurements. In addition to this work, we reviewed the current state of circulating miRNA analysis in Supplement IV.

In a biological setting, we apply miRNA profiling and complementary technologies (mRNA-Seq, single-cell RNA-Seq, high-throughput RT-qPCR) to study gene expression changes and gene regulatory networks underlying CNS injuries, such as spinal cord injury and ischemic stroke. For similar biological questions, miRNA profiling alone provides only limited picture, considering that miRNAs have pleiotropic roles, and changes in their mRNA targets have to be inferred indirectly. Depending on the goal of biological experiments, direct mRNA profiling or integrative profiling of mRNA and miRNA expression from the same samples can provide more accurate picture of the gene expression changes. Therefore, we first focused on the mRNA analysis with the perspective of integrated analysis in the next study.

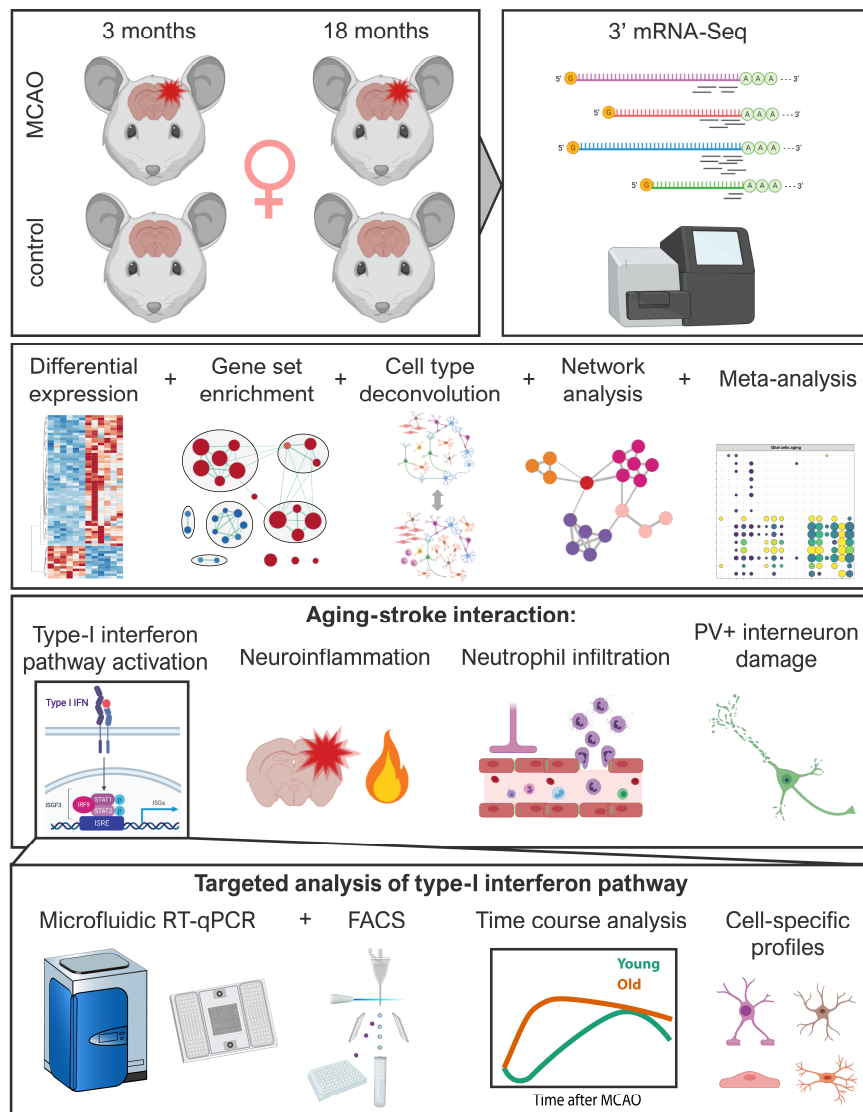


Figure 10. Design, analysis workflow and major findings of the study.

Ischemic stroke is one of the leading causes of mortality and major healthcare and economic burden. It is a well-recognized disease of aging, yet it is unclear how the age-dependent vulnerability occurs and what are the underlying mechanisms. To address these issues, we performed a comprehensive mRNA-Seq analysis of aging, ischemic stroke and their interaction (Figure 10 and Suppl. V). Briefly, we have modelled ischemic stroke in 3-month old and 18-month old mice and RNA-sequenced the brain of injured (at 3 days after stroke) and age-matched control mice. We have analyzed the gene expression changes with normal aging (Supp. V, Fig. 1) and after ischemic stroke (Supp. V, Fig. 2 and 3) on the level of genes, gene ontology (GO) terms, pathways and custom gene sets. These analyses revealed activation of glial subpopulations with normal aging and increased inflammatory environment after stroke in aged mice. In order to provide cell-specific context to the observed transcriptional changes, we assessed relative changes of the cell type proportions by computational deconvolution of the RNA-Seq data (Supp. V, Fig. 4). This revealed increased infiltration of peripheral leukocytes and greater damage to Parvalbumin-positive interneurons in aged mice after stroke. To capture the full extent of expression trends from systems perspective, we complemented our results with weighted co-expression network analysis (WGCNA, Supp. V, Fig. 5). WGCNA recapitulated results of cell type proportion estimates in an unsupervised manner and highlighted amplified activation of module of inflammatory and interferon-stimulated genes (ISGs).

It has been reported that activation of type-I interferon (IFN-I) signaling is detrimental to stroke outcome in young mice [267]. To explore the IFN-I signaling in detail, we mapped our RNA-Seq data to the published IFN-I regulatory network, which revealed age-dependent activation of one of the signaling submodules and suggested that age-dependent amplification of ISG expression occurs through action of canonical regulators including *Stat1* and *Irf9* (Supp. V, Fig. 6 A-D). To reveal temporal changes of interferon pathway, we analyzed expression of ISGs at several time-points after stroke by microfluidic RT-qPCR in both age groups. This revealed activation of IFN-I pathway around 1 day post-stroke, which prevailed at least until 14 days and was higher in aged mice (Supp. V, Fig. 6E). Because little was known about cell types that contribute to the IFN-I signaling post-stroke, we FACS-purified endothelial and three glial cell populations and analyzed their expression of ISGs. Our cell-specific analysis revealed that not only microglia, but also oligodendrocytes heavily induce IFN-I signaling following stroke and all cell types converge on higher ISG expression in aged mice, although their individual responses differ (Supp. V, Fig. 6F). In summary, this study provides detailed insights into transcriptional response to stroke in young and aged mice and may contribute to our understanding of the interplay between stroke pathology and aging.

5.5 mRNA-miRNA regulatory networks in CNS injury (unpublished results)

Several studies have implicated *in vitro* and *in vivo* effects of miRNAs after CNS injury (see chapter 3.3.1). After previously exploring global mRNA changes at single time-point after experimental stroke, we focused on understanding of temporal alterations of mRNA-miRNA regulatory networks following CNS injury. Two models of CNS injuries have been employed: middle cerebral artery occlusion (MCAo) on mice to model ischemic stroke and spinal cord compression lesion on rat, to model spinal cord injury (SCI). Groups of injured and sham operated control animals have been sacrificed at different time points after injury: 3h, 7h, 24h, 3d and 7d. RNA from each sample was used for preparation of mRNA and small-RNA sequencing libraries, which were sequenced on

Illumina NextSeq instrument. The detailed analysis is ongoing and here we report the initial results from the SCI model. Samples were clustered using PCA based on mRNA (Figure 11A) and miRNA expression (Figure 11B) separately. In both cases, naive controls cluster together with sham samples and injured samples clearly separate from controls indicating widespread changes in expression of both RNA modalities after injury. mRNA expression more clearly separated individual time-points, while miRNA expression separated injured samples into early (3h, 12h, 24h) and later (3d, 7d) time-points, indicating that overall post-transcriptional regulatory response is lacking behind mRNA changes and is lower in magnitude during early phase.

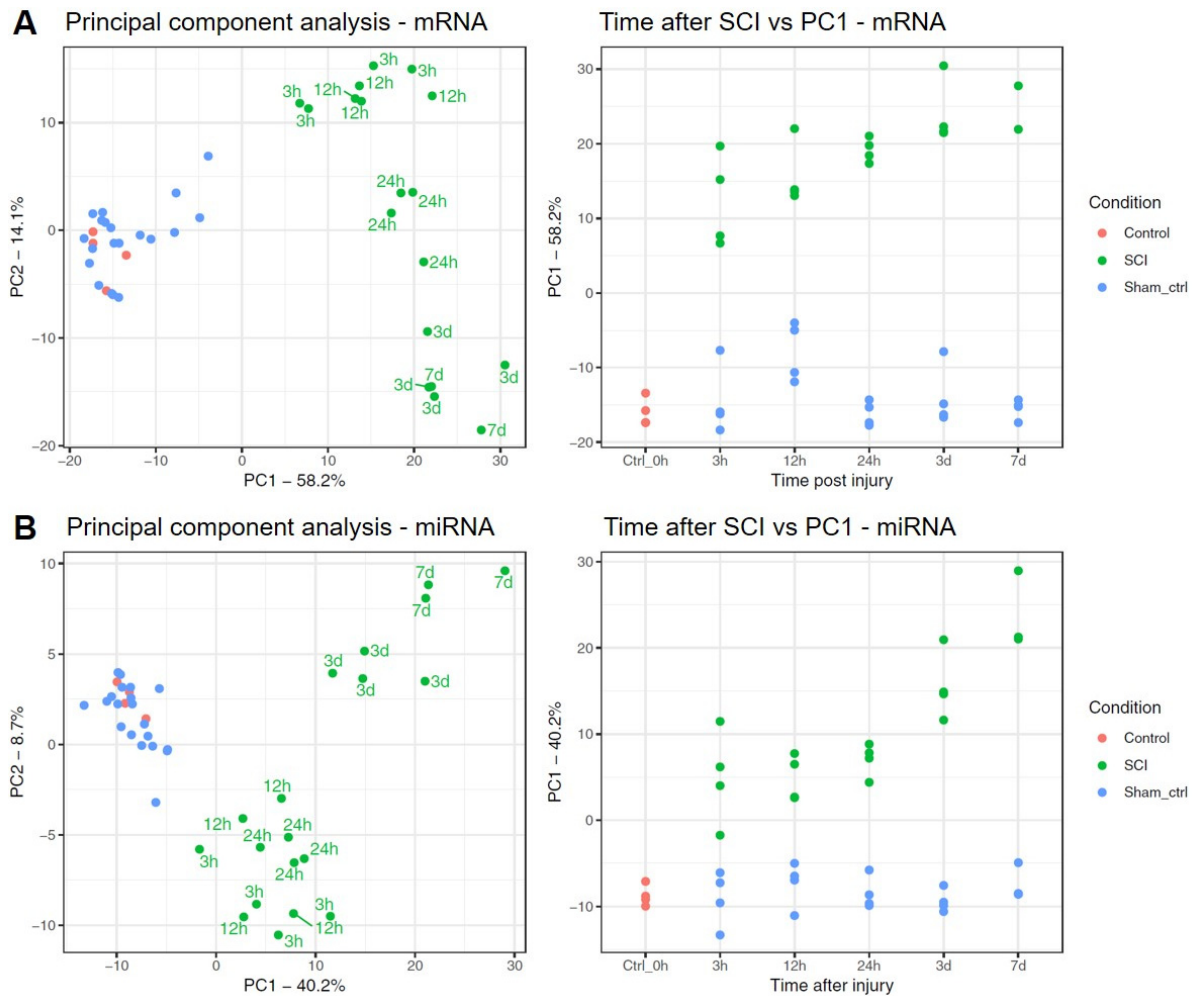


Figure 11. Principal component analysis of mRNA and miRNA expression after spinal cord injury. A) PCA clustering based on mRNA expression shows well-separated clusters of control and injured samples at different time-points after SCI. B) PCA clustering based on miRNA expression shows separation of control and injured samples and progressively greater global miRNA changes.

Next, we used WGCNA to cluster genes with similar expression profiles into modules. Several gene modules showing various patterns of differential expression with respect to control were identified (Figure 12A). We then searched for significant enrichment of cell-specific markers in the modules and found that several of them are associated with single cell type, and likely reflect temporal changes of the particular cell type or cell-specific expression (Figure 12B). We also annotated selected modules functionally by GO and pathway enrichment analysis. This revealed time-dependent changes in wound healing processes upon SCI (Figure 13). Positive regulation of wound

healing was associated mainly with brown expression module with continuously increasing expression (Figure 13A), while green-yellow module associated with negative regulation of wound healing was downregulated in early time-points and then its expression started to increase after 12h post injury (Figure 13B).

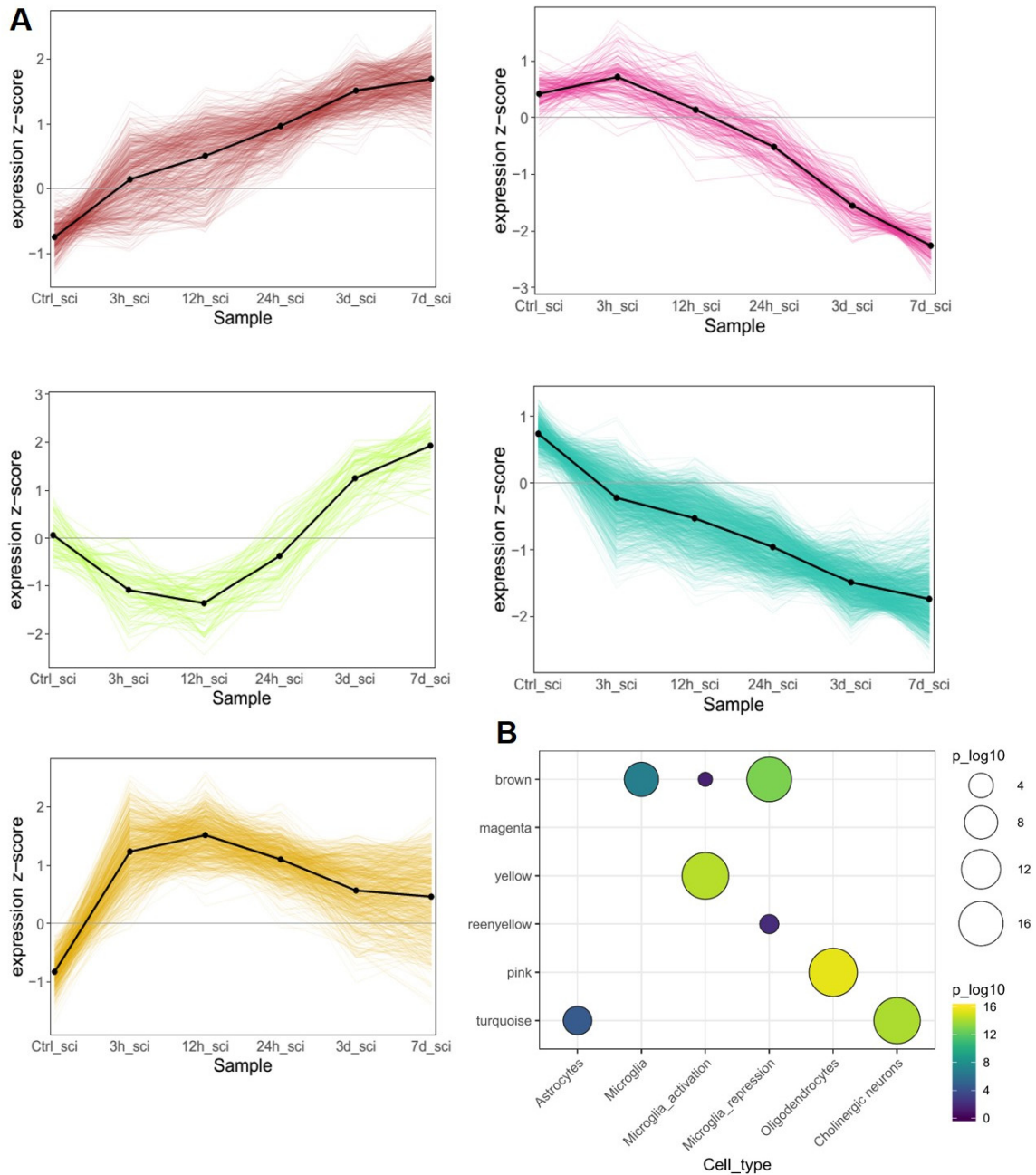


Figure 12. WGCNA identified modules of co-expressed genes associated with progression of SCI. A) Selected modules with various patterns of temporal differential expression are shown. B) Several modules are associated with specific cell-type, as revealed by significant enrichment of cell-specific markers.

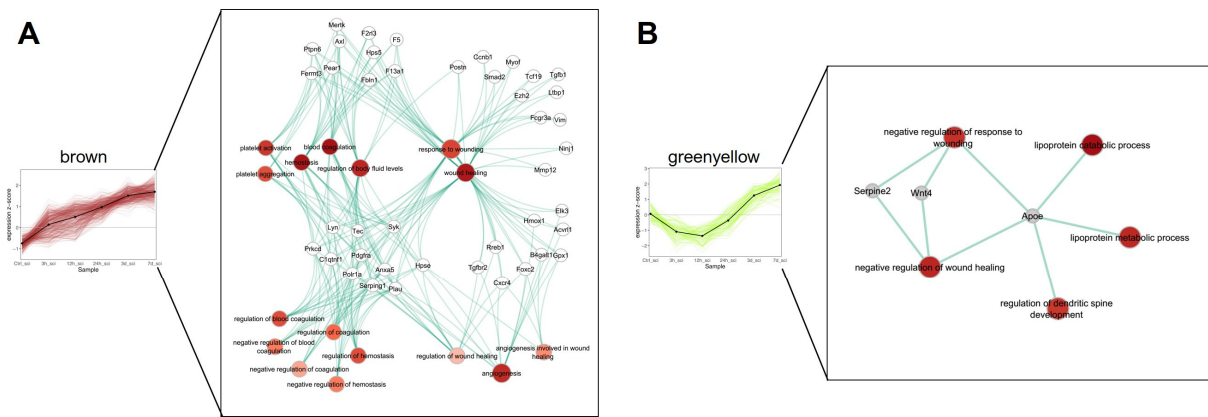


Figure 13. Gene ontology and pathway enrichment analysis revealed time-dependent changes in wound healing processes after SCI. Positive regulation of wound healing (A) was associated mainly with brown expression module identified by WGCNA, whereas negative regulation (B) with greenyellow module. Enriched GO terms and genes involved are visualized as network, with nodes representing enriched GO terms and genes, and lines representing association between GO term and genes.

Clustering of the top differently expressed miRNAs identified in total 135 miRNAs with distinctive profiles in each time-point (Figure 14). Similarly to mRNA expression, the miRNA expression variance is driven partly by cellular composition. We identified several cell-specific miRNAs corresponding to the distinct profiles including miR-142 (microglia), miR-124 and miR-129-2 (neurons), miR-144 and miR-451 (hematopoietic lineage), miR-221 and miR-27a (endothelial cells) or miR-92b (astrocytes). These initial results point to significant reorganization of both cellular composition and expression of mRNAs and miRNAs after SCI and highlight several interesting avenues for further exploration such as expression changes of wound-healing-associated genes.

As a next step, we aim to perform *in silico* integrative analysis of miRNA-mRNA expression to elucidate roles of miRNAs on their respective targets. However, the miRNA-mRNA correlative relationship may be confounded by the expression variation due to changing cellular composition. We are therefore developing procedures to correct bulk tissue gene expression as well as miRNA expression data for cellular composition before parsing them to integrative analysis algorithms in order to better reveal true targeting relationships. These analyses are subject to ongoing research.

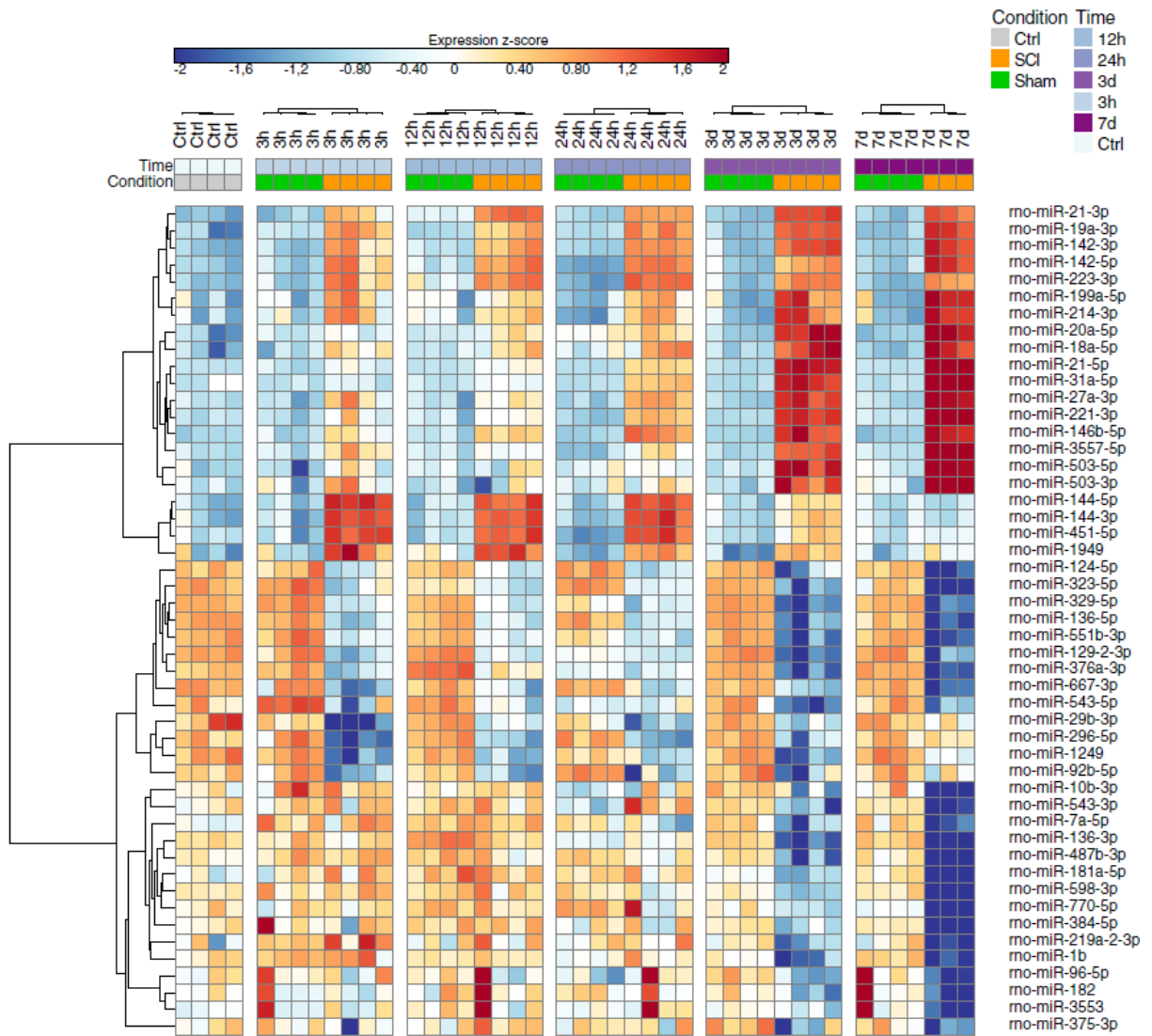


Figure 14. Heatmap showing top differentially expressed miRNAs after SCI. Differential gene expression analysis identified in total 135 microRNA having distinctive profiles in 3h, 12h, 24h, 3d and 7 day after injury.

6 Conclusion

This thesis deals with methods and workflows for targeted as well as global quantitative miRNA analysis and applies them together with global measurements of mRNA expression to understand pathophysiological responses after acute CNS injuries.

The most important outcomes of the work are summarized below:

- Novel RT-qPCR method called “Two-tailed RT-qPCR” for accurate measurement of miRNA expression in cells, tissues and biofluids was developed (Supp. I).
- Two-tailed RT-qPCR was utilized to develop quality control protocol allowing monitoring of technical workflow of miRNA expression studies, particularly studies of circulating miRNAs from biofluid samples (Supp. II).
- Methods for small RNA-Seq library preparation were comprehensively evaluated in a benchmarking study with focus on biofluid samples and understanding of technical biases distorting the small RNA-Seq data (Supp. III).
- Current state of the art of technical aspects of circulating miRNA analysis was reviewed (Supp. IV).
- Global RNA-Seq analysis provided detailed insights into the impact of stroke, aging and their interaction on genome-wide expression profiles (Supp. V).
- Paired mRNA and miRNA profiling of neural tissue revealed temporal changes to the transcriptome and miRNome after SCI.
- Future detailed integrative analysis will reveal impact of post-transcriptional regulation of miRNAs on the mRNA targets after SCI and stroke.

In summary, this thesis provides novel tools for the field of miRNA analysis and contributes to the detailed understanding of the technical performance of small-RNA-Seq methods including various biases that hamper the accurate quantification. These results may serve for the miRNA research community to obtain reliable results from the quantitative studies and guide the choice of the small RNA-Seq platform. They can be utilized for the development of improved small RNA-Seq protocols and computational methods aimed at correction of small RNA-Seq data. In addition, global analyses of mRNA and miRNA expression reported in this thesis may contribute to our understanding of molecular and cellular mechanisms underlying CNS injuries such as stroke and spinal cord injury and open new avenues for the search for future therapeutic strategies.

7 References

1. Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M. C., & Caracausi, M. (2019). Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes*, *12*(1), 315. <https://doi.org/10.1186/s13104-019-4343-8>
2. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., ... Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. <https://doi.org/10.1038/nature11247>
3. Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, *75*(5), 843–854. [https://doi.org/10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y)
4. Bartel, D. P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. *Cell*, *136*(2), 215–233. <https://doi.org/10.1016/j.cell.2009.01.002>
5. Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell*, *173*(1), 20–51. <https://doi.org/10.1016/j.cell.2018.03.006>
6. Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, *47*(D1), D155–D162. <https://doi.org/10.1093/nar/gky1141>
7. Friedman, R. C., Farh, K. K. H., Burge, C. B., & Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*, *19*(1), 92–105. <https://doi.org/10.1101/gr.082701.108>
8. Ivey, K. N., & Srivastava, D. (2015). microRNAs as Developmental Regulators. *Cold Spring Harbor Perspectives in Biology*, *7*(7), a008144. <https://doi.org/10.1101/cshperspect.a008144>
9. Rodriguez, A., Griffiths-Jones, S., Ashurst, J. L., & Bradley, A. (2004). Identification of mammalian microRNA host genes and transcription units. *Genome Research*, *14*(10 A), 1902–1910. <https://doi.org/10.1101/gr.2722704>
10. de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., ... de Hoon, M. J. L. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nature Biotechnology*, *35*(9), 872–878. <https://doi.org/10.1038/nbt.3947>
11. Olena, A. F., & Patton, J. G. (2009). Genomic organization of microRNAs. *Journal of Cellular Physiology*, *222*(3), n/a-n/a. <https://doi.org/10.1002/jcp.21993>
12. Fromm, B. (2016). microRNA Discovery and Expression Analysis in Animals. In *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing* (pp. 121–142). Springer International Publishing. https://doi.org/10.1007/978-3-319-31350-4_6
13. Ha, M., & Kim, V. N. (2014, July 16). Regulation of microRNA biogenesis. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. <https://doi.org/10.1038/nrm3838>
14. Treiber, T., Treiber, N., & Meister, G. (2019). Regulation of microRNA biogenesis and its crosstalk with other cellular pathways. *Nature Reviews Molecular Cell Biology*, *20*(1), 5–20. <https://doi.org/10.1038/s41580-018-0059-1>
15. Ruby, J. G., Jan, C. H., & Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature*, *448*(7149), 83–86. <https://doi.org/10.1038/nature05983>
16. Meng, Y., & Shao, C. (2012). Large-scale identification of mirtrons in arabidopsis and rice. *PLoS ONE*, *7*(2), e31163. <https://doi.org/10.1371/journal.pone.0031163>
17. Okamura, K., Hagen, J. W., Duan, H., Tyler, D. M., & Lai, E. C. (2007). The Mirtron Pathway Generates microRNA-Class Regulatory RNAs in *Drosophila*. *Cell*, *130*(1), 89–100. <https://doi.org/10.1016/j.cell.2007.06.028>
18. Wen, J., Ladewig, E., Shenker, S., Mohammed, J., & Lai, E. C. (2015). Analysis of Nearly One Thousand Mammalian Mirtrons Reveals Novel Features of Dicer Substrates. *PLoS Computational Biology*, *11*(9), e1004441. <https://doi.org/10.1371/journal.pcbi.1004441>
19. Ender, C., Krek, A., Friedländer, M. R., Beitzinger, M., Weinmann, L., Chen, W., ... Meister, G. (2008). A

- Human snoRNA with MicroRNA-Like Functions. *Molecular Cell*, 32(4), 519–528.
<https://doi.org/10.1016/j.molcel.2008.10.017>
20. Rnas, D., Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P., & Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other. *Genes & Development*, 2773–2785.
<https://doi.org/10.1101/gad.1705308>.)
 21. Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M. Di, Yartseva, V., ... Steitz, J. A. (2013). Mammalian 5'-capped microRNA precursors that generate a single microRNA. *Cell*, 155(7), 1568–1580.
<https://doi.org/10.1016/j.cell.2013.11.027>
 22. Patrick, D. M., Zhang, C. C., Tao, Y., Yao, H., Qi, X., Schwartz, R. J., ... Olson, E. N. (2010). Defective erythroid differentiation in miR-451 mutant mice mediated by 14-3-3 ζ . *Genes and Development*, 24(15), 1614–1619. <https://doi.org/10.1101/gad.1942810>
 23. Cheloufi, S., Dos Santos, C. O., Chong, M. M. W., & Hannon, G. J. (2010). A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature*, 465(7298), 584–589.
<https://doi.org/10.1038/nature09092>
 24. Cifuentes, D., Xue, H., Taylor, D. W., Patnode, H., Mishima, Y., Cheloufi, S., ... Giraldez, A. J. (2010). A novel miRNA processing pathway independent of dicer requires argonaute2 catalytic activity. *Science*, 328(5986), 1694–1698. <https://doi.org/10.1126/science.1190809>
 25. Yoda, M., Cifuentes, D., Izumi, N., Sakaguchi, Y., Suzuki, T., Giraldez, A. J., & Tomari, Y. (2013). Poly(A)-specific ribonuclease mediates 3'-end trimming of argonaute2-cleaved precursor micrornas. *Cell Reports*, 5(3), 715–726. <https://doi.org/10.1016/j.celrep.2013.09.029>
 26. Ameres, S. L., & Zamore, P. D. (2013, August 26). Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group. <https://doi.org/10.1038/nrm3611>
 27. Jones-Rhoades, M. W., Bartel, D. P., & Bartel, B. (2006). MicroRNAs AND THEIR REGULATORY ROLES IN PLANTS. *Annual Review of Plant Biology*, 57(1), 19–53.
<https://doi.org/10.1146/annurev.arplant.57.032905.105218>
 28. McGeary, S. E., Lin, K. S., Shi, C. Y., Pham, T., Bisaria, N., Kelley, G. M., & Bartel, D. P. (2019). The biochemical basis of microRNA targeting efficacy. *Science*, eaav1741.
<https://doi.org/10.1126/science.aav1741>
 29. Jonas, S., & Izaurralde, E. (2015, July 19). Towards a molecular understanding of microRNA-mediated gene silencing. *Nature Reviews Genetics*. Nature Publishing Group. <https://doi.org/10.1038/nrg3965>
 30. Guo, H., Ingolia, N. T., Weissman, J. S., & Bartel, D. P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*, 466(7308), 835–840. <https://doi.org/10.1038/nature09267>
 31. Eichhorn, S. W., Guo, H., McGeary, S. E., Rodriguez-Mias, R. A., Shin, C., Baek, D., ... Bartel, D. P. (2014). mRNA Destabilization Is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Molecular Cell*, 56(1), 104–115. <https://doi.org/10.1016/j.molcel.2014.08.028>
 32. Hendrickson, D. G., Hogan, D. J., McCullough, H. L., Myers, J. W., Herschlag, D., Ferrell, J. E., & Brown, P. O. (2009). Concordant regulation of translation and mRNA abundance for hundreds of targets of a human microRNA. *PLoS Biology*, 7(11). <https://doi.org/10.1371/journal.pbio.1000238>
 33. Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., & Rajewsky, N. (2008). Widespread changes in protein synthesis induced by microRNAs. *Nature*, 455(7209), 58–63.
<https://doi.org/10.1038/nature07228>
 34. Bazzini, A. A., Lee, M. T., & Giraldez, A. J. (2012). Ribosome Profiling Shows That miR-430 Reduces Translation Before Causing mRNA Decay in Zebrafish. *Science*, 336(6078), 233–237.
<https://doi.org/10.1126/science.1215704>
 35. Subtelny, A. O., Eichhorn, S. W., Chen, G. R., Sive, H., & Bartel, D. P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature*, 508(7494), 66–71.
<https://doi.org/10.1038/nature13007>
 36. Djuranovic, S., Nahvi, A., & Green, R. (2012). miRNA-Mediated Gene Silencing by Translational Repression Followed by mRNA Deadenylation and Decay. *Science*, 336(6078), 237–240.
<https://doi.org/10.1126/science.1215691>

37. Axtell, M. J., Westholm, J. O., & Lai, E. C. (2011). Vive la différence: biogenesis and evolution of microRNAs in plants and animals. *Genome Biology*, *12*(4), 221. <https://doi.org/10.1186/gb-2011-12-4-221>
38. Rajagopalan, R., Vaucheret, H., Trejo, J., & Bartel, D. P. (2006). A diverse and evolutionarily fluid set of microRNAs in *Arabidopsis thaliana*. *Genes and Development*, *20*(24), 3407–3425. <https://doi.org/10.1101/gad.1476406>
39. Zhu, Q. H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., & Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. *Genome Research*, *18*(9), 1456–1465. <https://doi.org/10.1101/gr.075572.107>
40. Li, A., & Mao, L. (2007). Evolution of plant microRNA gene families. *Cell Research*, *17*(3), 212–218. <https://doi.org/10.1038/sj.cr.7310113>
41. Wang, J., Mei, J., & Ren, G. (2019). Plant microRNAs: Biogenesis, Homeostasis, and Degradation. *Frontiers in Plant Science*, *10*. <https://doi.org/10.3389/fpls.2019.00360>
42. Budak, H., & Akpinar, B. A. (2015). Plant miRNAs: biogenesis, organization and origins. *Functional & Integrative Genomics*, *15*(5), 523–531. <https://doi.org/10.1007/s10142-015-0451-2>
43. Li, J., Yang, Z., Yu, B., Liu, J., & Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in *Arabidopsis*. *Current Biology*, *15*(16), 1501–1507. <https://doi.org/10.1016/j.cub.2005.07.029>
44. Ren, G., Chen, X., & Yu, B. (2014, September 1). Small RNAs meet their targets: When methylation defends miRNAs from uridylation. *RNA Biology*. Landes Bioscience. <https://doi.org/10.4161/rna.36243>
45. Munafo, D. B., & Robb, G. B. (2010). Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA*, *16*(12), 2537–2552. <https://doi.org/10.1261/rna.2242610>
46. Androvic, P., Valihrach, L., Elling, J., Sjoback, R., & Kubista, M. (2017). Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Research*, *45*(15), e144–e144. <https://doi.org/10.1093/nar/gkx588>
47. Dard-Dascot, C., Naquin, D., D'Aubenton-Carafa, Y., Alix, K., Thermes, C., & van Dijk, E. (2018). Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics*, *19*(1), 118. <https://doi.org/10.1186/s12864-018-4491-6>
48. Iwakawa, H. oki, & Tomari, Y. (2013). Molecular insights into microRNA-mediated translational repression in plants. *Molecular Cell*, *52*(4), 591–601. <https://doi.org/10.1016/j.molcel.2013.10.033>
49. Brodersen, P., Sakvarelidze-Achard, L., Bruun-Rasmussen, M., Dunoyer, P., Yamamoto, Y. Y., Sieburth, L., & Voinnet, O. (2008). Widespread translational inhibition by plant miRNAs and siRNAs. *Science*, *320*(5880), 1185–1190. <https://doi.org/10.1126/science.1159151>
50. Fang, W., & Bartel, D. P. (2015). The Menu of Features that Define Primary MicroRNAs and Enable De Novo Design of MicroRNA Genes. *Molecular Cell*, *60*(1), 131–145. <https://doi.org/10.1016/j.molcel.2015.08.015>
51. Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. A., Holstein, T. W., Heber, S., & Peterson, K. J. (2009). The deep evolution of metazoan microRNAs. *Evolution and Development*, *11*(1), 50–68. <https://doi.org/10.1111/j.1525-142X.2008.00302.x>
52. Desvignes, T., Batzel, P., Berezikov, E., Eilbeck, K., Eppig, J. T., McAndrews, M. S., ... Postlethwait, J. H. (2015). miRNA Nomenclature: A View Incorporating Genetic Origins, Biosynthetic Pathways, and Sequence Variants. *Trends in Genetics*, *31*(11), 613–626. <https://doi.org/10.1016/j.tig.2015.09.002>
53. Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L., ... Peterson, K. J. (2015). A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. *Annual Review of Genetics*, *49*(1), 213–242. <https://doi.org/10.1146/annurev-genet-120213-092023>
54. Wang, X., & Liu, X. S. (2011). Systematic curation of mirbase annotation using integrated small RNA high-throughput sequencing data for *C. Elegans* and *Drosophila*. *Frontiers in Genetics*, *2*(MAY), 25. <https://doi.org/10.3389/fgene.2011.00025>

55. Meng, Y., Shao, C., Wang, H., & Chen, M. (2012). Are all the miRBase-registered microRNAs true? A structure- and expression-based re-examination in plants. *RNA Biology*, *9*(3), 249–253. <https://doi.org/10.4161/rna.19230>
56. 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. *Scientific Reports*, *7*(1), 1–11. <https://doi.org/10.1038/s41598-017-05070-0>
57. Alles, J., Fehlmann, T., Fischer, U., Backes, C., Galata, V., Minet, M., ... Meese, E. (2019). An estimate of the total number of true human miRNAs. *Nucleic Acids Research*, *47*(7), 3353–3364. <https://doi.org/10.1093/nar/gkz097>
58. Londin, E., Loher, P., Telonis, A. G., Quann, K., Clark, P., Jing, Y., ... Rigoutsos, I. (2015). Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. *Proceedings of the National Academy of Sciences*, *112*(10), E1106–E1115. <https://doi.org/10.1073/pnas.1420955112>
59. Isakova, A., & Quake, S. R. (2020). A Mouse Tissue Atlas of Small Non-Coding RNA. *bioRxiv*, 430561. <https://doi.org/10.1101/430561>
60. Engkvist, M. E., Stratford, E. W., Lorenz, S., Meza-Zepeda, L. A., Myklebost, O., & Munthe, E. (2017). Analysis of the miR-34 family functions in breast cancer reveals annotation error of miR-34b. *Scientific Reports*, *7*(1), 9655. <https://doi.org/10.1038/s41598-017-10189-1>
61. Zhong, X., Heinicke, F., & Rayner, S. (2019). miRBaseMiner, a tool for investigating miRBase content. *RNA Biology*, *16*(11), 1534–1546. <https://doi.org/10.1080/15476286.2019.1637680>
62. Van Peer, G., Lefever, S., Anckaert, J., Beckers, A., Rihani, A., Van Goethem, A., ... Vandesompele, J. (2014). miRBase Tracker: keeping track of microRNA annotation changes. *Database*, 2014. <https://doi.org/10.1093/database/bau080>
63. Fromm, B., Domanska, D., Høy, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., ... Peterson, K. J. (2020). MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Research*, *48*(D1), D132–D141. <https://doi.org/10.1093/nar/gkz885>
64. Kozomara, A., & Griffiths-Jones, S. (2014). miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Research*, *42*(D1), D68–D73. <https://doi.org/10.1093/nar/gkt1181>
65. Neilsen, C. T., Goodall, G. J., & Bracken, C. P. (2012). IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends in Genetics*, *28*(11), 544–549. <https://doi.org/10.1016/j.tig.2012.07.005>
66. Han, B. W., Hung, J.-H., Weng, Z., Zamore, P. D., & Ameres, S. L. (2011). The 3′-to-5′ Exoribonuclease Nibbler Shapes the 3′ Ends of MicroRNAs Bound to Drosophila Argonaute1. *Current Biology*, *21*(22), 1878–1887. <https://doi.org/10.1016/j.cub.2011.09.034>
67. Liu, N., Abe, M., Sabin, L. R., Hendriks, G.-J., Naqvi, A. S., Yu, Z., ... Bonini, N. M. (2011). The Exoribonuclease Nibbler Controls 3′ End Processing of MicroRNAs in Drosophila. *Current Biology*, *21*(22), 1888–1893. <https://doi.org/10.1016/j.cub.2011.10.006>
68. Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A. G., & Nishikura, K. (2007). Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science*, *315*(5815), 1137–1140. <https://doi.org/10.1126/science.1138050>
69. Tan, G. C., Chan, E., Molnar, A., Sarkar, R., Alexieva, D., Isa, I. M., ... Dibb, N. J. (2014). 5′ isomiR variation is of functional and evolutionary importance. *Nucleic Acids Research*, *42*(14), 9424–9435. <https://doi.org/10.1093/nar/gku656>
70. Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P., & Lai, E. C. (2007). Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. *Genome Research*, *17*(12), 1850–1864. <https://doi.org/10.1101/gr.6597907>
71. Chiang, H. R., Schoenfeld, L. W., Ruby, J. G., Auyeung, V. C., Spies, N., Baek, D., ... Bartel, D. P. (2010). Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. *Genes and Development*, *24*(10), 992–1009. <https://doi.org/10.1101/gad.1884710>
72. Agarwal, V., Bell, G. W., Nam, J.-W., & Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *eLife*, *4*. <https://doi.org/10.7554/eLife.05005>

73. Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., ... Grimmond, S. M. (2011). MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biology*, *12*(12), R126. <https://doi.org/10.1186/gb-2011-12-12-r126>
74. Koppers-Lalic, D., Hackenberg, M., Bijnsdorp, I. V., van Eijndhoven, M. A. J., Sadek, P., Sie, D., ... Pegtel, D. M. (2014). Nontemplated Nucleotide Additions Distinguish the Small RNA Composition in Cells from Exosomes. *Cell Reports*, *8*(6), 1649–1658. <https://doi.org/10.1016/j.celrep.2014.08.027>
75. Boele, J., Persson, H., Shin, J. W., Ishizu, Y., Newie, I. S., Sokilde, R., ... de Hoon, M. J. L. (2014). PAPD5-mediated 3' adenylation and subsequent degradation of miR-21 is disrupted in proliferative disease. *Proceedings of the National Academy of Sciences*, *111*(31), 11467–11472. <https://doi.org/10.1073/pnas.1317751111>
76. McCall, M. N., Kim, M.-S., Adil, M., Patil, A. H., Lu, Y., Mitchell, C. J., ... Halushka, M. K. (2017). Toward the human cellular microRNAome. *Genome Research*, *27*(10), 1769–1781. <https://doi.org/10.1101/gr.222067.117>
77. Koppers-lalic, D., Hackenberg, M., & Menezes, R. De. (2014). Non-invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget*, *7*(16). <https://doi.org/10.18632/oncotarget.8124>
78. Nejad, C., Pillman, K. A., Siddle, K. J., Pépin, G., Änkö, M.-L., McCoy, C. E., ... Gantier, M. P. (2018). miR-222 isoforms are differentially regulated by type-I interferon. *RNA*, *24*(3), 332–341. <https://doi.org/10.1261/rna.064550.117>
79. Yu, F., Pillman, K. A., Neilsen, C. T., Toubia, J., Lawrence, D. M., Tsykin, A., ... Bracken, C. P. (2017). Naturally existing isoforms of miR-222 have distinct functions. *Nucleic Acids Research*, *45*(19), 11371–11385. <https://doi.org/10.1093/nar/gkx788>
80. McGahon, M. K., Yarham, J. M., Daly, A., Guduric-Fuchs, J., Ferguson, L. J., Simpson, D. A., & Collins, A. (2013). Distinctive Profile of IsomiR Expression and Novel MicroRNAs in Rat Heart Left Ventricle. *PLoS ONE*, *8*(6), e65809. <https://doi.org/10.1371/journal.pone.0065809>
81. Baran-Gale, J., Fannin, E. E., Kurtz, C. L., & Sethupathy, P. (2013). Beta Cell 5'-Shifted isomiRs Are Candidate Regulatory Hubs in Type 2 Diabetes. *PLoS ONE*, *8*(9), e73240. <https://doi.org/10.1371/journal.pone.0073240>
82. Blondal, T., Brunetto, M. R., Cavallone, D., Mikkelsen, M., Thorsen, M., Mang, Y., ... Mouritzen, P. (2017). Genome-Wide Comparison of Next-Generation Sequencing and qPCR Platforms for microRNA Profiling in Serum (pp. 21–44). https://doi.org/10.1007/978-1-4939-6866-4_3
83. Telonis, A. G., Magee, R., Loher, P., Chervoneva, I., Londin, E., & Rigoutsos, I. (2017). Knowledge about the presence or absence of miRNA isoforms (isomiRs) can successfully discriminate amongst 32 TCGA cancer types. *Nucleic Acids Research*, *45*(6), 2973–2985. <https://doi.org/10.1093/nar/gkx082>
84. Telonis, A. G., Loher, P., Jing, Y., Londin, E., & Rigoutsos, I. (2015). Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Research*, *43*(19), 9158–9175. <https://doi.org/10.1093/nar/gkv922>
85. Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K. S., & Zhao, P. X. (2014). Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in *Arabidopsis thaliana*. *RNA Biology*, *11*(11), 1414–1429. <https://doi.org/10.1080/15476286.2014.996474>
86. Chim, S. S. C., Shing, T. K. F., Hung, E. C. W., Leung, T. Y., Lau, T. K., Chiu, R. W. K., & Lo, Y. M. D. (2008). Detection and characterization of placental microRNAs in maternal plasma. *Clinical Chemistry*, *54*(3), 482–490. <https://doi.org/10.1373/clinchem.2007.097972>
87. Lawrie, C. H., Gal, S., Dunlop, H. M., Pushkaran, B., Liggins, A. P., Pulford, K., ... Harris, A. L. (2008). Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *British Journal of Haematology*, *141*(5), 672–675. <https://doi.org/10.1111/j.1365-2141.2008.07077.x>
88. Mitchell, P. S., Parkin, R. K., Kroh, E. M., Fritz, B. R., Wyman, S. K., Pogossova-Agadjanyan, E. L., ... Tewari, M. (2008). Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(30), 10513–10518.

- <https://doi.org/10.1073/pnas.0804549105>
89. da Silveira, J. C., Veeramachaneni, D. N. R., Winger, Q. A., Carnevale, E. M., & Bouma, G. J. (2012). Cell-Secreted Vesicles in Equine Ovarian Follicular Fluid Contain miRNAs and Proteins: A Possible New Form of Cell Communication Within the Ovarian Follicle. *Biology of Reproduction*, *86*(3). <https://doi.org/10.1095/biolreprod.111.093252>
 90. Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., How Huang, K., Jen Lee, M., ... Wang, K. (2010). The MicroRNA Spectrum in 12 Body Fluids. *Clinical Chemistry*, *56*(11), 1733–1741. <https://doi.org/10.1373/clinchem.2010.147405>
 91. Zhou, Q., Li, M., Wang, X., Li, Q., Wang, T., Zhu, Q., ... Li, X. (2011). Immune-related microRNAs are abundant in breast milk exosomes. *International Journal of Biological Sciences*, *8*(1), 118–123. <https://doi.org/10.7150/ijbs.8.118>
 92. Gallo, A., Tandon, M., Alevizos, I., & Illei, G. G. (2012). The Majority of MicroRNAs Detectable in Serum and Saliva Is Concentrated in Exosomes. *PLoS ONE*, *7*(3), e30679. <https://doi.org/10.1371/journal.pone.0030679>
 93. Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., ... Zhang, C. Y. (2008). Characterization of microRNAs in serum: A novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Research*, *18*(10), 997–1006. <https://doi.org/10.1038/cr.2008.282>
 94. Arroyo, J. D., Chevillet, J. R., Kroh, E. M., Ruf, I. K., Pritchard, C. C., Gibson, D. F., ... Tewari, M. (2011). Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(12), 5003–5008. <https://doi.org/10.1073/pnas.1019055108>
 95. Théry, C., Witwer, K. W., Aikawa, E., Alcaraz, M. J., Anderson, J. D., Andriantsitohaina, R., ... Zuba-Surma, E. K. (2018). Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *Journal of Extracellular Vesicles*, *7*(1). <https://doi.org/10.1080/20013078.2018.1535750>
 96. Murillo, O. D., Thistlethwaite, W., Rozowsky, J., Subramanian, S. L., Lucero, R., Shah, N., ... Milosavljevic, A. (2019). exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell*, *177*(2), 463–477.e15. <https://doi.org/10.1016/j.cell.2019.02.018>
 97. O'Brien, J., Hayder, H., Zayed, Y., & Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. *Frontiers in Endocrinology*, *9*(AUG), 402. <https://doi.org/10.3389/fendo.2018.00402>
 98. Kibel, A. S. (2009, July 1). Commentary on Circulating microRNAs as stable blood-based markers for cancer detection. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Briant KC, Allen A, Lin DW, Urban N, Drescher CW, Knudsen, Stirewalt DL, Gentleman R, Vessella RL, Nelson PS, Martin DB, Tewari M, Divisions of Human Biology, Clinical Research, and Public Health Services. *Urologic Oncology: Seminars and Original Investigations*. Elsevier. <https://doi.org/10.1016/j.urolonc.2009.04.001>
 99. Cui, M., Wang, H., Yao, X., Zhang, D., Xie, Y., Cui, R., & Zhang, X. (2019). Circulating MicroRNAs in Cancer: Potential and Challenge. *Frontiers in Genetics*, *10*. <https://doi.org/10.3389/fgene.2019.00626>
 100. Grasedieck, S., Schöler, N., Bommer, M., Niess, J. H., Tumani, H., Rouhi, A., ... Kuchenbauer, F. (2012, November 16). Impact of serum storage conditions on microRNA stability. *Leukemia*. Nature Publishing Group. <https://doi.org/10.1038/leu.2012.106>
 101. Bayraktar, R., Roosbroeck, K. Van, & Calin, G. A. (2017). Cell-to-cell communication: microRNAs as hormones. *Molecular Oncology*, *11*(12), 1673–1686. [https://doi.org/10.1002/1878-0261.12144@10.1002/\(ISSN\)1878-0261.REVIEWS](https://doi.org/10.1002/1878-0261.12144@10.1002/(ISSN)1878-0261.REVIEWS)
 102. Mori, M. A., Ludwig, R. G., Garcia-Martin, R., Brandão, B. B., & Kahn, C. R. (2019). Extracellular miRNAs: From Biomarkers to Mediators of Physiology and Disease. *Cell Metabolism*, *30*(4), 656–673. <https://doi.org/10.1016/j.cmet.2019.07.011>
 103. Anfossi, S., Babayan, A., Pantel, K., & Calin, G. A. (2018). Clinical utility of circulating non-coding RNAs — an update. *Nature Reviews Clinical Oncology*, *15*(9), 541–563. <https://doi.org/10.1038/s41571-018-0035-x>

104. He, Y., Lin, J., Kong, D., Huang, M., Xu, C., Kim, T.-K., ... Wang, K. (2015). Current State of Circulating MicroRNAs as Cancer Biomarkers. *Clinical Chemistry*, *61*(9), 1138–1155. <https://doi.org/10.1373/clinchem.2015.241190>
105. Lee, I., Baxter, D., Lee, M. Y., Scherler, K., & Wang, K. (2017). The Importance of Standardization on Analyzing Circulating RNA. *Molecular Diagnosis and Therapy*. <https://doi.org/10.1007/s40291-016-0251-y>
106. Witwer, K. W. (2015). Circulating MicroRNA Biomarker Studies: Pitfalls and Potential Solutions. *Clinical Chemistry*, *61*(1), 56–63. <https://doi.org/10.1373/clinchem.2014.221341>
107. Moldovan, L., Batte, K. E., Trgovcich, J., Wisler, J., Marsh, C. B., & Piper, M. (2014). Methodological challenges in utilizing miRNAs as circulating biomarkers. *Journal of Cellular and Molecular Medicine*, *18*(3), 371–390. <https://doi.org/10.1111/jcmm.12236>
108. Chugh, P., & Dittmer, D. P. (2012). Potential pitfalls in microRNA profiling. *Wiley Interdisciplinary Reviews: RNA*. <https://doi.org/10.1002/wrna.1120>
109. Buschmann, D., Haberberger, A., Kirchner, B., Spornraft, M., Riedmaier, I., Schelling, G., & Pfaffl, M. W. (2016). Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow. *Nucleic Acids Research*, *44*(13), 5995–6018. <https://doi.org/10.1093/nar/gkw545>
110. Takahashi, K., Yokota, S. ichi, Tatsumi, N., Fukami, T., Yokoi, T., & Nakajima, M. (2013). Cigarette smoking substantially alters plasma microRNA profiles in healthy subjects. *Toxicology and Applied Pharmacology*, *272*(1), 154–160. <https://doi.org/10.1016/j.taap.2013.05.018>
111. Flowers, E., Won, G. Y., & Fukuoka, Y. (2015). MicroRNAs associated with exercise and diet: a systematic review. *Physiological Genomics*, *47*(1), 1–11. <https://doi.org/10.1152/physiolgenomics.00095.2014>
112. Ameling, S., Kacprowski, T., Chilukoti, R. K., Malsch, C., Liebscher, V., Suhre, K., ... Völker, U. (2015). Associations of circulating plasma microRNAs with age, body mass index and sex in a population-based study. *BMC Medical Genomics*, *8*(1), 61. <https://doi.org/10.1186/s12920-015-0136-7>
113. Margue, C., Reinsbach, S., Philippidou, D., Beaume, N., Walters, C., Schneider, J. G., ... Kreis, S. (2015). Comparison of a healthy miRNome with melanoma patient miRNomes: Are microRNAs suitable serum biomarkers for cancer? *Oncotarget*, *6*(14), 12110–12127. <https://doi.org/10.18632/oncotarget.3661>
114. Max, K. E. A., Bertram, K., Akat, K. M., Bogardus, K. A., Li, J., Morozov, P., ... Tuschl, T. (2018). Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proceedings of the National Academy of Sciences*, *115*(23), E5334–E5343. <https://doi.org/10.1073/pnas.1714397115>
115. Tichopad, A., Kitchen, R., Riedmaier, I., Becker, C., & Ståhlberg, A. (2009). Design and Optimization of Reverse-Transcription Quantitative PCR Experiments METHOD :, *1823*. <https://doi.org/10.1373/clinchem.2009.126201>
116. Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics (Oxford, England)*, *30*(3), 301–4. <https://doi.org/10.1093/bioinformatics/btt688>
117. Hart, S. N., Therneau, T. M., Zhang, Y., Poland, G. A., & Kocher, J. P. (2013). Calculating sample size estimates for RNA sequencing data. *Journal of Computational Biology*, *20*(12), 970–978. <https://doi.org/10.1089/cmb.2012.0283>
118. Williams, A. G., Thomas, S., Wyman, S. K., & Holloway, A. K. (2014). RNA-seq Data: Challenges in and Recommendations for Experimental Design and Analysis. *Current protocols in human genetics*, *83*, 11.13.1-11.13.20. <https://doi.org/10.1002/0471142905.hg1113s83>
119. Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., ... Barton, G. J. (2016). How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, *22*(6), 839–851. <https://doi.org/10.1261/rna.053959.115>
120. Jung, M., Schaefer, A., Steiner, I., Kempkensteffen, C., Stephan, C., Erbersdobler, A., & Jung, K. (2010). Robust MicroRNA stability in degraded RNA preparations from human tissue and cell samples. *Clinical Chemistry*, *56*(6), 998–1006. <https://doi.org/10.1373/clinchem.2009.141580>
121. Hall, J. S., Taylor, J., Valentine, H. R., Irlam, J. J., Eustace, A., Hoskin, P. J., ... West, C. M. L. (2012).

- Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *British Journal of Cancer*, 107(4), 684–694. <https://doi.org/10.1038/bjc.2012.294>
122. Sidova, M., Tomankova, S., Abaffy, P., Kubista, M., & Sindelka, R. (2015). Effects of post-mortem and physical degradation on RNA integrity and quality. *Biomolecular Detection and Quantification*, 5, 3–9. <https://doi.org/10.1016/j.bdq.2015.08.002>
 123. Ibberson, D., Benes, V., Muckenthaler, M. U., & Castoldi, M. (2009). RNA degradation compromises the reliability of microRNA expression profiling. *BMC Biotechnology*, 9(1), 102. <https://doi.org/10.1186/1472-6750-9-102>
 124. Becker, C., Hammerle-Fickinger, A., Riedmaier, I., & Pfaffl, M. W. (2010). mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, 50(4), 237–243. <https://doi.org/10.1016/j.ymeth.2010.01.010>
 125. Balzano, F., Deiana, M., Dei Giudici, S., Oggiano, A., Baralla, A., Pasella, S., ... Deiana, L. (2015). miRNA Stability in Frozen Plasma Samples. *Molecules*, 20(10), 19030–19040. <https://doi.org/10.3390/molecules201019030>
 126. Glinge, C., Clauss, S., Boddum, K., Jabbari, R., Jabbari, J., Risgaard, B., ... Tfelt-Hansen, J. (2017). Stability of circulating blood-based microRNAs-Pre-Analytic methodological considerations. *PLoS ONE*, 12(2). <https://doi.org/10.1371/journal.pone.0167969>
 127. Kirschner, M. B., Edelman, J. J. B., Kao, S. C., Michael, P., & Reid, G. (2013). The impact of hemolysis on cell-free microRNA biomarkers, 4(May), 1–13. <https://doi.org/10.3389/fgene.2013.00094>
 128. Pritchard, C. C., Kroh, E., Wood, B., Arroyo, J. D., Dougherty, K. J., Miyaji, M. M., ... Tewari, M. (2012). Blood cell origin of circulating microRNAs: A cautionary note for cancer biomarker studies. *Cancer Prevention Research*, 5(3), 492–497. <https://doi.org/10.1158/1940-6207.CAPR-11-0370>
 129. Valihrach, L., Androvic, P., & Kubista, M. (2019). Circulating miRNA analysis for cancer diagnostics and therapy. *Molecular Aspects of Medicine*, 100825. <https://doi.org/10.1016/j.mam.2019.10.002>
 130. Kim, Y. K., Yeo, J., Kim, B., Ha, M., & Kim, V. N. (2012, June 29). Short Structured RNAs with Low GC Content Are Selectively Lost during Extraction from a Small Number of Cells. *Molecular Cell*. Cell Press. <https://doi.org/10.1016/j.molcel.2012.05.036>
 131. Duy, J., Koehler, J. W., Honko, A. N., & Minogue, T. D. (2015). Optimized microRNA purification from TRIzol-treated plasma. *BMC Genomics*, 16(1), 1–9. <https://doi.org/10.1186/s12864-015-1299-5>
 132. Clerget, G., Bourguignon-Igel, V., & Rederstorff, M. (2015). Alcoholic precipitation of small non-coding RNAs. *Methods in Molecular Biology*, 1296, 11–16. https://doi.org/10.1007/978-1-4939-2547-6_2
 133. Redshaw, N., Wilkes, T., Whale, A., Cowen, S., Huggett, J., & Foy, C. A. (2013). A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability. <https://doi.org/10.2144/000114002>
 134. Guo, Y., Vickers, K., Xiong, Y., Zhao, S., Sheng, Q., Zhang, P., ... Flynn, C. R. (2017). Comprehensive evaluation of extracellular small RNA isolation methods from serum in high throughput sequencing. *BMC Genomics*, 18(1), 50. <https://doi.org/10.1186/s12864-016-3470-z>
 135. McAlexander, M. A., Phillips, M. J., & Witwer, K. W. (2013). Comparison of methods for miRNA extraction from plasma and quantitative recovery of RNA from cerebrospinal fluid. *Frontiers in Genetics*, 4(MAY), 1–8. <https://doi.org/10.3389/fgene.2013.00083>
 136. Androvic, P., Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., & Valihrach, L. (2019). Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Scientific Reports*, 9(1), 4255. <https://doi.org/10.1038/s41598-019-40513-w>
 137. El-khoury, V., Pierson, S., Kaoma, T., Bernardin, F., & Berchem, G. (2016). Assessing cellular and circulating miRNA recovery : the impact of the RNA isolation method and the quantity of input material. *Nature Publishing Group*, (December 2015), 1–14. <https://doi.org/10.1038/srep19529>
 138. Kim, D. J., Linnstaedt, S., Palma, J., Park, J. C., Ntrivalas, E., Kwak-Kim, J. Y. H., ... Duelli, D. M. (2012). Plasma components affect accuracy of circulating cancer-related microRNA quantitation. *Journal of Molecular Diagnostics*, 14(1), 71–80. <https://doi.org/10.1016/j.jmoldx.2011.09.002>

139. Andreasen, D., Fog, J. U., Biggs, W., Salomon, J., Dahlsveen, I. K., Baker, A., & Mouritzen, P. (2010). Improved microRNA quantification in total RNA from clinical samples. *Methods*, *50*(4), S6–S9. <https://doi.org/10.1016/j.ymeth.2010.01.006>
140. Van Deun, J., Mestdagh, P., Agostinis, P., Akay, Ö., Anand, S., Anckaert, J., ... Hendrix, A. (2017, February 28). EV-TRACK: Transparent reporting and centralizing knowledge in extracellular vesicle research. *Nature Methods*. Nature Publishing Group. <https://doi.org/10.1038/nmeth.4185>
141. Taylor, D. D., & Shah, S. (2015, October 1). Methods of isolating extracellular vesicles impact downstream analyses of their cargoes. *Methods*. Academic Press Inc. <https://doi.org/10.1016/j.ymeth.2015.02.019>
142. Das, S., Abdel-Mageed, A. B., Adamidi, C., Adelson, P. D., Akat, K. M., Alsop, E., ... Saugstad, J. A. (2019, April 4). The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research. *Cell*. Cell Press. <https://doi.org/10.1016/j.cell.2019.03.023>
143. Srinivasan, S., Yeri, A., Cheah, P. S., Chung, A., Danielson, K., De Hoff, P., ... Laurent, L. C. (2019). Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. *Cell*, *177*(2), 446–462.e16. <https://doi.org/10.1016/j.cell.2019.03.024>
144. Buschmann, D., Kirchner, B., Hermann, S., Märte, M., Wurmser, C., Brandes, F., ... Reithmair, M. (2018). Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *Journal of Extracellular Vesicles*. <https://doi.org/10.1080/20013078.2018.1481321>
145. Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., ... Shipley, G. L. (2009). The MIQE Guidelines : M inimum I nformation for Publication of Q uantitative Real-Time PCR E xperiments SUMMARY :, *622*, 611–622. <https://doi.org/10.1373/clinchem.2008.112797>
146. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., ... Ragg, T. (2006). The RIN: An RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology*, *7*, 3. <https://doi.org/10.1186/1471-2199-7-3>
147. Jiang, L., Schlesinger, F., Davis, C. A., Zhang, Y., Li, R., Salit, M., ... Oliver, B. (2011). Synthetic spike-in standards for RNA-seq experiments. *Genome Research*, *21*(9), 1543–1551. <https://doi.org/10.1101/gr.121095.111>
148. Yang, I. V. (2006). Use of External Controls in Microarray Experiments. In *Methods in Enzymology* (Vol. 411, pp. 50–63). Academic Press. [https://doi.org/10.1016/S0076-6879\(06\)11004-6](https://doi.org/10.1016/S0076-6879(06)11004-6)
149. Blondal, T., Nielsen, S. J., Baker, A., Andreasen, D., Mouritzen, P., Teilum, M. W., & Dahlsveen, I. K. (2013). Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods*, *59*(1), S1–S6. <https://doi.org/10.1016/j.ymeth.2012.09.015>
150. Kirschner, M. B., Kao, S. C., Edelman, J. J., Armstrong, N. J., Vallety, M. P., van Zandwijk, N., & Reid, G. (2011). Haemolysis during Sample Preparation Alters microRNA Content of Plasma. *PLoS ONE*, *6*(9), e24145. <https://doi.org/10.1371/journal.pone.0024145>
151. Vliet, E. A. Van, Puhakka, N., Mills, J. D., Srivastava, P. K., Michael, R., Roncon, P., ... Simonato, M. (2017). Standardization procedure for plasma biomarker analysis in rat models of epileptogenesis : Focus on circulating microRNAs, 1–12. <https://doi.org/10.1111/epi.13915>
152. Benes, V., & Castoldi, M. (2010). Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods*, *50*(4), 244–249. <https://doi.org/10.1016/j.ymeth.2010.01.026>
153. Pritchard, C. C., Cheng, H. H., & Tewari, M. (2012). MicroRNA profiling: approaches and considerations. *Nature Reviews Genetics*, *13*(5), 358–369. <https://doi.org/10.1038/nrg3198>
154. Hunt, E. A., Broyles, D., Head, T., & Deo, S. K. (2015). MicroRNA Detection: Current Technology and Research Strategies. *Annual Review of Analytical Chemistry*, *8*(1), 217–237. <https://doi.org/10.1146/annurev-anchem-071114-040343>
155. Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., ... Vandesompele, J. (2014). Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study, (june). <https://doi.org/10.1038/nmeth.3014>
156. Honda, S., & Kirino, Y. (2015). Dumbbell-PCR: a method to quantify specific small RNA variants with a

- single nucleotide resolution at terminal sequences. *Nucleic Acids Research*, *43*(12), e77–e77. <https://doi.org/10.1093/nar/gkv218>
157. Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonák, J., Lind, K., ... Zoric, N. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, *27*(2–3), 95–125. <https://doi.org/10.1016/j.mam.2005.12.007>
 158. Raymond, C. K. (2005). Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA*, *11*(11), 1737–1744. <https://doi.org/10.1261/rna.2148705>
 159. Chen, C. (2005). Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Research*, *33*(20), e179–e179. <https://doi.org/10.1093/nar/gni178>
 160. Huang, T., Yang, J., Liu, G., Jin, W., Liu, Z., Zhao, S., & Yao, M. (2015). Quantification of Mature MicroRNAs Using Pincer Probes and Real-Time PCR Amplification. *PLOS ONE*, *10*(3), e0120160. <https://doi.org/10.1371/journal.pone.0120160>
 161. Schamberger, A., & Orbán, T. I. (2014). 3' IsomiR Species and DNA Contamination Influence Reliable Quantification of MicroRNAs by Stem-Loop Quantitative PCR. *PLoS ONE*, *9*(8), e106315. <https://doi.org/10.1371/journal.pone.0106315>
 162. Balcells, I., Cirera, S., & Busk, P. K. (2011). Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. *BMC Biotechnology*, *11*(1), 70. <https://doi.org/10.1186/1472-6750-11-70>
 163. Benes, V., Collier, P., Kordes, C., Stolte, J., Rausch, T., Muckentaler, M. U., ... Castoldi, M. (2015). Identification of cytokine-induced modulation of microRNA expression and secretion as measured by a novel microRNA specific qPCR assay. *Scientific Reports*, *5*(1), 11590. <https://doi.org/10.1038/srep11590>
 164. Ballantyne, K. N., Oorschot, R. A. H. Van, & Mitchell, R. J. (2008). Locked nucleic acids in PCR primers increase sensitivity and performance, *91*, 301–305. <https://doi.org/10.1016/j.ygeno.2007.10.016>
 165. Zipper, H. (2004). Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research*, *32*(12), e103–e103. <https://doi.org/10.1093/nar/gnh101>
 166. Tam, S., de Borja, R., Tsao, M.-S., & McPherson, J. D. (2014). Robust global microRNA expression profiling using next-generation sequencing technologies. *Laboratory Investigation*, *94*(3), 350–358. <https://doi.org/10.1038/labinvest.2013.157>
 167. Persson, H., Søkilde, R., Pirona, A. C., & Rovira, C. (2017). Preparation of highly multiplexed small RNA sequencing libraries. *BioTechniques*, *63*(2), 57–64. <https://doi.org/10.2144/000114574>
 168. Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., ... Tuschl, T. (2008). Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods*, *44*(1), 3–12. <https://doi.org/10.1016/j.ymeth.2007.09.009>
 169. Hardigan, A. A., Roberts, B. S., Moore, D. E., Ramaker, R. C., Jones, A. L., & Myers, R. M. (2019). CRISPR/Cas9-targeted removal of unwanted sequences from small-RNA sequencing libraries. *Nucleic Acids Research*, *47*(14). <https://doi.org/10.1093/nar/gkz425>
 170. Kawano, M., Kawazu, C., Lizio, M., Kawaji, H., Carninci, P., Suzuki, H., & Hayashizaki, Y. (2010). Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *BioTechniques*, *49*(4), 751–755. <https://doi.org/10.2144/000113516>
 171. Alon, S., Vigneault, F., Eminaga, S., Christodoulou, D. C., Seidman, J. G., Church, G. M., & Eisenberg, E. (2011). Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Research*, *21*(9), 1506–1511. <https://doi.org/10.1101/gr.121715.111>
 172. Pease, J. (2011). Small-RNA sequencing libraries with greatly reduced adaptor-dimer background. *Nature Methods*, *8*(3), iii–iv. <https://doi.org/10.1038/nmeth.f.336>
 173. Shore, S., Henderson, J. M., Lebedev, A., Salcedo, M. P., Zon, G., McCaffrey, A. P., ... Hogrefe, R. I. (2016). Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation. *PLoS ONE*, *11*(11). <https://doi.org/10.1371/journal.pone.0167009>
 174. Raabe, C. A., Tang, T., Brosius, J., & Rozhdestvensky, T. S. (2014). Biases in small RNA deep sequencing

- data, 42(3), 1414–1426. <https://doi.org/10.1093/nar/gkt1021>
175. Zhuang, F., Fuchs, R. T., Sun, Z., Zheng, Y., & Robb, G. B. (2012). Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Research*, 40(7). <https://doi.org/10.1093/nar/gkr1263>
 176. Jayaprakash, A. D., Jadoo, O., Brown, B. D., & Sachidanandam, R. (2011). Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research*, 39(21), 1–12. <https://doi.org/10.1093/nar/gkr693>
 177. Baran-Gale, J., Lisa Kurtz, C., Erdos, M. R., Sison, C., Young, A., Fannin, E. E., ... Sethupathy, P. (2015). Addressing bias in small RNA library preparation for sequencing: A new protocol recovers microRNAs that evade capture by current methods. *Frontiers in Genetics*, 6(DEC), 1–9. <https://doi.org/10.3389/fgene.2015.00352>
 178. Barberán-Soler, S., Vo, J. M., Hogans, R. E., Dallas, A., Johnston, B. H., & Kazakov, S. A. (2018). Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biology*, 19(1), 105. <https://doi.org/10.1186/s13059-018-1488-z>
 179. Baroin-Tourancheau, A., Jaszczyszyn, Y., Benigni, X., & Amar, L. (2019). Evaluating and Correcting Inherent Bias of microRNA Expression in Illumina Sequencing Analysis. *Frontiers in Molecular Biosciences*, 6(APR). <https://doi.org/10.3389/fmolb.2019.00017>
 180. Argyropoulos, C., Etheridge, A., Sakhanenko, N., & Galas, D. (2017). Modeling bias and variation in the stochastic processes of small RNA sequencing. *Nucleic Acids Research*, 45(11), 104. <https://doi.org/10.1093/nar/gkx199>
 181. Fuchs, R. T., Sun, Z., Zhuang, F., & Robb, G. B. (2015). Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS ONE*, 10(5). <https://doi.org/10.1371/journal.pone.0126049>
 182. Hafner, M., Renwick, N., Brown, M., Mihailović, A., Holoch, D., Lin, C., ... Tuschl, T. (2011). RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*, 17(9), 1697–1712. <https://doi.org/10.1261/rna.2799511>
 183. Fu, Y., Wu, P. H., Beane, T., Zamore, P. D., & Weng, Z. (2018). Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics*, 19(1), 1–14. <https://doi.org/10.1186/s12864-018-4933-1>
 184. Wright, C., Rajpurohit, A., Burke, E. E., Williams, C., Collado-Torres, L., Kimos, M., ... Shin, J. H. (2019). Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics*, 20(1), 513. <https://doi.org/10.1186/s12864-019-5870-3>
 185. Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., ... Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. *Nature Methods*, 11(2), 163–166. <https://doi.org/10.1038/nmeth.2772>
 186. Coenen-Stass, A. M. L., Magen, I., Brooks, T., Ben-Dov, I. Z., Greensmith, L., Hornstein, E., & Fratta, P. (2018). Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA biology*, 15(8), 1133–1145. <https://doi.org/10.1080/15476286.2018.1514236>
 187. Giraldez, M. D., Spengler, R. M., Etheridge, A., Godoy, P. M., Barczak, A. J., Srinivasan, S., ... Tewari, M. (2018). Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nature Biotechnology*, 36(8), 746–757. <https://doi.org/10.1038/nbt.4183>
 188. Yeri, A., Courtright, A., Danielson, K., Hutchins, E., Alsop, E., Carlson, E., ... Van Keuren-Jensen, K. (2018). Evaluation of commercially available small RNAseq library preparation kits using low input RNA. *BMC Genomics*, 19(1), 331. <https://doi.org/10.1186/s12864-018-4726-6>
 189. Tsang, H. F., Xue, V. W., Koh, S. P., Chiu, Y. M., Ng, L. P. W., & Wong, S. C. C. (2017, January 2). NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics. *Expert Review of Molecular Diagnostics*. Taylor and Francis Ltd. <https://doi.org/10.1080/14737159.2017.1268533>
 190. Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., ... Hood, L. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*,

- 26(3), 317–325. <https://doi.org/10.1038/nbt1385>
191. Jin, J., Vaud, S., Zhelkovsky, A. M., Posfai, J., & McReynolds, L. A. (2016). Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Research*, *44*(13), e116–e116. <https://doi.org/10.1093/nar/gkw399>
 192. Li, J., Yao, B., Huang, H., Wang, Z., Sun, C., Fan, Y., ... Xi, J. (2009). Real-time polymerase chain reaction microRNA detection based on enzymatic stem-loop probes ligation. *Analytical Chemistry*, *81*(13), 5446–5451. <https://doi.org/10.1021/ac900598d>
 193. Ma, F., Liu, M., Tang, B., & Zhang, C. Y. (2017). Sensitive Quantification of MicroRNAs by Isothermal Helicase-Dependent Amplification. *Analytical Chemistry*, *89*(11), 6182–6187. <https://doi.org/10.1021/acs.analchem.7b01113>
 194. Deng, R., Zhang, K., & Li, J. (2017). Isothermal Amplification for MicroRNA Detection: From the Test Tube to the Cell. *Accounts of Chemical Research*, *50*(4), 1059–1068. <https://doi.org/10.1021/acs.accounts.7b00040>
 195. Kilic, T., Erdem, A., Ozsoz, M., & Carrara, S. (2018, January 15). microRNA biosensors: Opportunities and challenges among conventional and commercially available techniques. *Biosensors and Bioelectronics*. Elsevier Ltd. <https://doi.org/10.1016/j.bios.2017.08.007>
 196. Johnson, B. N., & Mutharasan, R. (2014). Biosensor-based microRNA detection: Techniques, design, performance, and challenges. *Analyst*, *139*(7), 1576–1588. <https://doi.org/10.1039/c3an01677c>
 197. Fu, X., & Dong, D. (2018). Bioinformatic analysis of MicroRNA sequencing data. In *Methods in Molecular Biology* (Vol. 1751, pp. 109–125). Humana Press Inc. https://doi.org/10.1007/978-1-4939-7710-9_8
 198. Motameny, S., Wolters, S., Nürnberg, P., & Schumacher, B. (2010). Next Generation Sequencing of miRNAs – Strategies, Resources and Methods. *Genes*, *1*(1), 70–84. <https://doi.org/10.3390/genes1010070>
 199. Norton, S. S., Vaquero-Garcia, J., Lahens, N. F., Grant, G. R., & Barash, Y. (2018). Outlier detection for improved differential splicing quantification from RNA-Seq experiments with replicates. *Bioinformatics*, *34*(9), 1488–1497. <https://doi.org/10.1093/bioinformatics/btx790>
 200. Schwarzenbach, H., da Silva, A. M., Calin, G., & Pantel, K. (2015). Data Normalization Strategies for MicroRNA Quantification. *Clinical Chemistry*, *61*(11), 1333–1342. <https://doi.org/10.1373/clinchem.2015.239459>
 201. Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., & Vandesompele, J. (2009). A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biology*, *10*(6), R64. <https://doi.org/10.1186/gb-2009-10-6-r64>
 202. Garmire, L. X., & Subramaniam, S. (2012). Evaluation of normalization methods in mammalian microRNA-Seq data. *RNA*, *18*(6), 1279–1288. <https://doi.org/10.1261/rna.030916.111>
 203. Dillies, M. A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., ... Jaffrézic, F. (2013). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, *14*(6), 671–683. <https://doi.org/10.1093/bib/bbs046>
 204. McCall, M. N., McMurray, H. R., Land, H., & Almudevar, A. (2014). On non-detects in qPCR data. *Bioinformatics*, *30*(16), 2310–2316. <https://doi.org/10.1093/bioinformatics/btu239>
 205. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. <https://doi.org/10.1186/s13059-014-0550-8>
 206. Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. <https://doi.org/10.1093/bioinformatics/btp616>
 207. Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. <https://doi.org/10.1186/gb-2014-15-2-r29>
 208. Sonesson, C., & Delorenzi, M. (2013). A comparison of methods for differential expression analysis of

- RNA-seq data. *BMC Bioinformatics*, 14(1), 1–18. <https://doi.org/10.1186/1471-2105-14-91>
209. Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., ... Betel, D. (2013). Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biology*, 14(9), R95. <https://doi.org/10.1186/gb-2013-14-9-r95>
 210. Riffo-Campos, Á., Riquelme, I., & Brebi-Mieville, P. (2016). Tools for Sequence-Based miRNA Target Prediction: What to Choose? *International Journal of Molecular Sciences*, 17(12), 1987. <https://doi.org/10.3390/ijms17121987>
 211. Singh, N. K. (2017). miRNAs target databases: developmental methods and target identification techniques with functional annotations. *Cellular and Molecular Life Sciences*, 74(12), 2239–2261. <https://doi.org/10.1007/s00018-017-2469-1>
 212. van Iterson, M., Bervoets, S., de Meijer, E. J., Buermans, H. P., 't Hoen, P. A. C., Menezes, R. X., & Boer, J. M. (2013). Integrated analysis of microRNA and mRNA expression: adding biological significance to microRNA target predictions. *Nucleic Acids Research*, 41(15), e146–e146. <https://doi.org/10.1093/nar/gkt525>
 213. Qureshi, I. A., & Mehler, M. F. (2012). Emerging roles of non-coding RNAs in brain evolution, development, plasticity and disease. *Nature Reviews Neuroscience*, 13(8), 528–541. <https://doi.org/10.1038/nrn3234>
 214. Cao, D.-D., Li, L., & Chan, W.-Y. (2016). MicroRNAs: Key Regulators in the Central Nervous System and Their Implication in Neurological Diseases. *International journal of molecular sciences*, 17(6). <https://doi.org/10.3390/ijms17060842>
 215. Liu, G., Mattick, J. S., & Taft, R. J. (2013). A meta-analysis of the genomic and transcriptomic composition of complex life. *Cell Cycle*, 12(13), 2061–2072. <https://doi.org/10.4161/cc.25134>
 216. Heimberg, A. M., Sempere, L. F., Moy, V. N., Donoghue, P. C. J., & Peterson, K. J. (2008). MicroRNAs and the advent of vertebrate morphological complexity. *Proceedings of the National Academy of Sciences of the United States of America*, 105(8), 2946–2950. <https://doi.org/10.1073/pnas.0712259105>
 217. Hu, H. Y., Guo, S., Xi, J., Yan, Z., Fu, N., Zhang, X., ... Khaitovich, P. (2011). MicroRNA expression and regulation in human, chimpanzee, and macaque brains. *PLoS Genetics*, 7(10), e1002327. <https://doi.org/10.1371/journal.pgen.1002327>
 218. Somel, M., Liu, X., Tang, L., Yan, Z., Hu, H., Guo, S., ... Khaitovich, P. (2011). MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. *PLoS Biology*, 9(12), e1001214. <https://doi.org/10.1371/journal.pbio.1001214>
 219. Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., ... Keller, A. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Research*, 44(8), 3865–3877. <https://doi.org/10.1093/nar/gkw116>
 220. Niu, C. S., Yang, Y., & Cheng, C. D. (2013). MiR-134 regulates the proliferation and invasion of glioblastoma cells by reducing Nanog expression. *International Journal of Oncology*, 42(5), 1533–1540. <https://doi.org/10.3892/ijo.2013.1844>
 221. Liu, C., Teng, Z. Q., Santistevan, N. J., Szulwach, K. E., Guo, W., Jin, P., & Zhao, X. (2010). Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation. *Cell Stem Cell*, 6(5), 433–444. <https://doi.org/10.1016/j.stem.2010.02.017>
 222. Stappert, L., Roese-Koerner, B., & Brüstle, O. (2015, January). The role of microRNAs in human neural stem cells, neuronal differentiation and subtype specification. *Cell and Tissue Research*. Springer Verlag. <https://doi.org/10.1007/s00441-014-1981-y>
 223. Hu, Z., & Li, Z. (2017, August 1). miRNAs in synapse development and synaptic plasticity. *Current Opinion in Neurobiology*. Elsevier Ltd. <https://doi.org/10.1016/j.conb.2017.02.014>
 224. Nowakowski, T. J., Rani, N., Golkaram, M., Zhou, H. R., Alvarado, B., Huch, K., ... Kosik, K. S. (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nature Neuroscience*, 21(12), 1784–1792. <https://doi.org/10.1038/s41593-018-0265-3>
 225. Rajman, M., & Schratt, G. (2017, July 1). MicroRNAs in neural development: From master regulators to fine-tuners. *Development (Cambridge)*. Company of Biologists Ltd. <https://doi.org/10.1242/dev.144337>

226. Merienne, N., Meunier, C., Schneider, A., Seguin, J., Nair, S. S., Rocher, A. B., ... Déglon, N. (2019). Cell-Type-Specific Gene Expression Profiling in Adult Mouse Brain Reveals Normal and Disease-State Signatures. *Cell Reports*, 26(9), 2477-2493.e9. <https://doi.org/10.1016/j.celrep.2019.02.003>
227. Jovicic, A., Roshan, R., Moiso, N., Pradervand, S., Moser, R., Pillai, B., & Luthi-Carter, R. (2013). Comprehensive Expression Analyses of Neural Cell-Type-Specific miRNAs Identify New Determinants of the Specification and Maintenance of Neuronal Phenotypes. *Journal of Neuroscience*, 33(12), 5127–5137. <https://doi.org/10.1523/JNEUROSCI.0600-12.2013>
228. He, M., Liu, Y., Wang, X., Zhang, M. Q., Hannon, G. J., & Huang, Z. J. (2012). Cell-Type-Based Analysis of MicroRNA Profiles in the Mouse Brain. *Neuron*, 73(1), 35–48. <https://doi.org/10.1016/j.neuron.2011.11.010>
229. Kaczmarczyk, L., Bansal, V., Rajput, A., Rahman, R., Krzyżak, W., Degen, J., ... Jackson, W. S. (2019). Tagger—A Swiss army knife for multiomics to dissect cell type-specific mechanisms of gene expression in mice. *PLOS Biology*, 17(8), e3000374. <https://doi.org/10.1371/journal.pbio.3000374>
230. Sasaki, Y., Gross, C., Xing, L., Goshima, Y., & Bassell, G. J. (2014). Identification of axon-enriched MicroRNAs localized to growth cones of cortical neurons. *Developmental Neurobiology*, 74(3), 397–406. <https://doi.org/10.1002/dneu.22113>
231. Tushev, G., Glock, C., Heumüller, M., Biever, A., Jovanovic, M., & Schuman, E. M. (2018). Alternative 3' UTRs Modify the Localization, Regulatory Potential, Stability, and Plasticity of mRNAs in Neuronal Compartments. *Neuron*, 98(3), 495-511.e6. <https://doi.org/10.1016/j.neuron.2018.03.030>
232. Kleaveland, B., Shi, C. Y., Stefano, J., & Bartel, D. P. (2018). A Network of Noncoding Regulatory RNAs Acts in the Mammalian Brain. *Cell*, 174(2), 350-362.e17. <https://doi.org/10.1016/j.cell.2018.05.022>
233. Packer, A. N., Xing, Y., Harper, S. Q., Jones, L., & Davidson, B. L. (2008). The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. *Journal of Neuroscience*, 28(53), 14341–14346. <https://doi.org/10.1523/JNEUROSCI.2390-08.2008>
234. Wu, J., & Xie, X. (2006). Comparative sequence analysis reveals an intricate network among REST, CREB and miRNA in mediating neuronal gene expression. *Genome Biology*, 7(9), 1–14. <https://doi.org/10.1186/gb-2006-7-9-r85>
235. Alkallas, R., Fish, L., Goodarzi, H., & Najafabadi, H. S. (2017). Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. *Nature Communications*, 8(1), 1–11. <https://doi.org/10.1038/s41467-017-00867-z>
236. Wu, Y. E., Parikshak, N. N., Belgard, T. G., & Geschwind, D. H. (2016). Genome-wide, integrative analysis implicates microRNA dysregulation in autism spectrum disorder. *Nature Neuroscience*, 19(11), 1463–1476. <https://doi.org/10.1038/nn.4373>
237. Alural, B., Genc, S., & Haggarty, S. J. (2017). Diagnostic and therapeutic potential of microRNAs in neuropsychiatric disorders: Past, present, and future. *Progress in Neuro-Psychopharmacology and Biological Psychiatry*, 73, 87–103. <https://doi.org/10.1016/j.pnpbp.2016.03.010>
238. Xu, B., Hsu, P. K., Karayiorgou, M., & Gogos, J. A. (2012, May). MicroRNA dysregulation in neuropsychiatric disorders and cognitive dysfunction. *Neurobiology of Disease*. NIH Public Access. <https://doi.org/10.1016/j.nbd.2012.02.016>
239. Jużwik, C. A., S. Drake, S., Zhang, Y., Paradis-Isler, N., Sylvester, A., Amar-Zifkin, A., ... Fournier, A. E. (2019, November 1). microRNA dysregulation in neurodegenerative diseases: A systematic review. *Progress in Neurobiology*. Elsevier Ltd. <https://doi.org/10.1016/j.pneurobio.2019.101664>
240. Ghibaudi, M., Boido, M., & Vercelli, A. (2017). Functional integration of complex miRNA networks in central and peripheral lesion and axonal regeneration. *Progress in Neurobiology*, 158, 69–93. <https://doi.org/10.1016/j.pneurobio.2017.07.005>
241. Li, P., Teng, Z.-Q., & Liu, C.-M. (2016). Extrinsic and Intrinsic Regulation of Axon Regeneration by MicroRNAs after Spinal Cord Injury. *Neural Plasticity*, 2016, 1–11. <https://doi.org/10.1155/2016/1279051>
242. Chen, R.-L., Balami, J. S., Esiri, M. M., Chen, L.-K., & Buchan, A. M. (2010). Ischemic stroke in the elderly: an overview of evidence. *Nature Reviews Neurology*, 6(5), 256–265.

<https://doi.org/10.1038/nrneuro.2010.36>

243. Liu, N., Wang, X., Lu, Q., & Xu, X. (2009). Altered microRNA expression following traumatic spinal cord injury. *Experimental Neurology*, 219(2), 424–429. <https://doi.org/10.1016/j.expneurol.2009.06.015>
244. Yunta, M., Nieto-Díaz, M., Esteban, F. J., Caballero-López, M., Navarro-Ruiz, R., Reigada, D., ... Maza, R. M. (2012). MicroRNA Dysregulation in the Spinal Cord following Traumatic Injury. *PLoS ONE*, 7(4), e34534. <https://doi.org/10.1371/journal.pone.0034534>
245. Motti, D., Lerch, J. K., Danzi, M. C., Gans, J. H., Kuo, F., Slepak, T. I., ... Lemmon, V. P. (2017). Identification of miRNAs involved in DRG neurite outgrowth and their putative targets. *FEBS Letters*, 591(14), 2091–2105. <https://doi.org/10.1002/1873-3468.12718>
246. Hu, J.-Z., Huang, J.-H., Zeng, L., Wang, G., Cao, M., & Lu, H.-B. (2013). Anti-apoptotic effect of microRNA-21 after contusion spinal cord injury in rats. *Journal of neurotrauma*, 30(15), 1349–60. <https://doi.org/10.1089/neu.2012.2748>
247. Almurshidi, B., Carver, W., Scott, G., & Ray, S. K. (2019). Roles of miRNAs in spinal cord injury and potential therapeutic interventions. *Neuroimmunology and Neuroinflammation*, 2019. <https://doi.org/10.20517/2347-8659.2019.19>
248. Izumi, B., Nakasa, T., Tanaka, N., Nakanishi, K., Kamei, N., Yamamoto, R., ... Ochi, M. (2011). MicroRNA-223 expression in neutrophils in the early phase of secondary damage after spinal cord injury. *Neuroscience Letters*, 492(2), 114–118. <https://doi.org/10.1016/j.neulet.2011.01.068>
249. Lindsay, M. A. (2008, July). microRNAs and the immune response. *Trends in Immunology*. <https://doi.org/10.1016/j.it.2008.04.004>
250. Tili, E., Michaille, J.-J., Cimino, A., Costinean, S., Dumitru, C. D., Adair, B., ... Croce, C. M. (2007). Modulation of miR-155 and miR-125b Levels following Lipopolysaccharide/TNF- α Stimulation and Their Possible Roles in Regulating the Response to Endotoxin Shock. *The Journal of Immunology*, 179(8), 5082–5089. <https://doi.org/10.4049/jimmunol.179.8.5082>
251. Hutchison, E. R., Kawamoto, E. M., Taub, D. D., Lal, A., Abdelmohsen, K., Zhang, Y., ... Mattson, M. P. (2013). Evidence for miR-181 involvement in neuroinflammatory responses of astrocytes. *Glia*, 61(7), 1018–1028. <https://doi.org/10.1002/glia.22483>
252. Iliopoulos, D., Jaeger, S. A., Hirsch, H. A., Bulyk, M. L., & Struhl, K. (2010). STAT3 Activation of miR-21 and miR-181b-1 via PTEN and CYLD Are Part of the Epigenetic Switch Linking Inflammation to Cancer. *Molecular Cell*, 39(4), 493–506. <https://doi.org/10.1016/j.molcel.2010.07.023>
253. Jee, M. K., Jung, J. S., Im, Y. Bin, Jung, S. J., & Kang, S. K. (2012). Silencing of miR20a is crucial for ngn1-mediated neuroprotection in injured spinal cord. *Human Gene Therapy*, 23(5), 508–520. <https://doi.org/10.1089/hum.2011.121>
254. Jee, M. K., Jung, J. S., Choi, J. I., Jang, J. A., Kang, K. S., Im, Y. Bin, & Kang, S. K. (2012). MicroRNA 486 is a potentially novel target for the treatment of spinal cord injury. *Brain : a journal of neurology*, 135(Pt 4), 1237–52. <https://doi.org/10.1093/brain/aws047>
255. Nieto-Díaz, M., Esteban, F. J., Reigada, D., Muñoz-Galdeano, T., Yunta, M., Caballero-López, M., ... Maza, R. M. (2014). MicroRNA dysregulation in spinal cord injury: causes, consequences and therapeutics. *Frontiers in Cellular Neuroscience*, 8(February), 1–19. <https://doi.org/10.3389/fncel.2014.00053>
256. Shi, Z., Zhou, H., Lu, L., Li, X., Fu, Z., Liu, J., ... Feng, S. (2017). The roles of microRNAs in spinal cord injury. *International Journal of Neuroscience*, 127(12), 1104–1115. <https://doi.org/10.1080/00207454.2017.1323208>
257. Jeyaseelan, K., Lim, K. Y., & Armugam, A. (2008). MicroRNA Expression in the Blood and Brain of Rats Subjected to Transient Focal Ischemia by Middle Cerebral Artery Occlusion. *Stroke*, 39(3), 959–966. <https://doi.org/10.1161/STROKEAHA.107.500736>
258. Ziu, M., Fletcher, L., Rana, S., Jimenez, D. F., & Digicaylioglu, M. (2011). Temporal Differences in MicroRNA Expression Patterns in Astrocytes and Neurons after Ischemic Injury. *PLoS ONE*, 6(2), e14724. <https://doi.org/10.1371/journal.pone.0014724>
259. Tan, K. S., Armugam, A., Sepramaniam, S., Lim, K. Y., Setyowati, K. D., Wang, C. W., & Jeyaseelan, K. (2009). Expression profile of microRNAs in young stroke patients. *PLoS ONE*, 4(11).

<https://doi.org/10.1371/journal.pone.0007689>

260. Wang, C., Pan, Y., Cheng, B., Chen, J., & Bai, B. (2014). Identification of Conserved and Novel microRNAs in Cerebral Ischemia-Reperfusion Injury of Rat Using Deep Sequencing. *Journal of Molecular Neuroscience*, *54*(4), 671–683. <https://doi.org/10.1007/s12031-014-0383-7>
261. Kobayashi, M., Benakis, C., Anderson, C., Moore, M. J., Poon, C., Uekawa, K., ... Darnell, R. B. (2019). AGO CLIP Reveals an Activated Network for Acute Regulation of Brain Glutamate Homeostasis in Ischemic Stroke. *Cell Reports*, *28*(4), 979–991.e6. <https://doi.org/10.1016/j.celrep.2019.06.075>
262. Saugstad, J. A. (2015). Non-Coding RNAs in Stroke and Neuroprotection. *Frontiers in Neurology*, *6*(MAR), 1–11. <https://doi.org/10.3389/fneur.2015.00050>
263. Abels, E. R., Maas, S. L. N., Nieland, L., Wei, Z., Cheah, P. S., Tai, E., ... Breakefield, X. O. (2019). Glioblastoma-Associated Microglia Reprogramming Is Mediated by Functional Transfer of Extracellular miR-21. *Cell reports*, *28*(12), 3105–3119.e7. <https://doi.org/10.1016/j.celrep.2019.08.036>
264. Huang, S., Ge, X., Yu, J., Han, Z., Yin, Z., Li, Y., ... Lei, P. (2018). Increased miR-124-3p in microglial exosomes following traumatic brain injury inhibits neuronal inflammation and contributes to neurite outgrowth via their transfer into neurons. *FASEB Journal*, *32*(1), 512–528. <https://doi.org/10.1096/fj.201700673R>
265. Zagrean, A. M., Hermann, D. M., Opris, I., Zagrean, L., & Popa-Wagner, A. (2018, November 6). Multicellular crosstalk between exosomes and the neurovascular unit after cerebral ischemia. therapeutic implications. *Frontiers in Neuroscience*. Frontiers Media S.A. <https://doi.org/10.3389/fnins.2018.00811>
266. Rusnakova, V., Honsa, P., Dzamba, D., Ståhlberg, A., Kubista, M., & Anderova, M. (2013). Heterogeneity of Astrocytes: From Development to Injury – Single Cell Gene Expression. *PLoS ONE*, *8*(8), e69734. <https://doi.org/10.1371/journal.pone.0069734>
267. Zhang, M., Downes, C. E., Wong, C. H. Y., Brody, K. M., Guio-Agulair, P. L., Gould, J., ... Crack, P. J. (2017). Type-I interferon signalling through IFNAR1 plays a deleterious role in the outcome after stroke. *Neurochemistry International*, *108*, 472–480. <https://doi.org/10.1016/j.neuint.2017.06.009>

8 Supplements

Supplement I

Androvc, P., Valihrach, L., Elling, J., Sjoback, R., and Kubista, M. (2017). Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Research* 45, e144–e144

Supplement II

Androvc, P., Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., and Valihrach, L. (2019). Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Scientific Reports* 9, 4255.

Supplement III

Androvc, P.*, Benesova, S. *, Kubista, M., Valihrach, L. (2020). Performance comparison of small RNA-Seq library preparation methods for biofluid samples. *Manuscript in preparation*. *These authors contributed equally

Supplement IV

Valihrach, L., **Androvc, P.**, Kubista, M. (2019). Circulating miRNA Analysis for Cancer Diagnostics and Therapy. *Molecular Aspects of Medicine*, <https://doi.org/10.1016/j.mam.2019.10.002>.

Supplement V

Androvc, P., Kirdajova, D., Tureckova, J., Zucha, D., Rohlova, E., Abaffy, P., Kriska, J., Anderova, M., Kubista, M., and Valihrach, L. (2019). Decoding the transcriptional response to ischemic stroke in young and aged mouse brain. BioRxiv 769331. *Online preprint. Manuscript is under review in an international journal.*

Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification

Peter Androvic^{1,2}, Lukas Valihrach¹, Julie Elling³, Robert Sjoback³ and Mikael Kubista^{1,3,*}

¹Laboratory of Gene Expression, Institute of Biotechnology CAS, Biocev, Vestec 252 50, Czech Republic,

²Laboratory of Growth Regulators, Faculty of Science, Palacky University, Olomouc 783 71, Czech Republic and

³TATAA Biocenter AB, Gothenburg 411 03, Sweden

Received December 06, 2016; Revised June 07, 2017; Editorial Decision June 24, 2017; Accepted June 28, 2017

ABSTRACT

MicroRNAs are a class of small non-coding RNAs that serve as important regulators of gene expression at the posttranscriptional level. They are stable in body fluids and pose great potential to serve as biomarkers. Here, we present a highly specific, sensitive and cost-effective system to quantify miRNA expression based on two-step RT-qPCR with SYBR-green detection chemistry called Two-tailed RT-qPCR. It takes advantage of novel, target-specific primers for reverse transcription composed of two hemiprobes complementary to two different parts of the targeted miRNA, connected by a hairpin structure. The introduction of a second probe ensures high sensitivity and enables discrimination of highly homologous miRNAs irrespectively of the position of the mismatched nucleotide. Two-tailed RT-qPCR has a dynamic range of seven logs and a sensitivity sufficient to detect down to ten target miRNA molecules. It is capable to capture the full isomiR repertoire, leading to accurate representation of the complete miRNA content in a sample. The reverse transcription step can be multiplexed and the miRNA profiles measured with Two-tailed RT-qPCR show excellent correlation with the industry standard TaqMan miRNA assays ($r^2 = 0.985$). Moreover, Two-tailed RT-qPCR allows for rapid testing with a total analysis time of less than 2.5 hours.

INTRODUCTION

MicroRNAs (miRNAs) are short non-coding RNA molecules (~19–24 nt long) that mediate regulation of gene expression at the post-transcriptional level (1,2). Production of miRNAs starts with the transcription of genomic DNA into long primary transcripts called pri-miRNAs. The pri-miRNAs are subsequently cleaved by RNase III Droscha into shorter precursor transcripts with

hairpin structure called pre-miRNAs. The pre-miRNAs are transported into the cytoplasm where they are processed by RNase III Dicer into ~22 nt double-stranded miRNA molecules. Both strands of this duplex may become functional mature miRNAs (3–5). MiRNAs may function as master regulators of numerous physiological and pathological processes and changes in their expression patterns are often observed in various diseases (6–10). Because of their remarkable stability in biofluids miRNAs have exciting potential to serve as minimally invasive diagnostic biomarkers (11–14).

MiRNA expression can be measured by many techniques; the three most common being microarrays, next generation sequencing (RNA-Seq), and reverse transcription quantitative PCR (RT-qPCR). In addition, non-PCR based isothermal amplification methods have also been proposed (15–18). Each of these methods has its advantages and limitations. Microarray analysis is generally more cost efficient than RNA-Seq and offers the possibility to monitor large number of targets, but, at least with conventional microarrays, specificity and dynamic range are limited. RNA-Seq is suitable for high-throughput and is the only platform capable of discovering new miRNAs. Disadvantages of RNA-Seq are the rather high cost per sample and the complexity of the workflow and data analysis. Also, the precision of quantitation is poor for the low abundant miRNAs. RT-qPCR is the method commonly used for the validation of results from screening experiments and when high accuracy and precision is required. It is also the method of choice when only a small number of targets is quantified, particularly when the amount of material is limiting. Another appealing aspect is the simple workflow easily set up in laboratories that have experience in RT-qPCR (19–22).

There are, however, significant technical challenges in miRNA expression profiling using RT-qPCR. MiRNA molecules are only 19–24 nt long, which is the length of a conventional PCR primer. The sequence of the mature miRNA is contained in its precursor molecules (pri-miRNAs and pre-miRNAs), however, only the mature miRNAs are believed to have effector functions and they usually

*To whom correspondence should be addressed. Tel: +46 733 928 168; Email: mikael.kubista@tataa.com

are the targets for quantification. The GC content of miRNAs is highly variable, which complicates assay design and protocol optimization, particularly when a common protocol is sought for multiple/all miRNA targets. The sequences of miRNAs within the same family, such as let-7, may be highly similar, differing only in a single base position. miRNAs are subject to various post-transcriptional modifications and may differ in sequence and nucleotide composition at either or both ends. This obscures specific quantification with many techniques (23–25).

Several methods to quantify miRNAs based on RT-qPCR have been developed to date, many of which are commercially available. Generally, the methods can be divided into two groups that are based on universal and specific reverse transcription, respectively. In addition, probe-ligation methods that do not require reverse transcription have also been proposed (26,27). In the universal RT approach, all miRNA molecules are elongated by an identical tail used to prime the reverse transcription into cDNA. These methods include the addition of poly-A tails to the 3'-end with the poly(A) polymerase (Exiqon's miRCURY LNA system, Qiagen's miScript PCR system, Quantabio's qScript microRNA system) (28,29), polyuridylation with poly(U) polymerase (30), ligation of a universal linker with T4 RNA ligase (31), and more recently combination of linker ligation and end tailing (Thermo Fisher's TaqMan Advanced miRNA assays). The main advantage of these approaches is that all miRNAs are converted into cDNA in the same tube. However, they may suffer from high background noise and they are often limited by the efficiency of the extra enzymatic steps needed. Some also require special reagents. Moreover, small RNAs possessing a 2'-O-methyl (2'-O-Me) modification on their 3' terminal nucleotide, such as plant mature miRNAs and piRNAs, are resistant to polyadenylation and cannot be efficiently reverse transcribed using a polyadenylation-based cDNA synthesis approach (32). The polyadenylation and ligation steps also introduce bias (33–35). The second group of methods includes the use of linear primers (36,37), pincer probes (38), and stem-loop RT primers (39), also known as TaqMan miRNA assays (Thermo Fisher). The stem-loop method is probably the most common today and is frequently used for benchmarking and validation of other methods (40).

The stem-loop primers of the TaqMan miRNA assays are composed of a short single stranded sequence at their 3'-ends that anneals to the 3'-end of the targeted miRNA, a double-stranded segment (the stem) and a loop. The stem-loop structure shifts the equilibrium to the formation of an RNA/DNA duplex and should prevent binding of the primer to pri- and pre-miRNAs and to any dsDNA that may be present. Nevertheless, it has been reported that presence of genomic and plasmid DNA containing sequences of the corresponding pre-miRNA give rise to significant background signal leading to false positive results (41). The method uses hydrolytic probes that are costly to produce and do not allow controlling the specificity of the reaction by melting curve analysis. Notably, the TaqMan probe does not contribute to the specificity of the reaction, as it binds to a site originating from the sequence of the RT primer. Another limitation of this design is the reduced ability to reverse transcribe isomiR variants (24,41,42). Although the

stem-loop approach employs target-specific RT primers, the reverse transcription step can be multiplexed using multiple RT primers in the same tube (43–45).

We have developed a novel specific and cost-effective approach to quantify miRNA expression that utilizes specific structured primers for reverse transcription and SYBR-green based qPCR named 'Two-tailed RT-qPCR'. The Two-tailed RT primers are composed of two hemiprobes complementary to separate regions of the target miRNA and of an oligonucleotide tether folded into a hairpin. This novel design increases the binding strength of the RT primer to its template leading to increased sensitivity. The 3'-hemiprobe can be short, providing high discriminatory power to mismatches in the 3'-region and leaving enough space for the design of miRNA specific qPCR primers with sufficient melting temperature (T_m). The 5'-hemiprobe improves the discrimination between highly similar sequences, particularly when the differing nucleotide is located in the center or close to the 5'-end of the miRNA sequence. Since Two-tailed RT primers do not interact with the ends of the miRNA they are able to detect all terminal variants of any miRNA (isomiRs) and therefore accurately reflect the true miRNA content in a sample.

MATERIALS AND METHODS

Primers, templates and synthetic oligonucleotides

Sequences of the miRNA oligonucleotides were obtained from the miRBase Release 21 (www.mirbase.org) (46). Sequences of primers and targets are listed in supplementary file. Secondary structure of the Two-tailed RT primers were predicted using the UNafold web server (<http://unafold.rna.albany.edu/>) (47). RNA oligonucleotides were synthesized and quantified by Integrated DNA Technologies. DNA primers were synthesized and quantified by Invitrogen. Precursor miRNAs were synthesized by *in vitro* transcription from corresponding PCR products using T7 RNA polymerase (New England Biolabs) according to the manufacturer's protocol (suppl. file). Reactions were treated with the Turbo DNA-free kit (Thermo Fisher), RNA was precipitated in 3M LiCl and quantified with the Qubit 2.0 fluorometer (Thermo Fisher). Correct size of the precursor miRNA products was verified using the Fragment Analyzer (Advanced Analytical).

cDNA synthesis

RT reactions were performed with the qScript flex cDNA kit (Quantabio) in a total reaction volume of 10 μ l. The reaction mixture contained either 10 ng of total RNA or synthetic miRNA template, 1 \times RT buffer, 0.05 μ M RT primer, 1 μ l GSP enhancer and 0.5 μ l RT enzyme. RT reactions were incubated in a 96-well plate in a Bio-Rad CFX 1000 thermocycler for 45 min at 25°C, 5 min at 85°C and then held at 4°C. Reactions using TaqMan miRNA assays (Thermo Fisher) and Quantabio qScript microRNA system (Quantabio) were performed according to the manufacturer's protocol except that the total reaction volume was scaled down to 10 μ l. Reactions using miQPCR method were performed as described in (31) according to

the protocol obtained from the corresponding author (personal communication): ligation of template miRNAs to the miLINKER adaptor was performed in a total reaction volume of 8 μ l containing 0.8 \times T4 buffer (New England Biolabs), 5 mM MgCl₂, 17% PEG 8000, 0.15 μ M miLINKER adaptor, 0.1 μ l RNaseOUT (40U/ μ l) (Thermo Fisher) and 0.18 μ l T4 RNA Ligase 2, truncated K227Q (New England Biolabs). The ligation reaction was incubated for 30 min at 25°C and then placed at 4°C. The ligated miRNAs were then incubated for 2 min at 85°C with 0.5 μ M dNTPs and 0.05 μ M universal mQ-RT primer in a total reaction volume of 14 μ l and then reverse transcribed in total reaction volume of 20 μ l containing 1 \times RT buffer, 5 mM DTT and 1 μ l SuperScript III (Thermo Fisher) for 30 min at 46°C, 5 min at 85°C and finally held at 4°C.

Quantitative PCR

qPCR was performed in a total reaction volume of 10 μ l containing 1 \times SYBR Grandmaster Mix (TATAA Biocenter), 0.4 μ M forward and reverse primer and the cDNA product diluted at least 10 \times . Reactions were performed in duplicates and incubated in a 96- or 384-well plate in a CFX 96 or CFX 384 Real Time Detection System (Bio-Rad) at 95°C for 30 s, followed by 45 cycles of 95°C for 5 s and 60°C for 15 s. Reaction specificity was assessed by melting curve analysis immediately after the qPCR. qPCR with TaqMan miRNA assays and Quantabio qScript microRNA system were performed according to manufacturers' protocols in a total reaction volume of 10 μ l. cDNA was diluted at least 15 \times or 10 \times for the TaqMan and Quanta reactions respectively. qPCR with the miQPCR method was performed in a total reaction volume of 10 μ l containing 1 \times SYBR Grandmaster Mix, 0.15 μ M forward and reverse primer and the cDNA product diluted 100 \times according to the recommended protocol (31). Reactions were incubated at 95°C for 30 s, followed by 45 cycles of 95°C for 10 s and 60°C for 35 s followed by melting curve analysis.

MiRNA profiling in mouse tissues

All procedures involving the use of laboratory animals were performed in accordance with the European Community Council Directive of 24 November 1986 (86/609/EEC) and animal care guidelines approved by the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic (Animal Care Committee decision on 17 April 2009; approval number 85/2009). Mouse tissue samples from brain, cerebellum, liver, lung, kidney, heart and skeletal muscle were dissected, placed into TRI Reagent (Sigma-Aldrich) and were immediately frozen on dry ice. Before use, samples were thawed, homogenized using the TissueLyser (Qiagen) and total RNA was extracted with TRI Reagent (Sigma-Aldrich) according to the manufacturer's protocol. RNA quantity and purity was assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher) and RNA integrity was assessed using the Fragment Analyzer (Advanced Analytical). Inhibition of the RT-qPCR workflow was tested for using an RNA spike control (Tataa Biocenter). Data were normalized to total amount of RNA. Same aliquots were used for all measurements. Cq values were transformed

to quantities relative to the sample with the lowest expression for each target miRNA separately and expression values were converted to log scale. Pearson correlation coefficients were calculated based on logarithmic expression values. GenEx 6 software (MultiD) was used for data preprocessing.

RESULTS

General assay design

A novel two step RT-qPCR system for the quantification of microRNAs is presented (Figure 1). Reverse transcription is performed with target-specific structured primers that are about 50 nucleotides long and contain two target specific hemiprobe complementary to the miRNA. The 3'-hemiprobe is about 6 nt long and binds to the 3'-region of the target miRNA. The 5'-hemiprobe is usually longer and binds within the 5'-region of the targeted microRNA. The two hemiprobe are connected by an oligonucleotide tether designed to fold into a hairpin to prevent nonspecific interactions (Figure 1A). After hybridization, the RT reaction is primed from the 3'-hemiprobe. The 5'-hemiprobe is displaced by the RT enzyme and the Two-tailed RT primer is elongated to produce cDNA with a sequence complementary to the targeted miRNA (Figure 1B and C). The cDNA is then quantified by conventional qPCR utilizing SYBR-Green chemistry with two target-specific PCR primers. The reverse PCR primer is specific for the miRNA target sequence while the forward primer is specific for the pre-designed region in the 5'-end of the Two-tailed RT primer (Figure 1D).

The Two-tailed RT primer has 3 functions: i) it primes specifically the reverse transcription of the target miRNA template ii) it contributes with additional sequence to the cDNA making it long enough for PCR amplification iii) it contains the sequence of the forward PCR primer. We reasoned that the introduction of a second binding element, the 5'-hemiprobe that binds within the 5'-end of the miRNA, will increase the sensitivity and specificity of the RT reaction, as more nucleotides can be interrogated. Also, the 3'-hemiprobe can be made shorter (5–6 nt), providing flexibility to design the reverse PCR primer with adequate T_m without overlapping with the 3'-hemiprobe sequence, thereby avoiding the risk of undesired self-priming and primer-dimer formation. The 5'-hemiprobe also contributes to increased discriminatory power between highly similar targets that differ only by 1 nt in the center or close to the 5'-end of the miRNA sequence. The reason is that the target miRNAs are subjected to sequence interrogation twice: first in the RT and then in the qPCR.

To test the concept we designed three Two-tailed RT primers with the same 3'-hemiprobe (5 nt), but different 5'-hemiprobe: one with 10 complementary nucleotides to the target, one with 10 non-complementary nucleotides, and one without 5'-hemiprobe. We found that Two-tailed RT primers that lacked complementary 5'-hemiprobe gave significantly higher qPCR Cq values demonstrating the contribution of the 5'-hemiprobe to the sensitivity of the system (Figure 2A).

To test the contribution of the 5'-hemiprobe to the specificity, we compared two Two-tailed RT primers for their

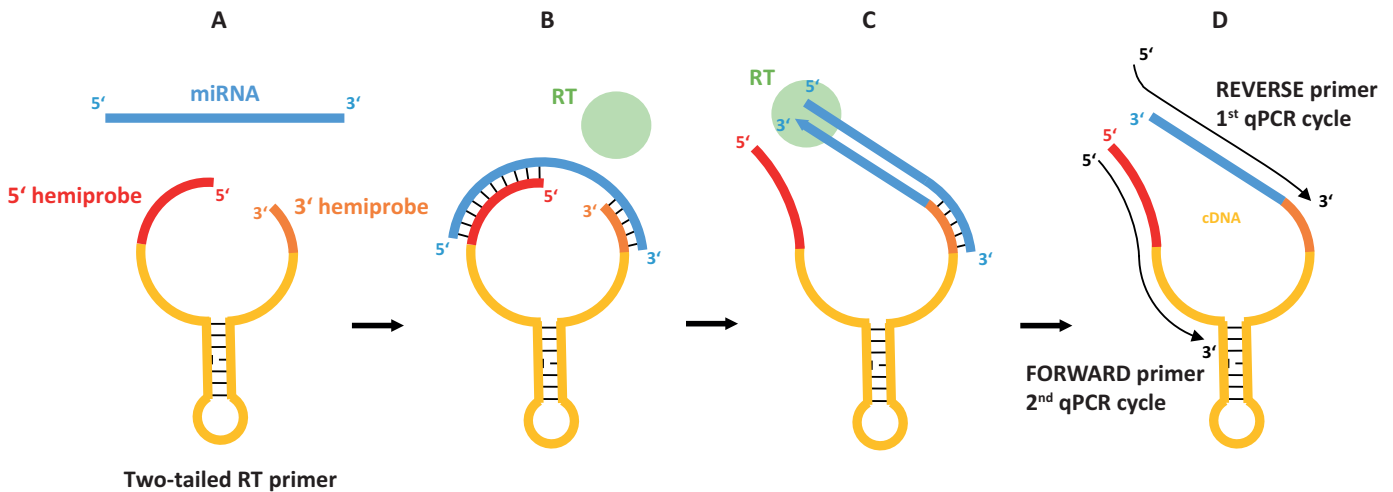


Figure 1. Schematic of Two-tailed RT-qPCR. (A) Two-tailed RT primer having two hemiprobe connected by a hairpin folding sequence. (B) The hemiprobe bind cooperatively, one at each end of the target miRNA, forming a stable complex. (C) Reverse transcriptase binds the 3'-end of the hybridized Two-tailed RT primer and elongates it to form tailed cDNA. (D) The cDNA is amplified by qPCR using two target-specific primers.

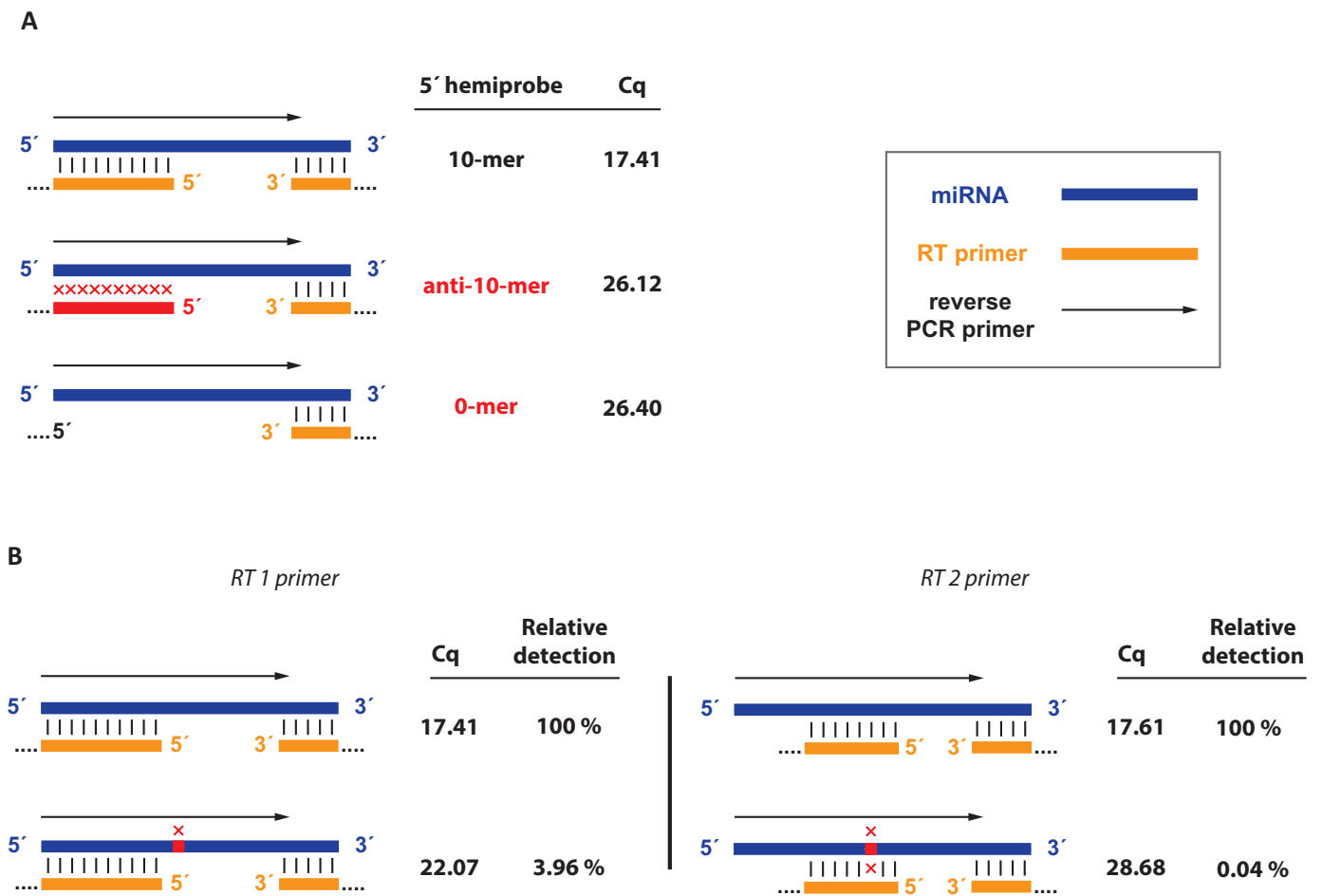


Figure 2. Importance of the 5'-hemiprobe for sensitivity and specificity. (A) Two-tailed RT primers with 5 nt complementary 3'-hemiprobe and either 10 nt complementary 5'-hemiprobe (top), 10 nt non-complementary 5'-hemiprobe (middle), or no 5'-hemiprobe at all (bottom) targeting let-7a. Cq values obtained with the Two-tailed RT primer having a complementary 5'-hemiprobe are about nine cycles lower than those obtained with the Two-tailed RT primers lacking 5'-complementarity. (B) Two-tailed RT primers used to assay targets that differ in one nucleotide. The variable nucleotide is in the non-interrogated region between the hybridization sites of the 3'- and 5'-hemiprobe (left) and the variable nucleotide is within the 5'-hemiprobe binding site (right).

ability to discriminate between two members of the Let-7 miRNA family: Let-7a and Let-7f. These differ only in one nucleotide located in the center of the miRNA sequence. RT 1 primer was designed to bind with its 5'-hemiprobe to the first ten nucleotides from the 5'-end of let-7a, while RT 2 primer had a 5'-hemiprobe that binds to 8 nucleotides in the center of the let-7a sequence (Figure 2B). With this design, the nucleotide distinguishing let-7a from let-7f is sensed by the 5'-hemiprobe of the RT 2 primer, but not by the RT 1 primer. The 5'-hemiprobe of the RT 2 primer was also shorter (8 instead of 10 nucleotides), as we reasoned the impact of a mismatch will be more prominent with a shorter probe sequence. The rest of the RT 1 and RT 2 primer sequences as well as the PCR primers used were identical. The measured ΔCq between the fully matched and the mismatched template was substantially larger when using the RT 2 primer, which overlapped the differing base with its 5'-hemiprobe ($\Delta Cq = 11.07$), than with the RT 1 primer, which did not ($\Delta Cq = 4.66$). This demonstrates the significant contribution from the 5'-hemiprobe to the specificity of the system. Notably, the Cq values for the let-7a microRNA, to which both Two-Tailed RT primers were fully complementary, were equal suggesting that the assay sensitivity remained the same even though the length of the 5'-hemiprobe differed by two bases.

Repeatability

To compare the performance of the Two-tailed RT-qPCR to other RT-qPCR methods for miRNA analysis we performed most of the experiments also with the TaqMan microRNA assays, the Quantabio qScript microRNA system, and the miQPCR method published in (31) (in the following text referred to as TaqMan, Quanta, and miQPCR). These methods use different strategies to produce cDNA for subsequent qPCR. TaqMan uses miRNA specific RT primers. Quanta uses poly(A) polymerase to add poly(A) tails to the 3' ends of miRNAs to allow for universal reverse transcription with an oligo(dT) primer, and miQPCR ligates a defined adaptor sequence to the 3' end of miRNAs prior to reverse transcription of ligated constructs with an universal RT primer.

We assessed the repeatability of the new Two-tailed RT-qPCR and the three known methods by measuring the imprecision, expressed as standard deviation of Cq values obtained from triplicate measurements having the same input RNA (Table 1). Three miRNAs expressed at high (let-7a), moderate (miR-21), and low levels (miR-193a) were measured in total RNA extracted from mouse cerebellum. First step in the workflow of every method was taken as the point for replication. Overall, all tested methods displayed very high repeatability as demonstrated by low standard deviations of the replicate measurements (Table 1).

Sensitivity and dynamic range

The sensitivity and dynamic range of the Two-tailed RT-qPCR were evaluated using synthetic let-7d miRNA as target. A dilution series spanning 8 orders of magnitude was prepared ranging from approx. 10 to 10^9 copies of let-7d miRNA molecules per RT reaction. 10% of the cDNA prod-

uct was used for qPCR. The Two-tailed RT-qPCR assay exhibited excellent linearity between the log of the miRNA input and Cq values over 7 orders of magnitude and accurately quantified down to 10 cDNA copies of let-7d corresponding to 100 miRNAs in the original sample (Figure 3A). Similar results were obtained when 100 ng of yeast total RNA was added to each RT reaction to simulate the complex background in biological samples (Figure 3A). To determine the limit of detection (LoD) more precisely, a 2-fold serial dilution of let-7d approaching zero-concentration was performed with each sample analyzed in hexaplicate. LoD is estimated at the lowest concentration that produces 95% positive replicates and can be roughly estimated by fitting the fraction of positive replicates to the logarithm of the concentration using GenEx software (MultiD) (48). LoD of the two-tailed RT PCR assay for let-7d microRNA was estimated to 111 miRNA molecules, which corresponds to 11 cDNA molecules in our workflow as only 10% of the cDNA was used as template for qPCR (suppl. file). The LoD estimate relies on the miRNA stock concentration provided by the oligonucleotide manufacturer (IDT), which was determined spectroscopically. This may have overestimated the concentration of intact full length miRNA to some degree as any contaminating byproducts contribute to absorption. Hence, the LoD we estimate is an upper limit.

To further compare the sensitivity and dynamic range of the two-tailed RT-qPCR to the other methods we prepared dilution series also of let-7a, let-7d and miR-21. Three of the methods (two-tailed, TaqMan, Quanta) typically detected hundreds of miRNA copies with linear response down to at least 10^3 miRNAs in the original sample or 100 cDNA per RT reaction (Figure 3B). miQPCR performed less good, producing significantly higher Cq values and a narrower linear dynamic range reaching down to 10^4 miRNA copies (Figure 3B).

Discrimination of highly similar sequences

The capability to discriminate between highly homologous sequences with the Two-tailed RT-qPCR was tested on the let-7 miRNA family. The members of this family are highly similar as four pairs of the let-7 miRNAs differ only in a single nucleotide in different positions, posing major challenge for specific quantification (Figure 4A). We started by identifying the optimal length of the 5'-hemiprobe to maximize specificity, without compromising too much on sensitivity, and then designed assays for the let-7 family members accordingly (Supplementary Figure S1, suppl. file). We assayed approximately 2×10^8 copies of each let-7 miRNA target with each let-7 Two-tailed RT-qPCR assay. Cross-reactivity was estimated for each of the assay-target pairs based on the Cq difference between the reactions with the perfectly matched target and with the mismatched target assuming 100% efficiency for the matched target. Only negligible levels of unspecific signal were observed ($<1\%$), and only for targets that differed from the perfect match by a single nucleotide: let-7a versus let-7c, let-7b versus let-7c, let-7a versus let-7e, let-7a versus let-7f (Figure 4B). Overall, all Two-tailed RT-qPCR assays exhibited exceptional specificity. None of the other tested methods reached simi-

Table 1. Average Cq values and standard deviations of triplicate measurements of three miRNAs quantified by four different methods. Sample was total RNA isolated from mouse cerebellum

	Two-tailed RT-qPCR		TaqMan		Quanta		miQPCR	
	Cq	St.dev	Cq	St.dev	Cq	St.dev	Cq	St.dev
let-7a	19.59	0.21	26.69	0.28	19.87	0.14	24.65	0.16
miR-21	23.73	0.22	28.92	0.24	23.36	0.02	29.39	0.16
miR-193a	29.12	0.17	35.95	0.85	31.13	0.29	35.15	0.12

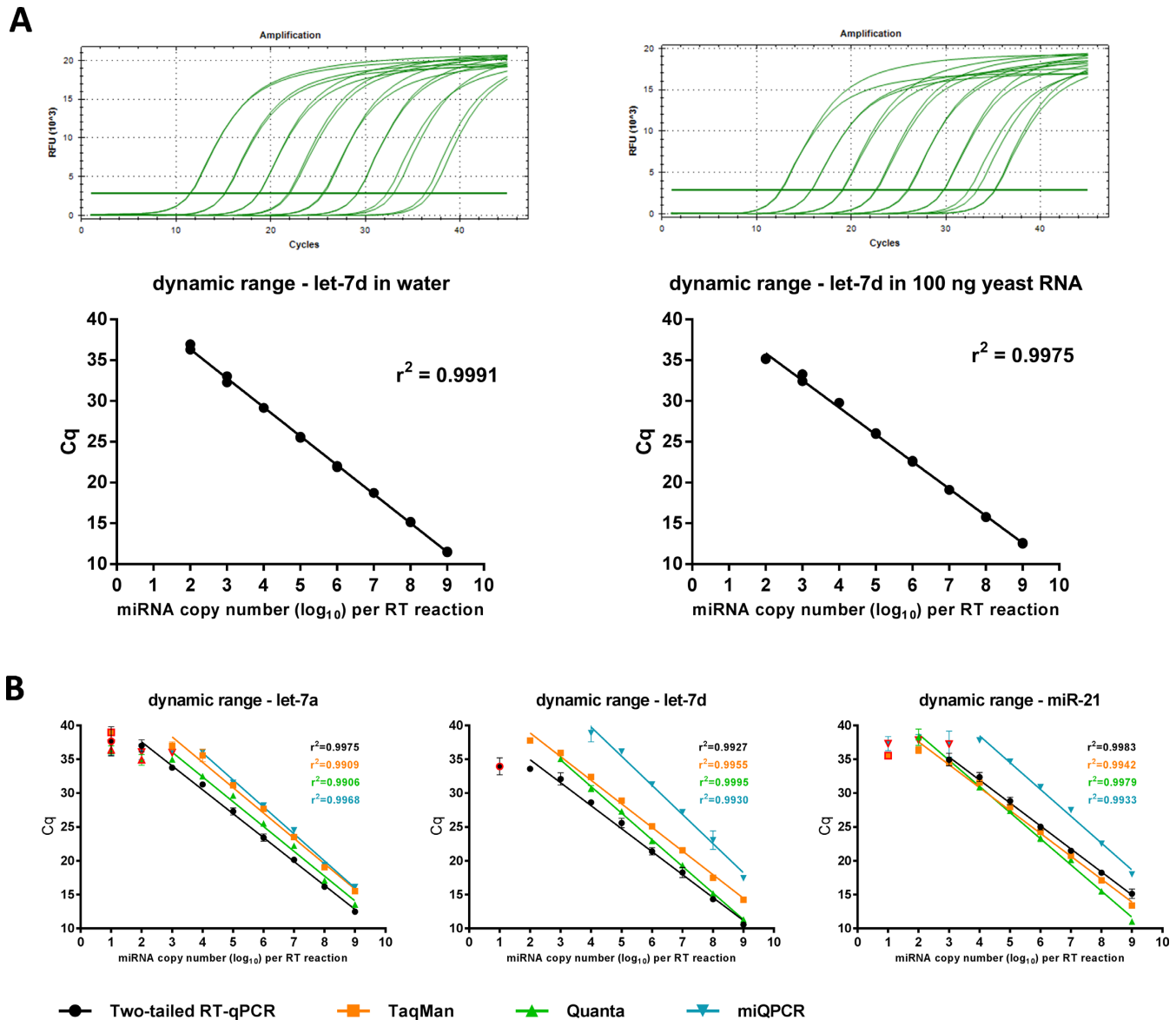


Figure 3. Dynamic range and sensitivity of Two-tailed RT-qPCR, TaqMan, Quanta and miQPCR. (A) Amplification plots and standard curves of let-7d assayed in water and against a background of 100 ng yeast RNA. The dynamic range is seven logs. (B) Standard curves of let-7a, let-7d, and miR-21 assayed with Two-tailed RT-qPCR, TaqMan, Quanta, and miQPCR. Cq values outside the linear range are indicated with red border.

lar level of specificity with false positive signals more than 200 times stronger, reaching as high as 22.48% (TaqMan), 50.71% (Quanta) and 122.47% (miQPCR) (Figure 4A).

Discrimination between mature and precursor miRNAs

To test if the Two-tailed RT-qPCR assays can distinguish mature miRNAs from their precursor molecules we individually assayed the same amount of mature let-7a, let-7b and let-7f miRNAs (~2 × 10⁸ copies) and their corresponding precursor miRNAs. Cross-reactivity with precu-

A

		Relative detection (%)							
		<i>let-7a</i>	<i>let-7b</i>	<i>let-7c</i>	<i>let-7d</i>	<i>let-7e</i>	<i>let-7f</i>	<i>let-7g</i>	<i>let-7i</i>
Assays	A	100.00	0.07	0.46	0.14	0.31	0.01	0.00	0.00
	B	0.00	100.00	0.61	0.00	0.00	0.00	0.00	0.00
	C	0.01	0.18	100.00	0.00	0.00	0.00	0.00	0.00
	D	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00
	E	0.15	0.00	0.00	0.01	100.00	0.00	0.00	0.00
	F	0.18	0.00	0.01	0.00	0.00	100.00	0.02	0.00
	G	0.00	0.00	0.00	0.00	0.00	0.01	100.00	0.00
	I	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

Two-tailed RT-qPCR

		Relative detection (%)							
		<i>let-7a</i>	<i>let-7b</i>	<i>let-7c</i>	<i>let-7d</i>	<i>let-7e</i>	<i>let-7f</i>	<i>let-7g</i>	<i>let-7i</i>
Assays	A	100.00	0.27	50.71	2.17	1.58	2.47	1.55	0.00
	B	0.09	100.00	32.84	0.00	0.00	0.00	0.00	0.02
	C	48.91	27.00	100.00	0.31	0.56	0.95	0.06	0.00
	D	0.12	0.33	0.07	100.00	0.00	0.00	0.00	0.00
	E	0.13	0.13	0.13	0.00	100.00	0.03	0.03	0.02
	F	0.73	0.85	0.72	0.02	0.00	100.00	0.05	0.04
	G	0.02	0.00	0.01	0.00	0.00	0.26	100.00	16.84
	I	0.00	0.00	0.00	0.00	0.00	0.00	0.38	100.00

Quanta

B

name	sequence
<i>let-7a</i>	UGAGGUAGUAGGUUGUAUAGUU
<i>let-7b</i>	UGAGGUAGUAGGUUGUGUGUU
<i>let-7c</i>	UGAGGUAGUAGGUUGUAUGUU
<i>let-7d</i>	AGAGGUAGUAGGUUGCAUAGUU
<i>let-7e</i>	UGAGGUAGGAGGUUGUAUAGUU
<i>let-7f</i>	UGAGGUAGUAGAUUGUAUAGUU
<i>let-7g</i>	UGAGGUAGUAGUUUGUACAGUU
<i>let-7i</i>	UGAGGUAGUAGUUUGUCUUGUU

		Relative detection (%)							
		<i>let-7a</i>	<i>let-7b</i>	<i>let-7c</i>	<i>let-7d</i>	<i>let-7e</i>	<i>let-7f</i>	<i>let-7g</i>	<i>let-7i</i>
Assays	A	100.00	0.44	20.89	2.20	3.68	8.38	0.37	0.00
	B	0.19	100.00	22.48	0.00	0.00	0.01	0.00	0.01
	C	0.09	1.77	100.00	0.00	0.00	0.01	0.00	0.00
	D	2.59	0.01	1.37	100.00	0.01	0.01	0.00	0.00
	E	9.88	0.07	7.87	0.09	100.00	0.40	0.03	0.00
	F	2.00	0.16	0.22	0.12	0.01	100.00	0.15	0.00
	G	0.96	0.00	0.32	0.01	0.01	2.72	100.00	0.02
	I	0.00	0.00	0.00	0.00	0.00	0.00	0.01	100.00

TaqMan

		Relative detection (%)							
		<i>let-7a</i>	<i>let-7b</i>	<i>let-7c</i>	<i>let-7d</i>	<i>let-7e</i>	<i>let-7f</i>	<i>let-7g</i>	<i>let-7i</i>
Assays	A	100.00	12.64	55.52	101.75	122.47	76.72	48.68	0.69
	B	7.78	100.00	45.46	1.08	0.06	0.08	0.01	1.39
	C	66.40	75.14	100.00	28.76	1.13	9.15	0.45	0.01
	D	14.84	0.00	0.09	100.00	0.21	0.19	0.03	0.00
	E	51.07	0.04	20.96	27.57	100.00	6.52	0.99	0.00
	F	54.28	0.01	0.56	11.85	3.28	100.00	14.45	0.05
	G	0.07	0.00	0.00	0.00	0.00	0.18	100.00	0.91
	I	0.00	0.00	0.00	0.00	0.00	0.00	7.43	100.00

miQPCR

C

	Mature	Precursor	Relative detection
<i>let-7a</i>	17.74	21.31	6.98%
<i>let-7b</i>	16.98	21.22	5.31%
<i>let-7f</i>	16.85	23.78	0.82%

Figure 4. Specificity of Two-tailed RT-qPCR, TaqMan, Quanta, and miQPCR. (A) Measured false-positive levels of *let-7* miRNA family members expressed relative to the level of the targeted member. (B) Sequences of eight members of the *let-7* family. Nucleotide variations relative to *let-7a* are indicated. (C) Cq values and relative detection levels of pre-miRNAs relative to the targeted mature microRNA measured with Two-tailed RT-qPCR.

sor molecules was estimated from the measured ΔCq values. It ranged from 0.82% for *let-7f* to 6.98% for *let-7a* (Figure 4C), demonstrating that the Two-tailed RT-qPCR assays specifically quantify the amounts of mature miRNAs.

Performance of the system with biological samples and comparison with independent platform

We validated the Two-tailed RT-qPCR assays on biological samples measuring the expression of 8 miRNAs across 7 mouse tissues and compared with measurements using commercially available TaqMan miRNA assays (Figure 5A). Relative expression levels across the tissues were calculated from the Cq values. The results were in agreement with previous reports (36,49) with miR-122-5p being highly expressed in liver, while miR-1a-3p having high expression

in heart and muscle. The other microRNA targets exhibited lower variation in expression levels across the tissues. *Let-7a* and miR-21a, which are thought to have housekeeping functions, were indeed expressed at high levels in all the examined tissues. MiR-615-5p was not detected by any of the methods, suggesting it is either not present at all or only at exceedingly low levels.

The correlation between the measured ΔCq ($= Cq_{\text{tissue, lowest expression}} - Cq_{\text{tissue, x}}$) with the Two-tailed RT-qPCR assays and with the TaqMan miRNA assays was excellent (Figure 5A). Pearson correlation coefficients (r) were 0.981 or larger for all the measured targets but miR-30c-1-3p, where r was 0.874. This lower correlation could be ascribed to the TaqMan miR-30c-1-3p assay, which generated very high and therefore uncertain Cq values (33.85 - 36.33). In the liver sample the TaqMan assay failed to de-

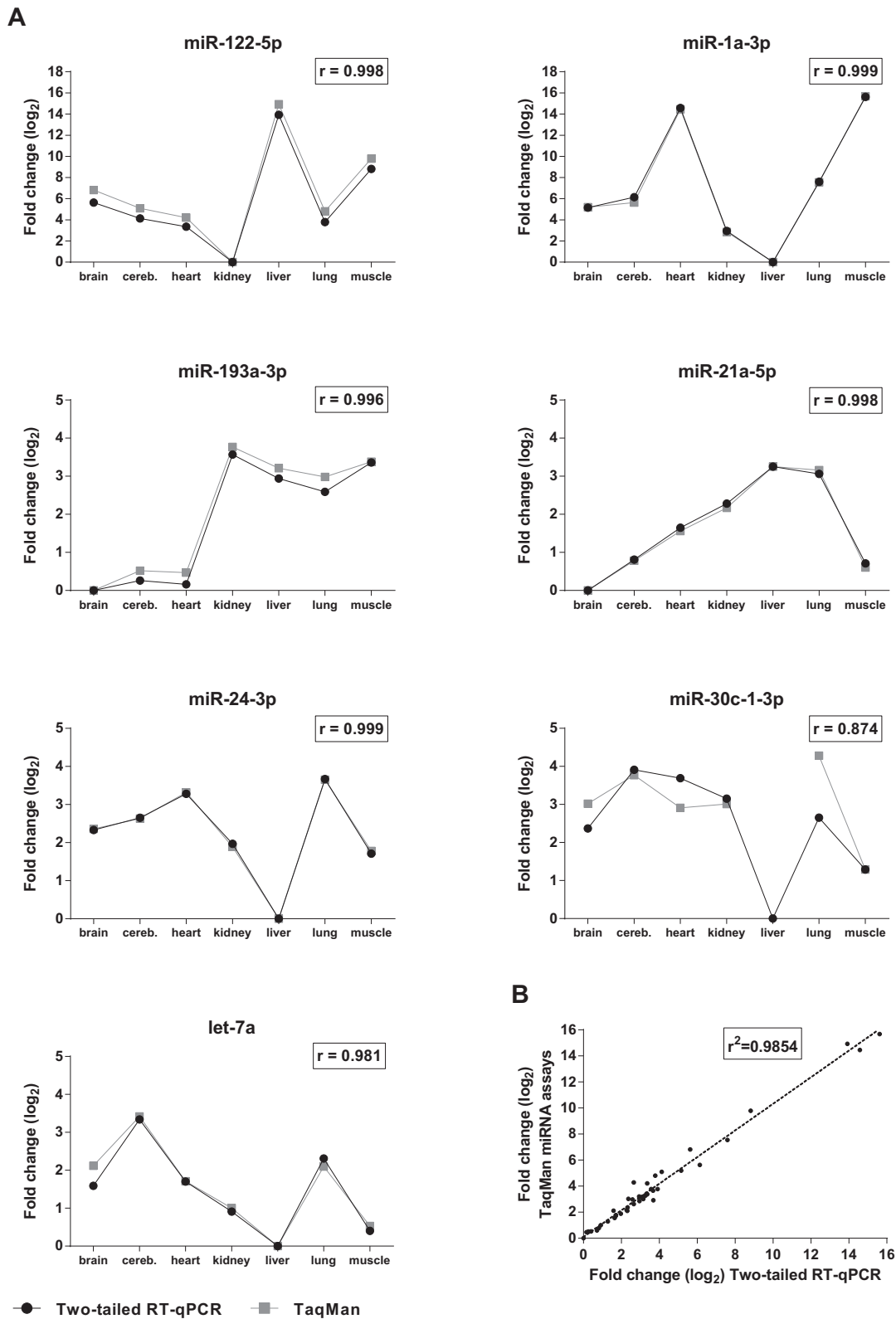


Figure 5. Comparison of expression profiles measured with Two-tailed RT-qPCR and TaqMan miRNA assays. (A) Relative fold changes of the expression of each target in seven tissues measured with Two-tailed RT-qPCR and TaqMan miRNA assays, respectively. (B) Overall correlation of the relative expression changes measured with the two methods.

tect miR-30c-1-3p, while the Two-tailed RT-qPCR assay showed clear positive signal with $C_q = 32.91$. Considering all the measured data the correlation between the Two-tailed RT-qPCR and the TaqMan miRNA assays was excellent ($R^2 = 0.985$, Figure 5B).

Multiplexing of the reverse transcription

Multiplexing the reverse transcription could significantly increase the analysis throughput, save on reagents costs, and reduce the amount of material required. We tested multiplexing the RT step with the Two-tailed RT-qPCR assays by measuring expression of eight miRNA targets across seven mouse tissues. Eight Two-tailed RT primers were pooled and 10 ng of total mouse RNA was reverse transcribed in multiplex. 5% of the cDNA produced was used for each qPCR with one set of PCR primer pair. The C_q values and calculated relative expression levels from singleplex and multiplex protocols were compared. Neither protocol detected miR-615-5p, as in the previous experiment. For miR-122, miR-24, and miR-30c-1 there was no significant difference in C_q values between the singleplex and multiplex protocol. For the remaining four assays there was a shift ranging from 0.68 cycles for let-7a to 2.30 cycles for miR-21a (Figure 6A). The shift was assay specific, but constant and reproducible across samples, and therefore did not influence the calculation of relative expression levels. This is analogous to the mRNA dependent RT yields we have reported before, which do not affect calculations of relative expression levels as they too are constant and reproducible across samples (50,51). The agreement between the relative quantities measured by the multiplex and singleplex protocols was excellent ($R^2 = 0.995$, Figure 6B).

Detection of isomiRs

We assessed the ability of the Two-tailed RT-qPCR assays to measure isomiR variants that differ from the canonical sequence in their length and nucleotide composition in the 3'-terminus. We modified the design of the miR-21 Two-tailed assay such that the 3'-hemiprobe binding site is shifted two nucleotides upstream from the 3'-end of the miR-21 canonical sequence. This allows the detection of isomiRs that are two nucleotides shorter at the 3'-end, as well as all isomiRs with extended 3'-ends. We tested the assay analyzing equal amounts ($\sim 2 \times 10^8$ copies) of five different synthetic variants of hsa-miR-21-5p that had 3'-terminus: (a) shortened by 2 nt, (b) shortened by 1 nt, (c) fully matching canonical sequence (d) extended by 1 nt (-C-3') and (e) extended by 2 nt (-CA-3') and also with the equimolar mixture of all five (suppl. file). For comparison, we also analyzed the samples with the miR-21 TaqMan, Quanta, and miQPCR assays.

We found that Two-tailed RT-qPCR along with the Quanta's miR-21 assay reflects the amounts of different isomiRs with much better precision than the TaqMan and miQPCR miR-21 assays (Figure 7). The C_q values of the different miR-21 isomiRs measured with the Two-tailed RT-qPCR and Quanta miScript assays were similar, as expected, since the initial amount of template had been the same. This demonstrates the ability of the two-tailed RT-qPCR to detect all 3'-isomiR variants with equal effi-

ciency. On the contrary, the TaqMan and miQPCR methods greatly underestimated the amounts of isomiRs that differ from the canonical sequence (Figure 7).

DISCUSSION

We present a new method for the quantification of miRNAs and other small RNAs by RT-qPCR (Figure 1). The new method is called Two-tailed RT-qPCR and is based on sequence specific RT primers with a novel design that allows the RT primer to hybridize to two regions of the miRNA target with separate complementary parts called hemiprobcs. This design offers several advantages over existing strategies for RT-qPCR based detection of miRNAs, including high sensitivity, improved discrimination between similar miRNAs, and ability to quantify isomiRs.

It is well known that specific detection of a nucleic acid sequence requires targeting it with two probe molecules, as a single unmodified standard probe does under most conditions not confer sufficient specificity. The short length of microRNAs limits the size of probes that can be used to interrogate the sequence. For example, two regular PCR primers cannot be used to amplify a regular cDNA copy of the microRNA, as they cannot be fitted without overlap. Reducing the primers' lengths in order to fit makes binding too weak. Binding strength can be increased by incorporation of modified bases such as the Locked Nucleic Acids (LNAs) into the primers (36). Modified primers may have higher melting temperature (T_m) and enhanced sequence discrimination (52). However, results are highly dependent on the design of the LNA oligomers, which often requires extensive trial-and-error optimization (40). LNA-containing primers are also more expensive than conventional primers and may exhibit lower amplification efficiencies (28,53).

We wanted to find a way to interrogate the sequence of a microRNA with two unmodified non-overlapping probes. To solve the thermal stability problem, we reasoned this should be possible by using two short hemiprobcs that are connected. This way each hemiprobe would bind with high specificity, as a single mismatch would greatly distort the rather short hybrid it forms, while overall high thermal stability is achieved through cooperativity. Connecting the hemiprobcs leads to cooperative binding as they drag each other to the binding site. As a consequence the overall binding strength is comparable to that of a long probe. Connecting the hemiprobcs with an oligonucleotide stretch that forms a hairpin protects it from undesired interactions. We call this new primer 'Two-tailed RT primer'. When used in reverse transcription, the Two-tailed RT primer is extended forming a tailed cDNA. The cDNA can then be amplified by PCR.

Another aspect of the Two-tailed RT primer is that the 3'-hemiprobe can be made rather short (5-6 nt). This leaves enough space to design an unmodified miRNA-specific qPCR primer without overlapping with the 3'-hemiprobe. Also, a short 3'-hemiprobe is more sensitive to mismatches in the target sequence, while sufficient binding strength is obtained through cooperative binding with the longer 5'-hemiprobe. Indeed, when we compared Two-tailed RT primers with the same length of the 3'-hemiprobe (5nt), but

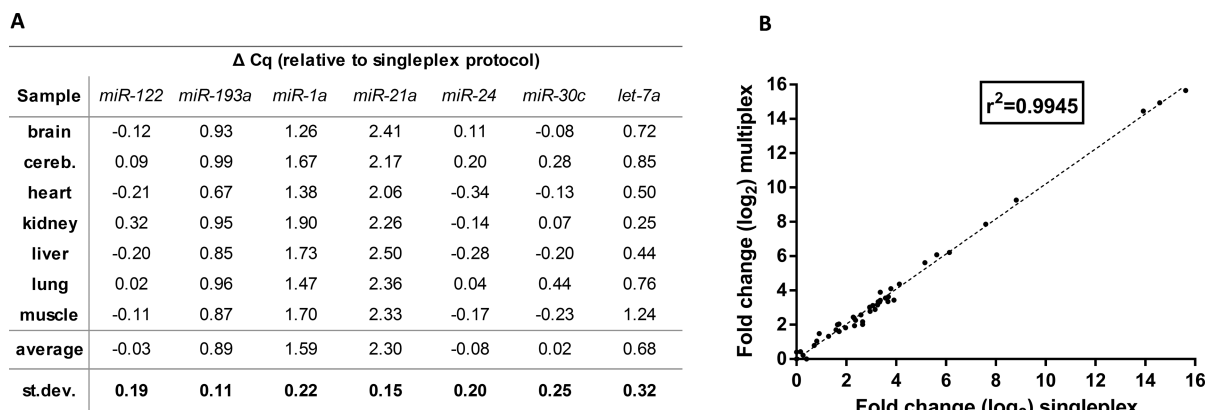


Figure 6. Comparison of singleplex and multiplex Two-tailed RT-qPCR. (A) $\Delta Cq = Cq_{\text{multiplex}} - Cq_{\text{singleplex}}$. (B) Overall correlation of the relative expression changes between tissues measured with the singleplex and multiplex protocol.

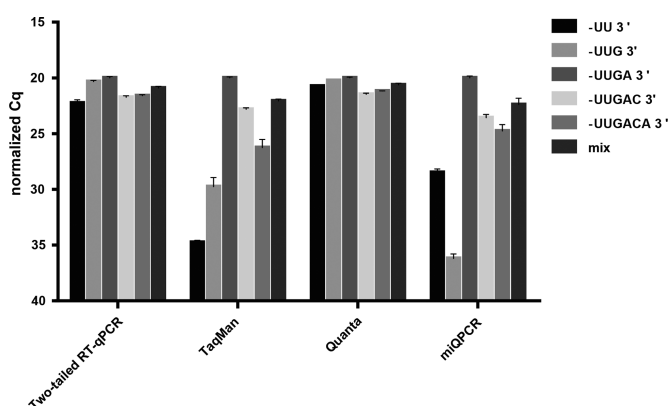


Figure 7. Relative sensitivities of Two-tailed RT-qPCR, TaqMan, Quanta, and miQPCR to miR-21 isomiRs. Cq values are normalized such that the Cq of the canonical form is set to 20. Error bars indicate SD of two independent cDNA syntheses.

different length of the 5'-hemiprobe we found significant differences (Figure 2A).

By strategic design of the 5'-hemiprobe the specificity of the Two-tailed RT-qPCR can be optimized for different cases. For example, closely related miRNAs often differ in base positions in their 5'-regions. Those variants are poorly distinguished with current methods for miRNA analysis, which use only on one of the qPCR primers for discrimination. This holds also for the specific-primer based TaqMan miRNA assays, as this part of the microRNA is not sensed by its stem-loop primer. With the Two-tailed RT-qPCR assays, the nucleotides in the 5'-regions, such as those distinguishing let-7a, let-7e, let-7f, and let-7g, are interrogated twice: first time in the reverse transcription by the 5'-hemiprobe of the RT primer and second time in the qPCR by the reverse PCR primer (Figure 1). Specificity can be maximized using a short 5'-hemiprobe designed to sense all the sequence variants in the variable region of the miRNA (Supplementary Figure S1). Using this strategy, we designed Two-tailed RT-qPCR assays that exhibit negligible cross-reactivity between the members of the let-7 miRNA family (Figure 4B). When these members were as-

sayed with the three other methods, substantial undesired cross-reactivity was observed (Figure 4A).

Since the sequence of the mature transcript is contained in its precursors (3), RT-qPCR assays designed to detect mature miRNAs may also amplify their precursor molecules. Although precursors are usually present in cells at much lower levels than the mature miRNAs (54–57), it may still be relevant to measure them separately. We decided to test if the two-tailed RT-qPCR can distinguish between mature miRNAs and the corresponding pre-miRNAs. Our results show that the Two-tailed RT-qPCR assays designed for mature miRNAs show minimal cross-reaction with pre-miRNAs (Figure 4C). One contributing factor is that reverse transcription is performed at rather low temperature (25°C) without any pre-heating step that would open the secondary structure of pre-miRNAs making them available for priming with the Two-tailed RT primer. We observed that even without the pre-heating step, sensitivity of the system is not influenced by potential microRNA-long RNA interactions (Supplementary Figure S2). The specificity of the two-tailed RT-qPCR assays for mature miRNAs is further confirmed by the excellent correlation with the expression profiles measured using TaqMan miRNA assays (Figure 5), which do not cross-react with precursor miRNAs because of its particular RT-priming mechanism (39,44).

Target-specific priming of the reverse transcription has several advantages including higher specificity and lower background, but a common disadvantage is that each target requires a separate RT reaction. This can be resolved by multiplexing the RT step using an RT primer pool. Multiplexing RT increases throughput, saves on reagent costs, minimizes labor, and reduces the sample amount needed. Usually, it is problematic to have many long oligomers, such as multiple RT primers, in the same reaction as it leads to unspecific amplification (58). The Two-tailed RT primers are, however, designed with a hairpin structure that prevents non-specific interactions. As demonstrated, multiplexing eight targets showed perfect agreement with the corresponding singleplex reactions (Figure 6). Importantly, the specificity of the multiplex measurement was not compromised, based on negative controls and melting curve analyses (suppl. file). This is likely due to the use of two spe-

Table 2. Comparison of parameters of four tested RT-qPCR methods for miRNA quantification. Cost estimate per assay with 20 RT and 60 qPCR reactions is indicated (details available in suppl. file)

	cDNA synthesis strategy	Sensitivity (miRNA copies in analyzed volume)	Linear dynamic range	Unspecific cross-reaction	Accurate detection of isomiRs	Melting curve for specificity control	Detection of piRNAs and plant miRNAs	Cost
Two-tailed RT-qPCR	specific RT primer	10 ² - 10 ³	6–7 logs	< 1%	yes	yes	yes	\$
TaqMan	specific RT primer	10 ² - 10 ³	6–7 logs	< 23%	no	no	yes	\$\$\$\$
Quanta	poly(A) tailing	10 ² - 10 ³	6–7 logs	< 51%	yes	yes	no	\$\$
miQPCR	linker ligation	10 ⁴	5 logs	> 100%	no	yes	yes, but efficiency is reduced	\$\$\$

cific PCR primers for each target in the downstream qPCR, while other methods for miRNA analysis commonly use only one miRNA specific primer combined with a universal primer. The Two-tailed RT-qPCR assays are designed with target specific PCR primers as follows: one miRNA specific reverse primer and one forward primer that is specific to an internal segment of the corresponding Two-tailed RT primer. This way both PCR primers are specific to one cDNA only. Although using a second specific PCR primer provides no advantage with the singleplex protocol, it adds specificity to the multiplex protocol, where many different cDNAs are present as qPCR templates. In our experiment, we performed reverse transcription in octaplex and then divided the cDNA into aliquots that were analyzed with target specific PCR primer pairs in singleplex qPCR using dye based detection. This workflow is, however, not limited to any particular number of targets and should be applicable to virtually any degree of multiplexing. When only a small number of microRNAs are targeted 1-step RT-qPCR can be employed, distinguishing the PCR products with fluorescent probes.

Another advantage of the Two-tailed RT-qPCR is its capability to reverse transcribe isomiRs with the same efficiency as the canonical sequence (Figure 7). IsomiRs are miRNA variants that differ in length and sequence composition in their 3'- and/or 5'-termini from the annotated canonical sequence (23,24). Growing evidence suggests isomiRs are expressed in a cell, tissue, and gender specific manner, possess relevant physiological functions, and are potential biomarkers in clinical diagnostics (59–63). In some cases, isomiRs, derived from the same precursor arm, bind to other targets than the canonical miRNA. Another function is that isomiRs cooperate with the canonical form to drive similar biology by targeting the same set of core biological networks while distributing the off-target effects and thus increasing the signal to noise ratio of gene silencing (62,64). Measuring full isomiR profiles may therefore be more valuable than targeting the canonical sequence only. This was recently demonstrated when clearly improved discrimination of cancerous and healthy tissues was obtained by inclusion of full isomiR profiles (59,65).

The heterogeneity of the miRNA sequences pose a substantial issue for many RT-qPCR methods. Those employing universal reverse transcription should generally be able to amplify all terminal sequence variants, but biochemical modifications of terminal nucleotides may interfere with

the enzymatic steps upstream of the RT-qPCR. Furthermore, the design of the stem-loop primers used in the TaqMan miRNA assays renders the method less sensitive to isomiRs. Due to the stem-part, which blocks annealing to longer sequences, only the particular sequence with the defined ends is amplified with optimal efficiency, and the method may completely miss on some variants (40–42). Particularly isomiRs that differ at the 3'-end pose a problem. Notably, largest variability across isomiRs is found in their 3'-ends, and in some cases the canonical sequence represents only a small fraction of the total amount of a miRNA (24,59,62,64,66,67).

To test the ability of the Two-tailed RT-qPCR to detect different 3'-isomiRs we measured five synthetic terminal variants of miR-21. We observed that the Two-tailed RT-qPCR accurately quantifies shorter as well as longer variants of the canonical sequence, reaching the level of a poly(A)-tail based method while providing all advantages of RT-specific priming. The full isomiR repertoire of the miRNA is thus measured and no potentially important isomiRs are missed (Figure 7). This is in difference from the TaqMan and miQPCR approaches, which exhibited high variation across the miR-21 isomiRs, greatly underestimating the amounts of those that differ from the canonical sequence (Figure 7).

In some applications only a certain isomiR is of interest. Although it should be possible to target it specifically with custom-designed TaqMan miRNA assays, our results (Figure 7) as well as reports in the literature suggest this is not always the case and various degree of cross-reactivity to other isomiRs is observed (41,42,68,69). Therefore, it seems the TaqMan miRNA assays are neither specific to a single 3'-end isomiR nor do they detect all the isomiRs with equal sensitivity, which may lead to an underestimation of the total amount of a given miRNA in a sample (25,42). To our knowledge, the only qPCR based method that can be used to distinctively quantify specific isomiRs with 1 nt resolution is the 'Dumbbell-PCR' (69). This method exploits the properties of T4 RNA ligase 2 to ligate stem-loop adapters to the ends of the targeted isomiR. The formed dumbbell-like structure is then quantified with TaqMan qPCR. Such extreme specificity is currently not achieved with the two-tailed RT-qPCR, but the Two-tailed RT primer can be designed to reverse transcribe all isomiRs of a particular miRNA to obtain a correct quantification of the total amount.

RT-qPCR technology remains unequalled tool in small-RNA expression profiling, vital for validation of genome-wide experiments and accurate measurement of challenging samples such as liquid biopsies. However, measurements of higher number of targets significantly increases the cost of such analyses. We have developed a highly sensitive and exceedingly specific method called Two-tailed RT-qPCR, suitable for rapid and cost-effective microRNA profiling. At the same time, Two-tailed RT-qPCR reflects on the current state of microRNA field and confers several advantages over current RT-qPCR methods, including increased specificity and ability to capture the full isomiR profile (Table 2). Two-tailed RT-qPCR uses only standard oligomers, can employ either dye or probe based detection and can be used for animal and plant small RNAs alike. The whole analysis can be performed in just 2.5 h.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Pavel Honsa and Martin Valny from the Institute of Experimental Medicine CAS for providing mouse tissue samples.

FUNDING

Czech Science Foundation [P303/16/10214S, P303/17/040 34S]; Swedish foundation for Strategic Research [ID14-0075]; BIOCEV CZ.1.05/1.1.00/02.0109 provided by ERDF and MEYS; Operational Programme Education for Competitiveness [CZ.1.07/2.3.00/30.0045] from the European Social Fund and the state budget of the Czech Republic. Funding for open access charge: Czech Science Foundation [P303/16/10214S].


Conflict of interest statement. M.K. and R.S. have shares in TATAA Biocenter AB. Two-tailed RT-qPCR is patent applied in PCT/US15/45966.

REFERENCES

- Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.
- Huntzinger,E. and Izaurralde,E. (2011) Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.*, **12**, 99–110.
- Kim,V.N., Han,J. and Siomi,M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.
- Winter,J., Jung,S., Keller,S., Gregory,R.I. and Diederichs,S. (2009) Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.*, **11**, 228–234.
- Krol,J., Loedige,I. and Filipowicz,W. (2010) The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.*, **11**, 597–610.
- Tan,L., Yu,J.T. and Tan,L. (2015) Causes and Consequences of MicroRNA Dysregulation in Neurodegenerative Diseases. *Mol. Neurobiol.*, **51**, 1249–1262.
- Croce,C.M. (2009) Causes and consequences of microRNA dysregulation in cancer. *Nat. Rev. Genet.*, **10**, 704–714.
- Xiao,C. and Rajewsky,K. (2009) MicroRNA control in the immune system: basic principles. *Cell*, **136**, 26–36.
- Lin,S. and Gregory,R.I. (2015) MicroRNA biogenesis pathways in cancer. *Nat. Rev. Cancer*, **15**, 321–333.
- Hatfield,S.D., Shcherbata,H.R., Fischer,K.A., Nakahara,K., Carthew,R.W. and Ruohola-Baker,H. (2005) Stem cell division is regulated by the microRNA pathway. *Nature*, **435**, 974–978.
- He,Y., Lin,J., Kong,D., Huang,M., Xu,C., Kim,T.K., Etheridge,A., Luo,Y., Ding,Y. and Wang,K. (2015) Current State of Circulating MicroRNAs as Cancer Biomarkers. *Clin. Chem.*, **61**, 1138–1155.
- Basak,I., Patil,K.S., Alves,G., Larsen,J.P. and Moller,S.G. (2016) microRNAs as neuroregulators, biomarkers and therapeutic agents in neurodegenerative diseases. *Cell. Mol. Life Sci.: CMLS*, **73**, 811–827.
- Moldovan,L., Batte,K.E., Trgovcich,J., Wisler,J., Marsh,C.B. and Piper,M. (2014) Methodological challenges in utilizing miRNAs as circulating biomarkers. *J. Cell. Mol. Med.*, **18**, 371–390.
- Schwarzenbach,H., Nishida,N., Calin,G.A. and Pantel,K. (2014) Clinical relevance of circulating cell-free microRNAs in cancer. *Nat. Rev. Clin. Oncol.*, **11**, 145–156.
- Jiang,H.X., Liang,Z.Z., Ma,Y.H., Kong,D.M. and Hong,Z.Y. (2016) G-quadruplex fluorescent probe-mediated real-time rolling circle amplification strategy for highly sensitive microRNA detection. *Anal. Chim. Acta*, **943**, 114–122.
- Liu,H., Li,L., Duan,L., Wang,X., Xie,Y., Tong,L., Wang,Q. and Tang,B. (2013) High specific and ultrasensitive isothermal detection of microRNA by padlock probe-based exponential rolling circle amplification. *Anal. Chem.*, **85**, 7941–7947.
- Ma,F., Liu,M., Tang,B. and Zhang,C.Y. (2017) Rapid and sensitive quantification of microRNAs by isothermal helicase-dependent amplification. *Anal. Chem.*, **89**, 6182–6187.
- Deng,R., Zhang,K. and Li,J. (2017) Isothermal amplification for microRNA detection: from the test tube to the cell. *Acc. Chem. Res.*, **50**, 1059–1068.
- Mestdagh,P., Hartmann,N., Baeriswyl,L., Andreasen,D., Bernard,N., Chen,C.F., Cheo,D., D’Andrade,P., DeMayo,M., Dennis,L. et al. (2014) Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study (vol 11, pg 809, 2014). *Nat. Methods*, **11**, 971–971.
- Svoboda,P. (2015) A toolbox for miRNA analysis. *FEBS Lett.*, **589**, 1694–1701.
- Pritchard,C.C., Cheng,H.H. and Tewari,M. (2012) MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.*, **13**, 358–369.
- Aldridge,S. and Hadfield,J. (2012) Introduction to miRNA profiling technologies and cross-platform comparison. *Methods Mol. Biol.*, **822**, 19–31.
- Guo,L. and Chen,F. (2014) A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene*, **544**, 1–7.
- Neilsen,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs—the overlooked repertoire in the dynamic microRNAome. *Trends Genet.: TIG*, **28**, 544–549.
- Dellett,M. and Simpson,D.A. (2016) Considerations for optimization of microRNA PCR assays for molecular diagnosis. *Expert Rev. Mol. Diagnost.*, **16**, 407–414.
- Jin,J., Vaud,S., Zhelkovsky,A.M., Posfai,J. and McReynolds,L.A. (2016) Sensitive and specific miRNA detection method using SplintR Ligase. *Nucleic Acids Res.*, **44**, e116.
- Li,J., Yao,B., Huang,H., Wang,Z., Sun,C., Fan,Y., Chang,Q., Li,S., Wang,X. and Xi,J. (2009) Real-time polymerase chain reaction microRNA detection based on enzymatic stem-loop probes ligation. *Anal. Chem.*, **81**, 5446–5451.
- Balcells,I., Cirera,S. and Busk,P.K. (2011) Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. *BMC Biotechnol.*, **11**, 70.
- Shi,R. and Chiang,V. (2005) Facile means for quantifying microRNA expression by real-time PCR. *BioTechniques*, **39**, 519–525.
- Mei,Q., Li,X., Meng,Y., Wu,Z., Guo,M., Zhao,Y., Fu,X. and Han,W. (2012) A facile and specific assay for quantifying microRNA by an optimized RT-qPCR approach. *PLoS One*, **7**, e46890.
- Benes,V., Collier,P., Kordes,C., Stolte,J., Rausch,T., Muckentaler,M.U., Haussinger,D. and Castoldi,M. (2015) Identification of cytokine-induced modulation of microRNA expression and secretion as measured by a novel microRNA specific qPCR assay. *Scientific Rep.*, **5**, 11590.
- Munafo,D.B. and Robb,G.B. (2010) Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA*, **16**, 2537–2552.

33. Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y. and Robb, G.B. (2012) Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.*, **40**, e54.
34. Sorefan, K., Pais, H., Hall, A.E., Kozomara, A., Griffiths-Jones, S., Moulton, V. and Dalmy, T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*, **3**, 4.
35. Yehudai-Resheff, S. and Schuster, G. (2000) Characterization of the E.coli poly(A) polymerase: nucleotide specificity, RNA-binding affinities and RNA structure dependence. *Nucleic Acids Res.*, **28**, 1139–1144.
36. Raymond, C.K., Roberts, B.S., Garrett-Engele, P., Lim, L.P. and Johnson, J.M. (2005) Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA*, **11**, 1737–1744.
37. Sharbati-Tehrani, S., Kutz-Lohroff, B., Bergbauer, R., Scholven, J. and Einspanier, R. (2008) miR-Q: a novel quantitative RT-PCR approach for the expression profiling of small RNA molecules such as miRNAs in a complex sample. *BMC Mol. Biol.*, **9**, 34.
38. Huang, T., Yang, J., Liu, G., Jin, W., Liu, Z., Zhao, S. and Yao, M. (2015) Quantification of mature microRNAs using pincer probes and real-time PCR amplification. *PLoS One*, **10**, e0120160.
39. Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R. *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.*, **33**, e179.
40. Benes, V. and Castoldi, M. (2010) Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods*, **50**, 244–249.
41. Schamberger, A. and Orban, T.I. (2014) 3' IsomiR species and DNA contamination influence reliable quantification of microRNAs by stem-loop quantitative PCR. *PLoS One*, **9**, e106315.
42. Soundara Pandi, S.P., Chen, M., Guduric-Fuchs, J., Xu, H. and Simpson, D.A. (2013) Extremely complex populations of small RNAs in the mouse retina and RPE/choroid. *Invest. Ophthalmol. Vis. Sci.*, **54**, 8140–8151.
43. Lao, K., Xu, N.L., Yeung, V., Chen, C., Livak, K.J. and Straus, N.A. (2006) Multiplexing RT-PCR for the detection of multiple miRNA species in small samples. *Biochem. Biophys. Res. Commun.*, **343**, 85–89.
44. Tang, F., Hajkova, P., Barton, S.C., O'Carroll, D., Lee, C., Lao, K. and Surani, M.A. (2006) 220-plex microRNA expression profile of a single cell. *Nat. Protoc.*, **1**, 1154–1159.
45. Tang, F., Hajkova, P., Barton, S.C., Lao, K. and Surani, M.A. (2006) MicroRNA expression profiling of single whole embryonic stem cells. *Nucleic Acids Res.*, **34**, e9.
46. Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
47. Markham, N.R. and Zuker, M. (2008) UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.*, **453**, 3–31.
48. Forootan, A., Sjöback, R., Björkman, J., Sjögreen, B., Linz, L. and Kubista, M. Methods to determine limit of detection and limit of quantification in quantitative real-time PCR (qPCR). *Biomol. Detect. Quantif.*, **12**, 1–6.
49. Liang, Y., Ridzon, D., Wong, L. and Chen, C. (2007) Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics*, **8**, 166.
50. Stahlberg, A., Hakansson, J., Xian, X., Semb, H. and Kubista, M. (2004) Properties of the reverse transcription reaction in mRNA quantification. *Clin. Chem.*, **50**, 509–515.
51. Stahlberg, A., Kubista, M. and Pfaffl, M. (2004) Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.*, **50**, 1678–1680.
52. Vester, B. and Wengel, J. (2004) LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry*, **43**, 13233–13241.
53. Veedu, R.N., Vester, B. and Wengel, J. (2007) Enzymatic incorporation of LNA nucleotides into DNA strands. *ChemBiochem*, **8**, 490–492.
54. Gan, L. and Denecke, B. (2013) Profiling pre-microRNA and mature microRNA expressions using a single microarray and avoiding separate sample preparation. *Microarrays*, **2**, 24–33.
55. Schmittgen, T.D., Lee, E.J., Jiang, J., Sarkar, A., Yang, L., Elton, T.S. and Chen, C. (2008) Real-time PCR quantification of precursor and mature microRNA. *Methods*, **44**, 31–38.
56. Li, N., You, X., Chen, T., Mackowiak, S.D., Friedlander, M.R., Weigt, M., Du, H., Gogol-Doring, A., Chang, Z., Dieterich, C. *et al.* (2013) Global profiling of miRNAs and the hairpin precursors: insights into miRNA processing and novel miRNA discovery. *Nucleic Acids Res.*, **41**, 3619–3634.
57. Faridani, O.R., Abdullayev, I., Hagemann-Jensen, M., Schell, J.P., Lanner, F. and Sandberg, R. (2016) Single-cell sequencing of the small-RNA transcriptome. *Nat. Biotechnol.*, **34**, 1264–1266.
58. Stahlberg, A., Krzyzanowski, P.M., Jackson, J.B., Egyud, M., Stein, L. and Godfrey, T.E. (2016) Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. *Nucleic Acids Res.*, **44**, e105.
59. Telonis, A.G., Lohrer, P., Jing, Y., Londin, E. and Rigoutsos, I. (2015) Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res.*, **43**, 9158–9175.
60. Lohrer, P., Londin, E.R. and Rigoutsos, I. (2014) IsomiR expression profiles in human lymphoblastoid cell lines exhibit population and gender dependencies. *Oncotarget*, **5**, 8790–8802.
61. Siddle, K.J., Tailleux, L., Deschamps, M., Loh, Y.H., Deluen, C., Gicquel, B., Antoniewski, C., Barreiro, L.B., Farinelli, L. and Quintana-Murci, L. (2015) bacterial infection drives the expression dynamics of microRNAs and their isomiRs. *PLoS Genet.*, **11**, e1005064.
62. Cloonan, N., Wani, S., Xu, Q., Gu, J., Lea, K., Heater, S., Barbacioru, C., Steptoe, A.L., Martin, H.C., Nourbakhsh, E. *et al.* (2011) MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.*, **12**, R126.
63. Wang, S., Xu, Y., Li, M., Tu, J. and Lu, Z. (2016) Dysregulation of miRNA isoform level at 5' end in Alzheimer's disease. *Gene*, **584**, 167–172.
64. Ahmed, F., Senthil-Kumar, M., Lee, S., Dai, X., Mysore, K.S. and Zhao, P.X. (2014) Comprehensive analysis of small RNA-seq data reveals that combination of miRNA with its isomiRs increase the accuracy of target prediction in Arabidopsis thaliana. *RNA Biol.*, **11**, 1414–1429.
65. Koppers-Lalic, D., Hackenberg, M., de Menezes, R., Misovic, B., Wachalska, M., Geldof, A., Zini, N., de Reijke, T., Wurdinger, T., Vis, A. *et al.* (2016) Noninvasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget*, **7**, 22566–22578.
66. McGahon, M.K., Yarham, J.M., Daly, A., Guduric-Fuchs, J., Ferguson, L.J., Simpson, D.A. and Collins, A. (2013) Distinctive profile of IsomiR expression and novel microRNAs in rat heart left ventricle. *PLoS One*, **8**, e65809.
67. Baran-Gale, J., Fannin, E.E., Kurtz, C.L. and Sethupathy, P. (2013) Beta cell 5'-shifted isomiRs are candidate regulatory hubs in type 2 diabetes. *PLoS One*, **8**, e73240.
68. Wu, H., Neilson, J.R., Kumar, P., Manocha, M., Shankar, P., Sharp, P.A. and Manjunath, N. (2007) miRNA profiling of naive, effector and memory CD8 T cells. *PLoS One*, **2**, e1020.
69. Honda, S. and Kirino, Y. (2015) Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res.*, **43**, e77.

SCIENTIFIC REPORTS



OPEN

Two-tailed RT-qPCR panel for quality control of circulating microRNA studies

Peter Androvic^{1,2}, Nataliya Romanyuk³, Lucia Urdzikova-Machova³, Eva Rohlova^{1,5}, Mikael Kubista^{1,4} & Lukas Valihrach¹

Circulating cell-free microRNAs are promising candidates for minimally invasive clinical biomarkers for the diagnosis, prognosis and monitoring of many human diseases. Despite substantial efforts invested in the field, the research so far has failed to deliver expected results. One of the contributing factors is general lack of agreement between various studies, partly due to the considerable technical challenges accompanying the workflow. Pre-analytical variables including sample collection, RNA isolation, and quantification are sources of bias that may hamper biological interpretation of the results. Here, we present a Two-tailed RT-qPCR panel for quality control, monitoring of technical performance, and optimization of microRNA profiling experiments from biofluid samples. The Two-tailed QC (quality control) panel is based on two sets of synthetic spike-in molecules and three endogenous microRNAs that are quantified with the highly specific Two-tailed RT-qPCR technology. The QC panel is a cost-effective way to assess quality of isolated microRNA, degree of inhibition, and erythrocyte contamination to ensure technical soundness of the obtained results. We provide assay sequences, detailed experimental protocol and guide to data interpretation. The application of the QC panel is demonstrated on the optimization of RNA isolation from biofluids with the miRNeasy Serum/Plasma Advanced Kit (Qiagen).

Circulating cell-free microRNAs have emerged in recent years as promising candidates for minimally invasive clinical biomarkers for diagnosis, prognosis, and monitoring of a multitude of human pathologies^{1–6}. After this recognition, a massive wave of research aiming at identifying disease-associated microRNAs followed. A search for the keywords “microRNA”, “biomarker” and “blood” returns over 5000 hits in the PubMed database (September 2018) with the number of studies increasing every year. Despite promising advances in the field (www.clinicaltrials.gov), there is still no microRNA test in clinical practice^{7,8}. There are many reasons behind the current unsatisfactory state and their comprehensive discussion is beyond the scope of this article (for reviews see^{8,9}). One issue is the poor agreement between studies, which may in part be attributed to the lack of standardization¹⁰ and technical difficulties associated with the workflow^{11,12}. Protocols for blood collection, sample processing, storage, RNA isolation, and microRNA quantification often vary across laboratories leading to discordant results^{11,13,14}. Efforts are ongoing to standardize the blood sampling and processing steps to improve the reproducibility of microRNA analyses^{15–18} (www.cancer-id.eu; www.spidia.eu). However, notable sources of variation remain. These include the RNA isolation, co-purification of inhibitors of enzymatic reactions, and cellular contamination of the biofluid samples^{19–21}. These factors may bias the measured microRNA profiles leading to false-positive discoveries of disease-associated biomarkers. Rigorous control of sample quality and technical workflow is therefore of highest importance.

An efficient way to monitor technical variation is the addition of exogenous spike-in molecules prior to RNA isolation^{22–24}. The signal from the spike-ins reflect yields and extraction efficiency, which identifies abnormal samples that should be reanalysed or disqualified. A second set of exogenous spike-ins can be added before the microRNA quantification to control for bias introduced downstream of this step, such as the inhibition of

¹Institute of Biotechnology of the Czech Academy of Sciences - BIOCEV, Vestec, 252 50, Czech Republic. ²Laboratory of Growth Regulators, Faculty of Science, Palacky University, Olomouc, 78371, Czech Republic. ³Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, 142 20, Czech Republic. ⁴TATAA Biocenter AB, Gothenburg, 411 03, Sweden. ⁵Department of Anthropology and Human Genetics, Faculty of Science, Charles University, Prague, 128 43, Czech Republic. Correspondence and requests for materials should be addressed to L.V. (email: lukas.valihrach@ibt.cas.cz)

enzymatic reactions. This concept has been described previously^{20,22,23} and is also available as a commercial product (e.g. RNA Spike-In Kit, for RT marketed by Qiagen). Yet, to our knowledge, there is no tool to perform such extended quality control on challenging experimental samples described in detail in literature.

Here, we present the Two-tailed quality control (QC) panel; a tool to assess the technical quality of RNA isolation, degree of inhibition and erythrocyte contamination, primarily in liquid biopsy samples, such as serum and plasma. The panel is based on Two-tailed RT-qPCR; a highly specific method for microRNA quantification²⁵. We provide detailed experimental protocol, guide to data interpretation, and sequences of the RNA oligonucleotides and RT-qPCR assays used (Supplementary file) that can be ordered from any licensed oligo manufacturer. The QC panel is based on standard reagents and is intended to provide researchers a convenient tool to assess technical performance and quality of the samples before investing resources into extensive quantification experiment, such as small-RNA sequencing or high-throughput RT-qPCR. We demonstrate its utility by optimizing an RNA isolation protocol, screening for haemolysis, and testing for outliers with compromised quality. We also report data obtained with the recently launched miRNeasy Serum/Plasma Advanced Kit (Qiagen) for two biofluids collected from human and rat.

Results

Design of the QC panel. The Two-tailed QC panel is composed of five synthetic spike-in microRNAs and eight Two-tailed assays targeting these synthetic spike-ins, and three endogenous microRNAs (Fig. 1). The spike-ins are based on *C. elegans* microRNAs and artificial sequences and have no significant homology to any known human, mouse or rat microRNA (Table 1). All spike-ins have 5' terminal phosphate to mimic endogenous microRNAs, and to allow incorporation into microRNA libraries for Next Generation Sequencing (NGS).

Three spike-in RNAs (cel-miR-54, spike-A and spike-B) comprise the isolation spike-in mix and are added to the samples at a known constant amount prior to RNA isolation, serving as control for the technical performance of the RNA isolation protocol (Fig. 1A). The three spike-ins have varying GC content (41.7–63.6%) and are present at concentrations reflecting high (cel-miR-54, $1e + 7$ copies/ μ l), moderate (spike-A, $2e + 5$ copies/ μ l), and low (spike-B, $4e + 3$ copies/ μ l) abundant microRNAs (Supplementary file). The Δ Cq's between the isolation spike-ins should, in absence of inhibition, be in the range 3.5–5.5 cycles (accounting for differences in RT-PCR efficiencies of the Two-tailed assays), however, these values may be influenced differently by individual isolation protocols due to various biases^{26,27}.

Two RNA spike-ins (cel-miR-76 and cel-miR-2) comprise the reverse transcription (RT) spike-in mix and are added to the RT reaction serving as controls for cDNA synthesis, PCR amplification and as general controls for the presence of inhibitors in RNA eluates (Fig. 1A). Cel-miR-76 ($1e + 7$ copies/ μ l) is added at 100x higher concentration than cel-miR-2 ($1e + 5$ copies/ μ l) and their Δ Cq should be 5.5–6.5 cycles (accounting for differences in PCR efficiency of the Two-tailed assays).

The QC panel also contains assays for the three endogenous microRNAs: let-7a, miR-23a and miR-451a. Let-7a is abundant in plasma and serum^{20,28,29} and serves as positive control. Mir-23a is also abundant in plasma/serum and its level is independent of haemolysis, while miR-451a is highly abundant in erythrocytes and its level increases dramatically upon haemolysis^{20,30}. The Δ Cq (mir-23a – mir-451a) indicates degree of haemolysis in the samples²⁰.

Optimization of sample input volume. A factor that is often neglected, but can have major impact on the quality of microRNA quantification data, is the initial input volume used for the RNA isolation^{13,26}. Liquid biopsy samples contain very low amounts of microRNAs and researchers may be tempted to use as much sample material as possible for RNA isolation. However, with increasing amount of starting material risk of carryover of contaminating substances and saturation of the purification column increases^{31,32}. Most commercial RNA isolation kit manufacturers recommend 200 μ l starting serum/plasma volume, however, optimum volume depends on the isolation protocol, sample type and also organism²⁶. Optimizing the sample volume is therefore recommended when setting up a new isolation protocol or extracting a new type of sample. For such optimization the Two-tailed QC panel is a tool to assess relative isolation efficiency, absolute yield, and test for the presence of inhibitors to decide the optimal input volume. With this strategy we optimized protocol based on the miRNeasy Serum/Plasma Advanced Kit (Qiagen) for RT-qPCR analysis of human plasma, human serum, and rat serum (Fig. 2).

We found a non-linear relation between the input sample volume and cDNA yield as reflected by RT-qPCR signal of endogenous microRNAs (Fig. 2). The non-linearity is caused neither by RT nor PCR inhibition, as the signals from the RT spike-ins were independent of volume. Rather the non-linear response is due to variations in RNA isolation efficiency, as reflected by the RT-qPCR response of the isolation spike-ins (Fig. 2). We observed poor isolation efficiency with low input volumes (<200 μ l for human, <100 μ l for rat), but also with higher input volumes (≥ 300 μ l for human, ≥ 200 μ l for rat), where the response was also more variable (Fig. 2). Based on our results, optimum starting sample volumes with our workflow are: 250 μ l for human plasma, 300–500 μ l for human serum, and 150 μ l for rat serum.

Assessing the effect of co-precipitants in the isolation procedure. Since biofluids like serum and plasma contain very low amounts of RNA, significant portion may be lost during the isolation procedure due to adsorption to the pipette tips, tube walls etc. Losses can be reduced by adding carriers such as MS2 phage RNA or yeast tRNA to the samples before RNA isolation^{33,34}. However, RNA-based carriers are less suited when NGS is used for downstream analysis as the exogenous RNAs may consume sequencing reads. Other carriers, such as linear acrylamide, BSA or glycogen may then be used instead²⁴. Using the Two-tailed QC panel we tested the impact of using glycogen as carrier in our isolation procedure (Fig. 3). In accordance with previous observations^{33,34}, we found that addition of glycogen significantly improved the reproducibility of isolation (F-test, $p < 0.001$) and significantly increased the yield (average Cq difference 1.25; paired T-test $p = 0.011$) with no negative effects on the

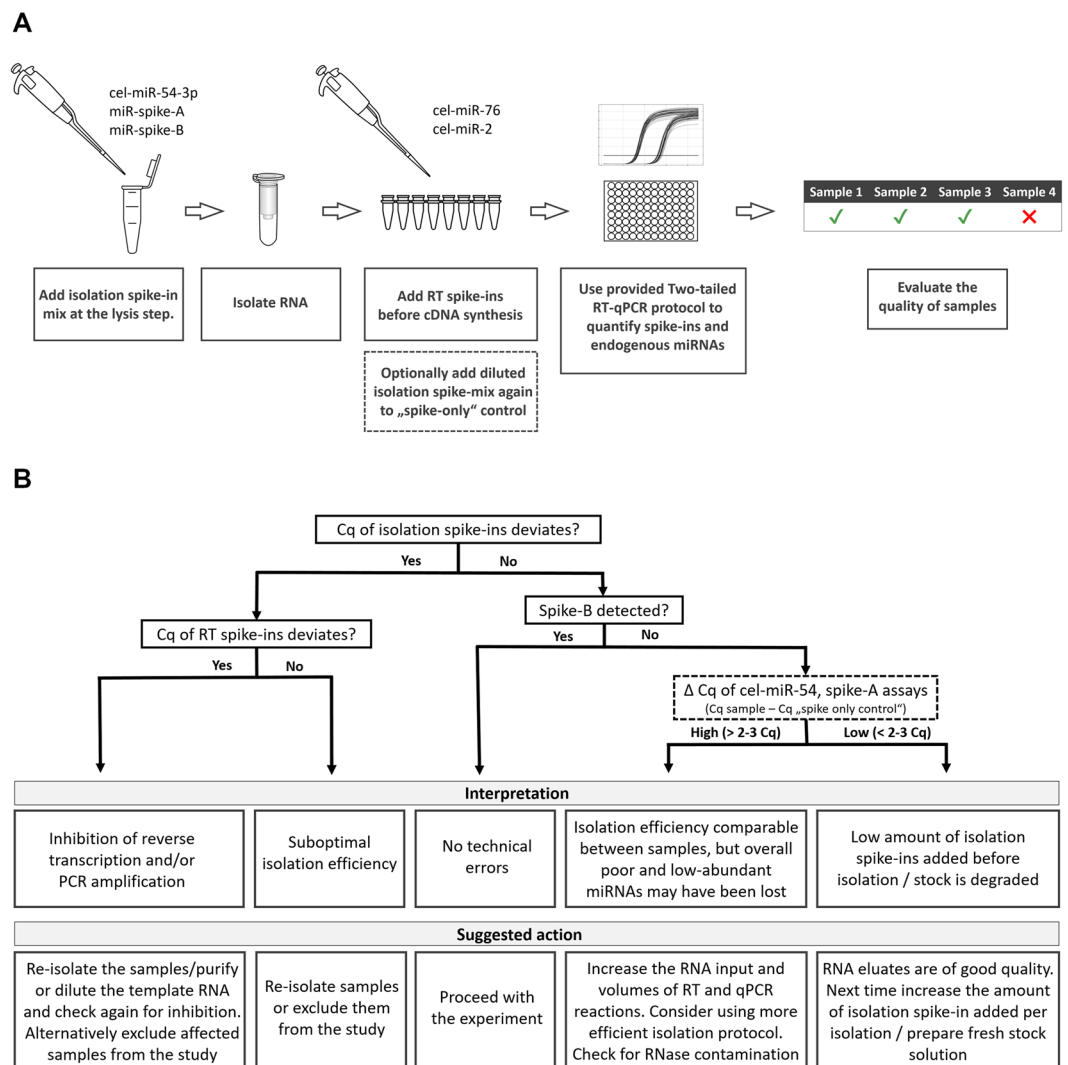


Figure 1. QC workflow with the Two-tailed QC panel. **(A)** A mix of three synthetic RNA spike-ins (cel-miR-54, spike-A, spike-B) is added prior to RNA isolation from the biofluid sample. A second mix of two spike-ins (cel-miR-76, cel-miR-2) is added before cDNA synthesis step. Optionally, a diluted isolation spike-in mix is used as a template in a “spike-only” control reaction to determine spike-in baseline signal (for details see Supplementary file section 3.2.2). Two-tailed RT-qPCR is used to quantify the spike-ins along with three endogenous microRNAs (let-7a, miR-23a and miR-451a) to evaluate the technical quality of RNA isolation, effect of inhibition and the level of haemolysis. **(B)** Decision chart for data interpretation and troubleshooting (see also Supplementary file section 4).

Usage	Name	Sequence	GC %	Origin
Isolation spike-ins	cel-miR-54-3p	/5Phos/UACCCGUAUCUUAUAUCCGAG	41.7	<i>C. elegans</i>
	miR-spike-A	/5Phos/UGCAGCCUACCGACACGUUCC	63.6	artificial
	miR-spike-B	/5Phos/ACUCAGGUUGUAGGAGCGGUCUU	52.2	artificial
RT spike-ins	cel-miR-76-3p	/5Phos/UUCGUUGUUGAUGAAGCCUUGA	40.9	<i>C. elegans</i>
	cel-miR-2-3p	/5Phos/UUUCACAGCCAGCUUUGAUGUGC	47.8	<i>C. elegans</i>

Table 1. Synthetic RNA spike-ins used in the Two-tailed QC panel.

downstream RT-qPCR analysis (Fig. 3). Based on these findings, we recommend addition of glycogen to increase the robustness and efficiency of microRNA isolation with the miRNeasy Serum/Plasma Advanced Kit (Qiagen).

Assessing the level of haemolysis in serum/plasma samples. A major complication in microRNA analysis of serum/plasma samples is contamination with microRNAs derived from lysed blood cells^{20,30,35} and in particular haemolysed erythrocytes. Plasma and serum samples should therefore be assessed for haemolysis.

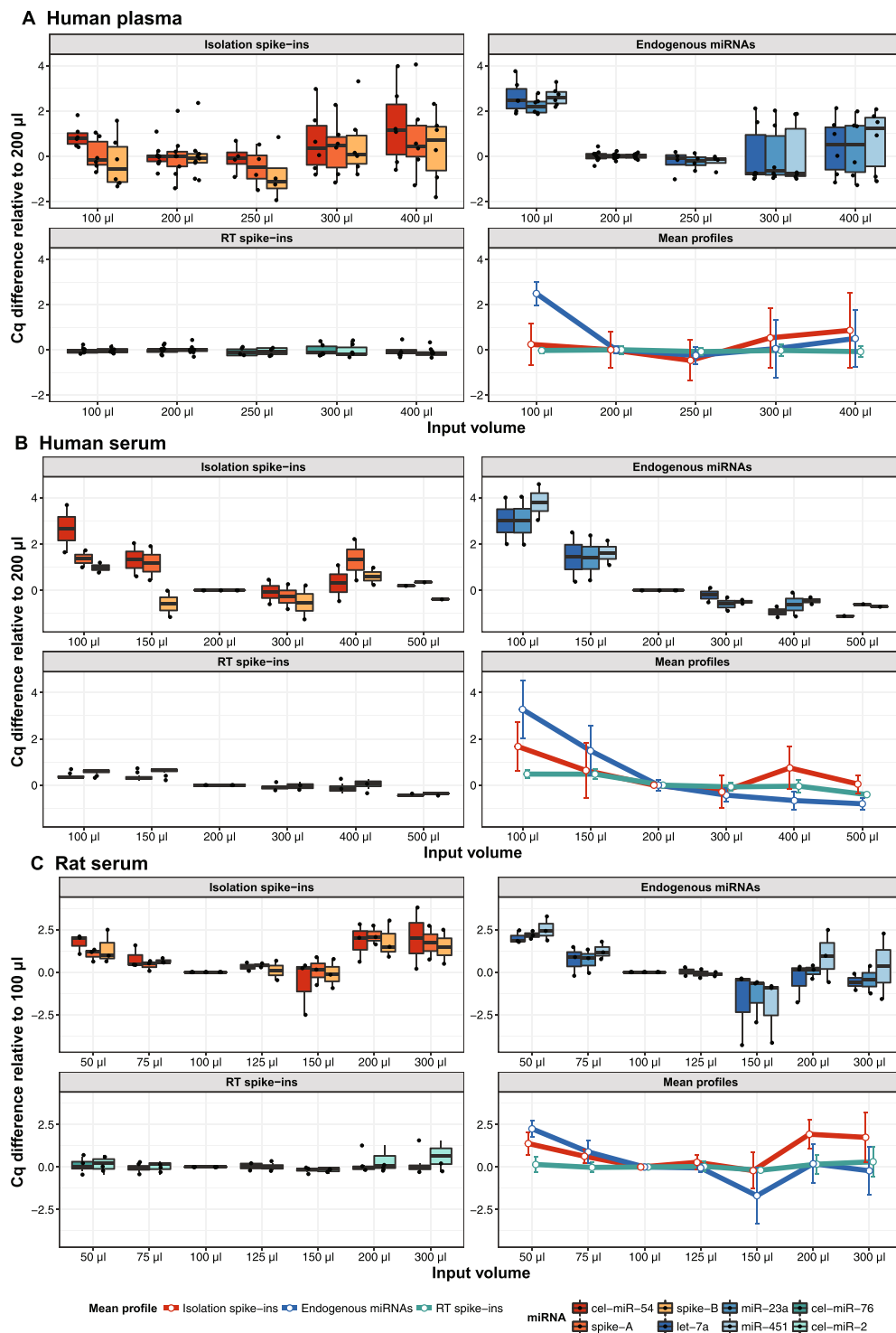


Figure 2. Optimizing input volumes of (A) human plasma, (B) human serum, and (C) rat serum for RNA isolation. Data are presented as ΔCq between Cqs obtained with the tested volume and an input volume of 200 μl (human) or 100 μl (rat). Each dot is one isolation replicate. Optimum starting serum/plasma volumes based on absolute endogenous microRNA yields are 250 μl for human plasma, 300–500 μl for human serum, and 150 μl for rat serum (blue mean profiles). Error bars on mean profiles panels indicate standard deviation (SD).

Standard method is to measure absorption at 414 nm, 540 nm and 578 nm, which are the absorption peaks of free oxyhemoglobin³⁶. An alternative approach, which is applicable also when the original sample is no longer available, is to measure the ratio of miR-23a, which is insensitive to haemolysis, and miR-451a, which is highly enriched in erythrocytes²⁰. Blondal *et al.*²⁰ established threshold ΔCq (miR-23a–miR-451a) values as quality indicators for

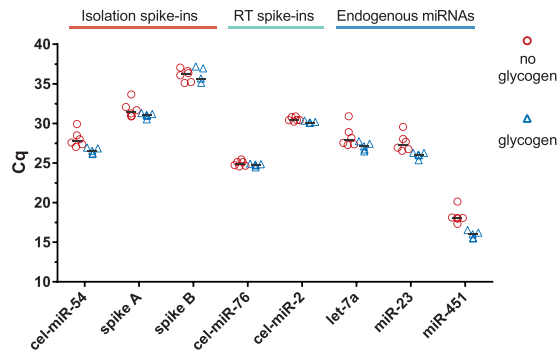


Figure 3. Effect of glycogen carrier on microRNA quantification in human plasma. Identical sample aliquots were isolated with (n = 5) or without (n = 6) addition of glycogen carrier starting from 200 μ l, and quantified with the Two-tailed QC panel. Extractions with glycogen had significantly higher yields (average difference between Cq means: 1.25 cycles; paired T-test $p = 0.011$) and higher reproducibility (F-test, $p < 0.001$).

human samples: $\Delta Cq > 5$ indicates there may be erythrocyte contamination, and $\Delta Cq > 7$ indicates high risk of haemolysis. A complication, however, is that the indicator is sensitive to the relative isolation yields of miR-23a and miR-451, as well as their relative RT yields and PCR efficiencies of the assays used to quantify them. Hence, the threshold ΔCq 's reported by Blondal *et al.*²⁰ are valid only for their particular workflow and protocol, and should not be used as general indicators. Here, we establish threshold ΔCq values for the Two-tailed QC panel and our recommended workflows (Supplementary file).

We prepared duplicate haemolysis dilution series for each sample type and constructed standard curves to correlate ΔCq (miR-23a – miR-451a) values to absorbance at 414 nm (Fig. 4B). To increase the number of data points, samples screened in other experiments with the same workflow were also included. Correlation between A_{414} and linear transformation of ΔCq ($2^{\Delta Cq}$) is significant for all three biofluids (Pearson $r \geq 0.80$, $p < 0.0001$). A_{414} for plasma sample 1 was outside the linear range of the absorption spectrophotometer and was estimated by interpolation. A_{540} and A_{578} nm dependences show the same trend, although those peaks are considerably less significant in the absorbance spectrum (Fig. 4A). $A_{414} \leq 0.2$ – 0.25 has previously been recommended as threshold for non-haemolysed samples^{13,20,35}. Based on our calibration this corresponds to a ΔCq of 15 cycles for human plasma, 11 cycles for human serum, and 6 cycles for rat serum for our workflow (Fig. 4B; Supplementary file).

Discussion

We present an RT-qPCR based protocol to assess the technical performance of workflows for analysis of microRNAs in body fluids such as serum and plasma. The QC panel developed is based on two sets of synthetic spike-in molecules and three endogenous microRNAs to assess RNA isolation yield, RT yield, PCR efficiency, and haemolysis (Fig. 1).

A highly error-prone step in microRNA analysis workflow is RNA isolation. Several studies have studied the effect of input volume on microRNA recovery, reporting varying results^{26,37,38}. However, consistent observations are substantial variability between replicate isolations and non-linear dependence of the input volume on the amount of microRNAs detected^{26,31}. Here, we studied the effect of input sample volume when extracting with the recently launched miRNeasy Serum/Plasma Advanced Kit from Qiagen and found that optimal input volume is different for the three sample types: human plasma, human serum, and rat serum (Fig. 2). We also found that higher input volumes ($>300 \mu$ l for human, $>150 \mu$ l for rat), although still in the range recommended by the manufacturer, lead to less reproducible Cq values compared to moderate input volumes (200–300 μ l for human, 100–150 μ l for rat). Using spike-in controls we showed this is due to inhibition of neither cDNA synthesis nor PCR, as suggested previously²⁰, but rather to impaired isolation efficiency, possibly because of saturation of the purification column. We confirm previous observations that adding a carrier improves extraction yield and reproducibility^{32–34,39}. We also show glycogen is a suitable alternative to RNA-based carriers when using the miRNeasy Serum/Plasma Advanced Kit (Qiagen) conferring advantage when samples shall be analysed with NGS (Fig. 3).

Another contribution to bias is microRNAs from leaking blood cells^{30,35}. While cellular contamination can be minimized by careful removal of the plasma fraction and dual centrifugation to efficiently remove platelets, haemolysis remains a problem. Haemolysis can occur during sampling and handling procedures and the released cellular microRNAs distort the measured microRNA profiles, which no longer reflect exclusively cell free microRNA^{30,35,36}. This not only hampers biological interpretation of the results, but can distort normalization or RT-qPCR data. For example, miR-16-5p is widely used as reference microRNA⁹, but it is also one of the most abundant microRNAs in erythrocytes³⁰ and its level is therefore perturbed even at low level of haemolysis^{35,40}.

Haemolysis can be assessed by visual inspection of the samples or, more precisely, spectroscopically. An alternative approach is to compare the levels of the erythrocyte-enriched miR-451a and the haemolysis-insensitive miR-23a²⁰. While visual inspection is rather subjective and not particularly sensitive, spectroscopic assessment and RT-qPCR quantification of miR-451a and miR-23a levels reveal even low degree of haemolysis³⁵. Shah *et al.*⁴¹, compared several methods to assess the level of haemolysis in human serum samples and found ΔCq (miR-23a – miR-451a) to be the most sensitive indicator⁴¹. In contrast, Vliet *et al.*¹³ reported that absorption measurement is more sensitive for rat plasma samples. Our results show that the approaches are comparable and correlate

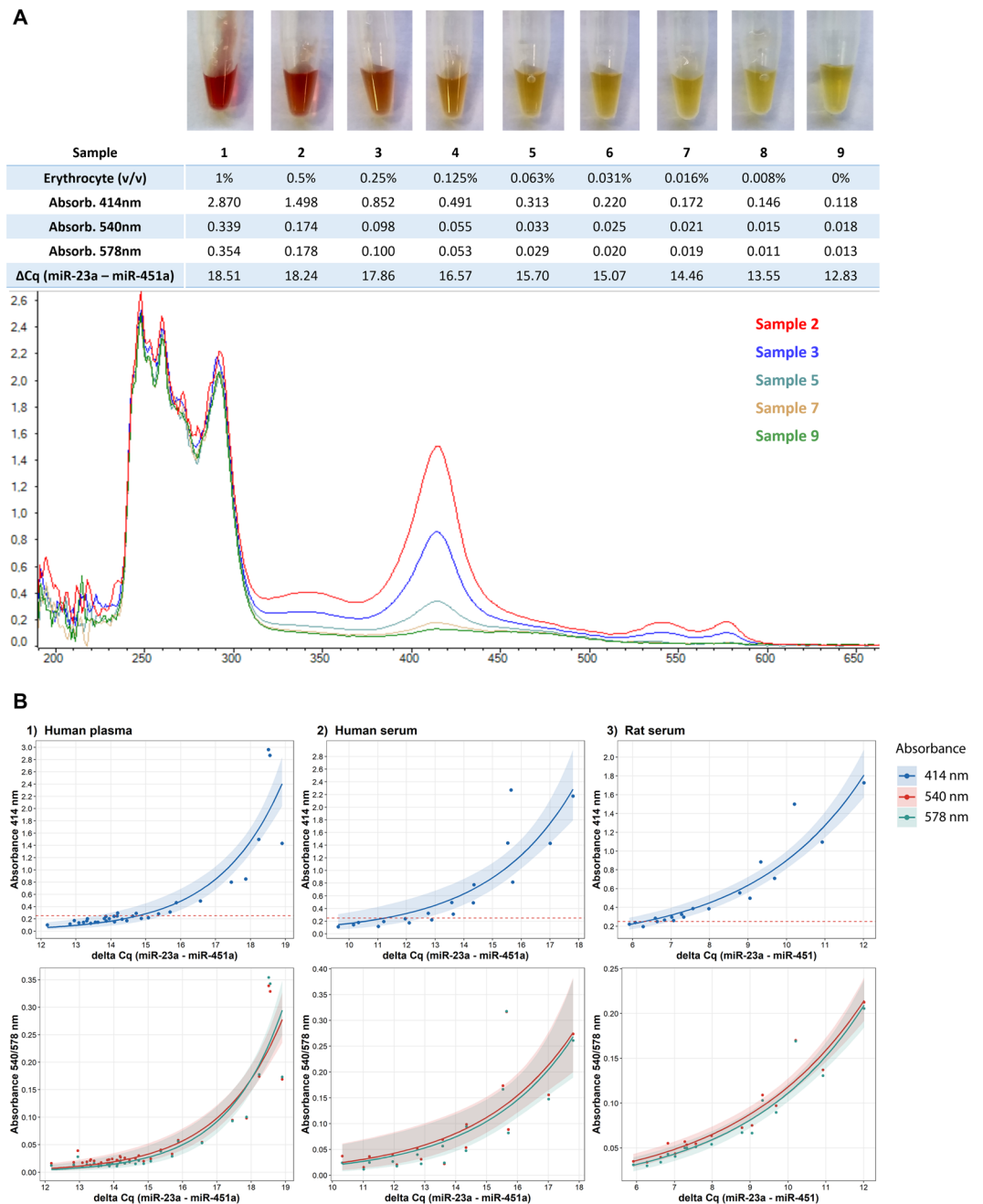


Figure 4. Assessing haemolysis in serum/plasma samples. **(A)** Human plasma samples with varying degree of haemolysis, corresponding A_{414} , ΔCq (miR-23a–miR-451a), and selected UV-Vis spectra. **(B)** Correlation between ΔCq (miR-23a–miR-451a) and A_{414} , A_{540} and A_{578} , respectively. Exponential regression line with 95% confidence interval is shown. Dashed red line indicates $A_{414} = 0.25$ as threshold for non-haemolysed samples. Corresponding ΔCq thresholds are ~ 15 (human plasma), ~ 11 (human serum), and ~ 6 (rat serum).

well even at very low levels of haemolysis for all three sample types we tested (Fig. 4). An advantage of the qPCR approach is that haemolysis can be assessed even when the original sample is no longer available. The same strategy can be used to assess contamination with other cell-types when needed. For example, miR-425 level may reflect contamination of platelets¹³. It is important to be aware that the ΔCq (miR-23a–miR-451) indicator must be calibrated for every new biofluid, isolation procedure and RT-qPCR method, as the ratio of the measured levels of miR-23a and miR-451 depend on the relative bias introduced by the methods used^{26,27,42}, but also the particular species and biofluids analysed. Indeed, in our study we concluded different threshold ΔCq values for the three sample types analysed. The ΔCq indicator should therefore be established for every workflow. Once calibrated the ΔCq indicator can be used to compare processed samples to identify outliers that should be reanalysed or discarded (see Supplementary file).

Despite several advances, circulating microRNA research has been hampered by inconsistency and poor reproducibility¹. The Two-tailed quality control panel developed here is a simple yet powerful tool for researchers to optimize new workflows, assess the technical performance of an analysis, identify outlier samples, and generally improve the reliability of circulating microRNA data.

Methods

Oligonucleotides. Sequences of mature microRNAs were obtained from the miRBase release 22 (www.mirbase.org). RNA oligonucleotides with 5'-phosphate were synthesized and quantified by Integrated DNA Technologies. Spike-in miRNA sequences were screened *in silico* for homology against human, mouse and rat miRBase records (Release 22) with the following parameters - search sequences: mature miRNAs, search method: SSEARCH, e-value cut-off: 100, max. no. of hits: 100. No significant homology was found. DNA oligonucleotides were synthesized and quantified by Invitrogen. Sequences are available in Supplementary file.

Samples. For the preparation of human serum, blood was collected from two healthy volunteers into 8.5 ml BD Vacutainer SST II Advance tubes (Beckman Dickinson) and allowed to clot for at least 30 min before centrifugation at 1500 g for 10 min at room temperature. The serum was then transferred to 2 ml tubes (Eppendorf) and stored at -80°C . For the preparation of human plasma, blood was collected from four healthy volunteers into K₂EDTA BD Vacutainer tubes (Beckman Dickinson) and centrifuged within 30 min at 1500 g for 15 min at room temperature. The plasma fraction was aspirated and transferred to 2 ml tubes (Eppendorf) and centrifuged again for 15 min at 3000 g. The supernatant was transferred to new 2 ml tubes and stored at -80°C until analysis. Informed consent was obtained from all volunteers participating in the study. All procedures involving the use of human samples were performed in accordance with the ethical standards of Institute of Experimental Medicine, Academy of Sciences of the Czech Republic and with the Declaration of Helsinki. All methods were approved by the Ethical committee of the Institute of Experimental Medicine (decision on 22 June 2018, approval number 04/2018). For the preparation of rat serum, animals were anesthetized using 2–4% isoflurane. One millilitre of blood was collected from orbital plexus into 2 ml tubes (Eppendorf) using glass capillary. Blood was allowed to clot for 1 hour at room temperature and then centrifuged at 1000 g for 10 min. The clot was mechanically retracted from the tube wall before the centrifugation. Serum was transferred to another 2 ml tube and centrifuged a second time at 3000 g for 10 min. The supernatant was then transferred to cryovials (Biologix) and stored at -80°C until analysis. All procedures involving the use of laboratory animals were performed in concordance with the European Community Council Directive of 24 November 1986 (86/609/EEC) and animal care guidelines approved by the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic (Animal Care Committee decision on 17 April 2009; approval number 85/2009).

Haemolysis dilution series. After whole-blood centrifugation, erythrocytes from the lower phase were collected into a separate tube and subjected to a freeze-thaw cycle followed by vigorous vortexing for at least 90 seconds to lyse the erythrocytes. The haemolysed test sample was prepared by adding 1% (v/v) of lysed erythrocytes into a non-haemolysed sample. A two-fold haemolysis dilution series was prepared by diluting the haemolysed sample sequentially with non-haemolysed sample. Dilution series from two subjects were prepared for each biofluid type (human serum, human plasma, and rat serum). Absorbance of free haemoglobin was measured at 414 nm, 540 nm, and 578 nm with a NanoDrop 2000 spectrophotometer (ThermoFisher) in duplicates. RNA was isolated from the serum and plasma samples as described below, starting with either 200 μl (human) or 150 μl (rat) input volume.

RNA isolation. Total RNA was isolated from human plasma, and human and rat serum samples using the miRNeasy Serum/Plasma Advanced Kit (Qiagen) according to the manufacturer's instructions. 1 μl of isolation spike-in mix containing synthetic cel-miR-54 (1e + 7 copies/ μl), spike-A (2e + 5 copies/ μl), spike-B (4e + 3 copies/ μl) and, when appropriate 1 μl of GlycoBlue Coprecipitant (15 mg/mL) (Invitrogen), per sample was added at the lysis step. RNA was eluted into 20 μl nuclease-free water and stored at -80°C .

Reverse transcription and quantitative PCR. Reverse transcription (RT) reactions were performed with the qScript flex cDNA kit (Quantabio) in a total reaction volume of 10 μl . One reaction contained 2 μl of template RNA, 1x buffer, mix of 0.05 μM Two-tailed RT primers, 1 μl of GSP enhancer and 0.5 μl of RT enzyme, and nuclease-free water up to 10 μl . RT reactions were incubated in a CFX 1000 thermocycler (Bio-Rad) for 45 min at 25 $^{\circ}\text{C}$, 5 min at 85 $^{\circ}\text{C}$ and then held at 4 $^{\circ}\text{C}$. Immediately after incubation, cDNA was diluted by addition of 50 μl nuclease-free water. Quantitative PCR (qPCR) was performed in a total volume of 10 μl . One reaction contained 1x SYBR Grandmaster Mix (Tataa Biocenter), forward and reverse primer (final concentration 0.4 μM), and 2 μl of diluted cDNA template (resulting in a final cDNA dilution of 15x). qPCR was performed in duplicates and incubated in a 384-well plate in a CFX 384 Real Time Detection System (Bio-Rad) at 95 $^{\circ}\text{C}$ for 30 s, 45 cycles of 95 $^{\circ}\text{C}$ for 5 s, and 60 $^{\circ}\text{C}$ for 15 s followed by melting-curve analysis.

Data Availability

Cq values were pre-processed with CFX Manager 3.1 (Bio-Rad). Missing values were replaced with maximum Cq per assay + 1 ($C_{q_{\text{max}}} + 1$). Paired two-tailed T-test was used to calculate significance of difference of mean Cq values between extractions with and without glycogen and F-test was used to calculate significance of difference of spread of replicates (Fig. 3). For the calculation of F-test, Cq values were transformed to achieve normal distribution as: $2^{\Delta} \Delta Cq$ ($Cq - C_{q_{\text{mean}}}$), where $C_{q_{\text{mean}}}$ represents mean Cq of particular assay.

References

- He, Y. *et al.* Current State of Circulating MicroRNAs as Cancer Biomarkers. *Clin Chem* **61**, 1138–1155, <https://doi.org/10.1373/clinchem.2015.241190> (2015).
- Anfossi, S., Babayan, A., Pantel, K. & Calin, G. A. Clinical utility of circulating non-coding RNAs - an update. *Nat Rev Clin Oncol* **15**, 541–563, <https://doi.org/10.1038/s41571-018-0035-x> (2018).
- Ojha, R., Nandani, R., Pandey, R. K., Mishra, A. & Prajapati, V. K. Emerging role of circulating microRNA in the diagnosis of human infectious diseases. *J Cell Physiol*, <https://doi.org/10.1002/jcp.27127> (2018).
- Blandford, S. N., Galloway, D. A. & Moore, C. S. The roles of extracellular vesicle microRNAs in the central nervous system. *Glia*, <https://doi.org/10.1002/glia.23445> (2018).
- Guay, C. & Regazzi, R. Circulating microRNAs as novel biomarkers for diabetes mellitus. *Nat Rev Endocrinol* **9**, 513–521, <https://doi.org/10.1038/nrendo.2013.86> (2013).
- Matsuzaki, J. & Ochiya, T. Circulating microRNAs and extracellular vesicles as potential cancer biomarkers: a systematic review. *Int J Clin Oncol* **22**, 413–420, <https://doi.org/10.1007/s10147-017-1104-3> (2017).
- Kreth, S., Hubner, M. & Hinske, L. C. MicroRNAs as Clinical Biomarkers and Therapeutic Tools in Perioperative Medicine. *Anesth Analg* **126**, 670–681, <https://doi.org/10.1213/ANE.0000000000002444> (2018).
- Lee, I., Baxter, D., Lee, M. Y., Scherler, K. & Wang, K. The Importance of Standardization on Analyzing Circulating RNA. *Mol Diagn Ther* **21**, 259–268, <https://doi.org/10.1007/s40291-016-0251-y> (2017).
- Witwer, K. W. Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin Chem* **61**, 56–63, <https://doi.org/10.1373/clinchem.2014.221341> (2015).
- Kirschner, M. B., van Zandwijk, N. & Reid, G. Cell-free microRNAs: potential biomarkers in need of standardized reporting. *Front Genet* **4**, 56, <https://doi.org/10.3389/fgene.2013.00056> (2013).
- Moldovan, L. *et al.* Methodological challenges in utilizing miRNAs as circulating biomarkers. *J Cell Mol Med* **18**, 371–390, <https://doi.org/10.1111/jcmm.12236> (2014).
- Chugh, P. & Dittmer, D. P. Potential pitfalls in microRNA profiling. *Wiley Interdiscip Rev RNA* **3**, 601–616, <https://doi.org/10.1002/wrna.1120> (2012).
- van Vliet, E. A. *et al.* Standardization procedure for plasma biomarker analysis in rat models of epileptogenesis: Focus on circulating microRNAs. *Epilepsia* **58**, 2013–2024, <https://doi.org/10.1111/epi.13915> (2017).
- Watson, A. K. & Witwer, K. W. Do platform-specific factors explain microRNA profiling disparities? *Clin Chem* **58**, 472–474; author reply 474–475, <https://doi.org/10.1373/clinchem.2011.175281> (2012).
- Tuck, M. K. *et al.* Standard Operating Procedures for Serum and Plasma Collection: Early Detection Research Network Consensus Statement Standard Operating Procedure Integration Working Group. *J Proteome Res* **8**, 113–117, <https://doi.org/10.1021/pr800545q> (2009).
- Ainsztein, A. M. *et al.* The NIH Extracellular RNA Communication Consortium. *J Extracell Vesicles* **4**, 27493, <https://doi.org/10.3402/jev.v4.27493> (2015).
- Laurent, L. C. & Alexander, R. Serum Collection Procedure (Small Scale) for the analysis of extracellular RNA (2015).
- Laurent, L. C. & Alexander, R. Plasma Collection Procedure (Small Scale) for the analysis of extracellular RNA (2015).
- McDonald, J. S., Milosevic, D., Reddi, H. V., Grebe, S. K. & Algeciras-Schimmich, A. Analysis of circulating microRNA: preanalytical and analytical challenges. *Clin Chem* **57**, 833–840, <https://doi.org/10.1373/clinchem.2010.157198> (2011).
- Blondal, T. *et al.* Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* **59**, S1–6, <https://doi.org/10.1016/j.jymeth.2012.09.015> (2013).
- Marzi, M. J. *et al.* Optimization and Standardization of Circulating MicroRNA Detection for Clinical Application: The miR-Test Case. *Clin Chem* **62**, 743–754, <https://doi.org/10.1373/clinchem.2015.251942> (2016).
- Li, Y. & Kowdley, K. V. Method for microRNA isolation from clinical serum samples. *Anal Biochem* **431**, 69–75, <https://doi.org/10.1016/j.ab.2012.09.007> (2012).
- Mitchell, P. S. *et al.* Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci USA* **105**, 10513–10518, <https://doi.org/10.1073/pnas.0804549105> (2008).
- Buschmann, D. *et al.* Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow. *Nucleic Acids Res* **44**, 5995–6018, <https://doi.org/10.1093/nar/gkw545> (2016).
- Androvic, P., Valihrach, L., Elling, J., Sjoback, R. & Kubista, M. Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Res* **45**, e144, <https://doi.org/10.1093/nar/gkx588> (2017).
- El-Khoury, V., Pierson, S., Kaoma, T., Bernardin, F. & Berchem, G. Assessing cellular and circulating miRNA recovery: the impact of the RNA isolation method and the quantity of input material. *Sci Rep* **6**, 19529, <https://doi.org/10.1038/srep19529> (2016).
- Li, X., Mauro, M. & Williams, Z. Comparison of plasma extracellular RNA isolation kits reveals kit-dependent biases. *Biotechniques* **59**, 13–17, <https://doi.org/10.2144/000114306> (2015).
- Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin Chem* **56**, 1733–1741, <https://doi.org/10.1373/clinchem.2010.147405> (2010).
- Max, K. E. A. *et al.* Human plasma and serum extracellular small RNA reference profiles and their clinical utility. *Proc Natl Acad Sci USA* **115**, E5334–E5343, <https://doi.org/10.1073/pnas.1714397115> (2018).
- Pritchard, C. C. *et al.* Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)* **5**, 492–497, <https://doi.org/10.1158/1940-6207.CAPR-11-0370> (2012).
- Brunet-Vega, A. *et al.* Variability in microRNA recovery from plasma: Comparison of five commercial kits. *Anal Biochem* **488**, 28–35, <https://doi.org/10.1016/j.ab.2015.07.018> (2015).
- McAlexander, M. A., Phillips, M. J. & Witwer, K. W. Comparison of Methods for miRNA Extraction from Plasma and Quantitative Recovery of RNA from Cerebrospinal Fluid. *Front Genet* **4**, 83, <https://doi.org/10.3389/fgene.2013.00083> (2013).
- Ramon-Nunez, L. A. *et al.* Comparison of protocols and RNA carriers for plasma miRNA isolation. Unraveling RNA carrier influence on miRNA isolation. *Plos One* **12**, e0187005, <https://doi.org/10.1371/journal.pone.0187005> (2017).
- Andreasen, D. *et al.* Improved microRNA quantification in total RNA from clinical samples. *Methods* **50**, S6–9, <https://doi.org/10.1016/j.jymeth.2010.01.006> (2010).
- Kirschner, M. B. *et al.* Haemolysis during sample preparation alters microRNA content of plasma. *Plos One* **6**, e24145, <https://doi.org/10.1371/journal.pone.0024145> (2011).
- Kirschner, M. B. *et al.* The Impact of Hemolysis on Cell-Free microRNA Biomarkers. *Front Genet* **4**, 94, <https://doi.org/10.3389/fgene.2013.00094> (2013).
- Kim, D. J. *et al.* Plasma components affect accuracy of circulating cancer-related microRNA quantitation. *J Mol Diagn* **14**, 71–80, <https://doi.org/10.1016/j.jmoldx.2011.09.002> (2012).
- Sourvinou, I. S., Markou, A. & Lianidou, E. S. Quantification of circulating miRNAs in plasma: effect of preanalytical and analytical parameters on their isolation and stability. *J Mol Diagn* **15**, 827–834, <https://doi.org/10.1016/j.jmoldx.2013.07.005> (2013).
- Duy, J., Koehler, J. W., Honko, A. N. & Minogue, T. D. Optimized microRNA purification from TRIzol-treated plasma. *BMC Genomics* **16**, 95, <https://doi.org/10.1186/s12864-015-1299-5> (2015).
- Pizzamiglio, S. *et al.* A methodological procedure for evaluating the impact of hemolysis on circulating microRNAs. *Oncol Lett* **13**, 315–320, <https://doi.org/10.3892/ol.2016.5452> (2017).

41. Shah, J. S., Soon, P. S. & Marsh, D. J. Comparison of Methodologies to Detect Low Levels of Hemolysis in Serum for Accurate Assessment of Serum microRNAs. *Plos One* **11**, e0153200, <https://doi.org/10.1371/journal.pone.0153200> (2016).
42. Kim, Y. K., Yeo, J., Kim, B., Ha, M. & Kim, V. N. Short structured RNAs with low GC content are selectively lost during extraction from a small number of cells. *Mol Cell* **46**, 893–895, <https://doi.org/10.1016/j.molcel.2012.05.036> (2012).

Acknowledgements

This study was supported by following grants: Czech Science Foundation P303/18/21942S; RVO 86652036; BIOCEV CZ.1.05/1.1.00/02.0109; EU H2020 SC1-HCO-02-2016 “Standardisation of pre-analytical and analytical procedures for *in vitro* diagnostics in personalised medicine”; SPIDIA4P and IMI call 11 “Blood-based biomarker assays for personalised tumour therapy: value of latest circulating biomarkers”; CANCER-ID.

Author Contributions

P.A. and L.V. conceived and designed the study. P.A. performed experiments, analysed the data, prepared figures and drafted the manuscript. N.R. and L.U.-M. provided and processed animal samples. E.R., L.V. and M.K. commented on the manuscript and produced the final version. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-019-40513-w>.

Competing Interests: M.K. has shares in TATAA Biocenter AB. P.A., N.R., L.U.-M., E.R. and L.V. declare no conflict of interest.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019

Performance Comparison of Small RNA-Seq Library Preparation Methods for Biofluid Samples

Peter Androvič^{1,2,✉*}, Šárka Benešová^{1,3,*}, Mikael Kubista^{1,4}, and Lukáš Valihrač¹

¹Laboratory of Gene Expression, Institute of Biotechnology CAS, BIOCEV, 252 50 Vestec, Czech Republic

²Laboratory of Growth Regulators, Faculty of Science, Palacky University, 783 71, Olomouc, Czech Republic

³Laboratory of Informatics and Chemistry, Faculty of Chemical Technology, University of Chemistry and Technology, 166 28 Prague, Czech Republic

⁴TATAA Biocenter AB, 411 03, Gothenburg, Sweden

Circulating microRNAs (miRNAs) represent a new group of promising biomarkers. Recently, small RNA-sequencing (RNA-seq) has been introduced for quantification of circulating miRNAs allowing their comprehensive analysis without need for prior knowledge of target sequences. Despite great promises, the miRNA profiling by RNA-seq has not delivered the expected outcomes, particularly for high level of biases inheritably related to the workflow. The issue has been addressed by several new approaches trying to minimize bias and provide more efficient workflow. Here, we compare the performance of all currently available approaches for RNA-seq based miRNA analysis in biofluids using a complex set of parameters. The protocols include traditional two-adapters ligation methods as well as methods developed to reduce bias via application of randomized adapters, polyadenylation, circularization or unique molecular identifiers (UMI). The new methods over-performed the traditional ones in the majority of parameters confirming their superiority. Among them, the technology based on randomized adapters showed the most balanced performance together with the protocol using UMI correction. On the contrary, the polyadenylation generated large proportion of wasted reads and hampered the analysis of isomiRs. The circularization approach failed in the reduction of ligation bias. To sum up, we provide the most comprehensive comparison of current RNA-seq based protocols for analysis of circulating miRNAs up to date. The study may be used as a guide for new users of the technology as well as a reference for further comparative studies or for developing of new technologies.

RNA-seq | microRNA | liquid biopsy | biofluid

Correspondence: peter.androvic@ibt.cas.cz

* Equal contribution

Introduction

MicroRNAs (miRNA) are class of short non-coding small RNAs playing a role in post transcriptional regulation of gene expression (1, 2). They repress translation or initiate mRNA degradation via complementary pairing with mRNA (3, 4) and thus influence various biological processes (5, 6). MiRNAs are primarily localized within the cell but they may be also found in various biofluids protected in extracellular vesicles or in complexes with RNA-binding proteins (7, 8). Due to the stability of extracellular miRNAs and easiness

of biofluids sampling, they have started to be studied as a new promising group of biomarkers (9, 10). Association of specific miRNAs with a disease or specific biological process requires precise and accurate quantification. Several approaches are applied for miRNA quantification in biofluids, including microarray, reverse transcription quantitative PCR (RT-qPCR) or small RNA-sequencing (RNA-seq). Small RNA-seq stands out as the most comprehensive method allowing for discovery of novel miRNAs, quantification of isomiRs or other classes of small RNA without need of probe designs and previous sequence information (11). As miRNAs are of short length, the first step in a standard RNA-seq workflow involves their extension by ligation of a pair of adapters. Unfortunately, the ligation reaction tend to incorporate bias in miRNA quantification as it may favor miRNAs with preferred sequences or secondary structures. This unequal ligation efficiency causes under- or over- representation of some miRNAs and is considered to be the main source of bias in small RNA-seq (12, 13). Another source of bias is introduced in the amplification step that may similarly to ligation bias prefer certain miRNAs. Although this source of bias has been considered negligible in small RNA-Seq workflows due to similar length of cDNA molecules, recent studies show its effect and efficient reduction of bias using unique molecular identifiers (UMIs) (14–16). There are ongoing efforts to overcome the ligation bias to make small RNA-seq more accurate and quantitative. Inclusion of randomized adapters has been shown to significantly improve sequence specific bias (12, 17). Another approach omits the 5' end adapter ligation and uses only 3' end ligation with subsequent circularization (18). The ligation step is completely omitted with polyadenylation-based protocols. However, the polyadenylation complicates the recognition of native adenine at the 3' end of miRNAs, thus impairs analysis of isomiRs (19). Moreover, there are also targeted RNA-seq approaches, such as as EdgeSeq (HTG Molecular Diagnostics) that also completely avoids ligation. Instead, EdgeSeq relies on large pool of capture probes that are consequently amplified and sequenced (20, 21). In the past years, there have been several efforts to compare different protocols for small RNA-seq (17, 22–24). Although they provided a valuable insight into the performance of selected protocols, a comprehensive benchmarking of all currently avail-

able methods for miRNA analysis in biofluids is still missing. In this study, we compared six commercially available protocols for small RNA-seq library preparation in terms of their yield, sensitivity, accuracy, precision, isomiR detection and correlation with RT-qPCR (Fig.1, Tab.1). The protocols utilized five different approaches to prepare libraries and minimize bias. We tested methods based on ligation of two adapters, single adapter ligation and subsequent circularization, randomized adapters, polyadenylation with template switching, and UMI correction. In order to assess various aspects of library preparation methods, we used two standardized types of samples, RNA isolated from human plasma and miRXplore Universal Reference containing equimolar mix of 962 miRNAs from human, mouse, rat and virus. The data we present may guide the informed selection of appropriate protocol for analysis of circulating miRNAs.

Results

In this study, we compared the performance of six different commercially available protocols for small RNA-seq library preparation using two standardized types of samples: human plasma and equimolar pool of miRNAs (Fig.1). The selected protocols utilized five different approaches to prevent the ligation and amplification bias (Fig.1), which are integral parts of inherent bias in small RNA-seq data. Protocols were evaluated in terms of their yield, isomiR detection, accuracy, precision, sensitivity of miRNAs detection and correlation with RT-qPCR data. For each metric, one or more parameters were selected as representatives for final evaluation.

Yield and IsomiR Detection. The first step in the data analysis is the mapping of raw sequencing reads to respective references. The percentage of successfully mapped reads (yield) provides the overall picture how efficiently a library preparation protocol captures molecules of interests and what proportion of reads is wasted. The information about protocol-specific yield helps to select appropriate sequencing depth and optimize the total cost of experiment. Mapping statistics of miRXplore, equimolar pool of miRNAs, showed consistently high mapping rate for all tested protocols ranging from 70 to 93% (Fig.2A). The best yield was reached by QIAseq and NEXTflex protocols which also had the highest proportion of reads that were not mapped to overrepresented miRNAs (top 10) or false isomiRs. Interestingly, most of the reads discarded in QIAseq and NEXTflex mapped to rRNA and UniVec (database of vector contamination occurring in NGS), while for the rest of the protocols, the main reason was the inappropriate length (Fig.2 C). As the miRXplore sample is composed of defined miRNA sequences and does not contain isomiRs, the mapping procedure in miRXplore sample allowed us to identify the rate of false isomiR detection, that are presumably introduced by protocol as a technical artefact. The highest level of false isomiRs was detected in SMARTer polyadenylation protocol, where they consumed up to 15% of all reads. The high incidence was caused by the technology of polyadenylation and template switching as shown by high frequency

of adenosines on 3' end of false isomiRs (70%) and cytosines and guanines (50, 35%) on 5' end of false isomiRs (Additional file 3, Fig.1). In other libraries the level of false isomiRs was low and contributed by 0.5% to 3% to mapping statistics (Additional file 2). Completely different results were observed in the plasma samples where additional factors such as various miRNA concentrations and presence of other RNA species play an important role. The overall mapping results showed that percentage of reads uniquely mapped to miRNAs was lower than in miRXplore for all protocols (Fig.2B). The highest percentage of discarded and unmapped reads was produced by SMARTer (70% and 23%, Fig.2B). Most of the SMARTer reads were discarded because of unsuitable length for miRNA which was probably caused by polyadenylation of fragments of other RNA species. SMARTer also showed most diverse mapping to various classes of small RNA (Fig.2E, Additional file 3, Fig.2). The highest percentage of mapped reads was detected by Norgen and RealSeq as both protocols showed considerably high percentage of detected isomiRs than other protocols (more than 15% compared to 1-7%). However, most of the detected isomiRs belonged to the most abundant miRNA detected by particular protocol (hsa-miR-451a in RealSeq and hsa-miR-486-5p in Norgen) which suggests that such high mapping ratio to isomiRs does not represent their truly higher detection rate and consequently wider spectrum of isomiRs captured (Additional file 2). Overall, the highest proportion of reads uniquely mapped to canonical miRNAs was reached by QIAseq (70%), closely followed by Norgen, RealSeq and NEXTflex. However, all these libraries resulted also in considerable proportion of reads mapped to ten most detected miRNAs (discussed below). As the true composition of the plasma sample is unknown, we can only hypothesize if higher mapping rate reflects high levels of particular miRNAs or the level of ligation bias. Considering this effect, the percentages of mapped reads excluding top 10 miRNAs may provide more relevant metrics to compare the protocols. In this comparison, the highest proportion of reads were captured by QIAseq and NEXTflex (5-6%) which is in accordance with the data measured in miRXplore sample (Fig.2A). Statistics of discarded reads showed balanced ratio of reads not passing length criteria and mapping to rRNA and UniVec database (Fig.2D). To sum up, QIAseq and NEXTflex protocols showed the best yield in both types of samples. The rest of the protocols showed comparable results, except SMARTer having considerably lower yield in both samples. On the other hand, the mapping statistics showed that SMARTer had the lowest proportion of reads mapping to the ten most abundant miRNAs which in the case of miRXplore sample indicates the lower level of ligation or PCR bias.

Accuracy and Precision. Precision is another important aspect which must be considered when choosing suitable protocol. High precision guarantees low deviation between technical replicates, high reproducibility of results and allows for inter-sample comparison. Here, we calculated Pearson correlation coefficient between technical replicates for both, miRXplore and plasma samples as a parameter reflect-

Table 1. **Evaluated small RNA-seq methods**

Information based on protocol manuals or webpages of commercial suppliers. Approximate price per reaction is without taxes and it was calculated from protocol with the highest possible number of reactions available for purchase. * Standardized volume used in this study was 5µl.

Commercial Supplier	Name	Name used in this study	Input RNA	Maximum input volume *	No. pipeting steps	Incubation time (hours)	Catalogue number	Approximate price per sample (no. reactions per protocol)
Bio Scientific Corp.	NEXTflex Small RNA-Seq Library Prep Kit v3	NEXTflex	1 ng - 2 µg	10.5 µl	55	6:00	5132-05	\$56 (48)
Lexogen	Small RNA-Seq Library Prep Kit	Lexogen	0.5 ng - 1 µg	6 µl	32	2:50	052	\$50 (96)
Norgen	Small RNA Library Prep Kit for Illumina	Norgen	50 ng - 500 ng	6 µl	25	3:40	63600	\$65 (24)
Qiagen	QIAseq miRNA Library Kit	QIAseq	1-500 ng	5 µl	46	5:30	331505	\$73 (96)
Somagenics	RealSeq Biofluids Plasma/Serum miRNA	RealSeq	0.2 ng-20 ng	9.5 µl	26	5:40	600-00048	\$71 (48)
Takara	SMARTer smRNA-Seq kit for Illumina	SMARTer	1 ng - 2 µg	7 µl	36	1:40	635031	\$77 (96)

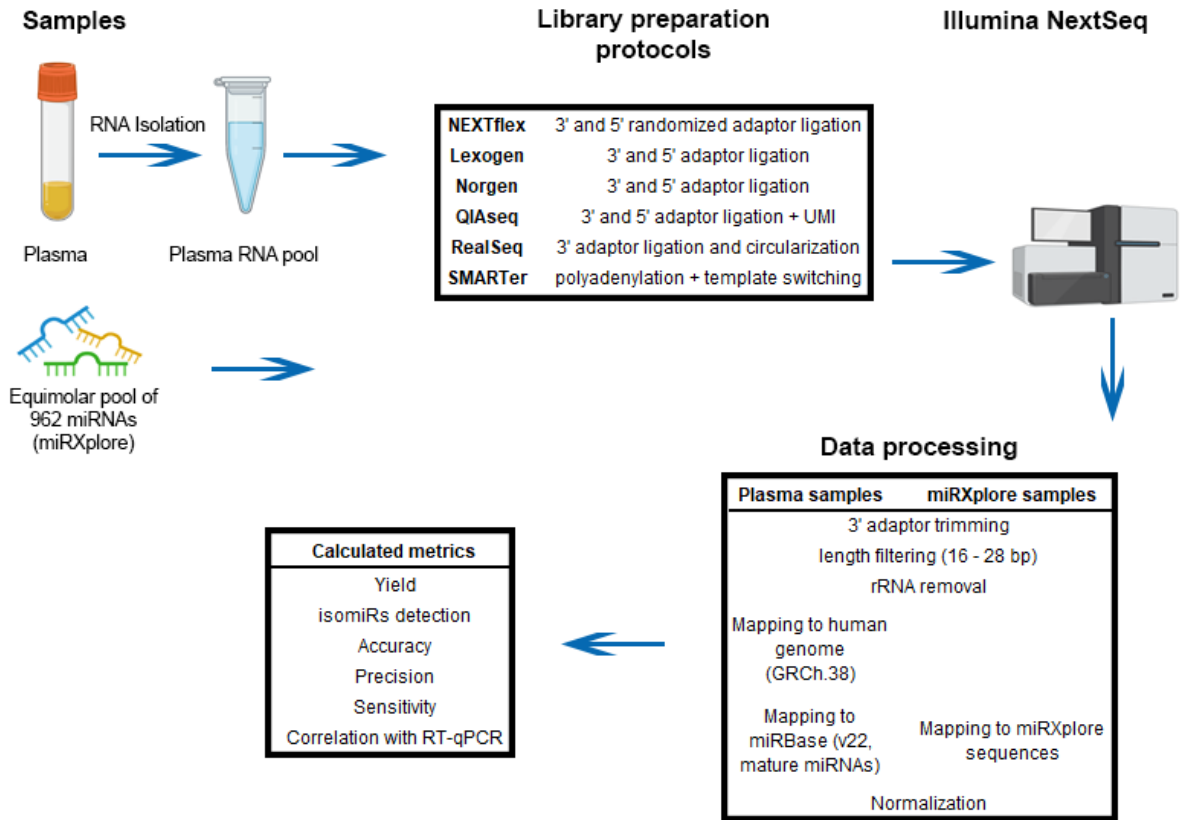


Fig. 1. **Workflow**

Experimental design. Two types of samples were used for comparison of selected approaches – human plasma and miRxplore universal reference. After RNA isolation, six different protocols were used for library preparation and resulting libraries were sequenced on Illumina NextSeq platform. Data were trimmed, filtered according to length and further mapped to the respective reference sequences. After normalization, various metrics were calculated.

ing the precision of measurement. The precision of measurement was very high for all protocols ($r > 0.98$, Additional file 3, Fig.5) which is in agreement with current knowledge considering RNA-Seq as highly reproducible type of analysis (25). Accuracy of measurement indicates how close are the measured values from the true values. Contrary to the plasma samples, the concentration of miRNAs in miRxplore sample are known, therefore allowed for the evaluation of each protocol accuracy. As the first parameter describing the accuracy, we used coefficient of variation (CV) expressing how measured expression varies between individual miRNAs. As the concentration of miRNAs in miRxplore is identical, the expected CV for a highly accurate protocol is zero. Contrary to the high precision of all protocols, the accuracy

of expression measurement was relatively poor. The lowest CVs (< 1) were shown by QIAseq after removal of PCR bias using UMI correction (“QIAseq_UMI” in Fig.3A) and by SMARTer, indicating the best accuracy of expression measurement. Other protocols showed higher CVs, with the highest values achieved in the protocols based on subsequent ligation of two adaptors (Lexogen, Norgen). Another parameter assessing the accuracy level was the distribution of fold-difference between measured and predicted expression values (Fig.3A). The visualization shows the overall pattern how each protocol overestimates or underestimates the expression of individual miRNAs. Unimodal distribution resembling Gauss distribution with zero mean indicates better accuracy than multimodal and skewed

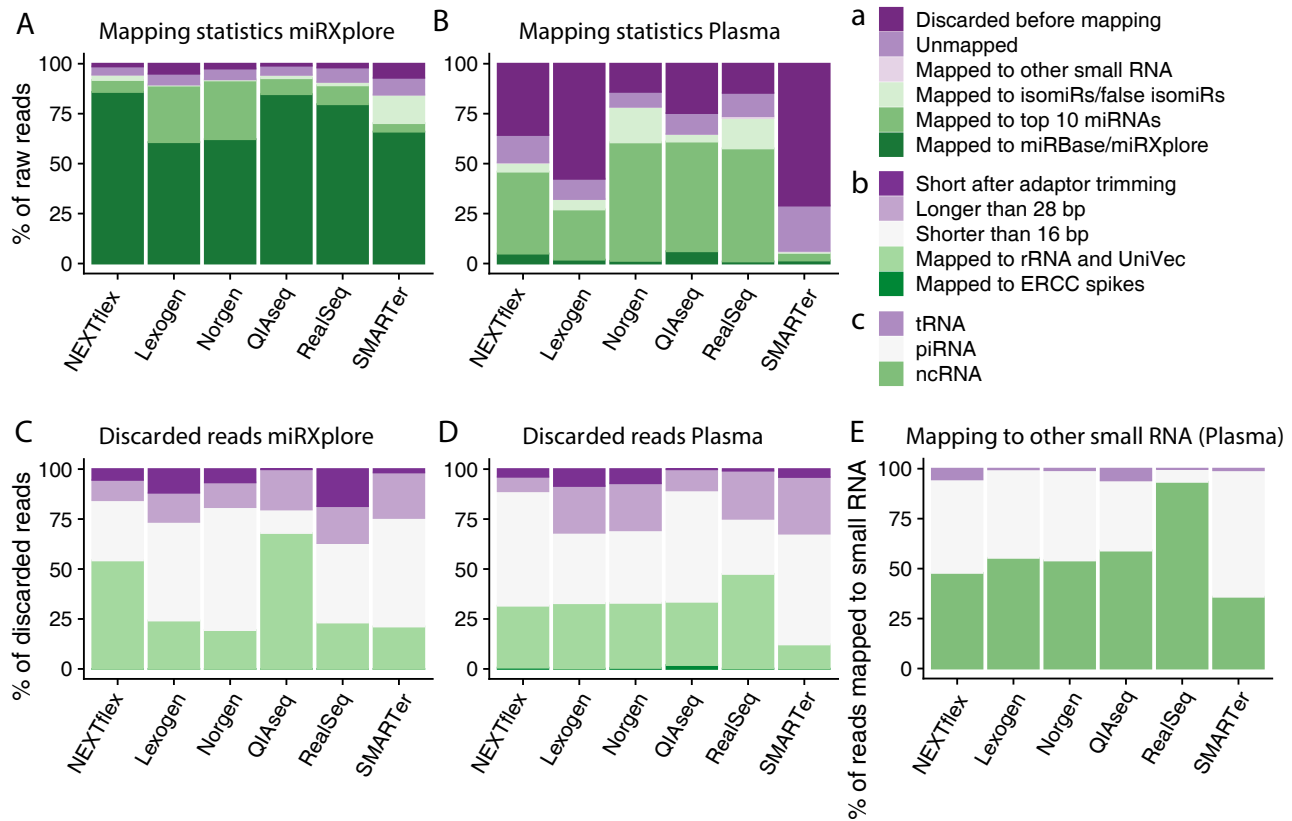


Fig. 2. Mapping Statistics

A, B, percent distribution of raw reads; Mapped reads, raw reads uniquely mapping to relevant reference; Unmapped reads, non-uniquely mapping reads and reads which did not map to any of used references; C, D, percent distribution of raw reads discarded before mapping; Short after adaptor trimming, reads containing only adapter sequences; Longer than 28 bp and Shorter than 16 bp; long or short reads after adaptor trimming; E, percent distribution of raw reads mapped to other small RNA than miRNA; ncRNA, RNA mapping to ensemble reference of non-coding RNA; a, legend to mapping statistics of miRXplore and plasma; b, legend to discarded reads of miRXplore and plasma; c, legend to mapping of other small RNA.

distribution that originates predominantly from ligation and PCR bias. Constantly with CVs, the highest level of accuracy was achieved by SMARTer protocol measuring more than 50% of miRNAs without bias (within two-fold change). The medium performance showed protocols employing approaches for bias reduction (NEXTflex, QIAseq) whereas traditional two-adapters ligation methods resulted in the highest proportion of miRNAs outside the optimal range (up to 86% for Norgen and Lexogen). Surprisingly, RealSeq protocol based on circularization displayed rather poor performance and traditional two-adapters based QIAseq protocol without UMI correction were similarly accurate as the NEXTflex method. Although SMARTer does not employ ligation at all and the distribution was closest to Gauss distribution, there was still substantial number of miRNAs over- or underestimated. To assess the contribution of ligation and PCR bias separately, we applied linear mixed model on QIAseq data and quantified the contribution of both type of biases (Fig. 3B). 518 miRNAs showed more than 75% contribution of ligation bias. However, over quarter (227) of miRNAs displayed more than 50% contribution of PCR bias, which indicates that PCR is also significant contributor to overall bias. As was seen from the distribution of ligation bias in Figure 3, the majority of the measured miRNAs were underestimated which suggested high consumption of reads by

small number of preferentially ligated miRNAs. The consumption of reads by individual miRNAs may be visualized as a dependence of cumulative frequency on number of miRNAs ordered according to increasing number of their counts (Fig. 4). Fast increase in cumulative frequency indicates that even miRNAs with the lowest abundance contribute significantly to the total number of counts per sample. Number of miRNAs at cumulative frequency of 50% (CF50) was selected as a metric for comparison on miRXplore samples. Ideally, CF50 value would be around 481 (half of 962 miRNAs consumes half of the reads) and the lower the number is the better. The best distribution of reads between all miRNA was shown by QIAseq with UMI correction and SMARTer (Fig. 4A), whereas Norgen and Lexogen resulted in highest CF50 indicating that the majority of miRNAs consumed only 50% of all reads and few miRNAs consumed the rest.

Curve of cumulative frequency calculated on plasma samples had considerably different progress than in miRXplore (Fig. 4B). Since the true concentration of miRNA in plasma was much less balanced than in miRXplore, the low abundant miRNAs consumed substantially less reads than the most abundant miRNA. Therefore, the number of miRNAs at cumulative frequency 1% (CF1) was selected as a metric for comparison (Fig. 4B). The cumulative frequency was plotted in \log_{10} scale in order to better visualize the differences be-

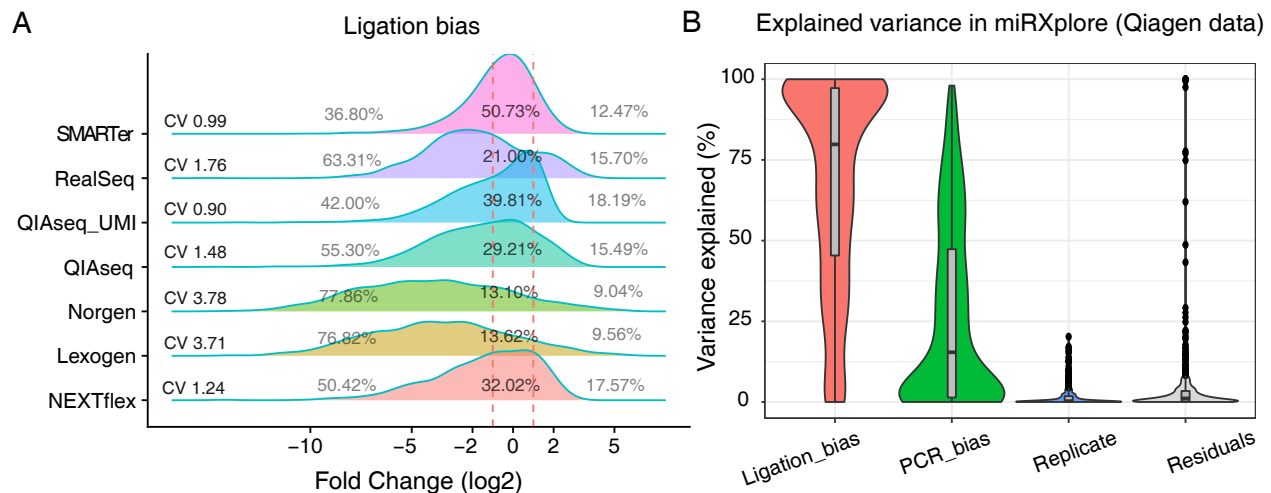


Fig. 3. **Ligation Bias**

A, Ligation bias for each small-RNA library preparation protocol was calculated on miRXplore samples for each miRNA as fold-deviation from expected value and was plotted as \log_2 . Dashed lines correspond to two-fold deviation from the theoretical expected value. Percentage of biased and unbiased miRNAs and coefficient of variation (CV) are shown for each protocol. B, Percentage of variance in miRXplore samples processed with QIAseq protocol explained by ligation bias and PCR bias. Replicate indicates percentage of variance explained by technical replicates. Residuals indicated percentage of variance unexplained by replicates or biases.

tween protocols. In agreement with percentage of unbiased miRNAs, SMARTer and QIAseq after UMI correction showed the best performance, whereas Norgen and RealSeq showed the worst performance.

Cumulative frequency is strongly affected by the percentage of reads consumed by the most abundant miRNAs. This was especially pronounced in the plasma, where the top 10 miRNAs consumed the majority of all reads mapped to miRNAs (Fig. 2B). This may reflect true biological abundance, but it might be also a consequence of the ligation and PCR biases (Fig. 3A). To visualize and potentially distinguish true biological abundance from the bias, we plotted ten most abundant miRNAs in plasma with the corresponding bias measured in miRXplore sample (Fig. 5). Surprisingly, in each protocol, except SMARTer and NEXTflex, there was always a single miRNA consuming more than 50% of all mapped reads. Moreover, the most abundant miRNA was not the identical in all protocols. In SMARTer, NEXTflex and RealSeq libraries hsa-miR-451a was the most abundant, whereas in Norgen and Lexogen samples the dominant miRNA was hsa-miR-486-5p. The inspection of \log_2 fold changes predicted based on miRXplore suggests that the relative abundance of hsa-miR-486-5p together with hsa-miR-10b-5p in Norgen and Lexogen is presumably strongly overestimated (up to 64x) and data are therefore distorted. On the other hand, hsa-miR-451a reflected rather true or underestimated biological concentration in the sample as its \log_2 fold change showed mostly negative values. The only exception was the RealSeq protocol. The observed high proportion of hsa miR-451a in this dataset (over 90% of all miRNA reads) may be therefore attributed to relatively higher ligation efficiency compared to other protocols. To sum up, although the precision of tested protocols was high, we observed dramatic differences in accuracy affecting many parameters of measurement. The accuracy was influenced mainly by the ligation bias, although the contribution of the PCR bias was not neg-

ligible. Omitting the ligation step via polyadenylation or the application of randomized adapters successfully reduced the ligation bias as well as the UMI correction diminished the bias originating from PCR step.

Sensitivity of miRNAs Detection. The important factor that needs to be consider is a number of detected miRNAs at a given sequencing depth (further called sensitivity). This parameter helps a researcher to predict the number of miRNAs that will be available for the analysis as well as optimize overall experimental design with respect to sequencing cost. To visualize the dependency of sensitivity on sequencing depth, the raw counts from miRXplore and plasma samples were down sampled (randomly from Binomial distribution) and plotted against number of detected miRNAs (Fig. 6). The plots may be used to compare library preparation protocols in terms of their requirements for sequencing output and consequently their cost efficiency. Ideally, the dependence will show steeply growing curve, which indicates that majority of miRNAs can be captured at low sequencing depth. In our settings, the sensitivity of detection for miRXplore samples was comparable between protocols resulting in 11-15 dropouts (undetected miRNAs) at detection threshold > 5 raw counts (Fig. 6A). The shape closest to the ideal state was shown by all protocols except Lexogen and Norgen in miRXplore samples which showed slowly growing curves suggesting that sequencing depth needs to be higher to capture full spectrum of miRNAs. The sensitivity of detection in plasma was much lower and differences between protocols more profound than in miRXplore samples, although the performance of each protocol relatively to others was retained. Highest sensitivity and the steepest curve were shown by QIAseq and NEXTflex (Fig. 6B), whereas Lexogen and Norgen achieved the lowest sensitivity. However, this was true only for higher sequencing depth ($> 5M$ reads). At lower depth, the sensitivity of RealSeq and SMARTer proto-

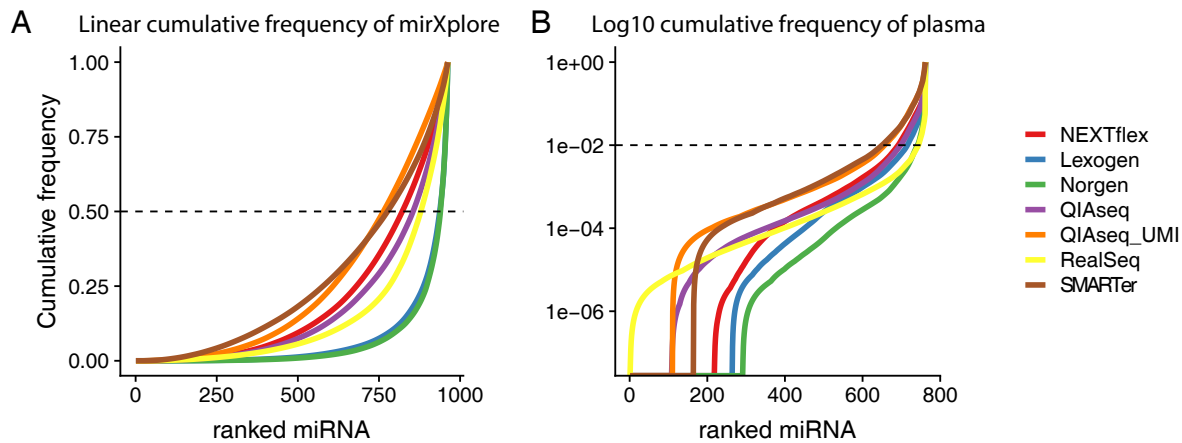


Fig. 4. **Cumulative Frequency**

A, Cumulative frequency of mirXplore samples in linear scale. Dashed line indicates CF50 – number of miRNAs ordered according to increasing number of counts at cumulative frequency of 50%. B, Cumulative frequency of plasma samples in log₁₀ scale. Dashed line indicates CF1 – cumulative frequency at 1%. Different scales in A and B were used for better visualization.

cols started to be similar to Lexogen and Norgen. Overall, the highest number of detected miRNAs were measured in QIAseq, 444 miRNAs, compared to 223 miRNAs in Norgen libraries. Interestingly, the number of miRNAs jointly detected by all protocols was relatively low – 90 (Additional file 3, Fig. 6). In conclusion, we observed a similar sensitivity between protocols in the sample with balanced miRNA concentration, but it varied profoundly in real plasma sample. While the best performing protocols detected over 200 miRNAs at sequencing depth below 1M reads, the worst protocol hardly reached this goal at 10-times higher sequencing depth. These differences highlight the importance of protocol selection, especially for analysis of samples where unbalanced proportion of miRNA concentrations may be expected.

Correlation with RT-qPCR. Despite many advantages, small RNA-seq is not capable to deliver absolute quantification of miRNA molecules in the sample. For this purpose, the method of choice is the RT-qPCR. In order to examine which protocols showed miRNA abundances in plasma samples closest to the truth, we have performed absolute quantification of 35 miRNAs in the plasma samples by Two tailed RT-qPCR (26). The targets involved miRNAs presented in different concentrations which were detected by all of the tested protocols. After rigorous quality control, the absolute concentration levels of 19 miRNAs were correlated with the RNA-Seq data.

The expression levels of selected miRNAs spread wide range of concentration from 4×10^2 to 3×10^7 molecules per μl of isolated plasma RNA (Additional file 3, Fig. 7). The most abundant miRNA was hsa-miR-451a, followed by hsa-miR-16-5p and hsa-miR-486-5p. Interestingly, the prime position of hsa-miR-451a agrees with the RNA-seq results of protocols preventing ligation bias - SMARTer, RealSeq and NEXTFlex (Fig. 5). Contrary, the third most abundant hsa-miR-486-5p accounts for majority of miRNA reads (>70%) in the libraries generated by traditional two-adapters ligation

procedure due to strong overestimation of its true abundance (\log_2 fold changes > 5). Correlation of qPCR data with RNA-seq data (Fig. 7) showed that SMARTer RNA-seq expression were closest to the true expression ($r = 0.94$), followed by RealSeq, QIAseq and NEXTFlex ($r > 0.88$). Lexogen and Norgen data showed the worse correlation ($r < 0.81$). The poor correlation of Lexogen and Norgen protocols with RT-qPCR may be explained by strong ligation bias (Fig. 3A). Interestingly, QIAseq, which did not use any prevention to ligation bias, showed good correlation with RT-qPCR data, which suggests that the application of UMI might lead to overall good accurate results. To assess the overall similarity of library preparation protocols, we calculated differentially expressed miRNAs between the protocols in mirXplore and plasma samples and clustered the data using 15 miRNAs with lowest adjusted p-value calculated by DESeq2 likelihood ratio test. Clustering based on mirXplore samples showed similar expression patterns in protocols utilizing the same principles for library preparation (Fig. 8). Lexogen and Norgen do not include any steps to overcome PCR and ligation bias and clustered closely together. Interestingly, RealSeq showed similar expression pattern even though the circularization approach should minimize ligation bias (18). QIAseq and SMARTer achieved the best performance in accuracy of measurement in mirXplore samples (Fig. 3A) and indeed clustering showed that their expression values were the most similar. Contrary to mirXplore, clustering of library preparation protocols based on top 15 differently expressed miRNAs in plasma samples displayed different patterns and were driven by the most abundant miRNAs that consumed the majority of reads (Fig. 5, Fig. 8B). To summarize, the RNA-seq data correlated with absolute numbers of miRNAs relatively well, although the protocols developed to minimize biases in the workflow achieved higher level of concordance. The differential analysis revealed similarity between protocols and created two groups of methods, clustering RealSeq protocol into the group of traditional two-adapter ligation-based methods.

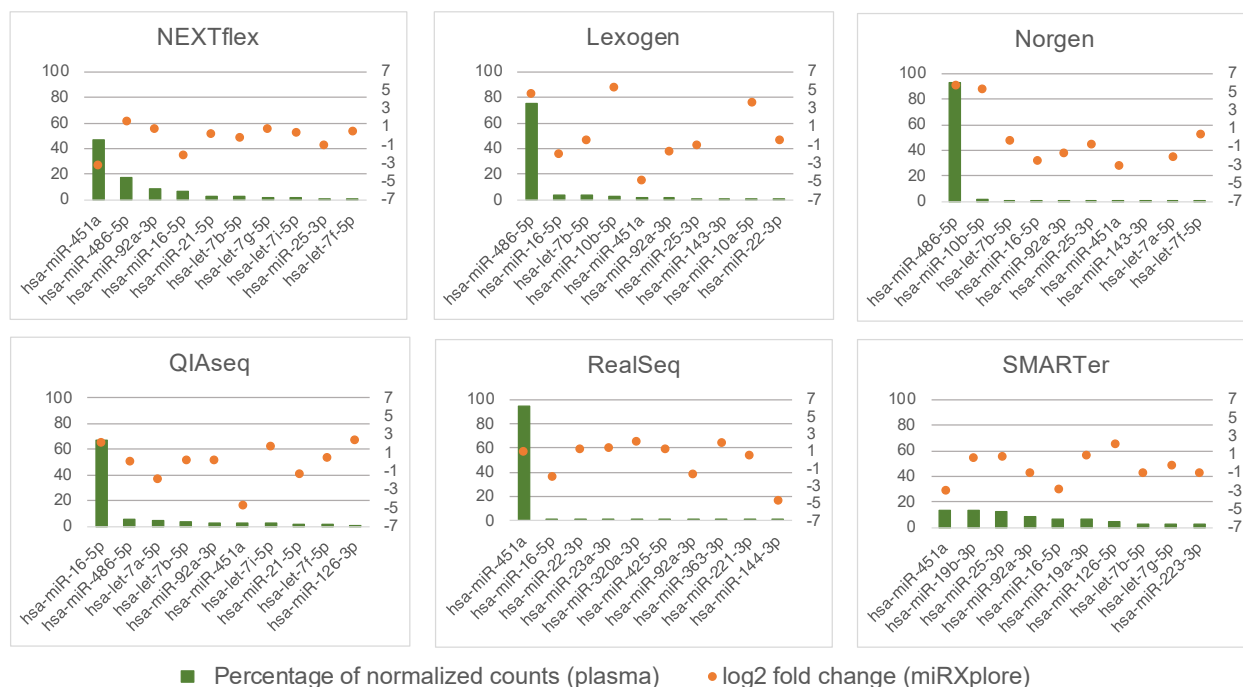


Fig. 5. **Most Abundant miRNAs in Plasma**

Percentage of normalized counts for ten most abundant miRNAs in plasma for each protocol. Orange dots indicate \log_2 fold change deviation of measured value from the predicted in miRxplore sample.

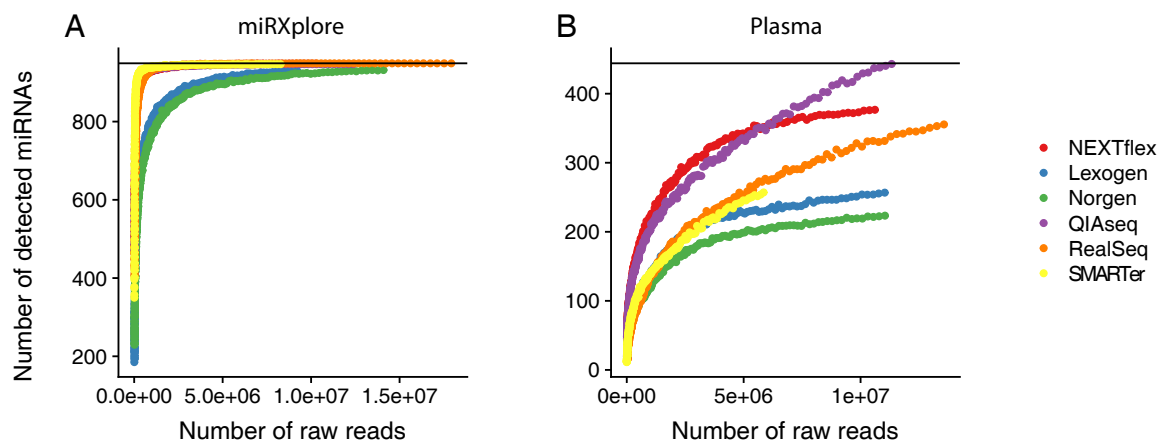


Fig. 6. **Sensitivity of miRNA Detection**

Dependence of number of detected miRNAs on sequencing depth (number of raw reads- before mapping). Black horizontal line indicates maximal number of detected miRNAs (detection threshold > 5 raw counts). A, miRxplore samples; B, plasma samples.

Evaluation Metrics. To summarize performance of protocols, all discussed metrics were assessed relatively to achieved maximum and minimum and three values representing relative performance were assigned (Good, Average, Poor). All absolute and relative values used in metrics are listed in Supplement1. Relative values were used for clustering of protocols to provide overall comparison (Fig.9). The heatmap and clustering clearly showed that Lexogen and Norgen had similar and worst performance which related to fact that ligation and PCR biases were not prevented. QIAseq and NEXTflex on the other hand showed the most consistent performance throughout all metrics. Interestingly, RealSeq which omits the second ligation and this way should reduce

ligation bias (18) clustered together with Lexogen and Norgen. It suggests that substantial proportion of the bias still remains in the process.

Discussion

Small RNA-seq is the most suitable method for discovery and global miRNA profiling, since it does not require any prior knowledge about sequences of interest. However, the library preparation workflow includes steps which introduce bias and affect sensitivity of detection and accuracy of quantification (12, 19). As a consequence, the expression of some miRNAs can be over- or under-estimated or completely undetected. To address this issue, several meth-

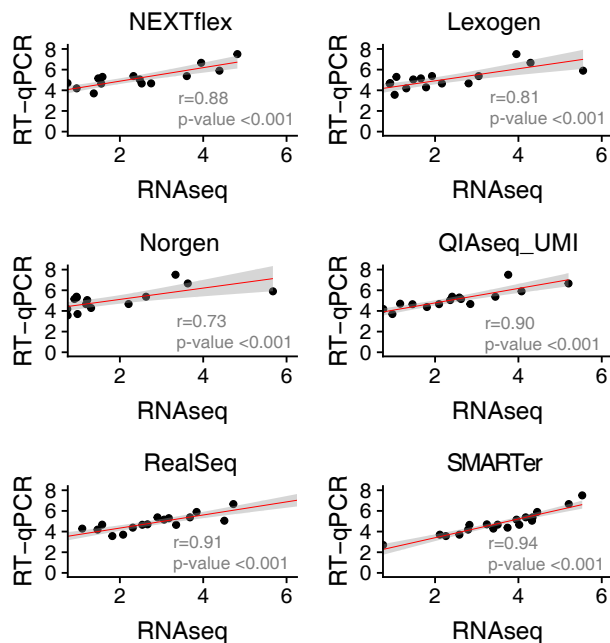


Fig. 7. **Correlation of small RNA-seq Data with RT-qPCR**
Correlation \log_{10} scaled normalized RNA-seq data with \log_{10} scaled qPCR absolute quantities (number of molecules per μ l) for each protocol.

ods to diminish the bias were developed and are currently commercially available. Here, we have compared six library preparation protocols using different approaches to minimize the bias (Tab.1). Using comprehensive set of metrics (Fig. 9), we assessed their performance in equimolar pool of synthetic miRNAs and RNA isolated from human plasma pooled from four volunteers (Fig. 1). The main conclusions of our study include i) superiority of new methods in minimizing ligation bias over the traditional ligation-based methods; ii) beneficial effect of UMI correction in small RNA sequencing experiments; iii) good performance of polyadenylation methods despite its low mapping rate and issues with isomiRs quantification; iv) balanced performance of methods using randomized adapters; v) lower accuracy of circularization methods and its unbalanced allocation of reads to miRNA spectrum; and vi) surprisingly consistent performance of the standard ligation-based method after data correction with UMIs. Takara SMARTer smRNA-seq protocol has been used in our comparison as a representative of polyadenylation methods. It is the only group of methods that completely avoids the ligation step which is believed to be the main source of bias in small RNA-seq (19, 27). As expected and also in agreement with previous studies (28, 29), our data showed poor mapping rate of reads generated by polyadenylation approach in plasma sample (Fig.2B). It is a predictable consequence of polyadenylation reaction that extends not only miRNA molecules, but also any other RNA species or fragments present in the reaction creating a substantial background. As this may represent a disadvantage in the studies focused on miRNA, it provides a chance for analysis of larger spectrum of various RNA species simultaneously. Contrary to plasma samples, the poor mapping rate was not observed in miRXplore sample (Fig. 2A). This is in

disagreement with recent work utilizing another polyadenylation protocol CATS Small RNA-seq Kit from Diagenode that documented low mapping rate even in miRXplore (22). We hypothesize that the differences may be caused by difference in library preparation protocols.

The poor mapping rate of polyadenylation methods is reflected in the sequencing requirements and total cost of an experiment. In plasma samples, only 5% of raw reads were mapped to miRNAs compared to 27-61% mapped reads in other libraries. This implicates that for robust differential expression analysis or discovery of novel miRNAs the sequencing cost will be inevitably higher. However, our results showed that even with low number of raw reads, SMARTer is still sensitive method detecting average number of miRNAs (Fig.6). As the ligation bias is absent, the accuracy of measurements achieved the best values among tested methods (Fig.3A) as well as there is no single miRNA dominating in the mapping statistics (Fig.5). The disadvantage of the polyadenylation is its inability to distinguish between naturally occurring adenines on the 3' end and adenines added during polyadenylation in the data analysis step. This may hamper quantification of miRNAs with adenine at the 3' end as well as increase proportion of false isomiRs (Fig.2A). The true importance of this bias is however questionable as it depends on sample composition and experimental goals. In our data, we observed relatively low proportion of isomiRs (up to 15%). A partial loss of them may be therefore acceptable and balanced by other advantages of the polyadenylation system. This is partially supported by the strongest correlation of RNA-Seq expression values achieved by SMARTer with RT-qPCR results (Fig.7) that is considered to be the gold standard for miRNA quantification (11). In conclusion, polyadenylation protocols are suitable only for experiments where high accuracy of measurement is required, and the sequencing cost together with hampered detection of isomiRs do not represent an issue.

Another promising approach for the reduction of ligation bias is based on the application of randomized adapters. The concept stands on the hypothesis that if a single adapter may prefer specific sequences, the pool of adapters having random sequences will contain all combinations that will efficiently and equally bind any miRNAs presented in a sample (12). Interestingly, the data soon showed that not the sequence itself, but the secondary structure is the key element influencing the efficiency of the ligation process (12). This approach was shown to be able to decrease the ligation bias and increase the sensitivity of detection (17, 24). Moreover, the effect was even stronger when the combination with UMIs were used (16). This agrees with our results that confirmed the good or average performance of the method in the majority of tested parameters including precision, accuracy and sensitivity (see Additional file 3, Fig.3A,5-6). Noteworthy, the protocol based on randomized adapters similarly to polyadenylation method showed a good agreement with qPCR data in the ability to quantify the most abundant miRNA - hsa miR-451a (Fig.7, Additional file 3, Fig.7). Although the mapping rate was moderate compared to other

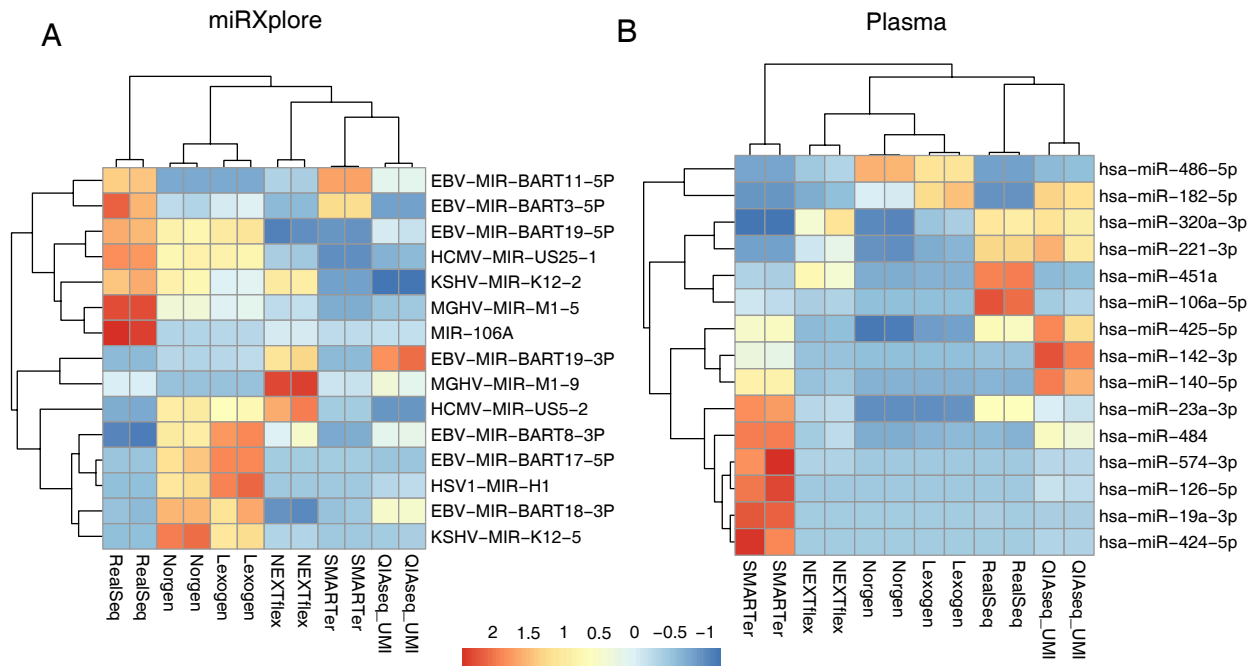


Fig. 8. Clustering of Protocols

Clustering of protocols based on 15 miRNAs with lowest adjusted p-value when tested for differential expression between protocols. **A**, miRXplore samples; **B**, plasma samples. Heatmap values are values of library size -normalized and scaled expression values.

methods, the fraction of reads mapped to miRNAs excluding top 10 mapped targets was together with QIAseq and SMARTer the highest among the tested protocols (Fig.2A-B). Similarly, to the mapping statistics, the correlation of small RNA-seq data with qPCR data was rather moderate, but still sufficiently high ($r = 0.88$, Fig.7). To sum up, the protocol based on randomized adapters is characterized by balanced performance and may be therefore recommended for routine applications in various experimental settings. The status of the well-established method will be probably even increasing in the future, as some of the large consortia selected this approach as a standard for their RNA-Seq based studies (17, 30). The latest approach to overcome the ligation bias comprises of ligation of single 3' pre-adenylated adapter and subsequent circularization (18). As the 5' ligation is intramolecular, rather than inter-molecular, the process should be more efficient and diminish a substantial part total bias. Its superior performance, especially the accuracy of measurement, was demonstrated with Somagenics RealSeq-AC protocol designed for RNA isolated from tissue (18). Noteworthy, in our study we used newer version of RealSeq library preparation protocol designed for serum or plasma samples. Since its introduction no other evaluation and comparison with other library preparation protocols was published. Only recently, (22) published benchmarking study which included beta version of Takara SMARTer miRNA-seq protocol utilizing single adapter ligation and circularization. However, the initial mapping statistics provided poor results, therefore the protocol was excluded from further detailed analysis. Our data are in agreement with this observation. Although in the miRXplore sample the mapping statistics were comparable to other protocols (Fig.2), the percent-

age of reads mapped to miRNAs excluding top 10 mapped miRNAs was the lowest (Fig.2, Additional file 2). Strikingly, among the top 10 mapped miRNAs the absolute majority of reads was consumed by only single miRNA - hsa miR-451a (Fig.5, Additional file 2). The unsatisfactory result was apparent also in the accuracy metric, where the circularization approach achieved similar values as traditional ligation-based methods (Fig.3A). The similarity is also highlighted by clustering of the circularization method with ligation methods in a single group (Additional file 3, Fig.8) and by clustering based on evaluation metrics (Fig.9). Our data show that circularization protocol is similarly prone to ligation bias as any other ligation-based methods. However, as hsa miR-451a was confirmed to be the most abundant miRNA quantified by RT-qPCR screen (Additional file 3, Fig.7), we further speculate if also other sources of bias that exaggerate the signal of most prominent miRNAs do not contribute to the process. To conclude, despite large promises during its introduction we saw rather poor performance of the circularization-based protocol. The reasons for this discrepancy are unclear and other more thorough studies focused on the individual steps of the workflow will be needed to provide some satisfactory conclusions. The last group of methods consists of traditional ligation-based methods utilizing subsequent ligation of 3' and 5' adapter. There are several vendors providing this type of solution. In our comparison we tested two of them (Lexogen and Norgen). As expected, they did not perform well in the majority of tested parameters (Fig.9) which is in agreement with the recent literature (17, 22). Recently, a new updated version of the classical ligation-based protocol was released by QIAseq. Contrary to others, their library protocol utilizes UMI to correct for PCR bias. Al-

though there has been some discussion about the importance of UMI correction in small RNA-library preparation protocols (12, 31, 32), recent studies showed their positive effect on the data quality (16, 33). The same effect we saw in our data. Although the UMI correction does not mitigate the ligation bias and contribution of PCR to the total bias is lower (Fig.3B), we observed a substantial improvement of the accuracy as well as cumulative frequency after UMI correction (Fig.3A, Fig.4). Surprisingly, the distribution of ligation bias closely resembled the profile of protocols based on randomized adapters or polyadenylation even before UMI correction and was not similar to traditional ligation-based methods (Fig. 3A). The higher similarity to the methods trying to minimize ligation biased was apparent also in the remaining metrics (Fig.2A-B, Fig. 4, Fig.5, Fig.6, Fig.7, Fig.9). The reason for this pattern is not clear as the protocol should resemble traditional ligation protocols, especially in the metrics where UMI correction is not utilized. Possible explanation may be the use of some proprietary additives, chemistry or careful optimization of the process (31, 34–36). As a similar data has been reported by other groups (22, 28), we may confirm a good performance of the protocol and advise its further use.

Conclusions

Here, we provide comprehensive comparison of all current approaches for high-throughput RNA-seq based analysis of small RNAs. We focused on miRNAs and their quantification in plasma samples as they represent promising candidates for application in diagnosis and prognosis of many human diseases and pathological states (37). Our data confirmed the large bias in the data generated by traditional two-steps ligation methods and highlighted the superiority of the methods using randomized adapters or polyadenylation and importance of UMI correction. At the same time, we documented some drawbacks that still exist and provide opportunities for further development and improvement of existing workflows. Taken together, our data provides a point of reference for an informed selection of library preparation method and contributes to the standardization of small RNA-seq for analysis of circulating RNAs.

Methods

Samples and RNA Isolation. Informed consent was obtained from all volunteers participating in the study. All procedures involving the use of human samples were performed in accordance with the ethical standards of Institute of Biotechnology of the Czech Academy of Sciences, and with the Declaration of Helsinki. Blood samples were collected from four healthy volunteers into K2EDTA BD Vacutainer tubes (Beckman Dickinson) and centrifuged within 30 min from collection at 1500 x g for 15 min at room temperature. The plasma fraction was aspirated and transferred into 2 ml tubes (Eppendorf) and centrifuged again for 15 min at 3000 x g. The supernatant was transferred into new 2 ml tubes and stored at -80 °C until analysis. Levels of hemolysis

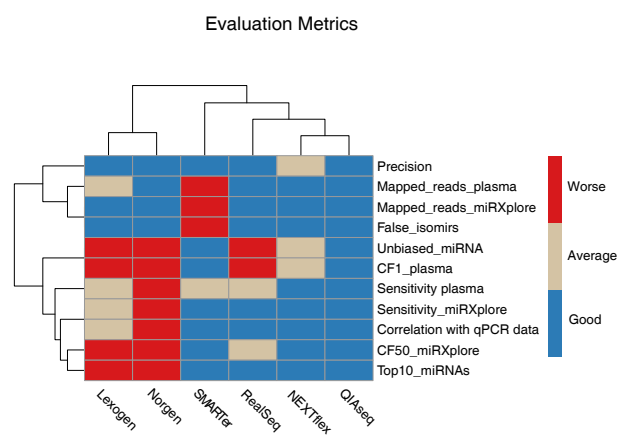


Fig. 9. **Performance Evaluation**

Clustering of protocols based on selected metrics for assessment of their performance. Performance categories (Good, Average and Worse) were assigned relatively according to the minimal and maximal values of all measured values at specific metric.

were assessed in each sample by measuring absorbance at 414 nm using NanoDrop 2000 (Thermo Fisher). Total RNA was isolated starting from plasma aliquots of 250 µl using the miRNeasy Serum/Plasma Advanced Kit (Qiagen) according to manufacturers instructions. 1 µl of isolation spike-in mix and 1 µl of GlycoBlue Coprecipitant (Invitrogen) were added at the lysis step as described in (38). RNA was eluted into 20 µl of nuclease-free water and stored at -80 °C. Each RNA eluate was assessed for quality of isolation, levels of hemolysis and presence of inhibitors by Two-tailed RT-qPCR panel, as described in (38). RNA eluates were then pooled together to produce standard plasma RNA sample used through the study. Finally, 497 µl of plasma RNA pool was mixed with 1µl ExiSeq spike-in mix (Exiqon) and 2 µl of the mix of cel-miR-76 and cel-miR-2 synthetic spike-ins (5x10³ copies/µl), aliquoted and stored at -80°C. miRXplore Universal Reference, an equimolar mixture of 962 synthetic microRNAs, was purchased from Miltenyi Biotec.

Library Preparation. Libraries were prepared in technical duplicates starting from 5 µl of plasma RNA pool and 5µl of miRXplore Universal Reference (2x10⁶ copies/µl) according to each manufacturers protocol using compatible set of sample indexes. The version of the protocol, adapter concentrations and number of PCR cycles for each protocol are listed in Additional file 1. Libraries were quantified on the Qubit 3 fluorometer (ThermoFisher) and Fragment Analyzer (Agilent). Libraries generated by the same protocol were pooled and run on 5% TBE-PAGE on Mini-PROTEAN tetra cell (BioRad). A region representing fragments with RNA inserts of length 22 nt ± ~10 nt (i.e. fragments originating from miRNAs) was excised from the gel, DNA was eluted into nuclease-free water and purified with SPRIselect reagent (Beckman Coulter). All libraries were sequenced in one sequencing run on NextSeq 500 high-output (Illumina) with 85bp single-end reads. 5.8 – 17.9 million reads per library were obtained with a median of 11 million reads (see Additional file 3).

Data Processing. Raw reads of all samples were trimmed with cutadapt tool (version 1.18) (39) according to the respective library preparation manual. After trimming, reads were filtered based on their length. Only reads longer than 16 and shorter than 29 bp were kept. Filtered reads were subsequently mapped with Bowtie aligner (40) to rRNA and UniVec databases obtained from sortmerna github repository. Reads which did not map to UniVec and rRNA sequences were further mapped to relevant references with STAR aligner (41) with usage of “end-to-end” mode and 5% of sequence was allowed to mismatch. Counting of reads was performed with featureCounts and only uniquely mapping reads were counted. Since QIAseq protocol utilizes UMI for correction of PCR bias, UMI-tools software was used for deduplication before counting of mapped reads in QIAseq samples. For comparability with other protocols non-deduplicated QIAseq samples were used for calculation of relevant metrics. Deduplicated QIAseq samples are referred to as “QIAseq_UMI”. Plasma samples were first mapped to human genome (GRCh38.95). Reads mapping to genome were further mapped to mature human miRNA sequences in miRBase v22 (42). Reads which were not mapped to miRBase were further mapped in descended order to tRNA database (435 mature tRNA sequences from gtrRNAdb), piRNA database (8 million sequences from piRBase v2) and ncRNA database (36 thousand non-coding sequences from ensemble GRCh38). Only reads which were not mapped were used for mapping to subsequent database. MiRxplore samples were mapped to miRxplore reference with same settings as plasma samples to miRBase. For counting of isomiRs detected in plasma samples isomiRROR tool was used with adjusted setting, when only longer and shorter isomiRs without mismatch within mature sequence were counted. Counting of false isomiRs detected in miRxplore was performed by custom bash-based script and only isomiRs longer than mature sequences were counted with no mismatch within the original miRxplore sequence.

Evaluation Metrics. Data were normalized to library size (counts per million CPM) for further calculations and if not stated otherwise all statistics were calculated separately for each technical replicate and their mean values are shown. Ligation bias was calculated for each miRNA as a fold change of mean value of two technical replicates from its predicted value. The predicted value was calculated as a number of normalized counts per sample divided by 962 (number of miRNAs in miRxplore). The contribution of PCR bias and ligation bias to overall bias in small RNA-seq was assessed on samples processed by QIAseq protocol with usage of variancePartition R package which employed linear mixed model to separate the variance of multiple variables (PCR bias, ligation bias and technical replicates). Dependence of number of detected miRNAs on sequencing depth was assessed by down sampling the raw counts. Down sampling was performed with usage of random generator for binomial distribution in R. The number of miRNAs was used as a number of observations and the number of raw counts belonging to individual miRNAs corresponded to number of

trials. The probability of success in each trial corresponded to proportion of raw reads at specific sequencing depth related to the number of raw reads at the original sequencing depth. For correlation of RT-qPCR data with RNA-seq data Pearson correlation was used and technical replicates were averaged before correlation.

RT-qPCR. Absolute quantification was performed for 35 pre-selected miRNAs using Two-tailed PCR technology as described in (26). Briefly, 4 µl of the standard sample (miRxplore sample) in different concentration (5-5x10⁷ copies/µl) were reverse transcribed in the 20-µl reaction containing pool of miRNA-specific primers. After cDNA synthesis, the total volume was diluted to 200 µl and 2 µl of 10-times diluted cDNA used as a template in 10-µl qPCR reaction. The data was processed in the Biorad CFX Manager software discarding Cq values generated by reactions which melting curve analysis indicated aberrant Tm values. For each assay, the standard curves were calculated using the available number of miRxplore standards. The absolute concentration of the standard plasma was calculated based on the parameters of the standard curve. The plasma sample was measured in four technical replicates and two replicates were used for miRxplore standards. After quality control, only 19 miRNAs measured with the high confidence were used for correlation analysis with RNA-Seq data.

Bibliography

- Rosalind C. Lee, Rhonda L. Feinbaum, and Victor Ambros. The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to lin-14. *Cell*, 75:843–854, 1993. ISSN 00319007. doi: 10.1103/PhysRevLett.99.179704.
- Bruce Wightman, Ilho Ha, and Gary Ruvkun. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell*, 75:855–862, 1993. ISSN 00928674. doi: 10.1016/0092-8674(93)90530-4.
- Stefan L. Ameres and Phillip D. Zamore. Diversifying microRNA sequence and function. *Nature Reviews Molecular Cell Biology*, 14(8):475–488, 2013. ISSN 14710072. doi: 10.1038/nrm3611.
- B. M. Engels and G. Hutvagner. Principles and effects of microRNA-mediated post-transcriptional gene regulation. *Oncogene*, 25(46):6163–6169, 2006. ISSN 09509232. doi: 10.1038/sj.onc.1209909.
- Wigard P. Kloosterman and Ronald H.A. Plasterk. The Diverse Functions of MicroRNAs in Animal Development and Disease. *Developmental Cell*, 11(4):441–450, 2006. ISSN 15345807. doi: 10.1016/j.devcel.2006.09.009.
- Giovanni Stefani and Frank J. Slack. Small non-coding RNAs in animal development. *Nature Reviews Molecular Cell Biology*, 9(3):219–230, 2008. ISSN 14710072. doi: 10.1038/nrm2347.
- Oscar D. Murillo, William Thistlethwaite, Joel Rozowsky, Sai Lakshmi Subramanian, and Lucero. exRNA Atlas Analysis Reveals Distinct Extracellular RNA Cargo Types and Their Carriers Present across Human Biofluids. *Cell*, 177(2):463–477.e15, 2019. ISSN 10974172. doi: 10.1016/j.cell.2019.02.018.
- Andrey Turchinovich, Ludmila Weiz, Anne Langheinz, and Barbara Burwinkel. Characterization of extracellular circulating microRNA. *Nucleic Acids Research*, 39(16):7223–7233, 2011. ISSN 03051048. doi: 10.1093/nar/gkr254.
- Bridget Martinez and Philip Peplow. MicroRNAs in blood and cerebrospinal fluid as diagnostic biomarkers of multiple sclerosis and to monitor disease progression. *Neural Regeneration Research*, 15(4):606–619, 2020. ISSN 18767958. doi: 10.4103/1673-5374.266905.
- Baohong Zhang, Xiaoping Pan, George P. Cobb, and Todd A. Anderson. microRNAs as oncogenes and tumor suppressors. *Developmental Biology*, 302(1):1–12, 2007. ISSN 00121606. doi: 10.1016/j.ydbio.2006.08.028.
- Lukas Valihrač, Peter Androvic, and Mikael Kubista. Circulating miRNA analysis for cancer diagnostics and therapy. *Molecular Aspects of Medicine*, (October):1–19, 2019. ISSN 00982997. doi: 10.1016/j.mam.2019.10.002.
- Ryan T. Fuchs, Zhiyi Sun, Fanglei Zhuang, and G. Brett Robb. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS ONE*, 10(5):1–24, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0126049.
- Sam E.V. Linsen, Elzo de Wit, Georges Janssens, Sheila Heater, Laura Chapman, Rachael K. Parkin, Brian Fritz, Stacia K. Wyman, Ewart de Bruijn, Emile E. Voest, Scott Kuersten, Muneesh Tewari, and Edwin Cuppen. Limitations and possibilities of small RNA digital gene expression profiling. *Nature Methods*, 6(7):474–476, 2009. ISSN 15487091. doi: 10.1038/nmeth0709-474.
- G. K. Fu, J. Hu, P.-H. Wang, and S. P. A. Fodor. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proceedings of the National Academy of Sciences*, 108(22):9026–9031, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1017621108.

15. Teemu Kivioja, Anna Vähärautio, Kasper Karlsson, Martin Bonke, Martin Enge, Sten Linnarsson, and Jussi Taipale. Counting absolute numbers of molecules using unique molecular identifiers. *Nature Methods*, 9(1):72–74, 2012. ISSN 15487091. doi: 10.1038/nmeth.1778.
16. Carrie Wright, Anandita Rajpurohit, Emily E. Burke, Courtney Williams, Leonardo Collado-Torres, Martha Kimos, Nicholas J. Brandon, Alan J. Cross, Andrew E. Jaffe, Daniel R. Weinberger, and Joo Heon Shin. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics*, 20(1):1–21, 2019. ISSN 14712164. doi: 10.1186/s12864-019-5870-3.
17. Maria D. Giraldez, Ryan M. Spengler, Alton Etheridge, Paula M. Godoy, Andrea J. Barczak, Srimeenakshi Srinivasan, Peter L. De Hoff, Kahraman Tanriverdi, Amanda Courtright, Shulin Lu, Joseph Khoory, Renee Rubio, David Baxter, Tom A.P. Driedonks, Henk P.J. Buermans, Esther N.M. Nolte-T Hoen, Hui Jiang, Kai Wang, Ionita Ghiran, Yaoyu E. Wang, Kendall Van Keuren-Jensen, Jane E. Freedman, Prescott G. Woodruff, Louise C. Laurent, David J. Erle, David J. Galas, and Muneesh Tewari. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nature Biotechnology*, 36(8):746–757, 2018. ISSN 15461696. doi: 10.1038/nbt.4183.
18. Sergio Barberán-Soler, Jenny M. Vo, Ryan E. Hogans, Anne Dallas, Brian H. Johnston, and Sergei A. Kazakov. Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biology*, 19(1):1–9, 2018. ISSN 1474760X. doi: 10.1186/s13059-018-1488-z.
19. Fanglei Zhuang, Ryan T. Fuchs, and G. Brett Robb. Small RNA expression profiling by high-throughput sequencing: Implications of enzymatic manipulation. *Journal of Nucleic Acids*, 2012, 2012. ISSN 20900201. doi: 10.1155/2012/360358.
20. Luc Girard, Jaime Rodriguez-Canales, Carmen Behrens, Debrah M. Thompson, Ihab W. Botros, Hao Tang, Yang Xie, Natasha Rekhtman, William D. Travis, Ignacio I. Wistuba, John D. Minna, and Adi F. Gazdar. An Expression Signature as an Aid to the Histologic Classification of Non-Small Cell Lung Cancer. *Clinical Cancer Research*, 22(19):4880–4889, 2016. doi: 10.1158/1078-0432.ccr-15-2900.
21. Daneida Lizarraga, Karen Huen, Mary Combs, Maria Escudero-Fung, Brenda Eskenazi, and Nina Holland. MiRNAs differentially expressed by next-generation sequencing in cord blood buffy coat samples of boys and girls. *Epigenomics*, 8(12):1619–1635, 2016. ISSN 1750192X. doi: 10.2217/epi-2016-0031.
22. Fatima Heinicke, Xiangfu Zhong, Manuela Zucknick, Johannes Breidenbach, Arvind Y.M. Sundaram, Siri T. Fläm, Magnus Leithaug, Marianne Dalland, Andrew Farmer, Jordana M. Henderson, Melanie A. Hussong, Pamela Moll, Loan Nguyen, Amanda McNulty, Jonathan M. Shaffer, Sabrina Shore, Hoichong Karen Yip, Jana Vitkowska, Simon Rayner, Benedicte A. Lie, and Gregor D. Gilfillan. Systematic assessment of commercially available low-input miRNA library preparation kits. *RNA Biology*, 17(1):75–86, 2020. ISSN 15558584. doi: 10.1080/15476286.2019.1667741.
23. Ryan K.Y. Wong, Meabh MacMahon, Jayne V. Woodside, and David A. Simpson. A comparison of RNA extraction and sequencing protocols for detection of small RNAs in plasma. *BMC Genomics*, 20(1):1–12, 2019. ISSN 14712164. doi: 10.1186/s12864-019-5826-7.
24. Ashish Yeri, Amanda Courtright, Kirsty Danielson, Elizabeth Hutchins, Eric Alsop, Elizabeth Carlson, Michael Hsieh, Olivia Ziegler, Avash Das, Ravi V. Shah, Joel Rozowsky, Saumya Das, and Kendall Van Keuren-Jensen. Evaluation of commercially available small RNAseq library preparation kits using low input RNA. *BMC Genomics*, 19(1):1–15, 2018. ISSN 14712164. doi: 10.1186/s12864-018-4726-6.
25. Yuwen Liu, Jie Zhou, and Kevin P. White. RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, 30(3):301–304, 2014. ISSN 13674803. doi: 10.1093/bioinformatics/btt688.
26. Peter Androvic, Lukas Valihrach, Julie Elling, Robert Sjoback, and Mikael Kubista. Two-tailed RT-qPCR: A novel method for highly accurate miRNA quantification. *Nucleic Acids Research*, 45(15), 2017. ISSN 13624962. doi: 10.1093/nar/gkx588.
27. Carsten A. Raabe, Thean Hock Tang, Juergen Brosius, and Timofey S. Rozhdstvensky. Biases in small RNA deep sequencing data. *Nucleic Acids Research*, 42(3):1414–1426, 2014. ISSN 03051048. doi: 10.1093/nar/gkt1021.
28. Anna M.L. Coenen-Stass, Iddo Magen, Tony Brooks, Iddo Z. Ben-Dov, Linda Greensmith, Eran Hornstein, and Pietro Fratta. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biology*, 15(8):1133–1145, 2018. ISSN 15558584. doi: 10.1080/15476286.2018.1514236.
29. Cloelia Dard-Dascot, Delphine Naquin, Yves D'Aubenton-Carafa, Karine Alix, Claude Thernes, and Erwin van Dijk. Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics*, 19(1):1–16, 2018. ISSN 14712164. doi: 10.1186/s12864-018-4491-6.
30. Saumya Das, Asim B. Abdel-Mageed, Catherine Adamidi, P. David Adelson, and Akat. The Extracellular RNA Communication Consortium: Establishing Foundational Knowledge and Technologies for Extracellular RNA Research. *Cell*, 177(2):231–242, 2019. ISSN 10974172. doi: 10.1016/j.cell.2019.03.023.
31. Markus Hafner, Neil Renwick, Miguel Brown, Aleksandra Mihailović, Daniel Holoch, Carola Lin, John T.G. Pena, Jeffrey D. Nusbaum, Pavel Morozov, Janus Ludwieg, Tolulope Ojo, Shujun Luo, Gary Schroth, and Thomas Tuschl. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *Rna*, 17(9):1697–1712, 2011. ISSN 13558382. doi: 10.1261/ma.2799511.
32. Anitha D. Jayaprakash, Omar Jabado, Brian D. Brown, and Ravi Sachidanandam. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Research*, 39(21):1–12, 2011. ISSN 03051048. doi: 10.1093/nar/gkr693.
33. Xiaonan Fu and Doyuan Dong. Bioinformatic Analysis of MicroRNA Sequencing Data. *Methods in molecular biology*, 1751(January 2018):109–125, 2018. doi: 10.1007/978-1-4939-7710-9.
34. Markus Hafner, Neil Renwick, Thalia A. Farazi, Aleksandra Mihailović, John T. G. Pena, and Thomas Tuschl. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods (San Diego, Calif.)*, 58(2):164–170, 2012. ISSN 15378276. doi: 10.1016/j.jymeth.2012.07.030.
35. David G.J. Mann, Zachary R. King, Wusheng Liu, Blake L. Joyce, Ryan J. Percifield, Jennifer S. Hawkins, Peter R. LaFayette, Barbara J. Artelt, Jason N. Burris, Mitra Mazarei, Jeffrey L. Benetzen, Wayne A. Parrott, and Charles N. Stewart. Switchgrass (*Panicum virgatum* L.) polyubiquitin gene (PvUbi1 and PvUbi2) promoters for use in plant transformation. *BMC Biotechnology*, 11(1):72, 2011. ISSN 14726750. doi: 10.1186/1472-6750-11-72.
36. Daniela B. Munafó and G. Brett Robb. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *Rna*, 16(12):2537–2552, 2010. ISSN 13558382. doi: 10.1261/rna.2242610.
37. Shailendra Dwivedi, Purvi Purohit, and Praveen Sharma. MicroRNAs and Diseases: Promising Biomarkers for Diagnosis and Therapeutics. *Indian Journal of Clinical Biochemistry*, 34(3):243–245, 2019. ISSN 09740422. doi: 10.1007/s12291-019-00844-x.
38. Peter Androvic, Nataliya Romanyuk, Lucia Urdzikova-Machova, Eva Rohlova, Mikael Kubista, and Lukas Valihrach. Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Scientific Reports*, 9(1):1–9, 2019. ISSN 20452322. doi: 10.1038/s41598-019-40513-w.
39. Marcel Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBNETjournal*, 17(1):10–12, 2011.
40. Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3), 2009. ISSN 14747596. doi: 10.1186/gb-2009-10-3-r25.
41. Alexander Dobin, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21, 2013. ISSN 13674803. doi: 10.1093/bioinformatics/bts635.
42. Ana Kozomara, Maria Birgaoanu, and Sam Griffiths-Jones. MiRBase: From microRNA sequences to function. *Nucleic Acids Research*, 47(1):D155–D162, 2019. ISSN 13624962. doi: 10.1093/nar/gky1141.

Supplementary Note 1: Additional file 1

Supplementary methods, S1.xlsx.

Supplementary Note 2: Additional file 2

Supplementary tables with data, S2.xlsx.

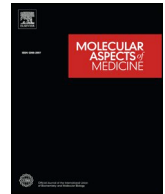
Supplementary Note 3: Additional file 3

Supplementary figures, S3.docx.



Contents lists available at ScienceDirect

Molecular Aspects of Medicine

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

Circulating miRNA analysis for cancer diagnostics and therapy

Lukas Valihrach^{a,*}, Peter Androvic^{a,b}, Mikael Kubista^{a,c}^a Laboratory of Gene Expression, Institute of Biotechnology CAS, BIOCEV, 252 50, Vestec, Czech Republic^b Laboratory of Growth Regulators, Faculty of Science, Palacky University, 783 71, Olomouc, Czech Republic^c TATAA Biocenter AB, 411 03, Gothenburg, Sweden

ARTICLE INFO

Keywords:

Circulating miRNA
RNA-Seq
RT-qPCR
Cancer
Liquid biopsies

ABSTRACT

Successful treatment of cancer depends on early diagnosis and effective monitoring of patients' response to therapy. Traditional tools based on tumor biopsies lack the sensitivity and specificity to capture cancer development in its early phases and are not applicable for continuous monitoring. To overcome these barriers, liquid biopsies have been introduced as a minimally invasive and cost-efficient means of diagnosis with high level of specificity and sensitivity. Traditionally, liquid biopsy markers include circulating tumor cells and circulating tumor DNA. During the last decade, a new promising group of biomarkers has appeared and its utilization for cancer diagnosis and monitoring is intensively studied – the microRNAs (miRNAs). In this review, we provide a comprehensive overview of circulating miRNA analysis. We highlight the importance of sampling and quality control, discuss the technical aspects of miRNA extraction and quantification, summarize recommendations for downstream analysis and conclude with future perspectives. Taken together, we present the current state of knowledge in the field of miRNA analysis in liquid biopsies and the expected development and standardization.

1. Introduction

miRNAs are a class of naturally occurring short non-coding RNA molecules containing about 22 nucleotides (He and Hannon, 2004). In total, over 48000 miRNAs have been identified so far (based on miR-Base Sequence Database release 22), including 2693 miRNAs of human origin (Kozomara et al., 2019). The function of most eukaryotic miRNAs is to regulate gene expression post-transcriptionally (Gebert and MacRae, 2019). After base-pairing with target mRNA molecules, miRNAs reduce translation rate by cleavage of the mRNA, blocking translation or regulating mRNA turnover (Ameres and Zamore, 2013). Since a single miRNA may target up to 400 different mRNAs, it is predicted that over half of the human genes are directly regulated by miRNAs (Friedman et al., 2009). miRNAs are therefore expected to have key regulatory roles in possibly every physiological and pathological aspect of biology. Dysfunctional expression of miRNAs is a feature of many pathological processes, including cancer, metabolic disorders, inflammatory, cardiovascular, neuro-developmental and autoimmune diseases (Rupaimoole and Slack, 2017).

The majority of miRNAs is found within the cells. The average number of individual miRNAs per cell has been estimated to about 500 copies (Liang et al., 2007). Low levels of miRNAs are also found in extracellular environments, including various biological fluids and cell

culture media - referred to as extracellular or circulating miRNAs (Pritchard et al., 2012a). In the extracellular space, miRNAs are attached to proteins or lipoproteins, or loaded inside extracellular vesicles (EVs - a generic term referring to all secreted vesicles independent of size and origin, Thery et al. (2018)), providing them high stability (Cortez et al., 2011). If protected by lipid or protein-based carriers, miRNAs are resistant to boiling, pH changes, repeated freeze-thawing cycles, and fragmentation by chemical agents and enzymes (Chen et al., 2008; Gilad et al., 2008; Mitchell et al., 2008). miRNAs are released into the extracellular space by either passive or active secretion. While the actively secreted miRNAs may act as signaling molecules, the passively released miRNAs are mostly artefacts of broken cells after injury, cells undergoing apoptotic or necrotic processes, cells exposed to chronic inflammation and cells with a short half-life such as platelets (Bayraktar et al., 2017). Interestingly, recent studies showed the possibility that tumor-derived miRNAs could modulate non-tumor cells to the benefit of the tumor (Ruivo et al., 2017). The facts that miRNAs are present in various body fluids, are stable and may reflect the pathophysiological condition of the tissue of origin brought them to attention as a new promising group of biomarkers.

The first evidence that miRNAs may have diagnostic and therapeutic potential was presented soon after the identification of the first miRNAs in human (Pasquinelli et al., 2000). The study showed

* Corresponding author.

E-mail addresses: lukas.valihrach@ibt.cas.cz (L. Valihrach), peter.androvic@ibt.cas.cz (P. Androvic), mikael.kubista@tataa.com (M. Kubista).<https://doi.org/10.1016/j.mam.2019.10.002>

Received 14 August 2019; Received in revised form 1 October 2019; Accepted 7 October 2019

0098-2997/ © 2019 Elsevier Ltd. All rights reserved.

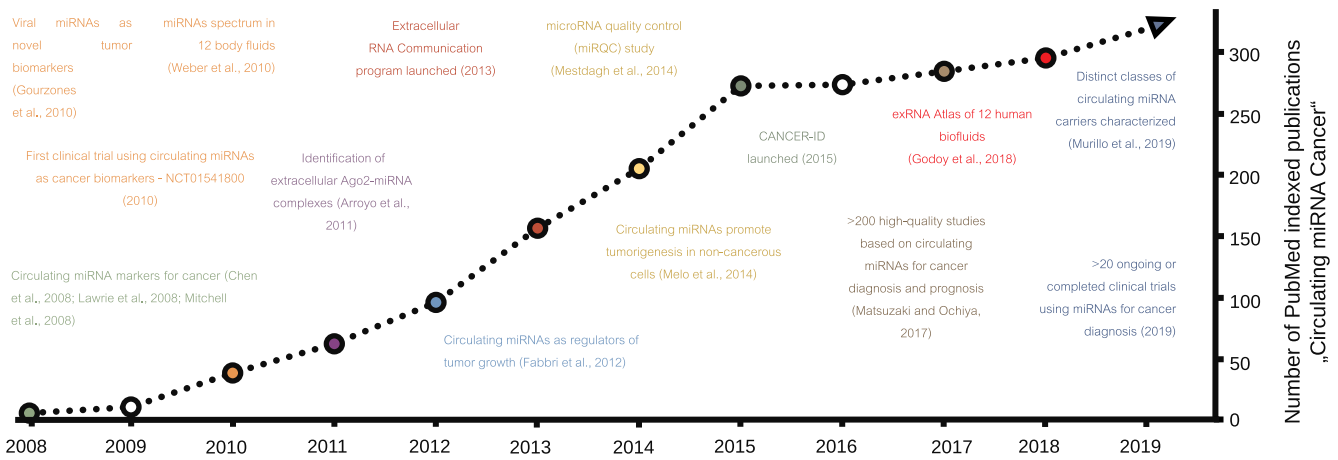


Fig. 1. Timeline showing the number of publications and historical milestones related to circulating miRNAs in cancer research. Data from PubMed using search term “circulating microRNA cancer” (Fabbri et al., 2012; Gourzoues et al., 2010; Melo et al., 2014).

correlation between the loss of miR-15 and miR-16 and the occurrence of B-cell leukemia (Calin et al., 2002). The first genetic evidence for the vital importance of miRNAs was described three years later. The homozygous deletion of the gene coding for Dicer, which is an enzyme essential for miRNA biogenesis, disrupted prenatal development of the murine embryo (Yang et al., 2005). The first indication that miRNAs may become easily accessible biomarkers for cancer diagnosis and prognosis came three years later when miRNAs were isolated from patient serum (Lawrie et al., 2008; Mitchell et al., 2008) and their profiling revealed specific patterns across different groups of diseases (Chen et al., 2008). Following studies confirmed specific miRNA signatures in many types of human diseases, including different cancers (Izzotti et al., 2016; Larrea et al., 2016; Matsuzaki and Ochiya, 2017), and showed that such signatures can be measured in various body fluids (Godoy et al., 2018; Weber et al., 2010). Increasing number of publications reporting applicability of circulating miRNAs for cancer diagnostics and prognostics (Fig. 1) led to initiation of several clinical trials (Anfossi et al., 2018) as well as international efforts aimed at deeper understanding of circulating miRNA function and standardization of the miRNA analysis field, e.g. Extracellular RNA Communication Consortium (ERCC; <https://exrna.org/>) or CANCER-ID (www.cancer-id.eu).

Circulating miRNAs have many features that make them attractive biomarker candidates for cancer diagnosis and monitoring of patients' responses to therapy. Contrary to standard tissue biopsies, sampling of biofluids is quick, minimally invasive and painless. There is low risk of associated complications and the tumor does not even have to be localized (Hayes et al., 2014). Individual miRNAs may be detected with a resolution down to single nucleotide (Hunt et al., 2015). There are indications that the power to distinguish normal from cancer samples may be higher than with traditional biomarkers such as proteins and mRNAs (Lu et al., 2005). Cost and time-effectiveness of the analysis and the exceptional stability of miRNAs in biofluids as well as in routinely prepared formalin-fixed paraffin-embedded (FFPE) materials are additional reasons miRNAs have become promising biomarkers (Andrews et al., 2015; Arata et al., 2012; Boisen et al., 2015; Hall et al., 2012).

Despite these advantages, reproducible and robust miRNA quantification is challenging and attention must be paid to the technological aspects of the measurements (Buschmann et al., 2016; Khan et al., 2017). Challenges in miRNA analysis include pre-analytical variables, usually limited amounts of the analyte, cellular contamination, risk of inhibition, inadequate standardization of methods, data analysis, normalization, and interpretation (Table 1). In this review, we focus on the impact of confounding factors and provide recommendations to minimize their effects.

2. Sampling

Proper experimental design and sampling protocol are critical to ensure the validity of the obtained results and eventually their translation to clinical practice. miRNA studies often suffer from poor reproducibility, since the natural variation of miRNA expression is underestimated (Becker and Lockwood, 2013; Jarry et al., 2014). It is known that the miRNA spectrum detected in various body fluids may vary not only due to pathophysiological processes, but also with the time of the day, diet, gender, age, alcohol consumption, medications, etc. (Ameling et al., 2015; de Boer et al., 2013; Flowers et al., 2015; Takahashi et al., 2013). Well-documented is also high inter-individual variability of the basal miRNA levels (Margue et al., 2015; McDonald et al., 2011). Taken together, these factors confound the measurement and large sample size are often needed to reach statistical significance. Confounding variation is also presumed cause of the poor consistency of many recent studies reporting contradictory results (Witwer and Halushka, 2016).

During the specimen collection measures have to be taken to stabilize the analytes. miRNAs are generally considered rather stable molecules, quite resistant to many potential biases such as storage conditions or inappropriate handling. Indeed, numerous studies have documented extraordinary stability of miRNAs stored at various temperatures and upon repetitive freeze-thawing (Balzano et al., 2015; Chen et al., 2008; Grasedieck et al., 2012; Mitchell et al., 2008). However, a recent study reported the opposite (Glinge et al., 2017). Using blood samples from 12 healthy individuals, authors investigated the stability of miRNAs under various conditions including different collection tubes, storage at different temperatures, physical disturbance, as well as serial freeze-thaw cycles. Contrary to previous reports, the majority of tested parameters influenced the measured miRNA levels. Although the study tested only selected miRNAs and data may not be generalized to the complete miRNA spectrum, the results point at several possible sources of bias that are easily overlooked.

Measures should be taken during specimen collection and transport to prevent cell lysis. Failure to do so may lead to contamination of the cell-free miRNA fraction with cellular miRNAs introducing severe bias into the measured profiles (Kirschner et al., 2011; McDonald et al., 2011). For example, hemolyzed plasma or serum show very high background of erythrocyte miRNAs and should not be used for analysis of circulating miRNAs (Androvic et al., 2019; Blondal et al., 2013). Issues caused by contaminating cellular miRNAs were reported as early as four years after the discovery of circulating miRNAs in human serum (Lawrie et al., 2008). Pritchard et al. (2012b) analyzed miRNA profiles in erythrocytes, myeloid cells, and lymphocytes and correlated the

Table 1
Challenges analyzing miRNAs in liquid biopsies.

Analysis workflow	Typical challenges
Study design	Establish suitable numbers of biological and technical replicates Estimate confounding variation due to genetic factors in the studied populations Identify environmental and biological pre-analytic variables Establish sampling and analysis protocols, and document those
Sampling and storage ^a	Select type of specimen (e.g. plasma or serum, extracellular vesicles or total cell-free miRNAs, total RNA or enriched small RNA fraction) Minimize variation in specimen collection and processing Select collection tubes and protocol Document deviations in the sampling protocol
RNA extraction ^a	Establish procedures to prevent contamination with intracellular miRNAs Select and validate appropriate extraction method Optimize and validate extraction protocol - volume of sample, use of carrier, etc.
Quality control ^a	Assess variability in extraction including batch effects Identify samples of compromised quality Assess the quality and quantity of the isolated RNA Control for contamination of intracellular miRNAs Test for activity of RNases, inhibitors and other interfering substances
Quantification	Select and validate method to quantify miRNAs Control for method-specific biases Detect and quantify low abundant miRNAs (sensitivity) Distinguish closely related miRNAs (specificity) Capture full isomiR spectrum Control for batch effects
Analysis	Establish method for normalization of measured data Establish annotation rules (isomiRs, degradation artefacts) Establish RNA-Seq data analysis pipeline Select miRNAs reflecting the studied conditions (biomarker identification)
Interpretation	Validation of potential biomarkers Inter-laboratory comparison Influence of pre-analytical variables

^a Identify technical specifications, standards and regulations. The protocol shall be compliant with recent CEN (the European Committee for Standardization) and ISO (the International Organization for Standardization) guidelines on the pre-analytical workflow and miRNA analysis. For diagnostic applications also compliance with the new CE-IVDR regulation - (EU) 2017/745 (European Community marking for *in vitro* diagnostic devices) will be requested from 2022 (Dagher et al., 2019).

abundant miRNAs with those previously identified as potential biomarkers for several types of solid tumors. Out of 79 frequently nominated miRNA biomarkers, 58% were highly expressed in one or more blood cells e.g., miR-223, miR-197, miR-574-3p, let-7a, which are prevalent in myeloid, miR-150 prevalent in lymphoid, and miR-486-5p, miR-451, miR-92a, miR-16 prevalent in erythrocytes. Even miniscule contamination with blood cells increased certain miRNA levels by up to 50-fold, which seriously confounded the analysis. Results that are interfered by contamination should be interpreted with great care (Pritchard et al., 2012b).

The first instance of the risk of blood cell miRNA contamination is the venipuncture itself (Lippi et al., 2012). Care should be taken to minimize the risk of damaging the cells by using a needle of appropriate size (typically ≥ 22 -gauge), avoiding a traumatic venipuncture, prolonged tourniquet application or fist clenching, discarding the first 1–2 mL of blood (skin plug), etc. (Becker and Lockwood, 2013; Khan et al., 2017). It is also advisable to discuss the detailed venipuncture protocol, including the follow-up collection tube handling with experienced staff directly at the blood collection site (Witwer et al., 2013). After venipuncture, the time interval between blood collection and processing should be kept short to minimize cell lysis and contamination risk (Page et al., 2013). Blood cells are typically removed by two sequential centrifugations, 800–2000 g for 10–15 min, optionally followed by a third bench top spin (1000 g for 5 min) to pellet any remaining cells and cellular debris (Page et al., 2006). Notably, the centrifugation protocol influences the spectrum of miRNAs detected (Page et al., 2013). Since leaking erythrocytes are the most frequent source of miRNA contamination (Kirschner et al., 2013b), plasma and serum should routinely be inspected for hemolysis either visually, by absorbance measurement at 414 nm - detecting oxyhemoglobin (Kirschner et al., 2011), or by the use of molecular markers (see section 4. Quality control) (Androvic et al., 2019; Blondal et al., 2013). Anti-coagulants used in plasma collection tubes, including EDTA, citrate and heparin,

may influence the measured miRNA spectrum (Fichtlscherer et al., 2010) and impact the downstream analysis, e.g. heparin-coated tubes inhibit reverse transcription quantitative PCR (RT-qPCR) (Glinge et al., 2017; Kroh et al., 2010). Blood should be processed within 2 h after collection, whereas processed plasma or serum could be stored at 4 °C for up to 24 h (Khan et al., 2017). For their long-term storage, –20 °C or preferably –80 °C freezers are recommended.

The processing of all samples should be documented and any deviation from the protocol recorded (Robb et al., 2014). For detailed instructions for specimen collection, obey relevant CEN and ISO guidelines when available (<https://www.cen.eu/>, <https://www.iso.org/>), and appropriate standard operating procedures (SOP) produced in dedicated initiatives such as SPIDIA4P (www.spidia.eu), CANCER-ID (www.cancer-id.eu), ERCC (<https://exrna.org/>), Early Detection Research Network (EDRN, <https://edrn.nci.nih.gov/>).

3. miRNA extraction

After sample collection and specimen pre-processing, miRNAs are extracted. General principles for the isolation of miRNAs are similar to total RNA, except that some protocols are modified to retain, alternatively enrich the small RNA fraction. The extraction methods may broadly be grouped into three categories - organic extraction (guanidine-phenol-chloroform based method), filter-based methods (derivatized silica) and magnetic particles-based methods. A popular choice represents hybrid methods that combine the effectiveness of organic extraction with the easiness of filter-based methods (e.g. miRNeasy Serum/Plasma Kit, Qiagen, Germany).

The choice of extraction method depends on several factors: type of sample, volume, interest in EVs or total cell-free miRNA content, need to enrich the small RNA fraction, expected RNA yield, required elution volume, miRNA purity, reproducibility, user-friendliness, turnaround time, health hazards of reagents, throughput, cost and need to

Table 2
Overview of studies comparing circulating miRNAs extraction procedures in biofluids.^a

Sample type	Volume ^b (µl)	Isolation methods ^{c,d}	Additional factors examined	Quantification	Number of targets	Reference ^e
Serum, Plasma	250	T, Q_M	Type of stabilizer (EDTA/citrate/heparin-plasma)	RT-qPCR	2 endogenous miRNAs	Fichtlscherer et al. (2010)
Serum, Plasma	400	TF_P, Q_M, S, R, T, LS	Centrifugation, stability, hemolysis, plasma vs serum, intra- and inter-assay variability, inter-individual variability	RT-qPCR	3 endogenous miRNAs + 1 spike-in	McDonald et al. (2011)
Serum	200	Q_M, TF_P, NB_R	Qiazol:serum ratio	Capillary electrophoresis, RT-qPCR	2 endogenous miRNAs + 3 spike-ins	Li and Kowdley (2012)
Plasma (CSF)	200	BIO, TF_M, TF_P, T, T_B, T_LS, T, Q_M, I, S	Different volumes of CSF, repeated phenol-chloroform extraction with NFW	Fluorometer, RT-qPCR	2 endogenous miRNAs + 3 spike-ins	Burgos et al. (2013)
Urine	20 ml	UC/Q_M, TF_P, T_LS + TF_P, E_B, NB_U	Modifications of original protocols, miRNA content in different urine fractions	Capillary electrophoresis, RNA-Seq	miRBase v19, Emsembl release 17	Cheng et al. (2014)
Plasma (CSF)	50, 100, 200	TF_M, E_R, E_B, Q_SP, T_LS	Addition of glycogen, different sample volumes	RT-qPCR	5 endogenous miRNAs + 1 spike-in	McAlexander et al. (2013)
Plasma	1000, 500, 200	TF_M, Q_SP, Q_CNA	Plasma processing delay, centrifugation, long-term stored samples (> 12 years)	Capillary electrophoresis, RT-qPCR, RT-qPCR array	2 endogenous miRNAs + 1 spike-in; 377 miRNAs	Page et al. (2013)
Plasma	200-1000 (200, 250, 300, 1000 for T_LS)	TF_P, Q_M, T_LS, NB_M	Storage conditions (temperature, time), different sample volume, precipitation time in T_LS protocol)	RT-PCR	2 endogenous miRNAs + 1 spike-in	Sourvinou et al. (2013)
Serum	200-600	Q_M, NB_M, MN	GC content bias, different sample volumes	Capillary electrophoresis, RT-qPCR array	754 miRNAs	Monieau et al. (2014)
Plasma	2000, 400	Q_CNA, E_B	Yield of co-isolated circulating DNA	RT-qPCR	2 endogenous miRNAs	Sedlackova et al. (2014)
TRIzol-inactivated plasma	400 (5, 10, 25, 50 µl of plasma)	TF_M, E_B, Q_R, Q_M, ZR_D	Addition of glycogen and linear acrylamide, different sample volumes	RT-qPCR	7 endogenous miRNAs + 3 spike-ins	Duy et al. (2015)
Plasma	200-300	E_B, NB_PS, MN, Q_SP, ZR_D	Inter-sample variability, PCR efficiencies per sample and amplicon	RT-qPCR	11 endogenous miRNAs + 5 spike-ins	Brunet-Vega et al. (2015)
Plasma	200-400	ZR_Q, ZR_D, BC, TF_P, Q_SP, TF_MM, E_B	Length-dependent bias (200-6000 bp), total RNA/DNA yield	Capillary electrophoresis, Fluorometer, RT-PCR	2 endogenous miRNAs + 6 synthetic RNAs	Li et al. (2015)
Plasma	MR	Q_SP, E_B, TF_M, MN, NB_PS	RT-qPCR systems; effect of pre-amplification, plasma vs serum, intra-individual variability within a period of one year	RT-qPCR	endogenous miRNAs + 2 spike-ins	Tan et al. (2015)
Plasma, urine	50, 100, 200	UC/Q_M, E_R, E_B	Different volume of plasma, urine (cell culture density down to 100 cells)	RT-qPCR, RT-qPCR array	7 exogenous miRNAs + 1 spike-in; 756 miRNAs	El-Khoury et al. (2016)
Serum	100-400	ZR_WB, ZR_D, Q_SP, T	Lysis reagents, storage time, serum input volume, second phenol extraction, second column elution	Spectrophotometer, Capillary electrophoresis	N/A	Unger et al. (2016)
Serum	200	Q_CNA, T_LS, Q_M, Q_SM, E_B	Kit-specific isolation bias (including other small RNA classes), RNA-Seq performance metrics	Fluorometer, Capillary electrophoresis, RNA-Seq	miRBase v21	Guo et al. (2017)
Plasma	200, 1000	TF_M, Q_SP, Q_A, NB_PS, E_B	Normalization of RT-qPCR data	RT-qPCR, RNA-Seq	6 endogenous miRNAs + 1 spike-in, miRBase v21	Kloten et al. (2019)

^a Comparative studies concerning EVs isolation are not included. For comprehensive reports see Buschmann et al. (2018) or Srinivasan et al. (2019).

^b Intervals indicate the use of volumes based on manufacturers' recommendation (MR), whereas individual values mean that the influence of different starting volume was examined.

^c The best performing protocols highlighted in bold (if applicable).

^d Individual protocols abbreviated as: BC - RNAAdvance (Beckman Coulter); BIO - MaxRecovery BiooPure RNA Isolation Reagent (BiooScientific); E_B/R - miRCURY RNA Isolation Kit - Biofluid/Cell and Plant (Exiqon); I - PureLink microRNA Isolation Kit (Invitrogen); MN - NucleoSpin miRNA Plasma (MachereyNagel); NB_M/PS/R/U - miRNA purification kit/Plasma-Serum Circulating and Exosomal RNA Purification Mini Kit/Total RNA isolation kit/Urine Exosome RNA Isolation kit (Norgen Biotek); Q_CNA/M/R/SM/SP/A - QIAamp Circulating Nucleic Acid Kit/miRNeasy/RNeasy/Qiasymphony RNA extraction kit/miRNeasy S/P/miRNeasy Advanced S/P (Qiagen); R - High Pure miRNA isolation kit (Roche); R - High Pure miRNA isolation kit (Roche); S - mirPremier (Sigma); T - TRI Reagent/Trizol/RNAzol (various vendors); T_B - TRI Reagent RT-Blood (Molecular Research Center); T_LS - TRIzol LS (Thermo Fisher Scientific); TF_M/MM/P - mirVana miRNA Isolation Kit/MagMAX/mirVana PARIS (Thermo Fisher Scientific); UC - Ultracentrifugation; ZR_D/Q/WB - Direct-zol RNA MiniPrep Kit/Quick-RNA Mini Prep/Whole-Blood RNA MiniPrep (Zymo Research).

^e Recommended studies for reading highlighted in bold.

automatize (Thatcher, 2015). The nature of specimen is usually the predominant factor in the selection process. For most common plasma and serum samples (Larrea et al., 2016), a large spectrum of commercial extraction kits is available. Next, the decision to isolate EVs or total cell-free miRNA content from plasma or serum is given purely by research purposes and technological considerations (Salehi and Sharifi, 2018; Schwarzenbach, 2015). In contrast, miRNAs in urine and cerebrospinal fluid (CSF) are typically analyzed via EVs extraction, which concentrates miRNAs to the levels suitable for quantification (Cheng et al., 2014; Whitehead et al., 2017). The concentration can also be increased by enriching for short RNA molecules. The enrichment reduces the background of other RNA molecules and may enhance sensitivity when using microarray-based technologies or northern blotting. However, the enrichment leads to losses reducing the total amount of some miRNAs, and it is likely to introduce bias. RT-qPCR and small RNA sequencing (RNA-Seq) analyses may therefore benefit from isolation of total RNA, which is usually preferred (Redshaw et al., 2013).

Since concentrations of most miRNAs are low in liquid biopsies, many efforts have been made to compare different isolation protocols (Table 2). Most studies focused on plasma/serum, although some less frequent specimens including CSF and urine have also been studied. The studies show no consensus, although kits from Qiagen (Germany) and Exiqon (a part of Qiagen now) usually ranked among the best judged by the yield, spectrum of detected miRNAs, or recovery of exogenous spike-in molecules. Most notably, majority of the studies uncovered unexpectedly large variability of repeated extractions, highlighting the need for rigorous quality control with consequences to data analysis and interpretation (Brunet-Vega et al., 2015; Burgos et al., 2013; El-Khoury et al., 2016; Klotten et al., 2019; McAlexander et al., 2013; McDonald et al., 2011). Similar results were reported already in one of the earliest studies comparing two commercially available kits against TRIzol-based extraction (Ach et al., 2008). Somewhat unexpectedly, TRIzol extraction showed significantly lower sample-to-sample variation than the kits.

Today the standard TRIzol protocol is not widely used for miRNA extraction from liquid biopsies, partly for its toxicity, but also due to a study which reported severe bias due to loss of miRNAs with stable secondary structures and low GC content (Kim et al., 2012b). The bias was documented in cell culture samples with low cell numbers and it is assumed that similar bias will arise in liquid biopsies that also have low RNA content. Possible ways to overcome this limitation is to extract the aqueous phase remaining after TRIzol lysis with commercially available spin column (Duy et al., 2015) or increase the concentration of ethanol (to 80%) in the purification step of the standard TRIzol protocol. Higher ethanol concentration reduces the solubility of the RNA pellet which consequently minimizes losses of small RNA molecules (Clerget et al., 2015). Still, studies investigating the effect of the bias introduced using TRIzol extraction are missing.

Regardless of the differences in extraction yield, all methods tend to introduce some biases because they give preference to certain miRNAs relative to others. A recent RNA-Seq-based study compared performance of five commonly used isolation kits using serum of a single healthy donor (Guo et al., 2017). The spectrum of miRNAs, but also other classes of RNAs, showed a kit-specific profile clearly clustering the samples by the extraction protocol. Moreover, up to 21% of miRNAs were uniquely identified in samples isolated by one of the tested kits, but not by the others. Even higher proportion of preferentially or exclusively isolated miRNAs (over 50%) was identified in another recent study comparing miRNA profiles in plasma, plasma EVs and urine, processed by two broadly utilized small RNA extraction kits (El-Khoury et al., 2016). These reports clearly demonstrate that each isolation kit displays inherent characteristics that introduce extraction bias and must be considered when comparing different studies.

Another layer of bias is added when EVs are purified. Noteworthy, the term "extracellular vesicles" include all extracellular particles of different size and origin released from cells that are delimited by a lipid

bilayer and cannot replicate, including exosomes, microvesicles, apoptotic vesicles, etc. (Thery et al., 2018). Depending on the isolation principle, different populations of EVs with varying cargo of miRNAs and other molecules are isolated, resulting in different profiles and clinical relevance (Chevillet et al., 2014; Taylor and Shah, 2015). Extraction techniques exploit particular features of EVs, such as density, shape, size, solubility, dispersibility or surface proteins for isolation (Li et al., 2017). A recent study compared five isolation methods for EVs and their suitability for miRNA-based biomarker discovery using RNA-Seq (Buschmann et al., 2018). The study revealed method-specific variation in the properties and miRNA composition of the isolated EVs. While profiling of EVs isolated by precipitation and membrane affinity separated patients with septic shock from controls, methods based on size-exclusion chromatography showed less successful separation. However, for other conditions, a different subpopulation of EVs may be more clinically relevant. The optimal method of EVs isolation may not be therefore the one with highest yield, nor purity, but rather the method that isolates the miRNAs that are biologically most informative of the studied condition.

An important step towards standardization of EVs-oriented research has recently been made by the ERCC program (Das et al., 2019). The consortium compared ten popular miRNA extraction methods in five specimens: plasma, serum, bile, cells and cell cultured medium (Srinivasan et al., 2019). Focus was put on methods for EVs isolation, although some non-EVs specific methods were tested as well. To minimize other confounding contributions, all samples were analyzed using the same RNA-Seq protocol (Giraldez et al., 2018) and data analysis pipeline (Rozowsky et al., 2019). The results confirmed previous observations that the spectrum of the miRNAs and also the reproducibility vary broadly across the isolation methods. Using deconvolution techniques, the authors explained the observed variation by differential preference of the methods to distinct classes of miRNAs (two classes of EVs, AGO2-associated structures, high-density lipoproteins and a lipoprotein-free fraction considered). An interactive web-based tool miRDaR (<https://exrna.shinyapps.io/mirdar/>) was launched, which provides a quality metrics for each extraction technique for any set of target miRNAs entered by the user for a given type of specimen. Of note, the deconvolution approach was further employed in another ERCC study to correct for the isolation bias and enable cross-study analysis (Murillo et al., 2019).

When a method for miRNA extraction is selected, it is advisable to perform a small pilot experiment to explore the performance of the protocol before a more complex study is initiated (McAlexander et al., 2013; Tichopad et al., 2009). The manufacturer's recommendations do not always specify the optimum volume of specimen for processing. User optimization of input material can significantly improve the yield, while maintaining the levels of inhibitors at non-interfering levels (Androvic et al., 2017; El-Khoury et al., 2016; Kim et al., 2012a; McAlexander et al., 2013; Sourvinou et al., 2013). Addition of carriers may improve isolation yield and is frequently recommended. Common carriers are MS2 phage RNA, yeast tRNA, glycogen, BSA and linear acrylamide (Ramon-Nunez et al., 2017). RNA-based carriers are not suitable for RNA-Seq based workflows, since they may consume sequencing reads (Buschmann et al., 2016). Another important effect of carriers is the reduction of extraction variability and improvement of measurement reproducibility (Andreassen et al., 2010; Androvic et al., 2017; McAlexander et al., 2013).

After extraction, the RNA pellet is resuspended or the RNA is released from the isolation columns and stored for downstream analysis. Methods using small elution volumes are preferred as RNA at higher concentration is more stable upon storage and more easily tested for quality by analytical techniques. Higher miRNA concentration also allows larger miRNA input in downstream analysis, which maximizes the spectrum of reliably quantified miRNAs (Moret et al., 2013). The compatibility of the elution medium with downstream steps must be validated. TE-buffer or nuclease-free water are common choices

(Schrader et al., 2012). Lastly, since miRNAs are unprotected after isolation and susceptible to degradation (Aryani and Denecke, 2015; Ludwig et al., 2017), general rules for RNA manipulation and storage must be followed, including the use of RNase-free reagents and consumables, the addition of RNase inhibitors, keeping samples on ice during manipulation and storage of samples in aliquots at -80°C .

To sum up, miRNA extraction is a key step in the miRNA analysis workflow. The careful selection and optimization of the extraction procedure is highly recommended to minimize variability. The application of the same extraction method throughout a study is mandatory. An alternative approach represents direct detection of miRNAs in liquid biopsies without RNA extraction. The concept is already utilized commercially, e.g. HTG EdgeSeq miRNA Whole Transcriptome Assay (HTG Molecular Diagnostics, USA - see section 5.4) or Firefly particle technology (Abcam, UK). The advantages of these platforms include reduced turnaround time, ease of use and no material loss during extraction. On the other hand, lower sensitivity, accuracy and in the case of Firefly technology also reduced reproducibility compared to RNA-Seq represent potential drawbacks (Godoy et al., 2019).

4. Quality control

Quality control is essential in any gene expression study (Bustin et al., 2009). Controls should test for confounding technical variation introduced at different experimental steps and may allow to mitigate their effect in the data analysis (Becker et al., 2010). As demonstrated, important technical variation is introduced in the sampling and extraction steps.

During specimen collection, care must be taken to avoid cell lysis that would contaminate samples with cellular miRNAs (Kirschner et al., 2013a). This is particularly important for specimens rich in cells, such as blood. Standard quality control for hemolysis includes visual inspection for pink color or spectrophotometric measurement of hemoglobin absorbance at 414 nm (Pizzamiglio et al., 2017). Since other substances may contribute to absorbance at 414 nm, additional spectrophotometric assessment may help, e.g. 385 nm is used for lipemia indication, which increases absorption at 414 nm (Tiberio et al., 2015). Extended hemolysis control is based on measuring miRNAs enriched in different blood cell types and normalization of their levels to stable, hemolysis-insensitive miRNAs. This ratio indicates the increased lysis of particular cell type relative to normal, e.g. the ratio miR-23a/miR-451 indicates lysis of erythrocytes and miR-23a/miR-425 lysis of platelets (van Vliet et al., 2017). Some authors even define cut-off values to identify hemolyzed specimens, e.g. absorbance at 414 nm > 0.2 (Kirschner et al., 2011; Shah et al., 2016), or delta of quantification

cycles ΔCq (miR23a-miR451) > 5 indicating possible hemolysis and ΔCq (miR23a-miR451) > 7 indicating high risk of hemolysis (Blondal et al., 2013). However, these values are species- and assay-dependent, and the ranges need to be calibrated for every new study (Andrović et al., 2019). Samples that do not meet pre-set quality criteria should be excluded from analysis (Pritchard et al., 2012b). However, some contamination may be missed by this assessment of quality. A recent study demonstrated release of vesicle-associated miR-16 and miR-21 from blood cells during inappropriate storage of whole blood, which was independent of hemolysis and may not be reflected by ΔCq (miR23a-miR451) (Koberle et al., 2016).

The efficiency of the extraction procedure can be tested using spike-in controls. A spike-in control is a foreign small RNA molecule of same length as native miRNAs without sequence homology to the endogenous miRNAs. It can have a completely artificial sequence or a sequence from an unrelated species, such as *Caenorhabditis elegans* and *Arabidopsis thaliana*. The spike-in is added to the sample prior to extraction and its levels are measured at different stages of the workflow to identify any steps of severe material losses (Li and Kowdley, 2012; Mitchell et al., 2008). The procedure may be expanded to control also for the presence of inhibitors by adding a second set of spike-ins before the reverse transcription (RT) step (Fig. 2). As the extraction efficiency may be sequence-dependent, single spike-in should not be used for post-hoc correction of the data, but only for identification of problematic samples and steps introducing technical variability. To minimize the issue, complex mixtures of miRNA spike-ins are used. They are typically applied for quality control of high-throughput data (Locati et al., 2015; Lutzmayer et al., 2017).

Methods for reliable quantification of total miRNA concentration from liquid biopsies are currently not available. Instead, selected endogenous miRNAs expected to be present at a stable level are measured to assess the extraction procedure (Brunet-Vega et al., 2015; Duy et al., 2015). Their analysis is typically combined with exogenous spike-ins using RT-qPCR before the samples are processed by costly high-throughput methods such as microarrays or RNA-Seq (Andrović et al., 2019; Blondal et al., 2013). Endogenous miRNA controls should be present at high level that is not appreciably affected by contamination with cellular material. miRNAs that may be suitable as endogenous positive controls are collected in databases, such as miRandola (<http://mirandola.iit.cnr.it/>) and in literature (Russo et al., 2018). An excellent source is a recently updated version of the Human miRNA Tissue Atlas including 253 whole blood samples, 66 fractionated blood cell isolates, 72 serum samples, 278 plasma samples, 29 urine samples, and 16 saliva samples from different collection and storage conditions (Fehlmann et al., 2016). An alternative approach to test the quality of the

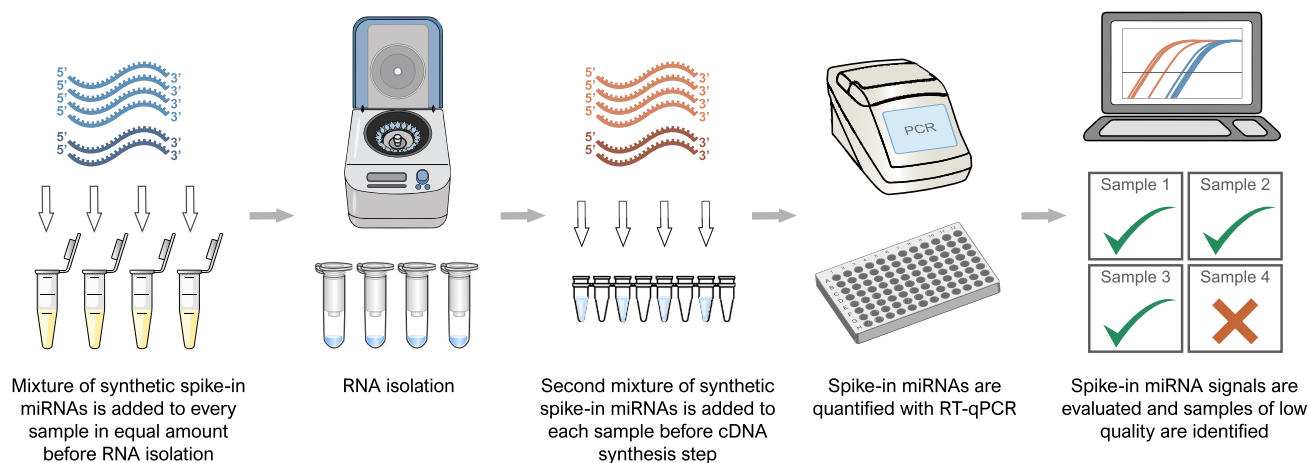


Fig. 2. Concept of miRNA spike-in quality control in a typical RT-qPCR workflow. Samples are spiked with known amounts of exogenous miRNAs before extraction and RT. Quantification of the spike-in levels identifies material losses, inhibition and steps introducing severe variability. The approach can be used in routine analysis to identify deviating samples.

extraction is to prepare a reference sample of high quality and compare levels of selected stable miRNAs in test samples to this standard (Ibberson et al., 2009).

Highly optimized and validated RT-qPCR assays are typically used to monitor the spike-ins, to assess hemolysis, presence of contaminating cellular miRNAs and positive endogenous miRNA controls (Androvic et al., 2019; Blondal et al., 2013). Such panels are excellent tools for the optimization and validation of sample processing and isolation protocols during method development and as tools to identify low quality samples during routine analysis (Fig. 2).

5. Quantification

Quantification of individual miRNAs possesses several challenges arising from their natural properties (Pritchard et al., 2012a). The highly variable GC content of miRNAs impedes efficient hybridization of detection probes and their length is too short to harbor a pair of standard PCR primers. miRNAs also lack a common sequence, such as the poly(A) tail of mRNAs, that could be used for selective capturing and priming of RT. For the analysis, miRNAs have to be extended by artificial sequences via specific primers, ligation of linkers or polyadenylation. However, the efficiency of these enzymatic steps is strongly dependent on the miRNA base composition and terminal sequence heterogeneity, referred to as isomiRs. This leads to severe method-dependent bias, with some miRNAs quantified preferentially, while other show low yields and some may even escape detection completely.

Ability to distinguish and quantify isomiRs is another challenge. IsomiRs comprise groups of miRNAs typically originating from the same precursor miRNA, but having different length and/or base composition, particularly at the 3' and less so at the 5' termini (Nielsen et al., 2012). In recent years, several reports acknowledged their importance as biomarkers (Koppers-Lalic et al., 2016; Telonis et al., 2015), therefore the ability to characterize the complete isomiR spectrum can be important for certain applications (Guo and Chen, 2014). The precise and discriminative quantification of closely related miRNAs, often differing in just a few nucleotides (e.g. let-7 and miR-302 family), or distinguishing mature miRNAs from their precursor (pri- and pre-miRNAs) represent additional challenges. Despite these complications, three major approaches have been adapted for miRNA quantification and are used with different frequency in cancer diagnostics: microarrays, RT-qPCR and RNA-Seq.

5.1. Microarrays

Microarray was the first technology that allowed high-throughput analysis of miRNAs (Liu et al., 2004). In the standard setting, microarray analysis consists of three steps: i) labelling of the miRNAs, ii) hybridization of the miRNAs to immobilized probes, and iii) scanning and quantification of the signal. Several variations of each step were developed during last two decades to improve the performance of the method (Love and Dave, 2013; Yin et al., 2008).

However, from the very beginning several inherent drawbacks of the microarray technology were evident. First, microarray hybridization is a semi-quantitative method, mostly used in a differential set-up (healthy vs cancer samples). Any absolute quantification is limited, or requires validation by other methods (typically by RT-qPCR). The method exhibits a limited linear dynamic range causing fold change compression and subsequent underestimation of real changes in miRNA levels (Dong et al., 2013; Pritchard et al., 2012a), although recent technology improvements, such as the elevated hybridization spots on the 3D-Gene platform (Toray, Japan) that allows for agitation, have increased the sensitivity and dynamic range dramatically (Nagino et al., 2006; Sato et al., 2009; Sudo et al., 2012). Microarrays require rather high RNA input (typically 100–2000 ng) which may be a limiting factor when analyzing liquid biopsies (Ono et al., 2015). Another drawback is

that hybridization is performed at one particular temperature, which makes the probe design challenging due to short length and variable GC content of miRNAs. The introduction of chemically modified nucleotides, such as locked nucleic acids (LNA) helped to mitigate this issue (Beuvink et al., 2007; Castoldi et al., 2006). Probes with LNA bases display enhanced hybridization properties including increased melting temperature (relative to probe length) and mismatch discrimination, allowing more flexible probe design (Castoldi et al., 2006; Vester and Wengel, 2004).

Hybridization strategies can only be used to analyze known miRNAs, as probes must be designed. Probes are typically based on the most recent version of the miRBase Sequence Database (Griffiths-Jones et al., 2006), which is periodically updated. This makes cross-study analysis complicated because the number of targets may differ, which may also impact data normalization if global strategies are used. Bias may also be introduced by unequal efficiencies of the labelling step of the different miRNAs (Hunt et al., 2015). Traditional enzymatic labelling using T4 RNA ligase or 3' tailing introduces significant sequence bias (Hafner et al., 2011; Jayaprakash et al., 2011; Zhuang et al., 2012), and alternative approaches such as chemical alkylation and platinum coordination chemistry have been developed (Pritchard et al., 2012a). All techniques to some degree label also other RNA molecules. These may bind to capture probes and increase background. Unspecific hybridization to other RNA species can be reduced by enriching for small RNAs using column or gel purification methods (Yin et al., 2008).

Despite many drawbacks, microarrays have been successfully applied in many studies and they remain popular even with the advent of other high-throughput methods. Ease of use, availability of equipment and expertise, and large amounts of reference data available for comparisons are the main reasons. The technology is quite standardized and various guidelines are available (Ball et al., 2002; Brazma et al., 2001; Chen et al., 2007). For these reasons, microarrays remain an attractive choice for large screening studies when the amounts of material are sufficient, there is no desire to detect novel miRNAs or isomiRs, and standardization and robustness are priorities rather than sensitivity and specificity.

5.2. RT-qPCR

RT-qPCR is considered gold standard for miRNA analysis. It offers high sensitivity and high specificity with large dynamic range (Chugh and Dittmer, 2012). Further, the complete analysis is fast, procedure is easily adaptable to any laboratory familiar with quantitative PCR (qPCR) and data are analyzed with well-established workflows (Kubista et al., 2006). It is also possible to analyze just selected miRNAs, which dramatically reduces cost, when only some are of interest. RT-qPCR results can be calibrated for absolute quantification and miRNAs differing in just a few nucleotides can be distinguished. On the other hand, high-throughput studies may be financially demanding and time consuming, although array-formats are also available. Similar to microarrays, only known miRNAs and limited spectrum of isomiRs can be quantified (Benes and Castoldi, 2010).

The essential steps of the RT-qPCR workflow are 1) RT producing cDNA and 2) cDNA amplification and quantification by qPCR. Since the short length of miRNAs does not allow the use of two standard PCR primers (each 18–20 bases long), the miRNAs must be extended. There are two strategies: i) extension of the miRNAs during cDNA synthesis using specific RT-primers and ii) extension of the miRNAs with a universal sequence prior to cDNA synthesis, which makes it possible to perform RT with a universal RT-primer (Table 3).

Specific RT-primers have 3' ends complementary to the miRNA sequence, whereas their 5' ends serve as the extensions. The specific RT-primers may be linear (Raymond et al., 2005), but more often are designed to form a secondary structure that improves sensitivity and specificity of the assay (Androvic et al., 2017; Chen et al., 2005; Honda and Kirino, 2015). When a universal sequence is added directly to the 3'

Table 3
Examples of available RT-qPCR chemistries for miRNA analysis.

Company/Reference	Name	RT approach	qPCR approach	Fluorescence detection
Thermo Fisher Scientific, USA/Chen et al. (2005)	TaqMan miRNA assays	Specific stem-loop RT primer	Specific FW primer + universal RV primer	Hydrolysis probe
Thermo Fisher Scientific, USA	TaqMan Advanced miRNA assays	Poly(A) tailing + 5' adapter ligation + oligo-dT with universal tag	Universal FW primer + universal RV primer	Hydrolysis probe
Exiqon (now part of Qiagen, Germany)	miRCURY LNA	Poly(A) tailing + oligo-dT with universal tag	Specific FW primer with LNA + universal oligo-dA primer	SYBR Green
Quantabio, USA	qScript microRNA cDNA Synthesis Kit	Poly(A) tailing + oligo-dT with universal tag	Specific FW primer + universal RV primer specific for the tag sequence	SYBR Green
Qiagen, Germany	miScript II RT Kit + miScript primer assays	Poly(A) tailing + oligo-dT with universal tag	Specific FW primer + universal RV primer specific for the tag sequence	SYBR Green
Takara Bio, Japan	Mir-X miRNA First-Strand Synthesis Kit	Poly(A) tailing + oligo-dT with universal tag	Specific FW primer + universal RV primer specific for the tag sequence	SYBR Green
Sigma-Aldrich (now part of Merck, Germany)	MystiCq microRNA qPCR System	Poly(A) tailing + oligo-dT with universal tag	Specific FW primer + universal RV primer specific for the tag sequence	SYBR Green
Benes et al. (2015)	miQPCR	3' adapter ligation + universal RT primer	Specific FW primer + universal RV primer	SYBR Green
TATAA Biocenter, Sweden/Androvic et al. (2017)	Two-tailed RT-qPCR	Specific dual probe RT primer	Specific FW primer + specific RV primer	SYBR Green
Honda and Kirino (2015)	Dumbbell-PCR	5' and 3' specific stem-loop adapter ligation	Specific FW primer + specific RV primer	Hydrolysis probe

end of miRNAs, either through polyadenylation by polyadenylate polymerase (PAP) or ligation of an adapter by T4 RNA ligase (Balcells et al., 2011; Benes et al., 2015; Shi and Chiang, 2005), an universal RT primer can be used. The combination of 5' end adapter ligation and 3' end polyadenylation has been also introduced (TaqMan Advanced miRNA Assays, Thermo Fisher Scientific, USA).

Each group of methods has its advantages and limitations. Specific RT-primers bind only one miRNA and increase the specificity of the assay. Secondary structures introduced into RT-primers reduce non-specific binding, which lowers the background (Benes and Castoldi, 2010). The main limitation of this approach is that only miRNAs that are targeted with RT-primers are quantified. Hence, if the researchers want to add new targets, they have to revert back to the RNA (rather than starting from the cDNA) and include additional primers in the RT reaction. The complex design of structured RT primers and higher cost of some commercial assays also represent a limiting factor, although cost-efficient alternatives were recently introduced (Androvic et al., 2017).

The possibility to reverse transcribe all miRNAs in a single reaction is the greatest advantage of universal RT-priming. On the other hand, the variable efficiency of the enzymatic step due to sequence preferences is a main issue of these methods. Particularly ligation introduces substantial sequence bias (Hafner et al., 2011; Jayaprakash et al., 2011; Zhuang et al., 2012). The efficiency of the reaction is also significantly lower for miRNAs with chemically modified terminal nucleotides (Munafu and Robb, 2010). As polyadenylation and ligation are not specific to miRNAs, other RNA molecules are also extended and reverse transcribed, which increases background. Since the universal RT-primers reverse transcribe all miRNAs, they do not contribute to the specificity of the analysis. Specificity is then conferred by a single qPCR primer, as the second qPCR primer is typically universal and binds to the extension sequence. As a consequence, specificity of microRNA assays relying on global cDNA synthesis is lower than of those based on specific RT priming, which are preferred when closely related miRNAs shall be distinguished. To improve specificity, primers containing LNA bases with improved hybridization properties were introduced (Raymond et al., 2005; Vester and Wengel, 2004). However, higher price and complicated design hinder their wide utilization (Benes and Castoldi, 2010).

Whichever the RT strategy, cDNA is subsequently amplified using standard qPCR with either dye or probe based detection. Using dyes lowers the cost and allows assessment of the reaction specificity by melting curve analysis (Benes and Castoldi, 2010). At the same time, there is a certain risk of false positivity, as the unspecific products may sometimes be undistinguishable from specific amplicons (Zipper et al., 2004). As an alternative, hydrolysis probes binding to a specific site in the target amplicon giving rise to specific fluorescence can be used. They were introduced for miRNA applications by Chen et al. (2005), commercially provided as TaqMan microRNA assays (Thermo Fisher Scientific, USA). However, in this system the hydrolysis probes bind to a universal sequence present in all RT primers and thus do not contribute to the specificity of the reaction (Androvic et al., 2017). The probe design is improved in the newer generation of TaqMan advanced microRNA assays (Thermo Fisher Scientific, USA).

In general, RT-qPCR-based miRNA analysis is preferred for smaller studies analyzing a predefined set of miRNAs, for validation, applications that require absolute quantification, and for routine diagnostics when high sensitivity and/or specificity is required. Several experiment-specific requirements shall be considered to select the most suitable RT-qPCR method. In our hands, polyadenylation combined with dye based qPCR provides a good balance between cost and performance, although optimization and occasionally primer redesign may be necessary to obtain good results. However, this is the case with any system currently available. If a high level of specificity is required or there is need to capture the complete isomiR repertoire of a given miRNA, the recently published Two-tailed RT-qPCR becomes the

Table 4
Overview of commercially available small RNA-Seq kits.

Company/Reference	Library prep kit	Technique used	Maximum no. of available barcodes
Illumina, USA	TruSeq Small RNA Library Prep Kit	Ligation of two adapters	48
New England Biolabs, USA	NEBNext Multiplex Small RNA Library Prep Set	Ligation of two adapters	48
SeqMatic, USA	TailorMix miRNA Sample Preparation Kit	Ligation of two adapters	96
Norgen Biotek Corp., Canada	Small RNA Library Prep Kit	Ligation of two adapters	48
Lexogen GmbH, Austria	Small RNA-Seq Library Prep Kit	Ligation of two adapters	96
TriLink BioTechnologies, Inc., USA/ Shore et al. (2016)	CleanTag Small RNA Library Prep Kit	Ligation of two adapters with chemical modifications	48
Bioo Scientific, USA/ Baran-Gale et al. (2015)	NEXTflex Small RNA Sequencing Kit	Ligation of two adapters with randomized nucleotides	48
Qiagen, Germany	QIAseq miRNA Library Kit	Ligation of two adapters and UMI correction	96
Takara Bio, Japan	SMARTer smRNA-Seq Kit	Poly(A) tailing and template switching	96
Diagenode, Belgium	CATS Small RNA-Seq Kit	Poly(A) tailing and template switching	24
Somagenics, USA/ Barberan-Soler et al. (2018)	RealSeq-AC and RealSeq-biofluids Kit	Single adapter ligation and circularization	48
Takara Bio, Japan	SMARTer microRNA-Seq Kit	Single adapter ligation and circularization	96

method of choice ([Androvic et al., 2017](#)). Finally, if only a single specific isomiR shall be detected, the Dumbbell-PCR is currently the only method available ([Honda and Kirino, 2015](#)).

5.3. RNA-Seq

RNA-Seq holds great promise to become a leading technology for miRNA research ([Hunt et al., 2015](#)). Unlike older technologies, RNA-Seq neither requires targets to be known nor probes to be designed. It can be used for *de novo* analysis including the discovery of novel miRNAs, isomiRs, edited miRNAs and even other classes of small RNAs. It is a high-throughput technology, currently allowing convenient multiplexing of up to 96 samples with commercially available kits ([Table 4](#)) and even more with in-house technologies ([Persson et al., 2017](#)). Sensitivity and specificity is higher and dynamic range is broader than that of classical microarrays (there is no fold change compression), though it may not match optimized RT-qPCR ([Chugh and Dittmer, 2012](#)). Drawbacks of the technology include complex and time-consuming library preparation, bias introduced during the workflow (see below), impossibility to perform absolute quantification, challenging data analysis requiring computational infrastructure and bioinformatics skills, and lack of standardization ([Pritchard et al., 2012a](#)). Cost may also be higher, although it depends on the number of samples and targets the users wish to analyze and scales favorably with increasing level of multiplexing. Another possible source of error is RNA degradation products, which may cause errors during data analysis ([Ludwig et al., 2017](#)).

Several protocols for miRNA library preparation are available. The most common protocol employs two sequential ligations of adapters to the 3' and 5' ends of the miRNAs ([Fig. 3A](#)) ([Hafner et al., 2008](#)). The protocol takes advantage of the 5' phosphate and the 3' hydroxyl groups on the miRNA termini to selectively target and enrich small RNA species using ligases that recognize these terminal groups. After ligation of the adapters, a universal RT-primer complementary to the 3' adapter is used for cDNA synthesis. The cDNA library is then PCR-amplified using primers complementary to the adapters. These primers also introduce the flow-cell binding sequences and sample-specific barcodes. The final amplified library typically consists of approximately 120 bp of adapter sequences plus an insert of the original miRNA sequence of 20–30 bp, which makes a total of 140–150 bp. Longer products are generated from adapter ligation to non-miRNA species, including tRNAs, snoRNAs, piRNAs and other RNAs having 3' hydroxyl and 5' phosphate termini. These can either be retained as part of the library or removed by bead-based size selection or polyacrylamide gel electrophoresis (PAGE) purification ([Buschmann et al., 2016](#)).

An issue with the ligation-based library preparation methods is

accumulation of adaptor dimers. This is an undesired side product formed by the direct ligation of a 3' and 5' adaptor without an RNA insert. The adaptor dimers amplify during the PCR step and may consume a substantial portion of the sequencing reads. This is particularly serious when RNA concentration is low, such as in liquid biopsy samples. There are strategies to reduce adapter-dimer formation. These include digestion of excess 3' adaptors with ssDNA-specific exonuclease RecJ ([Pease, 2011](#); [Xu et al., 2015](#)), using chemically modified adaptors that do not ligate without an insert ([Shore et al., 2016](#)), binding the excess of 3' adaptors to the RT-primer forming double stranded product that is less effective substrate for ligation to single stranded 5' adaptor ([Vigneault et al., 2012](#)), adding blocking oligonucleotide specific to adaptor dimers that interferes with their RT ([Kawano et al., 2010](#)), CRISPR/Cas9-targeted removal of adaptor dimers ([Hardigan et al., 2019](#)) or traditional PAGE or bead-based purification ([Head et al., 2014](#)).

Library preparation procedure introduces substantial bias. Ligation is considered the dominant source ([Raabe et al., 2014](#)). Ligation bias arises due to the variable efficiency of the adapter ligation, which is dependent on the sequence composition and secondary structure ([Hafner et al., 2011](#); [Jayaprakash et al., 2011](#); [Zhuang et al., 2012](#)). Consequently, many miRNAs are over or under represented in the data, and some may even completely drop out. The measured values provide therefore a distorted picture of the biological miRNA levels and the quantification must be performed relative to a reference or standard sample (i.e. case vs control). Another issue is that heavily over-represented miRNAs may consume substantial portion of the sequencing reads and higher sequencing depth may be needed to detect low-abundant and underrepresented miRNAs. The problem is particularly pronounced in liquid biopsies samples. Here, often a small number of miRNAs are highly abundant and dominate the sequencing reads ([Godoy et al., 2018](#)) and it is hard to separate the contribution of ligation bias from true variation in the levels of miRNAs ([Baroin-Tourancheau et al., 2019](#)).

To reduce the ligation bias, three approaches have so far been developed. One method uses adaptors with a stretch of random nucleotides at the ends ([Fig. 3B](#)) ([Baran-Gale et al., 2015](#)). The rationale is that a preferred pair of random adaptor sequences is present for each miRNA and the ligation efficiency for the different miRNAs will be equalized. The second method omits the ligation step and employs polyadenylation and a template-switching mechanism to incorporate the 3' and 5' adaptor sequences ([Fig. 3C](#)). As all poly-adenylated RNAs are processed into libraries, purification of the short-length fragments is mandatory with this approach. A problem also appears in the analysis of the sequencing data, as it is not possible to distinguish native adenines from those introduced by polyadenylation ([Barberan-Soler et al., 2018](#);

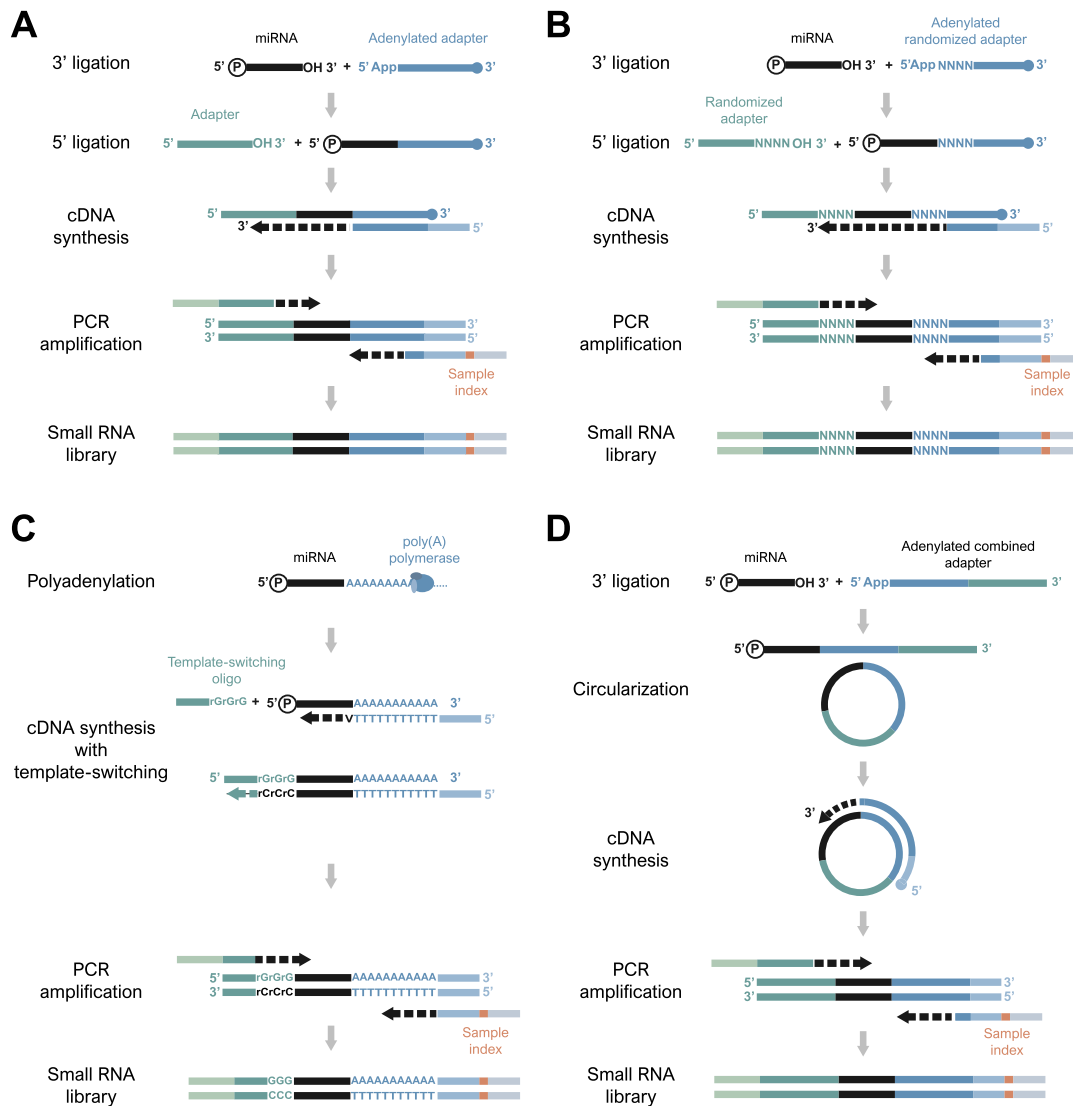


Fig. 3. Schematic representation of alternative protocols for small RNA library preparation. A) Sequential ligations of two defined adapters (Hafner et al., 2008). B) Sequential ligation of two adapters with randomized nucleotides (Baran-Gale et al., 2015). C) Poly(A) tailing and template switching. D) Single adapter ligation and circularization (Barberan-Soler et al., 2018).

Coenen-Stass et al., 2018; Dard-Dascot et al., 2018). The third method relies on the ligation of a single 3' adaptor and subsequent circularization of the molecule (Fig. 3D). The intramolecular circularization reaction is more efficient than intermolecular 5' ligation and introduces much less bias (Barberan-Soler et al., 2018).

Other sources of bias, though considered less important than the ligation bias, are the RT and in particular the reverse transcriptase used, number of PCR cycles, variation in gel size selection, library quantification, flow cell loading, sequencer-specific bias, lane and flow cell effects - reviewed in Buschmann et al. (2016). PCR bias, which is considered important in most RNA-Seq workflows, has been claimed negligible in small RNA-Seq protocols by several authors, possibly due to even length of cDNA library for PCR amplification (Fuchs et al., 2015; Hafner et al., 2011; Jayaprakash et al., 2011). However, recent studies call this into question (Fu et al., 2018; Wright et al., 2019). These studies assessed the impact of PCR bias via usage of unique molecular identifiers (UMIs), which allow identification of PCR duplicates. Both studies demonstrated increased quality of UMI-corrected data, advocating for their routine use in small RNA-Seq experiments.

Another source of error associated particularly with RNA-Seq, but

also with microarrays or RT-qPCR, is between-run technical variation, also known as batch-effects. Because even slight deviations from the workflow can have significant impact on the measured values, users should take great care to minimize batch variations with proper experimental design including balanced representation of sample groups between run batches (Buschmann et al., 2016).

Several recent studies have compared RNA-Seq protocols (Barberan-Soler et al., 2018; Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Giraldez et al., 2018; Wright et al., 2019; Yeri et al., 2018). In general, all protocols showed similar performance in terms of sensitivity, reproducibility, accuracy or diversity of detected miRNAs, but strongly differed in the ability to capture true representation of miRNA levels, proportion of miRNA reads in the libraries, or detection of isomiRs. According to a recent study the largest part of ligation bias was reduced with the library preparation based on intramolecular circularization (Barberan-Soler et al., 2018). By reducing ligation bias the method allowed detection of highest number of miRNAs as well as provided high accuracy across several logs of dilution. However, the comparison was made on tissue samples rich in miRNAs not being representative for liquid biopsy samples. The polyadenylation-based techniques did

mitigate ligation bias, but introduced other artefacts, including low mapping rates to miRNAs and false positive isomiR detection (Barberan-Soler et al., 2018; Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Wright et al., 2019). The protocol using adapters with random nucleotides at the ligation boundary showed consistently good results across the majority of comparison studies. It did reduce ligation bias, allowed detection of similar number of miRNAs as other methods, provided high proportion of miRNA reads or showed low level of false isomiR detection (Barberan-Soler et al., 2018; Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Giraldez et al., 2018; Wright et al., 2019; Yeri et al., 2018). In conclusion, at present time the protocols using adapters with random nucleotides represent the well-performing, established and validated approach. Noteworthy, this approach has been selected by the ERCC program as standard for their RNA-Seq based studies (Das et al., 2019; Giraldez et al., 2018).

Small RNA-Seq is a rapidly developing technology still undergoing important improvements. Currently, small RNA-Seq finds its application mainly for large screens and discovery studies that benefit from its power to cover the complete miRNA diversity, including the detection of novel molecules with single nucleotide resolution. Once the technology becomes standardized and user-friendly tools, including software for data analysis, become more common, small RNA-Seq will grow even more in popularity.

5.4. Other techniques

The traditional high-throughput techniques for miRNA detection are optimal for discovery studies focused on identification of candidate diagnostic and prognostic miRNAs. For routine applications, particularly if fast turnaround time is requested, they are less suitable due to their rather complicated setup and data analysis (Anfossi et al., 2018).

New approaches more suitable for the standardization required in routine applications are emerging. Examples include the HTG EdgeSeq miRNA Whole Transcriptome Assay (abbr. EdgeSeq; HTG Molecular Diagnostics, USA) and nCounter system (NanoString Technologies, USA). Both are based on hybridization for recognition of targets, but the hybridization is performed in solution which increases its efficiency (Geiss et al., 2008; Tsang et al., 2017). EdgeSeq then performs targeted RNA-Seq for readout. A large pool of capture probes is used to bind target miRNAs (> 2000) by hybridization and single strand specific nuclease is used to removes all unbound probes. The remaining probes are then amplified, barcoded and sequenced. Data analysis, which is performed in GenEx software (MultiD, Sweden), is standardized and includes several quality controls leading to reliable results that are comparable across laboratories. Compared to conventional small RNA-Seq, the Edgeseq workflow completely avoids ligation bias (Girard et al., 2016; Lizarraga et al., 2016) and allows analysis of crude biofluids (down to 15 μ l), which also eliminates bias in the extraction procedure.

nCounter relies on multi-step hybridization of miRNAs to probes followed by splinted ligation. Once targets are ligated to miRNA-specific extension sequences, streptavidin capture probes co-hybridize with color-coded reporter probes (> 800) to create target-probe complexes. These are immobilized and, after several washing steps, the number of fluorescent barcodes, which corresponds to the number of miRNA targets, is counted. Therefore, unlike microarrays, nCounter offers digital counting of the miRNA copies, which leads to higher precision and wider dynamic range (Pritchard et al., 2012a). The method requires neither RT nor amplification, and miRNAs, mRNAs, DNA and proteins can be measured simultaneously (3D Biology Technology). A recent publication compared the EdgeSeq and nCounter platforms (Godoy et al., 2019). Although both platforms showed lower bias than traditional RNA-Seq, their performance varied in several aspects that may affect their utility as assay systems for clinical applications.

Instruments for routine diagnostic use must be user-friendly, standardized, include controls and compliant with regulations such as IVDR

in Europe (*In Vitro* Diagnostic Regulation) and FDA (Food and Drug Administration) in the US. These instruments analyze a limited number of validated analytes, without upstream labeling and other laborious processes, provide quick response and robust results (Hunt et al., 2015). A number of methods and detection systems are being developed, most based on biosensors that can roughly be categorized as electrochemical or optical-based platforms and are described in several recent reviews (Degliangeli et al., 2014; Hamidi-Asl et al., 2013; Johnson and Mutharasan, 2014; Kilic et al., 2018).

5.5. Inter-platform comparison

In last two decades, several miRNA quantification platforms have been developed based on hybridization, RT-qPCR and sequencing. Since the very beginning there have been efforts to compare their performance and concordance (Ach et al., 2008; Blondal et al., 2017; Git et al., 2010; Jensen et al., 2011; Kelly et al., 2013; Kolbert et al., 2013; Leshkowitz et al., 2013; Meyer et al., 2012; Pradervand et al., 2010; Wang et al., 2011). However, these reports compared only some of the techniques using rather limited test material and data was evaluated with poorly defined performance metrics that cannot be extrapolated for comparison with other studies or for other applications. The microRNA quality control (miRQC) study was therefore initiated to provide a comprehensive evaluation of quantitative miRNA expression platforms (Mestdagh et al., 2014). The miRQC involved all major vendors of miRNA profiling technologies, who all received the same set of 20 samples that were analyzed for > 196 miRNAs. In total, 12 platforms (seven RT-qPCR, three hybridization-based arrays and two RNA-Seq workflows) from nine vendors were included.

Considering all the tested parameters, there was no obvious single superior technology. Still, some metrics were technology related. RT-qPCR showed superior sensitivity, particularly for low RNA input samples. Higher sensitivity was accompanied by better accuracy and more reliable results. Hybridization based methods suffered from low sensitivity even when RNA input was not limited. RNA-Seq technology demonstrated rather concentration dependent sensitivity. It was not compromised at high input RNA concentration, but decreased with low RNA input amount. The most striking variation between platforms was the poor agreement on differently expressed microRNAs detected. On average, only 54% of the differently expressed targets were in concordance between any two quantification systems. This underscores the necessity to validate results achieved on one platform with an alternative platform or technology. Of note, there was a considerable variability in specificity and reproducibility even between platforms based on same technology. Thus some platforms are not suitable to measure small expression changes, particularly when the number of samples is limited. Negative correlation was observed between the sensitivity and the specificity, i.e. platforms that detected higher number of miRNAs also showed higher number of false positives.

In summary, each technology has its advantages and limitations. The platform should therefore be selected based on the particular study requirements and questions addressed (Fig. 4). The validation of results by an independent platform or technology is always recommended.

6. Data analysis

Data analysis represents an important step in any miRNA study. Depending on the technology, data analysis has a varying level of complexity and different critical points that need to be addressed to obtain reliable data (Gao and Jiang, 2016; Pritchard et al., 2012a; Witwer and Halushka, 2016). Although each platform includes technology-specific data processing steps, these can be broadly categorized as: data pre-processing and quality control, data normalization, and secondary analysis.

The first assessment of data quality is performed during data pre-processing. It includes analysis of internal controls, that may be probes

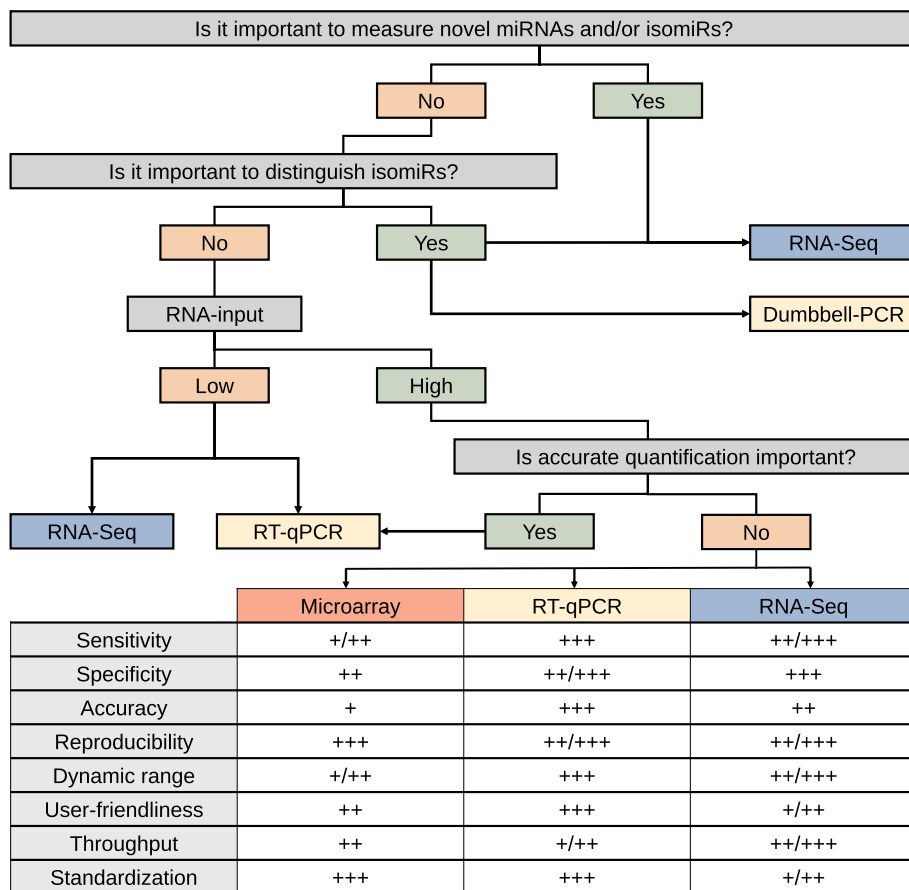


Fig. 4. Selection of miRNA profiling platform. Decision tree to guide informed choice of miRNA quantification platform. The structure was adapted from Pritchard et al. (2012a). The metrics are based on the data from the miRQC study (Mestdagh et al., 2014), recent comparison studies in the RNA-Seq field (Barberan-Soler et al., 2018; Coenen-Stass et al., 2018; Dard-Dascot et al., 2018; Giraldez et al., 2018; Wright et al., 2019; Yeri et al., 2018) and data obtained in the CANCER-ID (www.cancer-id.eu) and SPIDIA4P (www.spidia.eu) consortia.

randomly distributed on the microarray chips, negative and positive samples, replicates, melting curves in RT-qPCR, or examination of various parameters reflecting the quality of a sequencing run such as the number of identified reads, quality score distribution, flow cell loading, presence of over-represented sequences, etc. Data are then pre-processed to remove technical artefacts, identify unreliable measurements, and produce a matrix of measured target quantities.

The complexity of data pre-processing correlates with the maturity of a given technology (Cristiano and Veltri, 2016). Instruments for well-established microarrays as well as for RT-qPCR are today provided with software that performs most of the procedures with minimal user intervention such as identification of spots, probe intensity extraction, and background subtraction in microarrays; amplification curve modelling, fluorescence normalization or baseline subtraction, threshold setting and Cq readout for qPCR. In contrast, RNA-Seq data-pre-processing is time-consuming, computationally demanding and not fully standardized. It includes the trimming of sequencing reads from adapters, removal of sequences with inadequate length or low quality, and finally alignment to a reference sequence (Fu and Dong, 2018). miRNA reads are typically mapped against a reference genome or miRBase (Buschmann et al., 2016). Mapping against the genome sequence provides the most comprehensive view on the data and allows for the discovery of novel miRNAs. Mapping against miRBase or similar databases is significantly faster and avoids issues with reads mapped to multiple genomic locations, which can introduce severe bias if handled inappropriately (Taub et al., 2010). The stringency of the mapping may influence the rate of false positives and introduce bias due to erroneous mapping to other classes of small RNAs, degraded mRNAs and rRNAs (Ludwig et al., 2017). A specific challenge is the mapping of isomiRs, as a comprehensive database covering those is yet to be developed (Desvignes et al., 2015). After mapping of the miRNAs, other classes of small RNAs may be examined against dedicated databases such as

piRNABank (Sai Lakshmi and Agrawal, 2008), piRBase (Zhang et al., 2014) and piRNAcluster (Rosenkranz, 2016) for piRNA; gtRNAdb (Chan and Lowe, 2016) for tRNA; GENCODE release 27 (Harrow et al., 2012) for snRNA and snoRNA; and circBase (Glazar et al., 2014) for circular RNA. An interesting option is to map all remaining reads to a nucleotide database to reveal any contamination such as exogenous miRNAs in the sample (Coordinators, 2013).

After data pre-processing, another round of quality control follows, focusing on the identification of samples of compromised quality. Typically, low quality samples show reduced number of detected miRNAs, reduced overall signal or total number of sequencing reads, deviations from typical expression profiles, or high proportion of degraded RNAs (Fu and Dong, 2018; Motameny et al., 2010). To avoid subjective calling of failed samples, pass/fail criteria may be set or statistical tests for outlier detection applied (Kwak and Kim, 2017; Norton et al., 2018). The sample quality may also be evaluated based on the quantification of typically expressed miRNAs, indicators of hemolysis or other cell type contamination markers, and the examination of spike-in controls, as discussed above.

Next step in the data analysis workflow is normalization. This is arguably one of the most critical steps for comparison of samples and may severely compromise the quality of the data if done inappropriately, in worst case leading to totally erroneous conclusions (Schwarzenbach et al., 2015). The goal of normalization is to reduce between-sample technical variation arising during the experimental procedure.

Microarray data are typically normalized using a global measure of miRNA expression (Pritchard et al., 2012a). It assumes that the levels of the majority of the genes is unchanged under the conditions studied and each sample should therefore have similar global signal. Options include global mean, trimmed mean, which removes the least and highest expressed fractions, quantile normalization and their variations.

Normalization to mean expression may also be applied to RT-qPCR, if a large number of miRNAs is quantified (Mestdagh et al., 2009). However, this approach is less practical as the data cannot be compared to another study that used a different set of microRNAs as the two means will be calculated on different sets. For RT-qPCR preferred normalization is to reference genes that have stable expression under conditions that is invariable to the treatments applied (Meyer et al., 2010). Although this normalization concept is widely popular for mRNA analysis, its application to miRNA hurdles on the identification of genes having stable expression. It is therefore critical to identify and validate a set of reference miRNAs for every study. Recommended strategy is to perform a pilot experiment with a small representative set of samples that is screened for all microRNAs using a global platform and from those select the most promising candidates for validation by RT-qPCR (Pritchard et al., 2012a). Standard tools such as GeNorm or NormFinder are typically used to validate candidate's stable expression (Andersen et al., 2004; Vandesompele et al., 2002).

A widely used alternative is normalization to other small RNA molecules such as RNU6, RNU6A, RNU44, and 5S rRNA (Schwarzenbach et al., 2015). Although these may be suitable for cell cultures and some tissues, they are inappropriate for normalization of biofluids due to their intracellular character (Chugh and Dittmer, 2012). Also, individual variations in microRNA levels may not be reflected by variations of other short RNAs. Other normalization strategies include the volume of biofluid, total amount of miRNAs, and spike-in molecules. Main problems of these approaches is that they do not account for variation in RNA quality, input quantity, and individual variation (Meyer et al., 2010; Schwarzenbach et al., 2015). Lastly, the quality of sample needs to be considered when performing normalization. If a liquid biopsy sample is affected by hemolysis or other cellular contamination, the measured profile may be severely distorted and global normalization would compromise the data further (Blondal et al., 2013; Faraldi et al., 2019). The same problem arises if normalization is based on microRNAs influenced by cellular contamination.

Normalization of RNA-Seq data aims to minimize differences in library size and composition caused by varying sequencing depth between samples. The general strategy is based on calculating a scaling factor for each sample, which is used to adjust for library size. Numerous strategies have been developed for the sequencing of long RNAs and many are applicable also to miRNAs (Buschmann et al., 2016). Popular methods include: normalization to library size, to upper quartile of reads, to the quantile with most similar gene distribution, to weighted trimmed mean of the log expression ratios (M-values), to the median of expression ratios of geometric means and artificial spike-in control. Recent publications that compare these options identified the median of expression ratios from geometric means being preferred under the studied conditions (Dillies et al., 2013; Zypych-Walczak et al., 2015). However, the most appropriate normalization strategy may depend on the experiment. To avoid unwanted data distortion, a selection procedure was recently described using a set of well-defined criteria such as introduction of bias, variance, sensitivity, specificity and error prediction (Zypych-Walczak et al., 2015).

Once raw data have undergone quality control, pre-processing, and normalization, secondary analysis takes place. Secondary analysis depends on the particular objectives of the experiment and whether it is an exploratory or hypothesis testing study (Buschmann et al., 2016; Cristiano and Veltri, 2016; Fu and Dong, 2018). First step is usually descriptive statistics to identify the number of positive and negative miRNAs. Further, miRNAs with prohibitively low readouts for reliable analysis are discarded. If complete data matrices are needed for analysis, imputation methods may be applied to replace missing values. As the goal of majority of experiments is to identify differences between groups of samples, statistical tests are used to identify likely differentially expressed miRNAs. Standard statistical methods can be usually applied directly on RT-qPCR data analysis. However, microarray and RNA-Seq data, where the number of miRNAs typically is much larger

than the number of samples, may require more sophisticated models to identify differential expression (Buschmann et al., 2016). The most popular tools are DESeq and DESeq2 (Anders and Huber, 2010; Love et al., 2014), edgeR (Robinson et al., 2010) and limma (Law et al., 2014; Ritchie et al., 2015), although plethora of others has been developed. The selection of proper tool is dependent on the feature of individual data set and number of biological replicates. While the majority of methods perform well when the number of samples is large, DESeq2 and edgeR handle cases with only few samples per group better than others (Rapaport et al., 2013; Sonesson and Delorenzi, 2013). We note, however, that these findings were derived primarily from bulk mRNA data and future work is needed to establish whether these observations hold for microRNA data from liquid biopsies.

The experimental data analysis represents an important step in the miRNA quantification workflow. Whereas the traditional methods such as RT-qPCR and microarrays offer standardized protocols for data processing and analysis, they lack the comprehensiveness of RNA-Seq technology. On the other hand, RNA-Seq requires substantial computational skills and resources that may be a limiting factor for many laboratories. Advantages and disadvantages of each technology as well as experiment-specific requirements on the data analysis need to be therefore taken into account at the early phases of each study.

7. Biomarker identification and characterization

An ultimate goal of the majority of studies focused on liquid biopsies is to find miRNA biomarkers that identify diseased samples, allow to predicting patient's response to therapy or estimate the probability of relapse. Candidate miRNA biomarkers are often selected directly from the set of differentially expressed miRNAs. However, miRNAs that show largest differential expression may not be optimal for sample classification as underlying statistical tests used for differential analysis consider each miRNA as independent variables and correlations between their levels are not taken into account (Robotti et al., 2014). In contrast, multivariate modelling considers these dependences.

Taking positive and negative correlations into account and also the binary nature of the data (key miRNAs being either present or absent in a sample), multivariate analytic tools can identify the optimal set of biomarker miRNAs that produces highest sensitivity and specificity (Buschmann et al., 2016). Multivariate methods can be divided into: i) unsupervised pattern recognition methods such as hierarchical clustering, principal component analysis, self-organizing maps, canonical correlation analysis; and ii) supervised classification techniques such as linear discriminant analysis, classification and regression tree, partial least squares discriminant analysis, deep learning and artificial intelligence. Whereas the unsupervised methods (also called clustering techniques) can be applied directly on data measured on unknown samples, supervised methods require a training set of independently classified samples to be available. For comprehensive overview of multivariate methods for biomarker identification, we recommend recent reviews from the related field of proteomics (Robotti et al., 2014; Smit et al., 2008).

Once candidate miRNA biomarkers are identified and a classifier developed, its performance can be evaluated by receiver operating characteristic (ROC) analysis (Lusted, 1971). ROC analysis informs on specificity and sensitivity of the miRNA classifier and alternative classifiers can be compared by the area under the curve (AUC) of a plot of the sensitivity vs. (1-specificity) or other parameters derived from the ROC analysis (Florkowski, 2008; Hajian-Tilaki, 2013).

Parallel to the development of diagnostic applications based on circulating miRNAs, large number of studies focus on their potential to predict treatment response (Chen et al., 2019; Guo et al., 2018; Sabarimurugan et al., 2019). Predictors are identified using standard statistical methods for survival analysis, including Kaplan-Meier estimator, the log-rank test, and the Cox regression model (Bewick et al., 2004; Clark et al., 2003; Johnson and Shih, 2007). These methods are

typically based on categorical features such as the presence or absence of a mutation, gene fusion and other genetic events. Contrary to that, cell-free miRNA level is a continuous variable and adjustments to the traditional methods are needed. The simplest approach is the conversion of measured levels to binary output based on a pre-defined threshold (so called dichotomization). However, based on a recent study comparing survival analysis methods for biomarker identification in RNA-Seq cancer studies (Raman et al., 2019), methods that preserve continuous data were superior to those using dichotomization. In this comparison the well-established Cox regression methods achieved the highest level of accuracy, reliability and robustness. One advantage of the Cox regression methods is the possibility to include multiple covariates, such as clinical data, batch effects, etc., to improve prediction. Deep learning and other artificial intelligence approaches and integrative analysis are emerging as new strategies for biomarker identification. Although still in their infancy, they show great promise for the future development of the field (Huang et al., 2019; Iuliano et al., 2016).

miRNAs identified as potential biomarkers may be further characterized by functional analysis. First step usually involves miRNA targets prediction using tools like miRanda (Betel et al., 2010), TargetScan (Agarwal et al., 2015) and DIANA-microT tools (Paraskevopoulou et al., 2013; Reczko et al., 2012); reviewed by Riffocampo et al. (2016). As each tool usually predicts a large number of targets, a consensus set can be generated with various algorithms to reduce false positive hits (Oliveira et al., 2017). The resulting set of targets can then be tested for functional correlation using tools such as gene ontology (Ashburner et al., 2000), pathway databases such as the Kyoto Encyclopedia of Genes and Genomes (Ogata et al., 1999), the Reactome (Fabregat et al., 2018) and other specialized databases (Bader et al., 2006). Wide spectrum of tools has been developed for the enrichment analysis and visualization, including DAVID (Dennis et al., 2003), g:Profiler (Raudvere et al., 2019), Gorilla (Eden et al., 2009) and others (Liu, 2017). Interaction networks (miRNA-mRNA, miRNA/mRNA-protein, protein-protein) can add to the amount of information extracted from the data by assigning possible functions to the identified biomarkers (Miryala et al., 2018). For the gene set enrichment and interaction network analyses, some commercial all-in-one packages are available including the CLCGenomics Workbench and Ingenuity Pathway Analysis (both Qiagen, Germany), and Genomatix Genome Analyzer (Genomatix, Germany). For comprehensive overview of the available tools and trends in miRNA analysis we recommend the recent review compiling information about 1000 miRNA bioinformatics tools developed since 2003 (Chen et al., 2018).

8. Conclusions and future perspectives

Measurements of circulating miRNAs are likely to have major impact on cancer management in near future. Hundreds of studies have already proved their potential as an ideal biomarker for early diagnostics, prognostics and support of clinical decisions. Substantial number of clinical trials focused on different cancer types and disease stages highlights the ongoing trend (<https://clinicaltrials.gov>). Still, there are many challenges to bring miRNA biomarkers into clinical practice. Poor experimental design, insufficient sample cohorts, pre-analytical variability, bias introduced during sample processing, data acquisition and analysis are the most prominent ones. The issues have been already recognized and are currently addressed by international efforts such as SPIDIA4P (www.spidia.eu) and CANCER-ID (www.cancer-id.eu) aiming the standardization of procedures for use of miRNA biomarkers in diagnostics.

In the future, we anticipate the understanding of the roles of circulating miRNAs in cancer development and communication to be improved, potentially uncovering new routes for cancer therapy. We predict further standardization and improvement of current technologies allowing more precise and reliable cancer diagnostics and

prognostics. As in other fields, the application of machine learning will improve data analysis and in combination with patient's medical history or other classes of biomarkers will accelerate the clinical utility of miRNAs. Finally, completely new technologies for fast and robust point-of-care miRNA or multi-analyte quantification will ensure smooth and effective transition of any new knowledge into clinics, thus substantially improve management of cancer patients and hopefully the disease outcome.

Declaration of competing interest

Mikael Kubista has shares in TATAA Biocenter AB and MultiD Analysis AB and is an inventor of Two-tailed RT-qPCR technology (US20170233801A1). Lukas Valihrach and Peter Androvic declare no conflict of interest.

Acknowledgment

The work was supported by Czech Science Foundation [P303/18/21942S]; Institutional support [RVO 86652036, BIOCEV CZ.1.05/1.1.00/02.0109]; Standardisation of pre-analytical and analytical procedures for *in vitro* diagnostics in personalised medicine - SPIDIA4P [EU H2020 SCI-HCO-02-2016]; and IMI call 11 - Blood-based biomarker assays for personalised tumour therapy: value of latest circulating biomarkers [CANCER-ID].

References

- Ach, R.A., Wang, H., Curry, B., 2008. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. *BMC Biotechnol.* 8, 69.
- Agarwal, V., Bell, G.W., Nam, J.W., Bartel, D.P., 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4.
- Ameling, S., Kacprowski, T., Chilukoti, R.K., Malsch, C., Liebscher, V., Suhre, K., Pietzner, M., Friedrich, N., Homuth, G., Hammer, E., Volker, U., 2015. Associations of circulating plasma microRNAs with age, body mass index and sex in a population-based study. *BMC Med. Genomics* 8, 61.
- Ameres, S.L., Zamore, P.D., 2013. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* 14 (8), 475–488.
- Anders, S., Huber, W., 2010. Differential expression analysis for sequence count data. *Genome Biol.* 11 (10), R106.
- Andersen, C.L., Jensen, J.L., Orntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64 (15), 5245–5250.
- Andreasen, D., Fog, J.U., Biggs, W., Salomon, J., Dahlsveen, I.K., Baker, A., Mouritzen, P., 2010. Improved microRNA quantification in total RNA from clinical samples. *Methods* 50 (4), S6–S9.
- Andrews, W.J., Brown, E.D., Dellelt, M., Hogg, R.E., Simpson, D.A., 2015. Rapid quantification of microRNAs in plasma using a fast real-time PCR system. *Biotechniques* 58 (5), 244–252.
- Androvic, P., Valihrach, L., Elling, J., Sjoback, R., Kubista, M., 2017. Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Res.* 45 (15).
- Androvic, P., Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., Valihrach, L., 2019. Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Sci. Rep.* 9 (1), 4255.
- Anfossi, S., Babayan, A., Pantel, K., Calin, G.A., 2018. Clinical utility of circulating non-coding RNAs - an update. *Nat. Rev. Clin. Oncol.* 15 (9), 541–563.
- Arata, H., Komatsu, H., Hosokawa, K., Maeda, M., 2012. Rapid and sensitive microRNA detection with laminar flow-assisted dendritic amplification on power-free microfluidic chip. *PLoS One* 7 (11), e48329.
- Aryani, A., Denecke, B., 2015. In vitro application of ribonucleases: comparison of the effects on mRNA and miRNA stability. *BMC Res. Notes* 8, 164.
- Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G., Consortium, G.O., 2000. Gene Ontology: tool for the unification of biology. *Nat. Genet.* 25 (1), 25–29.
- Bader, G.D., Cary, M.P., Sander, C., 2006. Pathguide: a pathway resource list. *Nucleic Acids Res.* 34, D504–D506 Database issue.
- Balcells, I., Cirera, S., Busk, P.K., 2011. Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers. *BMC Biotechnol.* 11, 70.
- Ball, C.A., Sherlock, G., Parkinson, H., Rocca-Serra, P., Brooksbank, C., Causton, H.C., Cavaliere, D., Gaasterland, T., Hingamp, P., Holstege, F., Ringwald, M., Spellman, P., Stoekert, C.J., Stewart, J.E., Taylor, R., Brazma, A., Quackenbush, J., Mged, 2002. Standards for microarray data. *Science* 298 (5593) 539–539.
- Balzano, F., Deiana, M., Dei Giudici, S., Oggiano, A., Baralla, A., Pasella, S., Mannu, A.,

- Pescatori, M., Porcu, B., Fanciulli, G., Zinellu, A., Carru, C., Deiana, L., 2015. miRNA stability in frozen plasma samples. *Molecules* 20 (10), 19030–19040.
- Baran-Gale, J., Kurtz, C.L., Erdos, M.R., Sison, C., Young, A., Fannin, E.E., Chines, P.S., Sethupathy, P., 2015. Addressing bias in small RNA library preparation for sequencing: a new protocol recovers MicroRNAs that evade capture by current methods. *Front. Genet.* 6, 352.
- Barberan-Soler, S., Vo, J.M., Hogans, R.E., Dallas, A., Johnston, B.H., Kazakov, S.A., 2018. Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biol.* 19.
- Baroin-Tourancheau, A., Jaszczyszyn, V., Benigni, X., Amar, L., 2019. Evaluating and correcting inherent bias of microRNA expression in illumina sequencing analysis. *Front. Mol. Biosci.* 6.
- Bayraktar, R., Van Roosbroeck, K., Calin, G.A., 2017. Cell-to-cell communication: microRNAs as hormones. *Mol. Oncol.* 11 (12), 1673–1686.
- Becker, N., Lockwood, C.M., 2013. Pre-analytical variables in miRNA analysis. *Clin. Biochem.* 46 (10–11), 861–868.
- Becker, C., Hammerle-Fickinger, A., Riedmaier, I., Pfaffl, M.W., 2010. mRNA and microRNA quality control for RT-qPCR analysis. *Methods* 50 (4), 237–243.
- Benes, V., Castoldi, M., 2010. Expression profiling of microRNA using real-time quantitative PCR, how to use it and what is available. *Methods* 50 (4), 244–249.
- Benes, V., Collier, P., Kordes, C., Stolte, J., Rausch, T., Muckenthaler, M.U., Haussinger, D., Castoldi, M., 2015. Identification of cytokine-induced modulation of microRNA expression and secretion as measured by a novel microRNA specific qPCR assay. *Sci. Rep. UK* 5.
- Betel, D., Koppal, A., Agius, P., Sander, C., Leslie, C., 2010. Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites. *Genome Biol.* 11 (8), R90.
- Beuvink, I., Kolb, F.A., Budach, W., Garnier, A., Lange, J., Natt, F., Dengler, U., Hall, J., Filipowicz, W., Weiler, J., 2007. A novel microarray approach reveals new tissue-specific signatures of known and predicted mammalian microRNAs. *Nucleic Acids Res.* 35 (7).
- Bewick, V., Cheek, L., Ball, J., 2004. Statistics review 12: survival analysis. *Crit. Care* 8 (5), 389–394.
- Blondal, T., Jensby Nielsen, S., Baker, A., Andreasen, D., Mouritzen, P., Wrang Teilm, M., Dahlsveen, I.K., 2013. Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods* 59 (1), S1–S6.
- Blondal, T., Brunetto, M.R., Cavallone, D., Mikkelsen, M., Thorsen, M., Mang, Y., Pinheiro, H., Bonino, F., Mouritzen, P., 2017. Genome-wide comparison of next-generation sequencing and qPCR platforms for microRNA profiling in serum. *Methods Mol. Biol.* 1580, 21–44.
- Boisen, M.K., Dehlendorf, C., Linnemann, D., Schultz, N.A., Jensen, B.V., Hogdall, E.V.S., Johansen, J.S., 2015. MicroRNA expression in formalin-fixed paraffin-embedded cancer tissue: identifying reference MicroRNAs and variability. *BMC Canc.* 15.
- Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M., 2001. Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat. Genet.* 29 (4), 365–371.
- Brunet-Vega, A., Pericay, C., Quilez, M.E., Ramirez-Lazaro, M.J., Calvet, X., Lario, S., 2015. Variability in microRNA recovery from plasma: comparison of five commercial kits. *Anal. Biochem.* 488, 28–35.
- Burgos, K.L., Javaherian, A., Bomprezzi, R., Ghaffari, L., Rhodes, S., Courtright, A., Tembe, W., Kim, S., Metpally, R., Van Keuren-Jensen, K., 2013. Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. *RNA* 19 (5), 712–722.
- Buschmann, D., Haberberger, A., Kirchner, B., Spornraft, M., Riedmaier, I., Schelling, G., Pfaffl, M.W., 2016. Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow. *Nucleic Acids Res.* 44 (13), 5995–6018.
- Buschmann, D., Kirchner, B., Hermann, S., Marte, M., Wurmser, C., Brandes, F., Kotschote, S., Bonin, M., Steinlein, O.K., Pfaffl, M.W., Schelling, G., Reithman, M., 2018. Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *J. Extracell. Vesicles* 7 (1).
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55 (4), 611–622.
- Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., Croce, C.M., 2002. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* 99 (24), 15524–15529.
- Castoldi, M., Schmidt, S., Benes, V., Noerholm, M., Kulozik, A.E., Hentze, M.W., Muckenthaler, M.U., 2006. A sensitive array for microRNA expression profiling (miChip) based on locked nucleic acids (LNA). *RNA* 12 (5), 913–920.
- Chan, P.P., Lowe, T.M., 2016. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* 44 (D1), D184–D189.
- Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., Guegler, K.J., 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res.* 33 (20), e179.
- Chen, J.J., Hsueh, H.M., Delongchamp, R.R., Lin, C.J., Tsai, C.A., 2007. Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. *BMC Bioinf.* 8, 1–14.
- Chen, X., Ba, Y., Ma, L.J., Cai, X., Yin, Y., Wang, K.H., Guo, J.G., Zhang, Y.J., Chen, J.N., Guo, X., Li, Q.B., Li, X.Y., Wang, W.J., Zhang, Y., Wang, J., Jiang, X.Y., Xiang, Y., Xu, C., Zheng, P.P., Zhang, J.B., Li, R.Q., Zhang, H.J., Shang, X.B., Gong, T., Ning, G., Wang, J., Zen, K., Zhang, J.F., Zhang, C.Y., 2008. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 18 (10), 997–1006.
- Chen, L., Heikkinen, L., Wang, C., Yang, Y., Sun, H., Wong, G., 2018. Trends in the Development of miRNA Bioinformatics Tools. *Brief Bioinform.*
- Chen, B., Xia, Z., Deng, Y.N., Yang, Y., Zhang, P., Zhu, H., Xu, N., Liang, S., 2019. Emerging microRNA biomarkers for colorectal cancer diagnosis and prognosis. *Open Biol.* 9 (1), 180212.
- Cheng, L., Sun, X., Scicluna, B.J., Coleman, B.M., Hill, A.F., 2014. Characterization and deep sequencing analysis of exosomal and non-exosomal miRNA in human urine. *Kidney Int.* 86 (2), 433–444.
- Chevillet, J.R., Kang, Q., Ruf, I.K., Briggs, H.A., Vojtech, L.N., Hughes, S.M., Cheng, H.H., Arroyo, J.D., Meredith, E.K., Galichotte, E.N., Pogosova-Agadjanyan, E.L., Morrissey, C., Stirewalt, D.L., Hladik, F., Yu, E.Y., Higano, C.S., Tewari, M., 2014. Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proc. Natl. Acad. Sci. U. S. A.* 111 (41), 14888–14893.
- Chugh, P., Dittmer, D.P., 2012. Potential pitfalls in microRNA profiling. *Wiley Interdiscip. Rev. RNA* 3 (5), 601–616.
- Clark, T.G., Bradburn, M.J., Love, S.B., Altman, D.G., 2003. Survival analysis part I: basic concepts and first analyses. *Br. J. Canc.* 89 (2), 232–238.
- Clerget, G., Bourguignon-Igel, V., Rederstorff, M., 2015. Alcoholic precipitation of small non-coding RNAs. *Small Non-Coding RNAs: Methods Protoc.* 1296, 11–16.
- Coenen-Stass, A.M.L., Magen, I., Brooks, T., Ben-Dov, I.Z., Greensmith, L., Hornstein, E., Fratta, P., 2018. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. *RNA Biol.* 15 (8), 1133–1145.
- Coordinators, N.R., 2013. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 41, D8–D20 Database issue.
- Cortez, M.A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A.K., Calin, G.A., 2011. MicroRNAs in body fluids-the mix of hormones and biomarkers. *Nat. Rev. Clin. Oncol.* 8 (8), 467–477.
- Cristiano, F., Veltri, P., 2016. Methods and techniques for miRNA data analysis. *Methods Mol. Biol.* 1375, 11–23.
- Dagher, G., Becker, K.F., Bonin, S., Foy, C., Gelmini, S., Kubista, M., Kungl, P., Oelmueller, U., Parkes, H., Pinzani, P., Riegman, P., Schroder, U., Stumptner, C., Turano, P., Sjoback, R., Wutte, A., Zatloukal, K., 2019. Pre-analytical processes in medical diagnostics: new regulatory requirements and standards. *Nat. Biotechnol.* 52, 121–125.
- Dard-Dascot, C., Naquin, D., d'Aubenton-Carafa, Y., Alix, K., Thermes, C., van Dijk, E., 2018. Systematic comparison of small RNA library preparation protocols for next-generation sequencing. *BMC Genomics* 19.
- Das, S., Extracellular, R.N.A.C.C., Ansel, K.M., Bitzer, M., Breakefield, X.O., Charest, A., Galas, D.J., Gerstein, M.B., Gupta, M., Milosavljevic, A., McManus, M.T., Patel, T., Raffai, R.L., Rozowsky, J., Roth, M.E., Saugstad, J.A., Van Keuren-Jensen, K., Weaver, A.M., Laurent, L.C., 2019. The extracellular RNA communication consortium: establishing foundational knowledge and technologies for extracellular RNA research. *Cell* 177 (2), 231–242.
- de Boer, H.C., van Solingen, C., Prins, J., Duijs, J.M.G.J., Huisman, M.V., Rabelink, T.J., van Zonneveld, A.J., 2013. Aspirin treatment hampers the use of plasma microRNA-126 as a biomarker for the progression of vascular disease. *Eur. Heart J.* 34 (44), 3451–3457.
- Degliangeli, F., Pompa, P.P., Fiammengio, R., 2014. Nanotechnology-based strategies for the detection and quantification of microRNA. *Chemistry* 20 (31), 9476–9492.
- Dennis, G., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., Lempicki, R.A., 2003. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4 (9).
- Desvignes, T., Batzel, P., Berezikov, E., Eilbeck, K., Eppig, J.T., McAndrews, M.S., Singer, A., Postlethwait, J.H., 2015. miRNA nomenclature: a view incorporating genetic origins, biosynthetic pathways, and sequence variants. *Trends Genet.* 31 (11), 613–626.
- Dillies, M.A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloe, D., Le Gall, C., Schaeffer, B., Le Crom, S., Guedj, M., Jaffrezic, F., French StatOmiq, C., 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings Bioinf.* 14 (6), 671–683.
- Dong, H., Lei, J., Ding, L., Wen, Y., Ju, H., Zhang, X., 2013. MicroRNA: function, detection, and bioanalysis. *Chem. Rev.* 113 (8), 6207–6233.
- Duy, J., Koehler, J.W., Honko, A.N., Minogue, T.D., 2015. Optimized microRNA purification from TRIzol-treated plasma. *BMC Genomics* 16.
- Eden, E., Navon, R., Steinfeld, I., Lipson, D., Yakhini, Z., 2009. GORILLA: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf.* 10, 48.
- El-Khoury, V., Pierson, S., Kaoma, T., Bernardin, F., Berchem, G., 2016. Assessing cellular and circulating miRNA recovery: the impact of the RNA isolation method and the quantity of input material. *Sci. Rep. UK* 6.
- Fabbri, M., Paone, A., Calore, F., Galli, R., Gaudio, E., Santhanam, R., Lovat, F., Fadda, P., Mao, C., Nuovo, G.J., Zanoni, N., Crawford, M., Ozer, G.H., Wernicke, D., Alder, H., Caligiuri, M.A., Nana-Sinkam, P., Perrotti, D., Croce, C.M., 2012. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. *Proc. Natl. Acad. Sci. U. S. A.* 109 (31), E2110–E2116.
- Fabregat, A., Jupp, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., Milacic, M., Roca, C.D., Rothfels, K., Sevilla, C., Shamovsky, V., Shorser, S., Varusai, T., Viteri, G., Weiser, J., Wu, G., Stein, L., Hermjakob, H., D'Eustachio, P., 2018. The reactome pathway knowledgebase. *Nucleic Acids Res.* 46 (D1), D649–D655.

- Faraldi, M., Gomarasca, M., Sansoni, V., Perego, S., Banfi, G., Lombardi, G., 2019. Normalization strategies differently affect circulating miRNA profile associated with the training status. *Sci. Rep.* 9 (1), 1584.
- Fehlmann, T., Ludwig, N., Backes, C., Meese, E., Keller, A., 2016. Distribution of microRNA biomarker candidates in solid tissues and body fluids. *RNA Biol.* 13 (11), 1084–1088.
- Fichtlscherer, S., De Rosa, S., Fox, H., Schwietz, T., Fischer, A., Liebetrau, C., Weber, M., Hamm, C.W., Rohe, T., Muller-Ardogan, M., Bonauer, A., Zeiher, A.M., Dimmeler, S., 2010. Circulating microRNAs in patients with coronary artery disease. *Circ. Res.* 107 (5), 677–684.
- Florkowski, C.M., 2008. Sensitivity, specificity, receiver-operating characteristic (ROC) curves and likelihood ratios: communicating the performance of diagnostic tests. *Clin. Biochem. Rev.* 29 (Suppl. 1), S83–S87.
- Flowers, E., Won, G.Y., Fukuoka, Y., 2015. MicroRNAs associated with exercise and diet: a systematic review. *Physiol. Genom.* 47 (1), 1–11.
- Friedman, R.C., Farh, K.K.H., Burge, C.B., Bartel, D.P., 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19 (1), 92–105.
- Fu, X., Dong, D., 2018. Bioinformatic analysis of MicroRNA sequencing data. *Methods Mol. Biol.* 1751, 109–125.
- Fu, Y., Wu, P.H., Beane, T., Zamore, P.D., Weng, Z., 2018. Elimination of PCR duplicates in RNA-seq and small RNA-seq using unique molecular identifiers. *BMC Genomics* 19 (1), 531.
- Fuchs, R.T., Sun, Z., Zhuang, F., Robb, G.B., 2015. Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One* 10 (5), e0126049.
- Gao, L., Jiang, F., 2016. MicroRNA (miRNA) profiling. *Methods Mol. Biol.* 1381, 151–161.
- Gebert, L.F.R., MacRae, I.J., 2019. Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20 (1), 21–37.
- Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveri, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webber, P.J., Davidson, E.H., Hood, L., Dimitrov, K., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat. Biotechnol.* 26 (3), 317–325.
- Gilad, S., Meiri, E., Yegorov, Y., Benjamin, S., Lebanony, D., Yerushalmi, N., Benjamin, H., Kushnir, M., Cholak, H., Melamed, N., Bentwich, Z., Hod, M., Goren, Y., Chajut, A., 2008. Serum MicroRNAs are promising novel biomarkers. *PLoS One* 3 (9).
- Giraldez, M.D., Spengler, R.M., Etheridge, A., Godoy, P.M., Barczak, A.J., Srinivasan, S., De Hoff, P.L., Tanriverdi, K., Courtright, A., Lu, S., Khoory, J., Rubio, R., Baxter, D., Driedonks, T.A.P., Buermans, H.P.J., Nolte-t Hoen, E.N.M., Jiang, H., Wang, K., Ghiran, I., Wang, Y.E., Van Keulen-Jensen, K., Freedman, J.E., Woodruff, P.G., Laurent, L.C., Erle, D.J., Galas, D.J., Tewari, M., 2018. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. *Nat. Biotechnol.* 36 (8), 746–757.
- Girard, L., Rodríguez-Canales, J., Behrens, C., Thompson, D.M., Botros, I.W., Tang, H., Xie, Y., Rektman, N., Travis, W.D., Wistuba II, J., Minna, J.D., Gazdar, A.F., 2016. An expression signature as an aid to the histological classification of non-small cell lung cancer. *Clin. Cancer Res.* 22 (19), 4880–4889.
- Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., Caldas, C., 2010. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. *RNA* 16 (5), 991–1006.
- Glazar, P., Papavasiliou, P., Rajewsky, N., 2014. circBase: a database for circular RNAs. *RNA* 20 (11), 1666–1670.
- Glinge, C., Claus, S., Boddum, K., Jabbari, R., Jabbari, J., Risgaard, B., Tomsits, P., Hildebrand, B., Kaab, S., Wakili, R., Jespersen, T., Tfelt-Hansen, J., 2017. Stability of circulating blood-based MicroRNAs - pre-analytic methodological considerations. *PLoS One* 12 (2).
- Godoy, P.M., Bhakta, N.R., Barczak, A.J., Cakmak, H., Fisher, S., MacKenzie, T.C., Patel, T., Price, R.W., Smith, J.F., Woodruff, P.G., Erle, D.J., 2018. Large differences in small RNA composition between human biofluids. *Cell Rep.* 25 (5), 1346–1358.
- Godoy, P.M., Barczak, A.J., DeHoff, P., Srinivasan, S., Das, S., Erle, D.J., Laurent, L.C., 2019. Comparison of miRNA Profiling Methods Using Synthetic miRNA Pools and Standardized exRNA Samples Reveals Substantial Performance Differences. *bioRxiv*, pp. 645762.
- Gourzoues, C., Gelin, A., Bombik, I., Klibi, J., Verillaud, B., Guigay, J., Lang, P., Temam, S., Schneider, V., Amiel, C., Baconnais, S., Jimenez, A.S., Busson, P., 2010. Extracellular release and blood diffusion of BART viral micro-RNAs produced by EBV-infected nasopharyngeal carcinoma cells. *Virology* 407 (1), 7–27.
- Grasedieck, S., Scholer, N., Bommer, M., Niess, J.H., Tumani, H., Rouhi, A., Bloehdorn, J., Liebisch, P., Mertens, D., Dohner, H., Buske, C., Langer, C., Kuchenbauer, F., 2012. Impact of serum storage conditions on microRNA stability. *Leukemia* 26 (11), 2414–2416.
- Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144 Database issue.
- Guo, L., Chen, F., 2014. A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene* 544 (1), 1–7.
- Guo, Y., Vickers, K., Xiong, Y.H., Zhao, S.L., Sheng, Q.H., Zhang, P., Zhou, W.D., Flynn, C.R., 2017. Comprehensive evaluation of extracellular small RNA isolation methods from serum in high throughput sequencing. *BMC Genomics* 18.
- Guo, S., Fesler, A., Wang, H., Ju, J., 2018. microRNA based prognostic biomarkers in pancreatic Cancer. *Biomark. Res.* 6, 18.
- Hafner, M., Landgraf, P., Ludwig, J., Rice, A., Ojo, T., Lin, C., Holoch, D., Lim, C., Tuschl, T., 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* 44 (1), 3–12.
- Hafner, M., Renwick, N., Brown, M., Mihailovic, A., Holoch, D., Lin, C., Pena, J.T., Nusbaum, J.D., Morozov, P., Ludwig, J., Ojo, T., Luo, S., Schroth, G., Tuschl, T., 2011. RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA* 17 (9), 1697–1712.
- Hajian-Tilaki, K., 2013. Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Casp. J. Intern. Med.* 4 (2), 627–635.
- Hall, J.S., Taylor, J., Valentine, H.R., Irlam, J.J., Eustace, A., Hoskin, P.J., Miller, C.J., West, C.M.L., 2012. Enhanced stability of MicroRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *J. Pathol.* 228 S15–S15.
- Hamidi-Asl, E., Palchetti, I., Hasheminejad, E., Mascini, M., 2013. A review on the electrochemical biosensors for determination of microRNAs. *Talanta* 115, 74–83.
- Hardigan, A.A., Roberts, B.S., Moore, D.E., Ramaker, R.C., Jones, A.L., Myers, R.M., 2019 Aug 22. CRISPR/Cas9-targeted removal of unwanted sequences from small-RNA sequencing libraries. *Nucleic Acids Res.* 47 (14), e84. <https://doi.org/10.1093/nar/gkz425>.
- Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., Barnes, I., Bignell, A., Boychenko, V., Hunt, T., Kay, M., Mukherjee, G., Rajan, J., Despacio-Reyes, G., Saunders, G., Steward, C., Harte, R., Lin, M., Howald, C., Tanzer, A., Derrien, T., Chrast, J., Walters, N., Balasubramanian, S., Pei, B., Tress, M., Rodriguez, J.M., Ezkurdia, I., van Baren, J., Brent, M., Haussler, D., Kellis, M., Valencia, A., Reymond, A., Gerstein, M., Guigo, R., Hubbard, T.J., 2012. GENCODE: the reference human genome annotation for the ENCODE project. *Genome Res.* 22 (9), 1760–1774.
- Hayes, J., Peruzzi, P.P., Lawler, S., 2014. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.* 20 (8), 460–469.
- He, L., Hannon, G.J., 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* 5 (7), 522–531.
- Head, S.R., Komori, H.K., LaMere, S.A., Whisenant, T., Van Nieuwerburgh, F., Salomon, D.R., Ordoukhanian, P., 2014. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* 56 (2), 61–64 66, 68, passim.
- Honda, S., Kirino, Y., 2015. Dumbbell-PCR: a method to quantify specific small RNA variants with a single nucleotide resolution at terminal sequences. *Nucleic Acids Res.* 43 (12).
- Huang, Z., Zhan, X., Xiang, S., Johnson, T.S., Helm, B., Yu, C.Y., Zhang, J., Salama, P., Rizkalla, M., Han, Z., Huang, K., 2019. SALMON: survival analysis learning with multi-omics neural networks on breast cancer. *Front. Genet.* 10, 166.
- Hunt, E.A., Broyles, D., Head, T., Deo, S.K., 2015. MicroRNA detection: current technology and research strategies. *Annu. Rev. Anal. Chem.* 8, 217–237.
- Ibberson, D., Benes, V., Muckenthaler, M.U., Castoldi, M., 2009. RNA degradation compromises the reliability of microRNA expression profiling. *BMC Biotechnol.* 9, 102.
- Iuliano, A., Occhipinti, A., Angelini, C., De Feis, I., Lio, P., 2016. Cancer markers selection using network-based Cox regression: a methodological and computational practice. *Front. Physiol.* 7.
- Izzotti, A., Carozzo, S., Pulliero, A., Zhabayeva, D., Ravetti, J.L., Bersimbaev, R., 2016. Extracellular MicroRNA in liquid biopsy: applicability in cancer diagnosis and prevention. *Am. J. Cancer Res.* 6 (7), 1461–1493.
- Jarry, J., Schadendorf, D., Greenwood, C., Spatz, A., van Kempen, L.C., 2014. The validity of circulating microRNAs in oncology: five years of challenges and contradictions. *Mol. Oncol.* 8 (4), 819–829.
- Jayaprakash, A.D., Jabado, O., Brown, B.D., Sachidanandam, R., 2011. Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.* 39 (21), e141.
- Jensen, S.G., Lamy, P., Rasmussen, M.H., Ostenfeld, M.S., Dyrskjot, L., Orntoft, T.F., Andersen, C.L., 2011. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. *BMC Genomics* 12, 435.
- Johnson, B.N., Mutharasan, R., 2014. Biosensor-based microRNA detection: techniques, design, performance, and challenges. *Analyst* 139 (7), 1576–1588.
- Johnson, L.L., Shih, J.H., 2007. An Introduction to Survival Analysis. Principles and Practice of Clinical Research, second ed. pp. 273–282.
- Kawano, M., Kawazu, C., Lizio, M., Kawaji, H., Carninci, P., Suzuki, H., Hayashizaki, Y., 2010. Reduction of non-insert sequence reads by dimer eliminator LNA oligonucleotide for small RNA deep sequencing. *Biotechniques* 49 (4), 751–755.
- Kelly, A.D., Hill, K.E., Correll, M., Hu, L., Wang, Y.E., Rubio, R., Duan, S., Quackenbush, J., Spentzos, D., 2013. Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance. *Genomics* 102 (1), 8–14.
- Khan, J., Lieberman, J.A., Lockwood, C.M., 2017. Variability in, variability out: best practice recommendations to standardize pre-analytical variables in the detection of circulating and tissue microRNAs. *Clin. Chem. Lab. Med.* 55 (5), 608–621.
- Kilic, T., Erdem, A., Ozsoz, M., Carrara, S., 2018. microRNA biosensors: opportunities and challenges among conventional and commercially available techniques. *Biosens. Bioelectron.* 99, 525–546.
- Kim, D.J., Linnstaedt, S., Palma, J., Park, J.C., Ntrivalas, E., Kwak-Kim, J.Y.H., Gilman-Sachs, A., Beam, K., Hastings, M.L., Martin, J.N., Duelli, D.M., 2012a. Plasma components affect accuracy of circulating cancer-related MicroRNA quantitation. *J. Mol. Diagn.* 14 (1), 71–80.
- Kim, Y.K., Yeo, J., Kim, B., Ha, M., Kim, V.N., 2012b. Short structured RNAs with low GC content are selectively lost during extraction from a small number of cells. *Mol. Cell* 46 (6), 893–895.
- Kirschner, M.B., Kao, S.C., Edelman, J.J., Armstrong, N.J., Valley, M.P., van Zandwijk, N., Reid, G., 2011. Haemolysis during sample preparation alters microRNA content of plasma. *PLoS One* 6 (9), e24145.
- Kirschner, M.B., Edelman, J.J., Kao, S.C., Valley, M.P., van Zandwijk, N., Reid, G., 2013a. The impact of hemolysis on cell-free microRNA biomarkers. *Front. Genet.* 4, 94.
- Kirschner, M.B., van Zandwijk, N., Reid, G., 2013b. Cell-free microRNAs: potential

- biomarkers in need of standardized reporting. *Front. Genet.* 4, 56.
- Kloten, V., Neumann, M.H.D., Di Pasquale, F., Sprenger-Hausuels, M., Shaffer, J.M., Schlumpberger, M., Herdean, A., Betsou, F., Ammerlaan, W., Af Hallstrom, T., Serkkola, E., Forsman, T., Lianidou, E., Sjöback, R., Kubista, M., Bender, S., Lampignano, R., Krahn, T., Schlaange, T., consortium, C.-I., 2019 Sep. Multicentric evaluation of circulating plasma MicroRNA extraction technologies for the development of clinically feasible reverse transcription quantitative PCR and next-generation sequencing analytical work flows. *Clin. Chem.* 65 (9), 1132–1140. <https://doi.org/10.1373/clinchem.2019.303271>.
- Koberle, V., Kakoschky, B., Ibrahim, A.A., Schmithals, C., Peveling-Oberhag, J., Zeuzem, S., Kronenberger, B., Waidmann, O., Pleli, T., Pijper, A., 2016. Vesicle-associated microRNAs are released from blood cells on incubation of blood samples. *Transl. Res.* 169, 40–46.
- Kolbert, C.P., Feddersen, R.M., Rakhshan, F., Grill, D.E., Simon, G., Middha, S., Jang, J.S., Simon, V., Schultz, D.A., Zschunke, M., Lingle, W., Carr, J.M., Thompson, E.A., Oberg, A.L., Eckloff, B.W., Wieben, E.D., Li, P., Yang, P., Jen, J., 2013. Multi-platform analysis of microRNA expression measurements in RNA from fresh frozen and FFPE tissues. *PLoS One* 8 (1), e52517.
- Koppers-Lalic, D., Hackenberg, M., de Menezes, R., Misovic, B., Wachalska, M., Geldof, A., Zini, N., de Reijke, T., Wurdinger, T., Vis, A., van Moorselaar, J., Pegtel, M., Bijnsdorp, I., 2016. Non-invasive prostate cancer detection by measuring miRNA variants (isomiRs) in urine extracellular vesicles. *Oncotarget* 7 (16), 22566–22578.
- Kozomara, A., Birgaoanu, M., Griffiths-Jones, S., 2019. miRBase: from microRNA sequences to function. *Nucleic Acids Res.* 47 (D1), D155–D162.
- Kroh, E.M., Parkin, R.K., Mitchell, P.S., Tewari, M., 2010. Analysis of circulating microRNA biomarkers in plasma and serum using quantitative reverse transcription-PCR (qRT-PCR). *Methods* 50 (4), 298–301.
- Kubista, M., Andrade, J.M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjöback, R., Sjögreen, B., Strombom, L., Stahlberg, A., Zoric, N., 2006. The real-time polymerase chain reaction. *Mol. Asp. Med.* 27 (2–3), 95–125.
- Kwak, S.K., Kim, J.H., 2017. Statistical data preparation: management of missing values and outliers. *Korean J. Anesthesiol.* 70 (4), 407–411.
- Larrea, E., Sole, C., Manterola, L., Goicoechea, I., Armesto, M., Arestin, M., Caffarel, M.M., Araujo, A.M., Araiz, M., Fernandez-Mercado, M., Lawrie, C.H., 2016. New concepts in cancer biomarkers: circulating miRNAs in liquid biopsies. *Int. J. Mol. Sci.* 17 (5).
- Law, C.W., Chen, Y., Shi, W., Smyth, G.K., 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* 15 (2), R29.
- Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K., Banham, A.H., Pezzella, F., Boultonwood, J., Wainscoat, J.S., Hatton, C.S.R., Harris, A.L., 2008. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br. J. Haematol.* 141 (5), 672–675.
- Leshkowitz, D., Horn-Saban, S., Parmet, Y., Feldmesser, E., 2013. Differences in microRNA detection levels are technology and sequence dependent. *RNA* 19 (4), 527–538.
- Li, Y., Kowdley, K.V., 2012. Method for microRNA isolation from clinical serum samples. *Anal. Biochem.* 431 (1), 69–75.
- Li, X., Mauro, M., Williams, Z., 2015. Comparison of plasma extracellular RNA isolation kits reveals kit-dependent biases. *Biotechniques* 59 (1), 13–17.
- Li, P., Kaslan, M., Lee, S.H., Yao, J., Gao, Z.Q., 2017. Progress in exosome isolation techniques. *Theranostics* 7 (3), 789–804.
- Liang, Y., Ridzon, D., Wong, L., Chen, C., 2007. Characterization of microRNA expression profiles in normal human tissues. *BMC Genomics* 8, 166.
- Lippi, G., Fontana, R., Avanzini, P., Aloe, R., Ippolito, L., Sandei, F., Favaloro, E.J., 2012. Influence of mechanical trauma of blood and hemolysis on PFA-100 testing. *Blood Coagul. Fibrinolysis* 23 (1), 82–86.
- Liu, Z., 2017. Gene ontology, enrichment analysis, and pathway analysis. *Bioinform. Aquacult.* *Princ. Methods* 150–168.
- Liu, C.G., Calin, G.A., Meloon, B., Gamlie, N., Seignani, C., Ferracin, M., Dumitru, C.D., Shimizu, M., Zupo, S., Dono, M., Alder, H., Bullrich, F., Negrini, M., Croce, C.M., 2004. An oligonucleotide microchip for genome-wide microRNA profiling in human and mouse tissues. *Proc. Natl. Acad. Sci. U. S. A.* 101 (26), 9740–9744.
- Lizarraga, D., Huen, K., Combs, M., Escudero-Fung, M., Eskenazi, B., Holland, N., 2016. miRNAs differentially expressed by next-generation sequencing in cord blood buffy coat samples of boys and girls. *Epigenomics* UK 8 (12), 1619–1635.
- Locati, M.D., Terpstra, I., de Leeuw, W.C., Kuzak, M., Rauwerda, H., Ensink, W.A., van Leeuwen, S., Nehrdich, U., Spaink, H.P., Jonker, M.J., Breit, T.M., Dekker, R.J., 2015. Improving small RNA-seq by using a synthetic spike-in set for size-range quality control together with a set for data normalization. *Nucleic Acids Res.* 43 (14).
- Love, C., Dave, S., 2013. MicroRNA expression profiling using microarrays. *Methods Mol. Biol.* 999, 285–296.
- Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15 (12), 550.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. *Nature* 435 (7043), 834–838.
- 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 (1), 5162.
- Lusted, L.B., 1971. Signal detectability and medical decision-making. *Science* 171 (3977), 1217–1219.
- Lutzmayr, S., Enugutti, B., Nodine, M.D., 2017. Novel small RNA spike-in oligonucleotides enable absolute normalization of small RNA-Seq data. *Sci. Rep.* UK 7.
- Margue, C., Reinsbach, S., Philippidou, D., Beaume, N., Walters, C., Schneider, J.G., Nashan, D., Behrmann, I., Kreis, S., 2015. Comparison of a healthy miRNome with melanoma patient miRNomes: are microRNAs suitable serum biomarkers for cancer? *Oncotarget* 6 (14), 12110–12127.
- Matsuzaki, J., Ochiya, T., 2017. Circulating microRNAs and extracellular vesicles as potential cancer biomarkers: a systematic review. *Int. J. Clin. Oncol.* 22 (3), 413–420.
- McAlexander, M.A., Phillips, M.J., Witwer, K.W., 2013. Comparison of methods for miRNA extraction from plasma and quantitative recovery of RNA from cerebrospinal fluid. *Front. Genet.* 4, 83.
- McDonald, J.S., Milosevic, D., Reddi, H.V., Grebe, S.K., Algeciras-Schimmich, A., 2011. Analysis of circulating microRNA: preanalytical and analytical challenges. *Clin. Chem.* 57 (6), 833–840.
- Melo, S.A., Sugimoto, H., O'Connell, J.T., Kato, N., Villanueva, A., Vidal, A., Qiu, L., Vitkin, E., Perelman, L.T., Melo, C.A., Lucci, A., Ivan, C., Calin, G.A., Kalluri, R., 2014. Cancer exosomes perform cell-independent MicroRNA biogenesis and promote tumorigenesis. *Cancer Cell* 26 (5), 707–721.
- Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., Vandesompele, J., 2009. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* 10 (6), R64.
- Mestdagh, P., Hartmann, N., Baeriswyl, L., Andreasen, D., Bernard, N., Chen, C., Cheo, D., D'Andrade, P., DeMayo, M., Dennis, L., Derveaux, S., Feng, Y., Fulmer-Smentek, S., Gerstmayr, B., Gouffon, J., Grimley, C., Lader, E., Lee, K.Y., Luo, S., Mouritzen, P., Narayana, A., Patel, S., Peiffer, S., Ruberg, S., Schroth, G., Schuster, D., Shaffer, J.M., Shelton, E.J., Silveria, S., Ulmanella, U., Veeramachaneni, V., Staedtler, F., Peters, T., Guettouche, T., Wong, L., Vandesompele, J., 2014. Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study. *Nat. Methods* 11 (8), 809–815.
- Meyer, S.U., Pfaffl, M.W., Ulbrich, S.E., 2010. Normalization strategies for microRNA profiling experiments: a 'normal' way to a hidden layer of complexity? *Biotechnol. Lett.* 32 (12), 1777–1788.
- Meyer, S.U., Kaiser, S., Wagner, C., Thirion, C., Pfaffl, M.W., 2012. Profound effect of profiling platform and normalization strategy on detection of differentially expressed microRNAs—a comparative study. *PLoS One* 7 (6), e38946.
- Miryal, S.K., Anbarasu, A., Ramaiah, S., 2018. Discerning molecular interactions: a comprehensive review on biomolecular interaction databases and network analysis tools. *Gene* 642, 84–94.
- Mitchell, P.S., Parkin, R.K., Kroh, E.M., Fritz, B.R., Wyman, S.K., Pogosova-Agadjanyan, E.L., Peterson, A., Noteboom, J., O'Briant, K.C., Allen, A., Lin, D.W., Urban, N., Drescher, C.W., Knudsen, B.S., Stirewalt, D.L., Gentleman, R., Vessella, R.L., Nelson, P.S., Martin, D.B., Tewari, M., 2008. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc. Natl. Acad. Sci. U.S.A.* 105 (30), 10513–10518.
- Monleau, M., Bonnel, S., Gostan, T., Blanchard, D., Courgnaud, V., Lecellier, C.H., 2014. Comparison of different extraction techniques to profile microRNAs from human sera and peripheral blood mononuclear cells. *BMC Genomics* 15, 395.
- Moret, I., Sanchez-Izquierdo, D., Iborra, M., Tortosa, L., Navarro-Puche, A., Nos, P., Cervera, J., Beltran, B., 2013. Assessing an improved protocol for plasma microRNA extraction. *PLoS One* 8 (12), e82753.
- Motameny, S., Wolters, S., Nurnberg, P., Schumacher, B., 2010. Next generation sequencing of miRNAs - strategies, resources and methods. *Genes (Basel)* 1 (1), 70–84.
- Munafò, D.B., Robb, G.B., 2010. Optimization of enzymatic reaction conditions for generating representative pools of cDNA from small RNA. *RNA* 16 (12), 2537–2552.
- Murillo, O.D., Thistlethwaite, W., Rozowsky, J., Subramanian, S.L., Lucero, R., Shah, N., Jackson, A.R., Srinivasan, S., Chung, A., Laurent, C.D., Kitchen, R.R., Galeev, T., Warrell, J., Diao, J.A., Welsh, J.A., Hanspers, K., Riutta, A., Burgstaller-Muehlbacher, S., Shah, R.V., Yeri, A., Jenkins, L.M., Ahsen, M.E., Cordon-Cardo, C., Dogra, N., Gifford, S.M., Smith, J.T., Stolovitzky, G., Tewari, A.K., Wunsh, B.H., Yadav, K.K., Danielson, K.M., Filant, J., Moeller, C., Nejad, P., Paul, A., Simonson, B., Wong, D.K., Zhang, X., Balaj, L., Gandhi, R., Sood, A.K., Alexander, R.P., Wang, L., Wu, C.L., Wong, D.T.W., Galas, D.J., van Keuren-Jensen, K., Patel, T., Jones, J.C., Das, S., Cheung, K.H., Pico, A.R., Su, A.L., Raffai, R.L., Laurent, L.C., Roth, M.E., Gerstein, M.B., Milosavljevic, A., 2019. exRNA Atlas analysis reveals distinct extracellular RNA cargo types and their carriers across human biofluids. *Cell* 177 (2), 463–+.
- Nagino, K., Nomura, O., Takii, Y., Myamoto, A., Ichikawa, M., Nakamura, F., Higasa, M., Akiyama, H., Nobumasa, H., Shiojima, S., Tsujimoto, G., 2006. Ultrasensitive DNA chip: gene expression profile analysis without RNA amplification. *J. Biochem.* 139 (4), 697–703.
- Neilsen, C.T., Goodall, G.J., Bracken, C.P., 2012. IsomiRs—the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 28 (11), 544–549.
- Norton, S.S., Vaquero-Garcia, J., Lahens, N.F., Grant, G.R., Barash, Y., 2018. Outlier detection for improved differential splicing quantification from RNA-Seq experiments with replicates. *Bioinformatics* 34 (9), 1488–1497.
- Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H., Kanehisa, M., 1999. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res.* 27 (1), 29–34.
- Oliveira, A.C., Bovolenta, L.A., Nachtigall, P.G., Herkenhoff, M.E., Lemke, N., Pinhal, D., 2017. Combining results from distinct microRNA target prediction tools enhances the performance of analyses. *Front. Genet.* 8.
- Ono, S., Lam, S., Nagahara, M., Hoon, D.S.B., 2015. Circulating microRNA biomarkers as liquid biopsy for cancer patients: pros and cons of current assays. *J. Clin. Med.* 4 (10), 1890–1907.
- Page, K., Powles, T., Slade, M.J., MT, D.E.B., Walker, R.A., Coombes, R.C., Shaw, J.A., 2006. The importance of careful blood processing in isolation of cell-free DNA. *Ann. N. Y. Acad. Sci.* 1075, 313–317.
- Page, K., Guttery, D.S., Zahra, N., Primrose, L., Elshaw, S.R., Pringle, J.H., Blighe, K., Marchese, S.D., Hills, A., Woodley, L., Stebbing, J., Coombes, R.C., Shaw, J.A., 2013. Influence of plasma processing on recovery and analysis of circulating nucleic acids. *PLoS One* 8 (10), e77963.
- Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Rezcko, M., Filippidis, C., Dalamagas, T., Hatzigeorgiou, A.G., 2013. DIANA-miCROT web server v5.0: service integration into miRNA functional analysis workflows.

- Nucleic Acids Res. 41, W169–W173 (Web Server issue).
- Pasquinelli, A.E., Reinhart, B.J., Slack, F., Martindale, M.Q., Kuroda, M.I., Maller, B., Hayward, D.C., Ball, E.E., Degnan, B., Muller, P., Spring, J., Srinivasan, A., Fishman, M., Finnerty, J., Corbo, J., Levine, M., Leahy, P., Davidson, E., Ruvkun, G., 2000. Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408 (6808), 86–89.
- Pease, J., 2011. Small-RNA sequencing libraries with greatly reduced adaptor-dimer background. *Nat. Methods* 8, 272.
- Persson, H., Sokilde, R., Pirona, A.C., Rovira, C., 2017. Preparation of highly multiplexed small RNA sequencing libraries. *Biotechniques* 63 (2), 57–64.
- Pizzamiglio, S., Zanutto, S., Ciniselli, C.M., Belfiore, A., Bottelli, S., Gariboldi, M., Verderio, P., 2017. A methodological procedure for evaluating the impact of hemolysis on circulating microRNAs. *Oncol. Lett.* 13 (1), 315–320.
- Pradervand, S., Weber, J., Lemoine, F., Consales, F., Paillusson, A., Dupasquier, M., Thomas, J., Richter, H., Kaessmann, H., Beaudoin, E., Hagenbuehle, O., Harshman, K., 2010. Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs. *Biotechniques* 48 (3), 219–222.
- Pritchard, C.C., Cheng, H.H., Tewari, M., 2012a. MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* 13 (5), 358–369.
- Pritchard, C.C., Kroh, E., Wood, B., Arroyo, J.D., Dougherty, K.J., Miyaji, M.M., Tait, J.F., Tewari, M., 2012b. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer Prev Res (Phila)* 5 (3), 492–497.
- Raabe, C.A., Tang, T.H., Brosius, J., Rozhdzhevskiy, T.S., 2014. Biases in small RNA deep sequencing data. *Nucleic Acids Res.* 42 (3), 1414–1426.
- Raman, P., Zimmerman, S., Rath, K.S., de Torrente, L., Sarmady, M., Wu, C., Leipzig, J., Taylor, D.M., Tozeren, A., Mar, J.C., 2019. A comparison of survival analysis methods for cancer gene expression RNA-Sequencing data. *Cancer Genet.* 235–236, 1–12.
- Ramon-Nunez, L.A., Martos, L., Fernandez-Pardo, A., Oto, J., Medina, P., Espana, F., Navarro, S., 2017. Comparison of protocols and RNA carriers for plasma miRNA isolation. Unraveling RNA carrier influence on miRNA isolation. *PLoS One* 12 (10), e0187005.
- Rapaport, F., Khanin, R., Liang, Y., Pirun, M., Krek, A., Zumbo, P., Mason, C.E., Socci, N.D., Betel, D., 2013. Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.* 14 (9), R95.
- Raudverre, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47 (W1), W191–W198.
- Raymond, C.K., Roberts, B.S., Garrett-Engle, P., Lim, L.P., Johnson, J.M., 2005. Simple, quantitative primer-extension PCR assay for direct monitoring of microRNAs and short-interfering RNAs. *RNA* 11 (11), 1737–1744.
- Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., Hatzigeorgiou, A.G., 2012. Functional microRNA targets in protein coding sequences. *Bioinformatics* 28 (6), 771–776.
- Redshaw, N., Wilkes, T., Whale, A., Cowen, S., Huggett, J., Foy, C.A., 2013. A comparison of miRNA isolation and RT-qPCR technologies and their effects on quantification accuracy and repeatability. *Biotechniques* 54 (3), 155–+.
- Riffo-Campos, A.L., Riquelme, I., Brebi-Mioville, P., 2016. Tools for sequence-based miRNA target prediction: what to choose? *Int. J. Mol. Sci.* 17 (12).
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47.
- Robb, J.A., Gulley, M.L., Fitzgibbons, P.L., Kennedy, M.F., Cosentino, L.M., Washington, K., Dash, R.C., Branton, P.A., Jewell, S.D., Lapham, R.L., 2014. A call to standardize preanalytic data elements for biospecimens. *Arch. Pathol. Lab Med.* 138 (4), 526–537.
- Robinson, M.D., McCarthy, D.J., Smyth, G.K., 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26 (1), 139–140.
- Robotti, E., Manfredi, M., Marengo, E., 2014. Biomarkers discovery through multivariate statistical methods: a review of recently developed methods and applications in proteomics. *J. Proteom. Bioinform.* S3, 003.
- Rosenkranz, D., 2016. piRNA cluster database: a web resource for piRNA producing loci. *Nucleic Acids Res.* 44 (D1), D223–D230.
- Rozowsky, J., Kitchen, R.R., Park, J.J., Gileev, T.R., Diao, J., Warrell, J., Thistlethwaite, W., Subramanian, S.L., Milosavljevic, A., Gerstein, M., 2019. exceRpt: a comprehensive analytic platform for extracellular RNA profiling. *Cell Syst.* 8 (4), 352–357 e353.
- Ruivo, C.F., Adem, B., Silva, M., Melo, S.A., 2017. The biology of cancer exosomes: insights and new perspectives. *Cancer Res.* 77 (23), 6480–6488.
- Rupaimoole, R., Slack, F.J., 2017. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. *Nat. Rev. Drug Discov.* 16 (3), 203–222.
- Russo, F., Di Bella, S., Vannini, F., Berti, G., Scocini, F., Cook, H.V., Santos, A., Nigita, G., Bonnici, V., Lagana, A., Geraci, F., Pulvirenti, A., Giugno, R., De Masi, F., Bellini, K., Jensen, L.J., Brunak, S., Pellegrini, M., Ferro, A., 2018. miRandola 2017: a curated knowledge base of non-invasive biomarkers. *Nucleic Acids Res.* 46 (D1), D354–D359.
- Sabarimurugan, S., Kumarasamy, C., Baxi, S., Devi, A., Jayaraj, R., 2019. Systematic review and meta-analysis of prognostic microRNA biomarkers for survival outcome in nasopharyngeal carcinoma. *PLoS One* 14 (2), e0209760.
- Sai Lakshmi, S., Agrawal, S., 2008. piRNABank: a web resource on classified and clustered piRNA-interacting RNAs. *Nucleic Acids Res.* 36, D173–D177 (Database issue).
- Salehi, M., Sharifi, M., 2018. Exosomal miRNAs as novel cancer biomarkers: challenges and opportunities. *J. Cell. Physiol.* 233 (9), 6370–6380.
- Sato, F., Tsuchiya, S., Terasawa, K., Tsujimoto, G., 2009. Intra-platform repeatability and inter-platform comparability of MicroRNA microarray technology. *PLoS One* 4 (5).
- Schrader, C., Schielke, A., Ellerbroek, L., Johne, R., 2012. PCR inhibitors - occurrence, properties and removal. *J. Appl. Microbiol.* 113 (5), 1014–1026.
- Schwarzenbach, H., 2015. The clinical relevance of circulating, exosomal miRNAs as biomarkers for cancer. *Expert Rev. Mol. Diagn.* 15 (9), 1159–1169.
- Schwarzenbach, H., da Silva, A.M., Calin, G., Pantel, K., 2015. Data normalization strategies for MicroRNA quantification. *Clin. Chem.* 61 (11), 1333–1342.
- Sedlackova, T., Repiska, G., Minarik, G., 2014. Selection of an optimal method for co-isolation of circulating DNA and miRNA from the plasma of pregnant women. *Clin. Chem. Lab. Med.* 52 (11), 1543–1548.
- Shah, J.S., Soon, P.S., Marsh, D.J., 2016. Comparison of methodologies to detect low levels of hemolysis in serum for accurate assessment of serum microRNAs. *PLoS One* 11 (4), e0153200.
- Shi, R., Chiang, V.L., 2005. Facile means for quantifying microRNA expression by real-time PCR. *Biotechniques* 39 (4), 519–525.
- Shore, S., Henderson, J.M., Lebedev, A., Salcedo, M.P., Zon, G., McCaffrey, A.P., Paul, N., Hogrefe, R.I., 2016. Small RNA library preparation method for next-generation sequencing using chemical modifications to prevent adapter dimer formation. *PLoS One* 11 (11), e0167009.
- Smit, S., Hoefloot, H.C., Smilde, A.K., 2008. Statistical data processing in clinical proteomics. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 866 (1–2), 77–88.
- Soneson, C., Delorenzi, M., 2013. A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinf.* 14, 91.
- Sourvinou, I.S., Markou, A., Lianidou, E.S., 2013. Quantification of circulating miRNAs in plasma: effect of preanalytical and analytical parameters on their isolation and stability. *J. Mol. Diagn.* 15 (6), 827–834.
- Srinivasan, S., Yeri, A., Cheah, P.S., Chung, A., Danielson, K., De Hoff, P., Filant, J., Laurent, C.D., Laurent, L.D., Magee, R., Moeller, C., Murthy, V.L., Nejad, P., Paul, A., Rigoutsos, I., Rodosthenous, R., Shah, R.V., Simonson, B., To, C., Wong, D., Yan, I.K., Zhang, X., Balaj, L., Brakefield, X.O., Daaboul, G., Gandhi, R., Lapidus, J., Londin, E., Patel, T., Raffai, R.L., Sood, A.K., Alexander, R.P., Das, S., Laurent, L.C., 2019. Small RNA sequencing across diverse biofluids identifies optimal methods for exRNA isolation. *Cell* 177 (2), 446–462 e416.
- Sudo, H., Mizoguchi, A., Kawauchi, J., Akiyama, H., Takizawa, S., 2012. Use of non-amplified RNA samples for microarray analysis of gene expression. *PLoS One* 7 (2).
- Takahashi, K., Yokota, S., Tatsumi, N., Fukami, T., Yokoi, T., Nakajima, M., 2013. Cigarette smoking substantially alters plasma microRNA profiles in healthy subjects. *Toxicol. Appl. Pharmacol.* 272 (1), 154–160.
- Tan, G.W., Khoo, A.S., Tan, L.P., 2015. Evaluation of extraction kits and RT-qPCR systems adapted to high-throughput platform for circulating miRNAs. *Sci. Rep.* 5, 9430.
- Taub, M., Lipson, D., Speed, T.P., 2010. Methods for Allocating Ambiguous Short-Reads. Taylor, D.D., Shah, S., 2015. Methods of isolating extracellular vesicles impact downstream analyses of their cargoes. *Methods* 87, 3–10.
- Telonis, A.G., Loher, P., Jing, Y., Londin, E., Rigoutsos, I., 2015. Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity. *Nucleic Acids Res.* 43 (19), 9158–9175.
- Thatcher, S.A., 2015. DNA/RNA preparation for molecular detection. *Clin. Chem.* 61 (1), 89–99.
- Thery, C., Witwer, K.W., Aikawa, E., Alcaraz, M.J., Anderson, J.D., Andriantsitohaina, R., Antoniou, A., Arab, T., Archer, F., Atkin-Smith, G.K., Ayre, D.C., Bach, J.M., Bacher, D., Baharvand, H., Balaj, L., Baldacchini, S., Bauer, N.N., Baxter, A.A., Bebawy, M., Beckham, C., Bedina Zavec, A., Benmoussa, A., Berardi, A.C., Bergese, P., Bielska, E., Blenkiron, C., Bobis-Wozowicz, S., Boilard, E., Boireau, W., Bongiovanni, A., Borrás, F.E., Bosch, S., Boulanger, C.M., Brakefield, X., Breglio, A.M., Brennan, M.A., Brigstock, D.R., Brisson, A., Broekman, M.L., Bromberg, J.F., Bryl-Gorecka, P., Buch, S., Buck, A.H., Burger, D., Busatto, S., Buschmann, D., Bussolati, B., Buzas, E.I., Byrd, J.B., Camussi, G., Carter, D.R., Caruso, S., Chamley, L.W., Chang, Y.T., Chen, C., Chen, S., Cheng, L., Chin, A.R., Clayton, A., Clerici, S.P., Cocks, A., Cocucci, E., Coffey, R.J., Cordeiro-da-Silva, A., Couch, Y., Coumans, F.A., Coyle, B., Crescitelli, R., Criado, M.F., D'Souza-Schorey, C., Das, S., Datta Chaudhuri, A., de Candia, P., De Santana, E.F., De Wever, O., Del Portillo, H.A., Demaret, T., Deville, S., Devitt, A., Dhondt, B., Di Vizio, D., Dieterich, L.C., Dolo, V., Dominguez Rubio, A.P., Dominici, M., Dourado, M.R., Driedonks, T.A., Duarte, F.V., Duncan, H.M., Eichenberger, R.M., Ekstrom, K., El Andaloussi, S., Elie-Caille, C., Erdbrugger, U., Falcon-Perez, J.M., Fatima, F., Fish, J.E., Flores-Bellver, M., Forsonits, A., Frelet-Barrand, A., Fricke, F., Fuhrmann, G., Gabrielsson, S., Gamez-Valero, A., Gardiner, C., Gartner, K., Gaudin, R., Gho, Y.S., Giebel, B., Gilbert, C., Gimona, M., Giusti, I., Gorderhan, D.C., Gorgens, A., Gorski, S.M., Greening, D.W., Gross, J.C., Guzelier, A., Gupta, G.N., Gustafson, D., Handberg, A., Haraszti, R.A., Harrison, P., Hegyesi, H., Hendrix, A., Hill, A.F., Hochberg, F.H., Hoffmann, K.F., Holder, B., Holthofer, H., Hosseinkhani, B., Hu, G., Huang, Y., Huber, V., Hunt, S., Ibrahim, A.G., Ikezu, T., Inal, J.M., Isin, M., Ivanova, A., Jackson, H.K., Jacobsen, S., Jay, S.M., Jayachandran, M., Jenster, G., Jiang, L., Johnson, S.M., Jones, J.C., Jong, A., Jovanovic-Talisman, T., Jung, S., Kalluri, R., Kano, S.I., Kaur, S., Kawamura, Y., Keller, E.T., Khamari, D., Khomyakova, E., Khvorova, A., Kierulff, P., Kim, K.P., Kislinger, T., Klingeborn, M., Klinke, D.J., Kornek, M., Kosanovic, M.M., Kovacs, A.F., Kramer-Albers, E.M., Krasemann, S., Krause, M., Kurochkin, I.V., Kusuma, G.D., Kuypers, S., Laitinen, S., Langevin, S.M., Languino, L.R., Lannigan, J., Lasser, C., Laurent, L.C., Lavieu, G., Lazaro-Ibanez, E., Le Lay, S., Lee, M.S., Lee, Y.X.F., Lemos, D.S., Lenassi, M., Leszczynska, A., Li, I.T., Liao, K., Libregts, S.F., Ligeti, E., Lim, R., Lim, S.K., Line, A., Linnemannstons, K., Llorente, A., Lombard, C.A., Lorenovic, M.J., Lorincz, A.M., Lotvall, J., Lovett, J., Lowry, M.C., Loyer, X., Lu, Q., Lukomska, B., Lunavat, T.R., Maas, S.L., Malhi, H., Marcilla, A., Mariani, J., Mariscal, J., Martens-Uzunova, E.S., Martin-Jaular, L., Martinez, M.C., Martins, V.R., Mathieu, M., Mathivanan, S., Maugeri, M., McGinnis, L.K., McVey, M.J., Meckes Jr., D.G., Meehan, K.L., Mertens, I., Minciacchi, V.R., Moller, A., Moller Jorgensen, M., Morales-Kastresana, A., Morhayim, J., Mullier, F., Muraca, M., Musante, L., Mussack, V., Muth, D.C., Myburgh, K.H., Najrana, T., Nawaz, M., Nazarenko, I., Nejsum, P., Neri, C., Neri, T., Nieuwland, R., Nimrichter, L., Nolan, J.P., Nolte-t'Hoen, E.N., Noren Hooten, N., O'Driscoll, L., O'Grady, T., O'Loghlen, A., Ochiya, T., Olivier, M., Ortiz, A., Ortiz, L.A.,

- Osteikoetxea, X., Ostergaard, O., Ostrowski, M., Park, J., Pegtel, D.M., Peinado, H., Perut, F., Pfaffl, M.W., Phinney, D.G., Pieters, B.C., Pink, R.C., Pisetsky, D.S., Pogge von Strandmann, E., Polakovicova, I., Poon, I.K., Powell, B.H., Prada, I., Pulliam, L., Quesenberry, P., Radeghieri, A., Raffai, R.L., Raimondo, S., Rak, J., Ramirez, M.I., Raposo, G., Rayyan, M.S., Regev-Rudzki, N., Ricklefs, F.L., Robbins, P.D., Roberts, D.D., Rodrigues, S.C., Rohde, E., Rome, S., Rouschop, K.M., Rughetti, A., Russell, A.E., Saa, P., Sahoo, S., Salas-Huenuleo, E., Sanchez, C., Saugstad, J.A., Saul, M.J., Schiffelers, R.M., Schneider, R., Schoyen, T.H., Scott, A., Shahaj, E., Sharma, S., Shatnyeva, O., Shekari, F., Shelke, G.V., Shetty, A.K., Shiba, K., Siljander, P.R., Silva, A.M., Skowronek, A., Snyder 2nd, O.L., Soares, R.P., Sodar, B.W., Soekmadji, C., Sotillo, J., Stahl, P.D., Stoorvogel, W., Stott, S.L., Strasser, E.F., Swift, S., Tahara, H., Tewari, M., Timms, K., Tiwari, S., Tixeira, R., Tkach, M., Toh, W.S., Tomasini, R., Torrecilhas, A.C., Tosar, J.P., Toxavidis, V., Urbanelli, L., Vader, P., van Balkom, B.W., van der Grein, S.G., Van Deun, J., van Herwijnen, M.J., Van Keuren-Jensen, K., van Niel, G., van Royen, M.E., van Wijnen, A.J., Vasconcelos, M.H., Vechetti Jr., I.J., Veit, T.D., Vella, L.J., Velot, E., Verweij, F.J., Vestad, B., Vinas, J.L., Visnovitz, T., Vukman, K.V., Wahlgren, J., Watson, D.C., Wauben, M.H., Weaver, A., Webber, J.P., Weber, V., Wehman, A.M., Weiss, D.J., Welsh, J.A., Wendt, S., Wheelock, A.M., Wiener, Z., Witte, L., Wolfram, J., Xagorari, A., Xander, P., Xu, J., Yan, X., Yanez-Mo, M., Yin, H., Yuana, Y., Zappulli, V., Zarubova, J., Zekas, V., Zhang, J.Y., Zhao, Z., Zheng, L., Zheutlin, A.R., Zickler, A.M., Zimmermann, P., Zivkovic, A.M., Zocco, D., Zuba-Surma, E.K., 2018. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles* 7 (1), 1535750.
- Tiberio, P., Callari, M., Angeloni, V., Daidone, M.G., Appierto, V., 2015. Challenges in using circulating miRNAs as cancer biomarkers. *Biomed Res. Int.* 2015, 731479. <https://doi.org/10.1155/2015/731479>.
- Tichopad, A., Kitchen, R., Riedmaier, I., Becker, C., Stahlberg, A., Kubista, M., 2009. Design and optimization of reverse-transcription quantitative PCR experiments. *Clin. Chem.* 55 (10), 1816–1823.
- Tsang, H.F., Xue, V.W., Koh, S.P., Chiu, Y.M., Ng, L.P., Wong, S.C., 2017. NanoString, a novel digital color-coded barcode technology: current and future applications in molecular diagnostics. *Expert Rev. Mol. Diagn* 17 (1), 95–103.
- Unger, L., Fouche, N., Leeb, T., Gerber, V., Pacholewska, A., 2016. Optimized methods for extracting circulating small RNAs from long-term stored equine samples. *Acta Vet. Scand.* 58 (1), 44.
- van Vliet, E.A., Puhakka, N., Mills, J.D., Srivastava, P.K., Johnson, M.R., Roncon, P., Das Gupta, S., Karttunen, J., Simonato, M., Lukasiuk, K., Gorter, J.A., Aronica, E., Pitkanen, A., 2017. Standardization procedure for plasma biomarker analysis in rat models of epileptogenesis: focus on circulating microRNAs. *Epilepsia* 58 (12), 2013–2024.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., Peleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* 3 (7) RESEARCH0034.
- Vester, B., Wengel, J., 2004. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* 43 (42), 13233–13241.
- Vigneault, F., Ter-Ovanesyan, D., Alon, S., Eminaga, S.D.C.C., Seidman, J.G., Eisenberg, E.G.M.C., 2012. High-throughput multiplex sequencing of miRNA. *Curr. Protoc Hum. Genet.* Chapter 11 Unit 11 12 11-10.
- Wang, B., Howel, P., Bruheim, S., Ju, J., Owen, L.B., Fodstad, O., Xi, Y., 2011. Systematic evaluation of three microRNA profiling platforms: microarray, beads array, and quantitative real-time PCR array. *PLoS One* 6 (2), e17167.
- Weber, J.A., Baxter, D.H., Zhang, S., Huang, D.Y., Huang, K.H., Lee, M.J., Galas, D.J., Wang, K., 2010. The microRNA spectrum in 12 body fluids. *Clin. Chem.* 56 (11), 1733–1741.
- Whitehead, C.A., Luwor, R.B., Morokoff, A.P., Kaye, A.H., Stylli, S.S., 2017. Cancer exosomes in cerebrospinal fluid. *Transl. Cancer Res.* 6, S1352–S1370.
- Witwer, K.W., Halushka, M.K., 2016. Toward the promise of microRNAs - enhancing reproducibility and rigor in microRNA research. *RNA Biol.* 13 (11), 1103–1116.
- Witwer, K.W., Buzas, E.I., Bemis, L.T., Bora, A., Lasser, C., Lotvall, J., Nolte-t Hoen, E.N., Piper, M.G., Sivaraman, S., Skog, J., Thery, C., Wauben, M.H., Hochberg, F., 2013. Standardization of sample collection, isolation and analysis methods in extracellular vesicle research. *J. Extracell. Vesicles* 2.
- Wright, C., Rajpurohit, A., Burke, E.E., Williams, C., Collado-Torres, L., Kimos, M., Brandon, N.J., Cross, A.J., Jaffe, A.E., Weinberger, D.R., Shin, J.H., 2019. Comprehensive assessment of multiple biases in small RNA sequencing reveals significant differences in the performance of widely used methods. *BMC Genomics* 20 (1), 513.
- Xu, P., Billmeier, M., Mohorianu, I., Green, D., Fraser, W., Dalmay, T., 2015. An improved protocol for small RNA library construction using High Definition adapters. *Methods Next Gener. Seq.* 2, 1–10.
- Yang, W.J., Yang, D.D., Na, S.Q., Sandusky, G.E., Zhang, Q., Zhao, G.S., 2005. Dicer is required for embryonic angiogenesis during mouse development. *J. Biol. Chem.* 280 (10), 9330–9335.
- Yeri, A., Courtright, A., Danielson, K., Hutchins, E., Alsop, E., Carlson, E., Hsieh, M., Ziegler, O., Das, A., Shah, R.V., Rozowsky, J., Das, S., Van Keuren-Jensen, K., 2018. Evaluation of commercially available small RNAseq library preparation kits using low input RNA. *BMC Genomics* 19 (1), 331.
- Yin, J.Q., Zhao, R.C., Morris, K.V., 2008. Profiling microRNA expression with microarrays. *Trends Biotechnol.* 26 (2), 70–76.
- Zhang, P., Si, X., Skogerbo, G., Wang, J., Cui, D., Li, Y., Sun, X., Liu, L., Sun, B., Chen, R., He, S., Huang, D.W., 2014. piRBase: a Web Resource Assisting piRNA Functional Study. *Database (Oxford)* 2014. bau110.
- Zhuang, F., Fuchs, R.T., Sun, Z., Zheng, Y., Robb, G.B., 2012. Structural bias in T4 RNA ligase-mediated 3'-adapter ligation. *Nucleic Acids Res.* 40 (7), e54.
- Zipper, H., Brunner, H., Bernhagen, J., Vitzthum, F., 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Res.* 32 (12), e103.
- Zyprych-Walczak, J., Szabelska, A., Handschuh, L., Gorczak, K., Klamecka, K., Figlerowicz, M., Siatkowski, I., 2015. The impact of normalization methods on RNA-seq data analysis. *BioMed Res. Int.* 2015, 621690.

Decoding the transcriptional response to ischemic stroke in young and aged mouse brain

Peter Androvic^{1,3,5*}, Denisa Belov Kirdajova^{2,4}, Jana Tureckova², Daniel Zucha¹, Eva Rohlova¹, Pavel Abaffy¹, Jan Kriska^{2,4}, Martin Valny², Miroslava Anderova², Mikael Kubista¹, Lukas Valihrach^{1*}

¹ Institute of Biotechnology of the Czech Academy of Sciences – BIOCEV, Vestec, Czech Republic

² Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic

³ Laboratory of Growth Regulators, Faculty of Science, Palacky University, Olomouc, Czech Republic

⁴ Second Medical Faculty, Charles University, Prague, Czech Republic

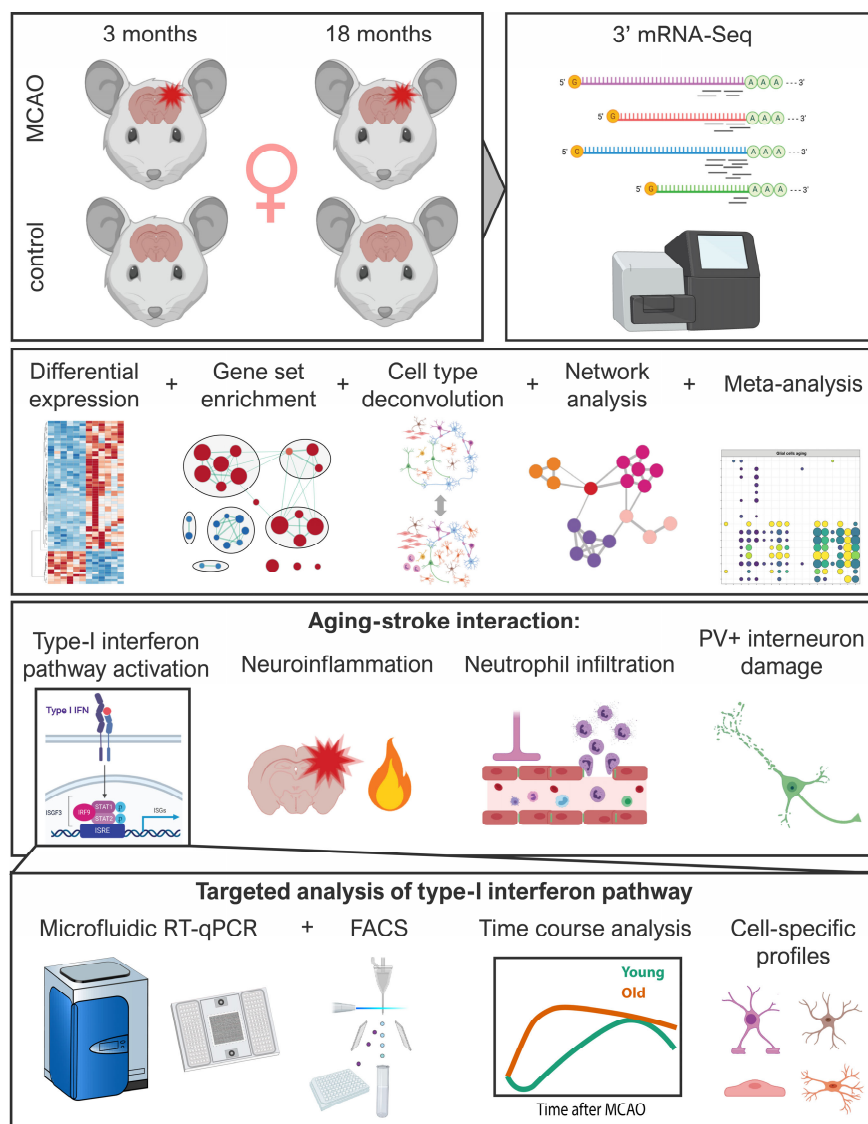
⁵ Lead Contact

*Correspondence to: peter.androvic@ibt.cas.cz or lukas.valihrach@ibt.cas.cz

Abstract

Ischemic stroke is one of the leading causes of mortality and major healthcare and economic burden. It is a well-recognized disease of aging, yet it is unclear how the age-dependent vulnerability occurs and what are the underlying mechanisms. To address these issues, we performed a comprehensive RNA-Seq analysis of aging, ischemic stroke and their interaction using a model of permanent middle cerebral artery occlusion (MCAO) in 3 and 18 month old female mice. We assessed differential gene expression across injury status and age, estimated cell type proportion changes, assayed the results against a range of transcriptional signatures from the literature and performed unsupervised co-expression analysis, identifying modules of genes with varying response to injury. We uncovered selective vulnerability of neuronal populations and increased activation of type-I interferon (IFN-I) signaling and several other inflammatory pathways in aged mice. We extended these findings via targeted expression analysis in tissue as well as acutely purified cellular populations to show differential temporal dynamics of IFN-I signaling between age groups and contribution of individual cell types. Together, these results paint a picture of ischemic stroke as a complex age-related disease and provide insights into interaction of aging and stroke on cellular and molecular level.

Graphical summary



Introduction

Cerebral stroke affects nearly 17 million people per year worldwide, with a death rate of 5.9 million¹, making it the second leading cause of death in the developed world². In addition, stroke is the primary cause of long-term disability³ and represents a major healthcare and economic burden⁴. Ischemic stroke is caused by a loss of blood flow to the brain and accounts for ~87% of all strokes. The remaining ~13% of stroke cases are hemorrhagic and are caused by blood leakage into the brain⁵.

Pathophysiology of ischemic stroke is complex, and involves various mechanisms including disruption of blood-brain barrier (BBB), excitotoxicity, inflammation, oxidative damage, ionic imbalances, apoptosis, angiogenesis and neuroprotection. The ultimate result of ischemic cascade is neuronal death along with an irreversible loss of neuronal function^{6,7}. Two main approaches considered to treat acute ischemic stroke are reperfusion (restoration of blood flow) and neuroprotection (protection of neurons from ischemic injury)^{8,9}. Despite substantial efforts invested into research and development of neuroprotective strategies, with more than 1000 drugs investigated and 100 tested in clinical trials¹⁰, early clot lysis with recombinant tissue plasminogen activator remains the sole approved therapy¹¹.

One of the reasons for this translational roadblock is that most preclinical studies have used only young animals, despite tremendous evidence that ischemic stroke is a disease of aging^{2,12-14}. 75-89% of strokes occur in people aged > 65 years, and for each decade after the age of 55 years the stroke rate doubles². Along with higher incidence, aged patients have higher mortality, suffer more severe deficits and recover slower than younger patients^{15,16}. In addition, sex modifies the influence of age; although ischemic burden is higher in men throughout most of the lifespan, elderly women suffer more strokes and have poorer functional outcomes and quality of life^{13,17-20}. These clinical features are largely recapitulated in experimental rodent models of acute ischemic stroke. Old animals of both sexes have higher mortality and more severe neurological deficits than young animals²¹⁻²³. In addition, sex-specific differences are seen throughout the lifespan in infarct volumes, inflammation, blood-brain barrier (BBB) permeability^{12,24} and glial cell reactivity^{25,26}.

With rapidly aging human population, the global stroke burden increases. There is a need to understand the age-related mechanisms of ischemic injury to improve our ability to discern which therapeutics can be translated from the bench to the bedside². Previous global gene expression studies of experimental stroke using microarrays²⁷⁻³⁴ and more recently RNA-Seq³⁵⁻³⁷ have provided useful insights into the pathophysiology of ischemic stroke and uncovered many altered molecular pathways³⁸. However, few studies included aged animals. One microarray report found that inflammation and synaptic plasticity-related genes had attenuated transcriptional response to stroke in aged mice³⁹. Another microarray study identified age-dependent response of DNA-damage and pro-apoptotic genes, but assayed only limited number of genes (442) in pre-selected pathways⁴⁰, while the follow-up study focused solely on angiogenesis⁴¹. In addition, these studies employed only male animals and clinically less relevant model of transient ischemia⁴²⁻⁴⁶. The complex factors

underlying worsened stroke outcome in the elderly thus remain poorly understood, particularly in females that are often underrepresented in both clinical and pre-clinical stroke research^{14,47-49}.

In this study, we aimed to dissect the interaction between stroke and aging at the genome-wide level. We used permanent middle cerebral artery occlusion (MCAO) – a clinically most relevant model of ischemic stroke⁴²⁻⁴⁶ – on young adult (3 months) and aged (18 months) female mice and analyzed the post-ischemic cortex at 3 days after MCAO using RNA-Seq. We combined differential gene expression and pathway analyses with network and cell type deconvolution approaches and intersected the results with relevant transcriptional signatures from the literature. Our results show age-dependent alterations in processes predominantly associated with inflammation and interferon signaling, as well as the selective vulnerability of specific neuronal subpopulations. We then complemented the results with targeted expression analysis on acutely isolated cell populations and provide detailed insight into temporal dynamics and cell-specific response of interferon signaling pathway in the young and aged post-ischemic brain. Our data provide new insights into the mechanisms of ischemic injury in the aged brain and serve as a publicly available resource for future studies.

Results

We hypothesized that there are two components leading to a more severe outcome of ischemic stroke in aged animals: i) changes in the brain environment during normal aging making the aged brain more susceptible to ischemic injury, and ii) the difference in the response to ischemic challenge between young and aged brain leading to secondary injury and impaired regeneration. To explore both of these components, we performed 3' mRNA sequencing of parietal cortex isolated from 3 month old and 18 month old female mice (representing young adult and aged animals) at 3 days after MCAO and from their age-matched controls (in total 4 groups, 6 animals per group).

Aging is accompanied with increased neuroinflammation involving primarily glial cells

Building on our hypotheses, we first explored factors that may contribute to the increased vulnerability of the aged brain to ischemic challenge. We compared differentially expressed (DE) genes between young and aged controls and analyzed them using Gene set enrichment analysis (GSEA)⁵⁰. We found 52 upregulated and only 13 downregulated genes ($\log_2FC > 1$, $p_{adj} < 0.05$; [Figure S1](#)). GSEA revealed upregulation of defense response-related processes, positive regulation of immune response, and increased secretion of cytokines and protein catabolism ([Figure 1A](#)). Concomitantly, downregulated processes mapped to positive regulation of protein polymerization, dendrite development and axon projection, altogether pointing towards increased inflammation and axonal degeneration in the aged brain. To gain further tissue- and cell-specific context of observed transcriptional changes, we searched the literature for transcriptomic datasets related to brain aging, neuroinflammation and stroke, and quantified overlap with our lists of differentially expressed genes ([Figures 1B, S2](#)). There was a significant overlap between the genes upregulated within the aged controls and signatures of aged astrocytes (such as *Gfap*, *Anln*, *Pcdhb6*,

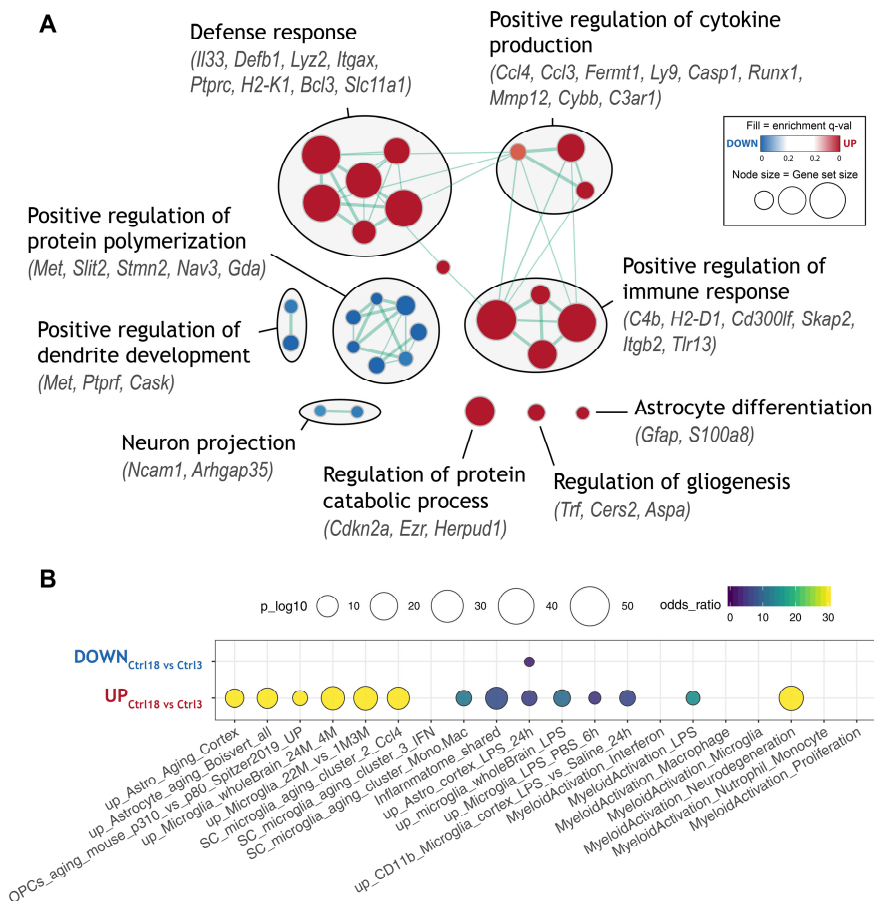


Figure 1. Gene sets with altered expression during normal aging

A) Enrichment map of significantly up- or down-regulated gene ontology (GO) terms between aged (18 months) and young adult (3 months) control mice. Nodes represent gene sets. Highly similar gene sets are connected by edges, grouped in sub-clusters and annotated manually.

B) Meta-analysis showing enrichment of selected transcriptional signatures from literature. Several inflammatory and glial cell activation-related states are significantly up-regulated. See [Table S1](#) for gene set descriptions.

See also [Figures S1](#) and [S2](#).

both young (2556) and aged (3435) mice, with a high prevalence of upregulation, both in terms of number of regulated genes and the fold-change ([Figures 2A, S3A](#)). There was a substantial overlap between DE genes in young and aged mice, although aged mice differentially regulated more genes, often with a greater magnitude ([Figures 2C, S4](#)).

C4b, *Serpina3n*, *Lyz2*, *Neat1*, *Plin4*^{51,52}, aged microglia (such as *Clec7a*, *Cst7*, *Cybb*, *Lgals3*, *Mmp12*, *Spp1*, *C4b*, *Ccl8*)^{53,54}, aging oligodendrocyte precursor cells (OPCs) (such as *Rab37*, *Tnfrsf25*)⁵⁵, in LPS-treated microglia (such as *Bcl3*, *C3ar1*, *Ccl3*, *Ccl4*, *Cst7*, *Cybb*, *Tnfrsf25*)^{56,57} and/or LPS-treated astrocytes (such as *Casp1*, *Flna*, *Mpeg1*, *Runx1*, *Serpina3n*)⁵⁷, as well as with genes that are part of the common inflammatory signature (such as *Ptprc*, *Rab32*, *Slc11a1*, *St14*, *Tep1*, *Trem2*, *Tyrbp*)⁵⁸. We also found significant upregulation of genes enriched in bone marrow-derived macrophages (BMDMs) versus brain-resident microglia (such as *Cdkn2a*, *Itgax*, *Tep1*), while microglia-enriched genes (such as *Cask*, *Gda*, *Nav3*, *Nrep*, *Sox4*) were downregulated, indicating the convergence of microglial and macrophagal signatures with aging, as previously suggested^{54,59}. Furthermore, aging-upregulated genes strongly overlapped with the neurodegeneration-related transcriptional profile of microglia⁵⁹ and recently identified markers of *Ccl4*-expressing subpopulation of microglia that expand during aging, injury⁶⁰ and neurodegeneration⁶¹. Overall, these results show that brain aging leads to subtle alterations in the neuroinflammatory environment, involving particularly glial cells. A subset of pro-inflammatory primed microglia and/or astrocytes may confer adverse milieu that may contribute to aggravation of the ischemic injury in aged animals.

Aging alters the magnitude of the transcriptional response to ischemic stroke

Next, we explored transcriptional changes at 3 days after stroke separately in young and aged mice relative to their age-matched controls. A large number of genes were DE in

revealed hundreds of significantly enriched gene ontology (GO) terms. To remove redundancy, we clustered overlapping, functionally related gene sets into the network using Enrichment Map⁶² ([Figure S5](#)). This analysis revealed several clusters of stroke-upregulated gene sets related to metabolic activity, reactive oxygen species (ROS) production, extracellular matrix, angiogenesis and two major clusters associated with the cell cycle and immune response ([Figure S5](#)). Stroke-downregulated genes were associated with ion channel and synaptic transporter activity, oxidative phosphorylation, neuronal projection and synaptic vesicles production and GO terms “learning and memory” and “locomotory behavior” ([Figure S5](#)). Although comparison of the significantly enriched gene sets between young and aged animals revealed substantial overlap, GO terms related to inflammatory response (such as type-I interferon signaling, cytokine production, neutrophil degranulation), cell-cell interactions (such as integrin cell-surface interaction, extracellular matrix organization) and cell-cycle (such as regulation of DNA replication) were upregulated to a greater extent in aged animals after injury ([Figures 2B, S3B](#)). Similarly, gene sets associated with synaptic signaling and plasticity, neurotransmitter transport and potassium ion channels were downregulated exclusively, or to a greater extent in aged animals after stroke ([Figures 2B, S3B](#)).

These results suggest that in addition to common genes, distinct cellular processes are activated and repressed in response to ischemic impact in aged brain. Surprisingly, almost no functional gene sets were induced or repressed exclusively in young animals ([Figures 2B, S3B](#)).

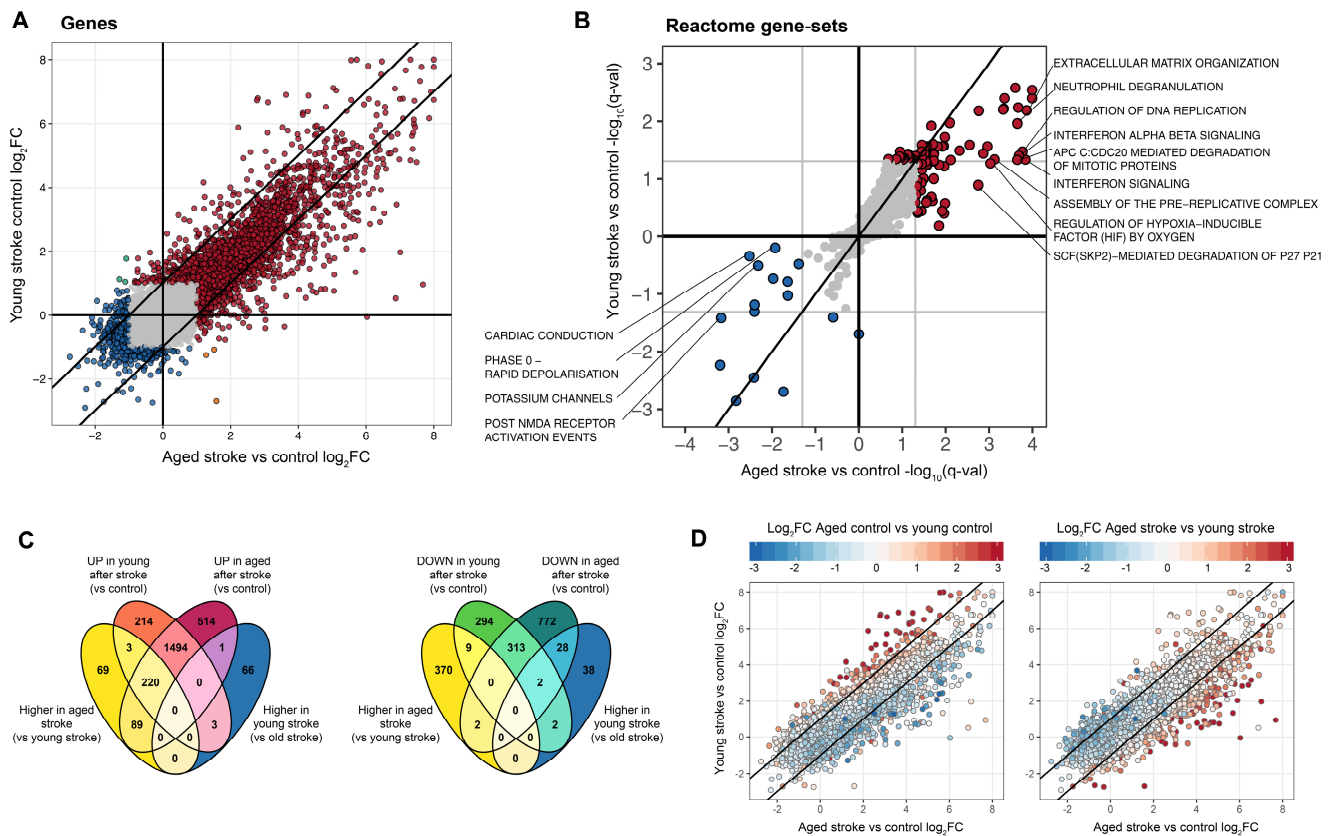


Figure 2. Comparison of differentially expressed genes and gene sets after stroke between young and aged mice

A) Scatter plot comparing stroke-induced \log_2 fold-change in young and aged mice. Genes with $|\log_2FC| > 1$ are highlighted in color. See also Figure S3A.

B) Scatter plot comparing stroke-induced alteration of Reactome pathways in young and aged mice. Pathways with $q\text{-val} < 0.05$ are highlighted in color. Sign depicts UP (+) or DOWN (-) regulation. See also Figure 4S3B and S5.

C) Venn diagrams showing overlaps of DE genes between selected pairwise comparisons. See also Figure S4.

D) Scatter plots of stroke-induced \log_2 fold-change in young and aged mice. Color maps to \log_2 fold-change between aged and young controls (left) or between aged and young stroke groups (right). Genes with larger stroke-induced upregulation in young mice tend to increase in expression during normal aging.

Stroke does not activate exclusive neuroprotective pathways in young compared to aged mice

Since we did not identify any uniquely responsive gene sets in young mice, we took a step back and focused on individual genes that showed a greater stroke-induced upregulation in young mice compared to the aged group (“more up MCAO3”). We postulated that such genes may display a neuroprotective effect (considering the better outcome of stroke in young animals), and serve as possible targets for pharmacological activation in aged patients. We found that although some genes do show greater upregulation upon MCAO in young mice ($\log_2FC_{\text{young}} \gg \log_2FC_{\text{aged}}$), almost all of them are also up-regulated in aged controls compared to young controls (Figure 2D, left) and their abundance levels in young stroke group do not rise above the levels in aged stroke group (Figure 2D, right). Considering the inflammatory nature of these genes, the strong overlap with markers of aged and activated microglia^{60,61,63} (Figure S2), as well as the presence of reactive astrocyte marker (*Gfap*), it is apparent that these genes reflect age-induced glial activation, which is partially saturated after stroke in aged mice.

These results suggest that the group of genes showing stronger activation in young animals (after stroke) does not involve exclusive neuroprotective pathways, but rather reflects the resting baseline level of microglia and astrocytes in

young control animals and a more polarized baseline state in aged control animals.

Combination of aging and stroke leads to massive activation of type-I interferon signaling and aggravated inflammatory response

We then explored the genes showing greater upregulation upon ischemic impact in aged mice, which are likely to exert detrimental effects (“more up MCAO18”). There were more than 400 such genes, of which a large proportion were immune-related (Figure 3A). GO and pathway enrichment analysis revealed T-cell activation, cell adhesion, chemotaxis and leukocyte migration, as the ones of the most strongly enriched functional clusters, suggesting an increased infiltration and activation of peripheral immune cells (Figure 3A). We also found strong enrichment of genes associated with antigen processing and presentation, MHC class I, and cytokine secretion. Several signaling pathways, namely ERK/MAPK signaling and cAMP/cGMP mediated signaling were also significantly enriched, as were the GO terms associated with lipid metabolism, transport of fatty acids and oxidative phosphorylation. Pathway enrichment revealed clusters of extracellular matrix organization and cluster of pathways mediating regulation of cell cycle, suggesting that cellular proliferation may be increased in aged post-stroke mice (Figure 3A). Genes that are more induced by stroke in aged mice also significantly

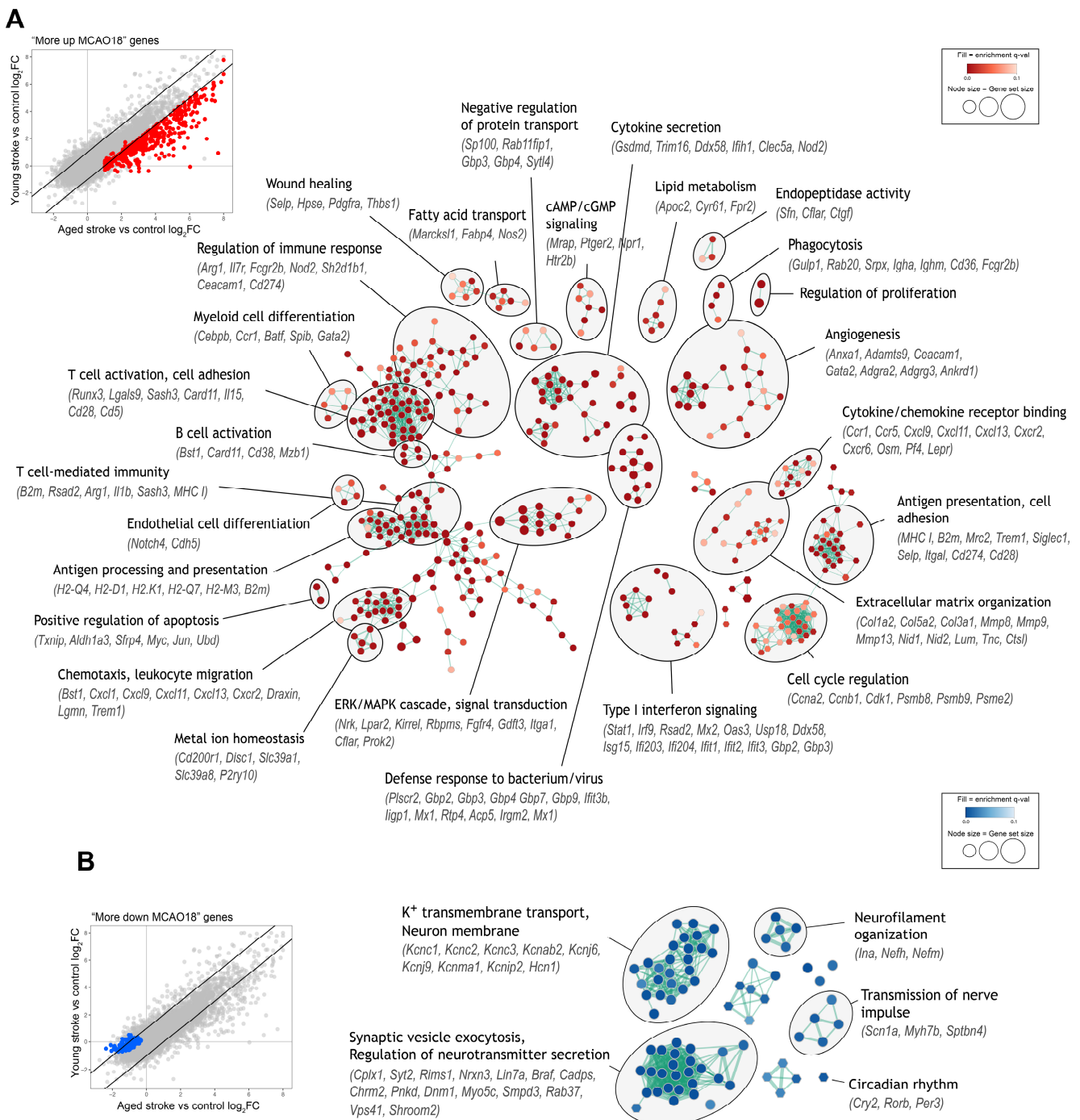


Figure 3. Functional annotation of genes with different quantitative response to stroke between young and aged mice

A) Enrichment map showing significantly enriched GO terms (circles) and pathways (hexagons) for genes with greater upregulation after stroke in aged mice (“more up MCAO18”). Similar gene sets are grouped into sub-clusters and annotated manually. See also Figure S2 for enrichment of gene sets from the literature.

B) Same as (A), but for genes with greater downregulation after stroke in aged mice (“more down MCAO18”).

overlapped with several inflammation and aging associated signatures from the literature (Figure S2). This was expected, considering the highly inflammatory nature of the genes in the “more up MCAO18” gene set. However, an interesting feature was the strong overlap with the LPS-induced / A1 pro-inflammatory astrocytic profile (Figure S2). Previously, it has been reported that MCAO induces a beneficial A2 astrocytic profile⁶⁴. Our result suggests this may not be the case in the aged brain, which would be consistent with reports that aging promotes inflammatory A1 profile of astrocytes^{51,52} and accelerates injury-induced astrocyte reactivity^{65–67}. Another striking feature was particularly strong overrepresentation of interferon-stimulated genes, indicating that specifically type-I

interferon signaling pathway is much strongly activated by stroke in aged animals (Figures 2A, S2).

In parallel to larger upregulation, aged mice also downregulated larger number of genes after stroke, often with a greater magnitude (“more down MCAO18”) (Figure 3B). Functional annotation of the “more down MCAO18” gene set revealed significant enrichment of K⁺ transmembrane transport (*Atp1b1*, *Hcn1*), voltage-gated K⁺ channels (*Kcnc1*, *Kcnc2*, *Kcnc3*, *Kcnj6*, *Kcnj9*, *Kcnma1*, *Kcnt1*) and their regulatory subunits (*Kcnab2*, *Kcnab3*, *Kcnjp2*), neurofilament proteins (*Nefh*, *Nefm*, *Ina*) and genes involved in synaptic vesicle exocytosis regulation and neurotransmitter release (such as *Cadps*, *Pnkd*, *Lin7a*, *Braf*, *Dnm1*, *Rims1*, *Cplx1*,

Syt2, *Nrxn3*) (Figure 3B). These genes mainly localize along the presynaptic and neuron projection membranes, indicating greater axonal damage and impaired synaptic communication in aged post-stroke mice. “More down MCAO18” genes were also enriched with genes involved in the regulation of circadian rhythm (such as *Cry2*, *Rorb*, *Per3*). An impact of ischemic stroke on circadian rhythm has been observed before⁶⁶ and our results suggest that aged animals may be more susceptible to its destabilization, which is linked to sleep, mood and post-stroke depression, and may therefore impact recovery^{69,70}. Overall, analysis of age-stroke interacting genes revealed an increased neuroinflammatory environment in aged animals, which is connected to higher infiltration and activation of peripheral immune cells, pro-inflammatory cytokine secretion and activation of signaling pathways (Erk/MAPK, type-I interferon) that may contribute to secondary injury. On the other hand, K⁺ transmembrane channels, neurofilament and synaptic communication proteins were specifically repressed in aged animals, likely reflecting increased axonal damage.

Transcriptome deconvolution reveals cell type composition changes during aging and after stroke

In order to provide cell-specific context to the observed transcriptional profiles, we assessed relative changes of cell type proportions by computational deconvolution. First, we built a reference of stable cell-specific genes for major CNS cell types (for details see Methods). We then used the *markerGeneProfile* R package⁷¹, which summarizes expression of multiple cell-specific genes into a single marker gene profile (MGP), serving as a surrogate for the cell type proportions. To validate the estimates, we used transcriptome deconvolution algorithm CIBERSORT⁷² (see Methods). Except for astrocytes, concordance analysis confirmed the robustness of the results (Figure S6). Closer inspection showed that astrocyte-specific genes do not significantly overlap with any DE geneset (Figure 4B), and cluster into different co-expression modules (see below), which may be due to heterogeneity of astrocytic response or intrinsic transcriptional regulation of large part of astro-specific genes. To ease the interpretation, we report the relative changes in cell type proportions as marker gene profiles (MGPs) selected by *markerGeneProfile* package (Figure 4A).

We found a significant increase in the MGPs of all non-neuronal cell types following stroke in both age groups ($p_{\text{adj}} < 2.2e-16$; Figure 4A). The largest increase was in microglial and endothelial marker genes ($\log_2\text{FC}$ 1.26-1.77) and the lowest in oligodendrocytic markers ($\log_2\text{FC}$ 0.47-0.60), which also significantly increased with normal aging ($p = 1.20e-19$; Figure 4A). Glial marker genes had generally higher expression in aged stroke group compared to young stroke group, although the magnitudes of their activation by stroke were relatively comparable between both ages. The most prominent difference was in endothelial cell markers ($\log_2\text{FC}_{\text{young}} = 1.26$, $\log_2\text{FC}_{\text{aged}} = 1.74$), which were also significantly overrepresented among “more up MCAO18” genes (Figure 4B). Furthermore, cell type proportion estimates revealed significant depletion of pyramidal/excitatory neurons during aging and following stroke in both age groups ($p_{\text{adj}} < 1.00e-06$,

$\log_2\text{FC}_{\text{young}} = -0.56$, $\log_2\text{FC}_{\text{aged}} = -0.63$) without effect of interaction ($p_{\text{interaction}} = 0.154$; Figure 4A).

Aged ischemic brain is characterized by selective vulnerability of PV+ interneurons and increased infiltration of peripheral leukocytes

Unlike pyramidal/excitatory neurons, parvalbumin-positive (PV+) GABAergic interneuron markers were downregulated after stroke to a greater degree in aged animals ($\log_2\text{FC}_{\text{young}} = -0.42$, $\log_2\text{FC}_{\text{aged}} = -0.89$, $p_{\text{interaction}} = 6.22e-08$; Figure 4A), which was supported by the significant overlap with “more down MCAO18” gene set (Figure 4B), suggesting that this neuronal class is particularly vulnerable to ischemia in aged mice. PV+ interneurons play key roles in cortical plasticity and thus may have profound effect on post-stroke recovery^{73,74}. Testing for enrichment of independent set of marker genes of 6 phenotypically well-defined interneuron populations⁷⁵ also revealed significant overlap of “more down MCAO18” genes with markers of PV+ fast-spiking basket cells, but not other interneuron populations (Figure 4C). We then searched the literature and found that additionally, at least 30 out of 122 genes in the “more down MCAO18” gene set are directly linked to PV+ interneurons and are often localized in their projections^{73,75,84-89,76-83} (Figure S7A). We then confirmed the enrichment of these genes in PV+ interneurons using two recent single-cell RNA-Seq datasets^{90,91} (Figure S7B, C), providing strong indication that the decreasing transcriptional signal indeed reflects the selective impairment of PV+ interneurons in aged post-stroke mice. Afterwards we analyzed selected marker genes of CNS cell types at several time-points after stroke by RT-qPCR in a new set of mice (see below), which again revealed a significant impact of stroke on PV+ interneurons in aged ($p = 6.95e-07$), but not young mice ($p = 0.156$; Figure S8).

Since peripheral immune cells can infiltrate the brain following the disruption of blood-brain barrier after stroke, we assessed their contribution by analyzing the overlap of DE gene sets with cell-specific genes of several leukocyte populations (Figure 4D). Stroke-induced genes in both age groups were highly enriched with macrophage-, monocyte- and neutrophil-specific genes ($p \leq 4.19e-14$) and less strongly with mast cell/basophil-, dendritic cell- and activated T-cell-specific genes ($p \leq 1.18e-5$). Interestingly, granulocyte-enriched genes (neutrophils, mast cells/basophils) were significantly overrepresented among “more up MCAO18” DE genes ($p \leq 1.67e-6$), as were monocyte-enriched genes ($p = 2.14e-6$) and, to a smaller degree, innate lymphocyte (NK cells; $p = 0.025$) and activated T-cell-enriched genes ($p = 0.003$), although T-cell pan markers were barely detectable in our sequencing data (Figure S9). These trends were recapitulated by expression of individual marker genes reported as highly specific and stable by at least two recent expression studies^{59,60,92,93} and by ImmGen database (www.immgen.org; Figure S9).

Overall, the analysis of relative cell type proportions highlighted both similarities and dissimilarities in cellular response to stroke between the two age groups. Similarities include proliferation of non-neuronal cells (particularly of microglia and endothelial cells), death of pyramidal neurons and temporal dynamics of cellular responses. Dissimilarities manifest

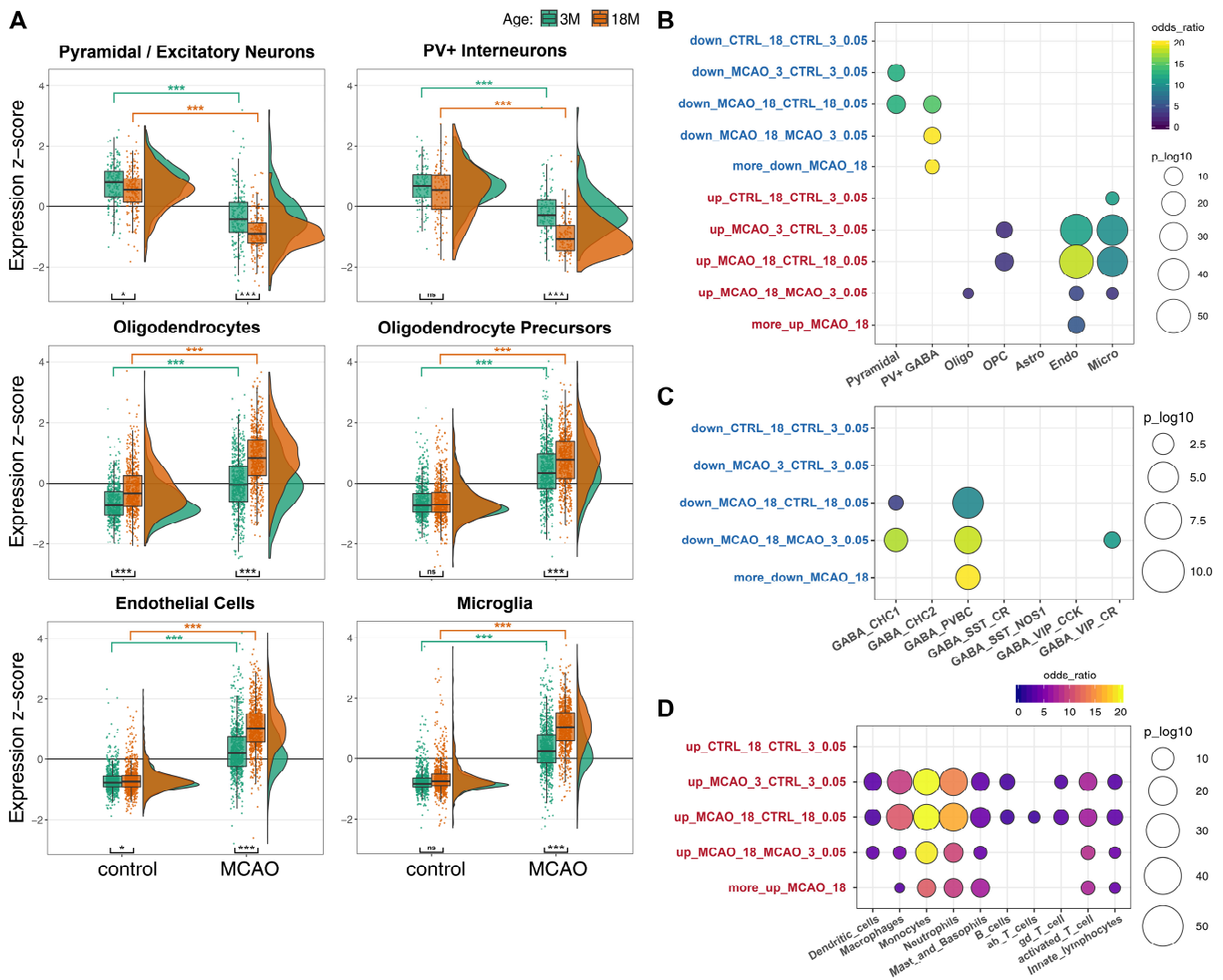


Figure 4. Cell type proportion estimation by transcriptome deconvolution

A) Raincloud plots showing z-scored expression of selected marker genes for major brain cell populations. Astrocytes not shown (see text and Figure S6). Asterisks show Holm-corrected post-hoc t-test p-values. Significance codes (***) <0.001; (**) <0.01; (*) <0.05; (ns) >0.05.

B) Overlap between cell marker gene sets and the sets of differentially expressed genes.

C) Overlap between sets of aging/stroke downregulated genes and the markers of seven phenotypically well-characterized GABAergic interneuron subpopulations. CHC = Chandelier cells; PVBC = PV+ fast-spiking basket cells; other subpopulations named after dominant markers (see ⁷⁵).

D) Overlap between sets of aging/stroke upregulated genes and the marker genes of the major leukocyte populations.

See also Figures S7 and S8.

in greater damage to PV+ interneurons, increased abundance of endothelial cells and increased infiltration of granulocytes/neutrophils in aged mice.

Network analysis provides systems perspective on aging, stroke and their interaction

To capture the full extent of expression trends from systems perspective and reveal relationships between the genes, we complemented our results with weighted co-expression network analysis (WGCNA). We found 27 modules of highly co-expressed genes. Ten modules were strongly associated with aging and/or injury status, of which nine modules corresponded to stroke-induced or stroke-repressed genes (Figure S10A, B). The remaining module (darkturquoise) was upregulated with age, contained many oligodendrocyte-specific genes and was enriched in GO terms like “myelin sheath” and “metabolism of lipids” (Figures 5A-D, S10A),

providing further support for positive correlation of oligodendrocyte-related expression with aging from previous analysis (Figure 4A). Among stroke-repressed modules, blue neuron-associated module was similarly downregulated in both young and aged mice, while green module showed greater downregulation in aged mice (Figure 5C). The green module was significantly enriched in markers of PV+ interneurons ($p = 9.29e-06$) and synaptic-transmission- and axon-related genes, showing that selective axonal damage of PV+ interneurons is embodied at the system-wide transcriptional level (Figure 5B, D). The turquoise module contained a large number of genes upregulated after stroke in both age groups, majority of which were immune/inflammation- and cell cycle-related and was significantly enriched with microglial and endothelial markers (Figure 5A-D). This module likely reflects the increased proportion and activation of these cell types after stroke.

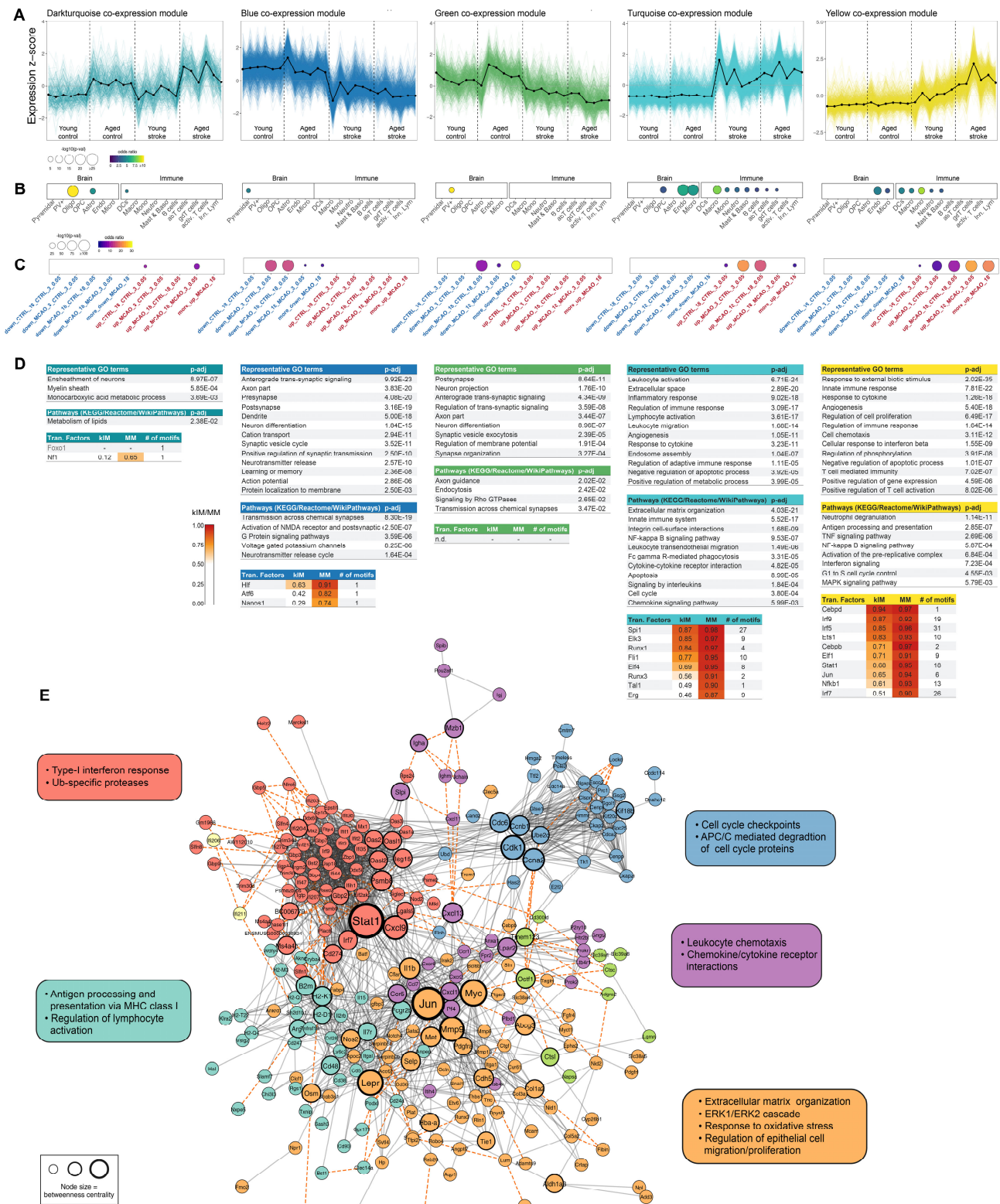


Figure 5. Gene co-expression and interaction network analysis.

A) Expression of selected WGCNA modules associated with aging and/or stroke.

B) Module enrichment with cell-specific genes of major CNS and immune cell populations.

C) Module enrichment with differentially expressed genes.

D) Functional characterization of modules via enrichment of gene ontology sets, cellular pathways and targets of transcription factors. kIM = intramodular connectivity. MM = module membership.

E) Protein interaction network constructed from "more up MCAO18" DE gene set. Network sub-clusters with functional annotations are shown in different colors. Key signaling hubs with high betweenness centrality highlighted with increasing node size.

See also Figure S10.

In addition, turquoise genes were significantly enriched with targets of transcription factors such as *Spi1*, *Runx1*, *Runx3*, *Tal1*, *Fli1*, *Elf4* - known master regulators of hematopoietic development^{94,95} and microglial homeostasis and activation^{96,97} (Figure 5D). The yellow co-expression module was more up-regulated by stroke in aged mice compared to young mice, significantly enriched in leukocyte and endothelial-specific genes, with many GO terms related to inflammation, cytokine production, antigen processing and presentation, leukocyte chemotaxis, lymphocyte activation and interferon signaling as well as regulation of gene expression, suggesting that the yellow module is largely subject to intrinsic transcriptional activation (Figure 5A-D). Supporting this, in the yellow module we detected the largest number of enriched transcription factors (relative to other modules) including *Stat1*, *Cebpd*, *Cebpb*, *Ets1*, *Elf1*, *Jun*, *Nfkb1* and many interferon regulatory factors (*Irf9*, *Irf5*, *Irf7*) (Figure 5D). In accordance, there was a striking overrepresentation of interferon-stimulated genes ($p = 1.25e-50$). Together, WGCNA largely recapitulated results of cell type proportion estimates in an unsupervised manner, supported our previous observation that no neuroprotective transcriptional program is activated specifically in young animals (as no such module was detected by WGCNA), and highlighted amplified activation of yellow inflammatory/interferon module by stroke in aged mice.

We noted that yellow co-expression module generally corresponds to the “more up MCAO18” DE gene set. Vast majority of the “more up MCAO18” DE genes either directly belonged to, or was highly correlated with the yellow module (Figure S10C, D). To investigate how these genes interact, we constructed a combined protein interaction network (Figure 5E). 227 out of 409 genes in the “more up MCAO18” DE set were highly connected within the network. Network clustering revealed six clusters enriched in related, but distinct functional terms including cluster of interferon-stimulated genes, cluster of genes involved in antigen presentation, cluster of cell-cycle regulatory genes, chemokine/cytokine and chemokine receptors cluster and one less rigid cluster of genes involved in extracellular matrix organization, ERK/MAPK signaling and response to oxidative stress (Figure 5E). Searching the network for signaling hubs with high betweenness centrality highlighted transcription factors *Stat1*, *Jun* and *Myc* as well as matrix metalloproteinase 9 (*Mmp9*), chemokine ligands *Cxcl9* and *Cxcl13*, leptin receptor (*Lepr*), and cyclin-dependent kinase 1 (*Cdk1*) as key hubs acting as crosstalks between functional clusters of the network (Figure 5E). These results suggest that “more up MCAO18” genes are part of the transcriptional program composed of several distinct, but molecularly connected gene modules.

Age-dependent activation of type-I IFN regulatory modules after stroke

A striking hallmark of the differential response of aged animals to ischemic stroke was activation of type-I interferon signaling pathway (Figures 2B, 3A, 5A-E, S2, S3B). Type I interferons (IFN-I) are key antiviral cytokines that elicit prototypical interferon-stimulated genes (ISGs) encoding antiviral and inflammatory mediators⁹⁸ and also activate other signaling pathways including MAPK cascades, other cytokines, chemokines and cell-cell interaction modifiers (MHC-I, *Lgals9*)⁹⁹. It has been reported that blocking the IFN-I signaling improves stroke outcome in young mice¹⁰⁰. IFN-I signaling

may therefore act as the central player in the increased neuroinflammatory signature that we detected in the aged animals post-stroke, aggravating the neuronal injury.

To explore this signaling component in greater detail, we mapped our expression data onto the recently published cross-species IFN-I regulatory network⁹⁹ (Figure 6B-D). IFN-I network is divided into five regulatory clusters (C1-C5) with functional differences and variable disease associations⁹⁹. Mapping our data against the network revealed that cluster C3 (composed mainly of antiviral effector genes) and C5 (enriched in inflammation mediators and regulators) were upregulated after stroke in both aged and young animals while the remaining three clusters were non-responsive (Figure 6B). Of the two responsive clusters only cluster C3 was differentially induced between young and aged animals ($p = 3.63e-36$). C3 represents a putative cluster containing predominantly genes under control of ISGF3 transcription factor complex composed of STAT1, STAT2 and IRF9⁹⁹. In accordance, several regulators of the C3 cluster including components of ISGF3 complex were part of “more up MCAO18” DE set (Figure 6C, D); altogether suggesting that age-dependent amplification of ISG expression occurs in a STAT-dependent manner.

Post-stroke temporal dynamics of IFN-I signaling in young and aged mice

IFN-I signaling is typically characterized by rapid ISG induction, which is afterwards quickly attenuated by negative feedback mechanisms⁹⁹. As the outcomes of the IFN-I signaling within the CNS may depend on a delicate balance^{98,101}, we asked if the increased IFN-I signature in the aged post-stroke mice is due to constitutively greater expression or rather altered expression dynamics relative to young mice. To answer this question, we analyzed the expression of IFN-I, IFN receptors, ISGs and other selected genes by microfluidic RT-qPCR at several time points following stroke (Figures 6E, S11).

As expected, all 23 ISGs changed in a largely coordinated fashion throughout the time course, albeit with varying fold-changes (Figure 6E). Expression of ISGs in young animals initially slightly decreased at 3h post MCAO and then started to increase, with peak at 7 days after MCAO (average $\log_2FC = 3.30$; $n = 23$ ISGs). In aged mice, the increase of ISG levels was on average four times greater at day 1 and day 3 compared to the increase in young animals. Unlike in young mice, the expression of ISGs did not further fluctuate, but remained at elevated levels until the latest investigated time point (14 days, average $\log_2FC = 2.66$). Together, these results show that activation of IFN-I signaling occurs early, but not sooner than 3h after stroke, and both timing and overall magnitude of ISG expression are important factors in the different response of aged brain to ischemic injury.

Cell-specific analysis of IFN-I signaling in young and aged mice after stroke

Stat1 activation in neurons¹⁰² and more recently increased IFN-I signaling in microglia¹⁰³ following tMCAO have been demonstrated in young animals. However, little is known about the IFN-I signaling in other cell types following ischemic stroke, and no cell type specific data are available in the aged brain. We have therefore focused on the key players in the brain inflammatory responses – glia and endothelial cells¹⁰⁴ – and aimed to identify how they contribute to IFN-I signaling

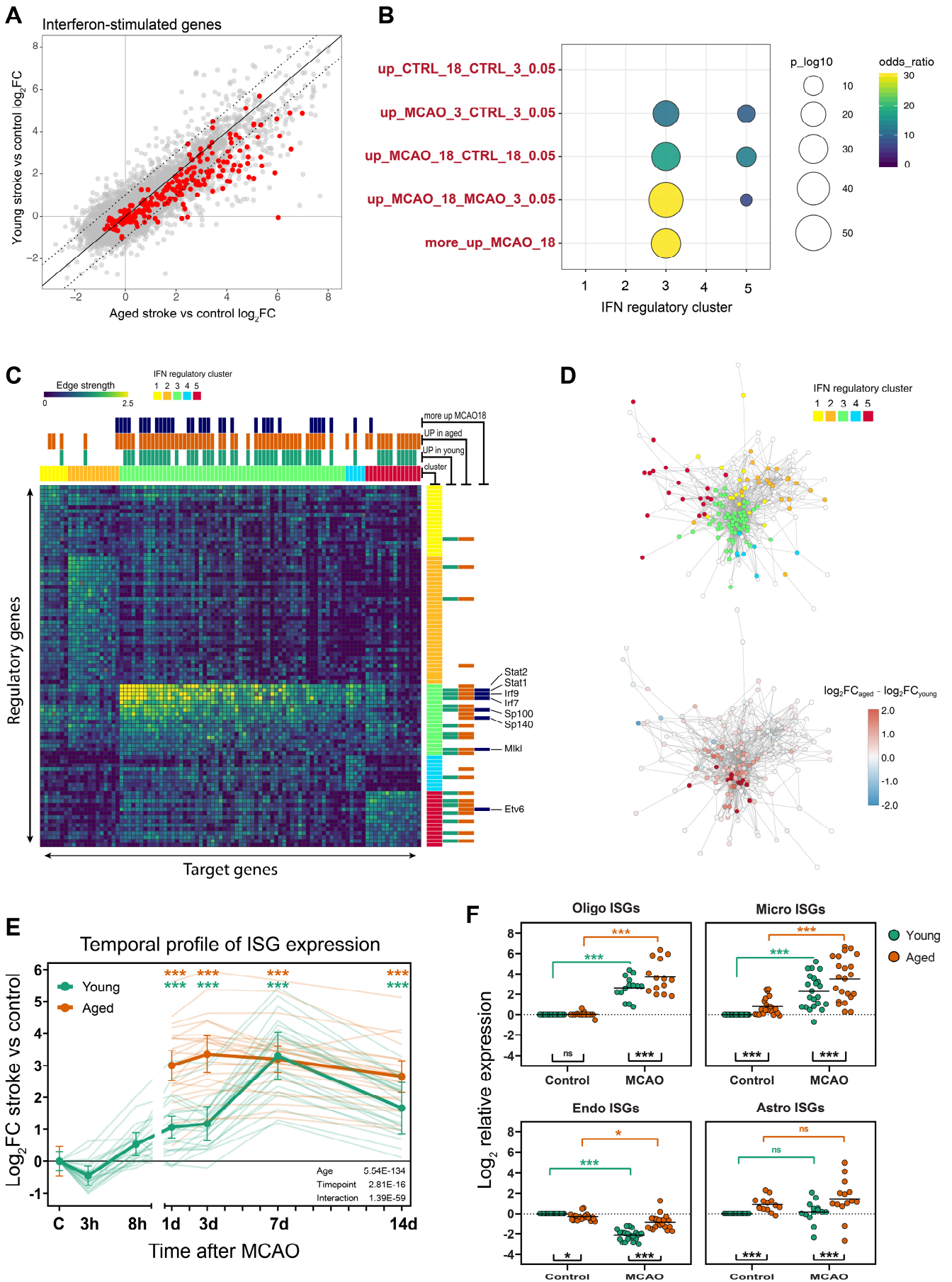


Figure 6. Transcriptional response of IFN-I signaling after stroke in young and aged mice

A) Scatter plot of stroke-induced \log_2 fold-change in young and aged mice. Highlighted is a set of 207 interferon-stimulated genes.

B) Enrichment of IFN-I regulatory network clusters from⁹⁹ in aging/stroke upregulated gene sets.

(figure legend continues on next page)

C) Heatmap of IFN-I network regulatory links between regulators (kinases, phosphatases, transcription factors) and target genes reconstructed from ⁹⁹. Binarized mapping of genes to stroke-upregulated gene sets in young/aged mice is shown on top and right.

D) Network visualization of IFN-I regulatory network; 1000 strongest edges are shown. Network is colored by regulatory cluster (top) or by difference in stroke-induced log₂FC between aged and young mice (bottom).

E) Time-course analysis of ISG expression after stroke in young and aged mice. Bold line shows average expression of 23 ISGs measured by RT-qPCR. Error bars show standard deviation of biological replicates (n = 2-5). Thin lines show average expression of each gene. Asterisks show significant difference relative to control for each age separately (mixed model with post-hoc t-tests). Mind the break in x-axis. Due to limited number of 18-month old animals, 3h and 8h time points were omitted for the aged group. See also [Figure S11](#) for expression of additional genes and [Table S2](#) for per-gene statistics.

F) ISG expression in purified cell populations 3 days after MCAO from young and aged mice. Asterisks show Holm-corrected post-hoc t-test p-values. See also [Figure S12](#).

Significance codes: (***) <0.001; (**) <0.01; (*) <0.05; (ns) >0.05.

following stroke. We have FACS-sorted populations of astrocytes (GFAP+), microglia (CD11b+), oligodendrocytes (O4+) and endothelial cells (CD31+) from young and aged mice 3 days after MCAO as well as from the age-matched controls and measured expression of ISGs and other selected genes by microfluidic RT-qPCR ([Figure 6F](#), [S12](#)).

Microglia and oligodendrocytes heavily upregulated ISG expression following stroke in both young and aged animals ([Figures 6F](#), [S12B](#)). Endothelial cells displayed opposite behavior and significantly downregulated vast majority of measured ISGs in young, and to a much lesser extent in the aged mice. In astrocytes, 9 out of 23 ISGs were undetected. The expression of the remaining 14 ISGs did not change in a synchronized fashion, altogether showing a limited response of astrocytes to IFN after stroke in both age groups. In addition, astrocytes and microglia, but not endothelial cells and oligodendrocytes, showed significant ISG upregulation with normal aging ([Figure 6F](#)).

Next, we assessed the relative differences in ISG expression between cell types for each experimental group separately ([Figure S12C](#)). This comparison revealed that in young and aged controls, endothelial cells are the main ISG expressors, both in terms of number of expressed ISGs and their expression levels, suggesting homeostatic role of IFN-I signaling in these cells. In contrast, microglia, oligodendrocytes and astrocytes expressed lower ISG levels under homeostasis and maintained predominantly expression of ISGs with regulatory roles, such as transcription factors (*Stat1*, *Stat2*, *Irf9*) and receptors (*Ifih1*, *Ddx58*). After stroke, the relative cell type contributions changed and microglia expressed similar levels of ISGs to endothelial cells ([Figure S12C](#)). Overall, our cell-specific analysis revealed that not only microglia, but also oligodendrocytes heavily induce IFN-I signaling following stroke. All cell types converge on higher ISG expression in aged post-stroke brain, although their individual contribution and response to stroke differ. Despite this trend, it is likely that effects in bulk tissue reflect also ISG induction in other cell types, such as non-microglial immune cells entering the brain through compromised BBB.

Discussion

In this study, we systematically analyzed the impact of aging, stroke and their interaction on genome-wide expression profiles. Several findings emerged from the analysis, including that i) brain aging is accompanied by increased inflammation driven by alterations of glial cells, ii) transcriptional response to stroke in young and aged brain is highly similar and differs primarily in magnitude, iii) aged PV+ GABAergic interneurons

are particularly vulnerable to stroke with potential implications for functional recovery, iv) differential stroke outcome is associated with over-activation of pro-inflammatory pathways and other potentially detrimental factors in aged mice, rather than activation of neuroprotective program in young mice, v) increased activation of IFN-I signaling represents a key difference in response to stroke between young and aged mice vi) subnetworks of IFN-I signaling are differentially sensitive to combination of stroke and aging, vii) temporal profile of IFN-I activity after stroke differs between young and aged mice and viii) microglia and oligodendrocytes, but not astrocytes and endothelial cells massively upregulate IFN-I signature following stroke.

First, we analyzed factors that may sensitize aged brain to greater stroke damage during normal aging process. A number of studies has documented that inflammation is increased in the aged brain¹⁰⁵⁻¹⁰⁷, and it has been reported that the changes occur predominantly in glial cells^{108,109}. Our results are well in line with these reports as we detected upregulation of a number of pro-inflammatory genes ([Figures 1A](#), [S1A](#)) that map to signatures of aged and reactive glial cells, including the signature of a recently identified subpopulation of *Ccl4+* pro-inflammatory microglia^{60,61} ([Figures 1B](#), [S2](#)). Activated microglia can promote astrocyte reactivity through secretion of modulatory factors¹¹⁰, and thus further amplify the neurotoxic environment. Together, these changes are likely to sensitize aged brain towards more severe stroke and secondary injury. In addition, we detected increase in oligodendrocyte-specific expression with normal aging ([Figures 4A](#), [5](#)), which has been also seen in human brain^{111,112} and can be possibly attributed to increased number of oligodendrocytes¹¹³⁻¹¹⁵.

Next, we investigated how adult and aged brain responds to ischemic challenge. The assessment of cellular differences by computational deconvolution revealed that PV+ GABAergic interneurons are particularly vulnerable to ischemic stroke in aged mice ([Figures 4A-C](#), [S8](#)). This result was recapitulated by RT-qPCR ([Figure S8](#)) and unsupervised WGCNA analysis where we detected PV+ interneuron-associated module ("green") with greater repression in aged post-stroke mice ([Figure 5](#)). Dysfunction or loss of PV+ interneurons is implicated in the pathology of numerous neuropsychiatric disorders, including schizophrenia^{116,117}, Alzheimer's disease¹¹⁸ and depression^{119,120}. Given their central role in the modulation of neuronal plasticity and cortical information processing^{73,74,121,122} – processes that underlie recovery after stroke¹²³ – the PV+ interneurons may represent novel therapeutic target to promote functional recovery in elderly stroke patients.

Surprisingly, we did not detect any exclusive or greater activation of neuroprotective genes in young mice compared to aged mice (Figure 2). Instead, stroke in young mice more highly induced similar pro-inflammatory signatures of activated glial cells that are already upregulated with normal aging (Figures 2, S2). On the other hand, in aged post-stroke mice we found markedly stronger upregulation of >400 genes, many involved in the processes of the inflammatory cascade (Figure 3A). This was accompanied by a greater influx of peripheral leukocytes, particularly neutrophils (Figure 4D), which are the strongest producers of reactive oxygen species (ROS) and matrix metalloproteinases (MMPs) and promote neuronal injury and BBB disruption¹²⁴⁻¹²⁶. In agreement, an increased number of neutrophils with altered phenotypic responses was previously seen in aged male mice and human stroke patients compared to their younger counterparts¹²⁷ and an increased neutrophil to lymphocyte ratio has been associated with stroke severity and outcome¹²⁸. Although the absence of neuroprotective signal in bulk tissue does not rule out the possibility of its presence in individual cell types, these results suggest that an increased neuroinflammation and infiltration of circulating immune cells are one of the primary drivers for the exacerbated pathology in aged mice.

An outstanding feature of differential response of aged animals to ischemic stroke was upregulation of IFN-I pathway (Figure 2B, 3A, 5A-E, S2, S3B), which persisted for at least 14 days (Figure 6E). IFN-I is an antiviral cytokine with pleiotropic roles¹²⁹ implicated in a number of CNS pathologies including multiple sclerosis^{130,131}, Aicardi-Goutières syndrome, amyotrophic lateral sclerosis¹³², Alzheimer's disease^{59,133,134}, spinal cord injury¹³⁵, traumatic brain injury^{136,137} and ischemic stroke^{100,103,138}.

Therefore, we have explored IFN-I signaling in more detail and found that two IFN-I regulatory modules are activated by stroke irrespective of age, but only the canonical STAT-dependent module is differentially activated in aged animals (Figure 6B-D), likely contributing to an increased neurotoxicity¹⁰⁰. Indeed, *Stat1* and *Irf9* have deleterious roles in stroke and can act directly on neurons^{102,138}. ISG activation typically requires both IFNAR1 and IFNAR2 receptor subunits¹³⁹. Recently the IFNAR2- and STAT-independent pathway, triggered by IFN β , was shown to be involved in systemic LPS-induced toxicity¹⁴⁰. While the genetic or pharmacological blocking of IFNAR1 leads to neuroprotection after transient MCAO¹⁰⁰, the same study reported no effect in IFNAR2-/- mice. This suggests the involvement of the compensatory IFNAR2- and STAT-independent pathway after stroke as well, although it has not been explored in aged animals. Our finding of a predominant increase in the IFNAR2- and STAT-dependent module in aged animals indicates that the detrimental effects of IFN-I signaling after stroke may be exerted by ISGs that are common to the IFNAR2-/STAT-dependent and independent pathways.

As the cell-specific context to the IFN-I signaling after stroke has not been well described in the literature, we profiled the responses of the glia and endothelial cells – known key players in brain neuroinflammatory responses¹⁰⁴ (Figures 6, S12). Our results support the perspective that not only microglia, but also oligodendrocytes are active players in the acute inflammatory response after stroke. On the other hand,

although endothelial cells expressed the highest levels of ISGs under control conditions, they downregulated IFN-I pathway after stroke. This disparity may be associated with different roles of IFN-I signaling in the endothelial cells, evident by the vastly different baseline ISG levels, in line with the role IFN β plays in the maintenance of the BBB integrity⁹⁸.

Despite the overall trend of increased ISG expression in the aged post-stroke brain in the analyzed cell populations, it is unlikely to explain alone the overall ISG increase we reproducibly detected in the bulk tissue. These results indicate the involvement of other contributors, such as the early-infiltrating peripheral leukocytes. Previously, the hematopoietic component has been identified as a major source of IFN-I signaling following traumatic brain injury¹³⁶ and aged mice with bone-marrow transplants from young mice have improved stroke outcome¹²⁷. In concordance, we detected greater upregulation of granulocyte signature genes in aged animals (Figure 4D). In addition, it has been suggested that IFN levels correlate with severity of injury and differently influence functional outcome, as IFN signaling is beneficial in context of mild ischemic preconditioning^{141,142}, but it is detrimental following more severe stroke¹⁰⁰. Our findings appear to be consistent with this notion, as aged animals generally suffer more severe strokes. One potential mechanism of this phenomenon could be the greater disruption of BBB, which in turn leads to the greater influx of peripheral leukocytes. Nonetheless, it is clear that the differential activation of IFN-I signaling pathway in aged animals is likely to contribute significantly to the exacerbated stroke outcome in aged mice and represents a potential target for therapeutic intervention that has been so far overlooked. Our results provide one of the first steps in this direction and open the door to future studies needed to address the mechanisms underlying IFN-I neurotoxicity following stroke in the aged brain.

As with the majority of studies, our results also need to be viewed in light of potential limitations. Since our RNA-Seq data are based on bulk tissue, the expression signal is partly confounded by the cell type composition and the power to detect genes altered in a cell-specific manner is lowered. To tackle this effect, we have employed cell type deconvolution techniques and assayed the results with a range of cell-specific signatures. Another limitation is that our RNA-Seq experiment assayed a single time-point (3 days) after experimental stroke relative to the control. Although the selected time-point can be considered well representative of the subacute phase, informative on early damage as well as initiation of repair processes^{143,144}, there is a space for future studies analyzing later subacute and chronic phases. Being aware of the aforementioned limitations, we provide direct cell-specific as well as time course data targeted at the most significant findings.

In conclusion, detailed insights into transcriptional response to stroke described in this study may contribute to our understanding of the interplay between stroke pathology and aging, and open new avenues for the future search for effective therapeutic approaches.

Methods

Animals

Experiments were performed on 3 and 18 month-old C57Black/6 or FVB female mice. FVB mice are GFAP/EGFP transgenic (line designation TgN(GFAP-EGFP), FVB background), in which the expression of enhanced green fluorescent protein (EGFP) is controlled by the human glial fibrillary acidic protein (GFAP) promoter (marker of astrocytes)¹⁴⁵. The mice were kept on a 12-hr light/dark cycle with access to food and water *ad libitum*. All procedures involving the use of laboratory animals were performed in accordance with the European Communities Council Directive 24 November 1986 (86/609/EEC) and animal care guidelines approved by the Institute of Experimental Medicine, Academy of Sciences of the Czech Republic (Animal Care Committee on April 7, 2011; approval number 018/2011). All efforts were made to minimize both the suffering and the number of animals used.

Induction of middle cerebral artery occlusion (MCAO)

Prior to the induction of MCAO, mice were anaesthetized with 3% isoflurane (Abbot) and maintained in 2% isoflurane using a vaporizer (Tec-3, Cyprane Ltd.). A skin incision between the orbit and the external auditory meatus was made, and a 1-2 mm hole was drilled through the frontal bone 1 mm rostral to the fusion of the zygoma and the squamosal bone and about 3.5 mm ventrally to the dorsal surface of the brain. The middle cerebral artery (MCA) was exposed after the dura was opened and removed. The MCA was occluded by short coagulation with bipolar tweezers (SMT) at a proximal location, followed by transection of the vessel to ensure permanent occlusion. During the surgery, body temperature was maintained at $37 \pm 1^\circ\text{C}$ using a heating pad. This MCAO model yields small infarct lesions in the parietal cortical region. Intact cortical tissue from 3 and 18 month-old mice was used as control.

Dissection of brain tissue from the mouse cortex

Mice were deeply anesthetized with pentobarbital (PTB) (100 mg/kg, i.p.), and perfused transcardially with cold ($4-8^\circ\text{C}$) isolation buffer containing (in mM): NaCl 136.0, KCl 5.4, Hepes 10.0, glucose 5.5, osmolality 290 ± 3 mOsmol/kg. To isolate the cerebral cortex, the brain (+2 mm to -2 mm from bregma) was sliced into 600 μm coronal sections using a vibrating microtome HM650V (MICROM International GmbH), and the uninjured or post-ischemic parietal cortex was carefully dissected out from the ventral white matter tracks.

Preparation of single cell suspensions

The collected tissues were incubated with continuous shaking at 37°C for 45 min in 2 ml of papain solution (20 U/ml) and 0.2 ml DNase (both Worthington) prepared in isolation buffer. After papain treatment, the tissue was mechanically dissociated by gentle trituration using a 1 ml pipette. Dissociated cells were layered on the top of 5 ml of Ovomuroid inhibitor solution (Worthington) and harvested by centrifugation ($140 \times g$ for 6 min). This method routinely yielded $\sim 2 \times 10^6$ cells per mouse brain. Cell aggregates were removed by filtering with 30 μm cell strainers (Becton Dickinson). The cell suspension was then labeled for oligodendrocyte marker 1:50 anti-O4-PE (Miltenyi Biotec), endothelial cell marker 1:50 anti-CD31-PE

(Miltenyi Biotec) and microglial marker 1:50 anti-CD11b-APC (Miltenyi Biotec), according to the standard manufacturer's protocol. To collect astrocytes, GFAP/EGFP mice were used. The cells were kept on ice until sorting.

Cell sorting and collection

Single cell suspensions were sorted using flow cytometry (FACS; BD Influx). Hoechst 33258 (Life Technologies, Carlsbad, CA, USA) was added to the suspension to check viability. 100 cells per well were collected into 96-well plates (Life Technologies) containing 5 μl nuclease-free water with bovine serum albumin (1 mg/ μl , Fermentas) and RNaseOut 20 U (Life Technologies). The plates were placed on a pre-cooled rack and stored at -80°C until analysis. After discarding samples of insufficient quality, we used 2-5 mice per cell type per experimental group for the RT-qPCR analysis (see below). We note that all samples were strongly enriched in respective cellular markers and depleted in markers of other cell types, confirming high purity of sorted suspensions (Figure S12A).

RNA isolation, library preparation and sequencing

Brain tissue samples were homogenized using the Tissue-Lyser (Qiagen). Total RNA was extracted with TRI Reagent (Sigma-Aldrich) according to the manufacturer's protocol and treated with TURBO DNA-free kit (Thermo Fisher). RNA quantity and purity was assessed using the NanoDrop 2000 spectrophotometer (Thermo Fisher) and RNA integrity was assessed using the Fragment Analyzer (Agilent). All samples had RQN > 8 . Libraries were prepared from 400 ng total RNA with QuantSeq 3' Library Prep Kit FWD (Lexogen) according to manufacturer's protocol. 1 μl of ERCC spike-in ($c = 0.01 \times$; Thermo Fisher) per library was included. This library preparation method generates stranded libraries predominantly covering the 3' end of the transcript, thus producing gene-centric expression values. Libraries were quantified on the Qubit 2 fluorometer (Thermo Fisher) and Fragment Analyzer (Agilent) and sequenced on the NextSeq 500 high-output (Illumina) with 85 bp single-end reads. 11.5 – 38 million reads were obtained per library with a median of 16 million reads.

RNA-Seq data processing, mapping and counting

Adaptor sequences and low quality reads were removed using TrimmomaticSE v0.36¹⁴⁶. Reads mapping to mtDNA and rRNA were filtered out using SortMeRNA v2.1 with default parameters¹⁴⁷. The remaining reads were aligned to GRCm38 and ERCC reference using STAR v2.5.2b with default parameters¹⁴⁸. Mapped reads were counted over Gencode vM8 gene annotation using htseq-count with union mode for handling of overlapping reads¹⁴⁹.

Differential expression analysis

Several comparisons of differential expression were generated using DESeq2 v1.16.1¹⁵⁰. For pairwise comparisons, we compared aged controls to young controls, young stroke group to young controls, aged stroke group to aged controls and aged stroke group to young stroke group ($p_{\text{adj}} < 0.05$, $\log_2\text{FC} > 1$ for upregulation and < -0.65 for downregulation). We also generated two-factor comparisons using injury (control/MCAO) and age (3m/18m) and their interaction as predictor variables. DESeq2 results are available in [Supplementary file 1](#). Searchable visualization table is available in [Supplementary file 2](#). During the initial analysis of the dataset, we noted that there was a relatively large number of genes

induced or repressed exclusively, or with a greater fold-change in aged animals, although only a subset of them reached statistically significant interaction term as outputted by DESeq2 analysis. In order to identify all of the genes that are likely subject to age-stroke interaction, we have prepared four additional sets of genes with age-dependent differential response to stroke containing genes that are significantly influenced by stroke in one age group and at the same time their fold-change (vs control) is at least doubled compared to second age group. That is, the set “more up MCAO18” is comprised of genes significantly upregulated in aged animals after stroke (compared to aged controls; $p_{\text{adj}} < 0.01$, $\log_2\text{FC}_{\text{aged}} > 1$), and at the same time having significant interaction term ($p_{\text{adj}} < 0.1$) and/or having at least double the fold change of young strokes (compared to young controls; $\log_2\text{FC}_{\text{aged}} - \log_2\text{FC}_{\text{young}} > 1$). The same rationale was applied for more highly upregulated genes in the young stroke group (“more up MCAO3”) and for downregulated genes in both age groups (“more down MCAO18”, “more down MCAO3”), with an exception that the $\log_2\text{FC}$ threshold was < -0.65 for downregulation. DE sets legend is also available in [Table S1](#). Gene sets composition can be found in [Supplementary file 3](#). Only genes with average expression ≥ 5 normalized counts in at least one experimental condition were considered for further analysis.

Gene set enrichment analysis (GSEA)

GSEA⁵⁰ was performed for pairwise differential expression comparisons. First, a gene score was calculated for every gene using DESeq2 output as $-\log_{10}(p_{\text{adj}})$ and assigned a positive or negative sign based on direction of regulation. Genes were ranked by their gene-scores and GSEA was run in a weighted pre-ranked mode with 1000 permutations. Gene sets were downloaded from http://download.baderlab.org/EM_Genesets/, and GSEA was run separately for two gene ontology (GO) categories (biological process – GOBP, cellular component – GOCC). Only gene sets containing between 15 and 1000 (for GOBP) or 5 to 1000 (for GOCC) genes were considered. Annotations with IEA (inferred from electronic annotation) evidence codes were excluded. For pathway enrichment, a gene set file integrating several pathway databases was used. Significantly overrepresented gene sets were visualized as a network using Enrichment Map⁶². In the network, each node represents gene set and highly overlapping gene sets are connected with edges, resulting in a tight clustering of highly redundant gene sets. For functional annotation of discrete sets of genes we used Cytoscape plugin ClueGO¹⁵¹ with the following parameters: no IEA codes, right-sided hypergeometric test with Benjamini-Hochberg correction for statistical testing and all genes after filtering (16048 genes) as the background set.

Cell-specific gene sets and cell type proportion estimation

Marker genes for major cell types specifically in the mouse cortex region were taken as an initial reference⁷¹ ([Supplementary File 3](#)). Unlike other marker databases that rely on a single data source, this marker set represents a consensus from several published studies and accounts for brain regional heterogeneity. The microglial marker genes in the reference marker set were already devoid of genes differentially expressed in activated microglia⁵³. In order to acquire marker genes with stable expression regardless of activation

states, we have further removed the genes previously found to be differentially expressed in microglia after tMCAO¹⁵²; in aged cortical microglia⁵⁴, and aged whole-brain microglia⁵³ compared to young microglia; and genes enriched in bone marrow-derived macrophages compared to microglia¹⁵³. We have also removed genes that were differentially expressed in astrocytes after tMCAO⁶⁴ and aged cortical astrocytes^{51,52}. Because peripheral immune cells may infiltrate the brain following stroke, we have also excluded genes enriched in the major leukocyte populations obtained from ImmGen database (www.immgen.org).

DESeq2-normalized gene expression data and the cell-specific gene lists were used as an input into the *marker-GeneProfile* R package v1.0.3⁷¹ for the estimation of marker gene profiles (MGPs), which serve as a proxy for relative cell type proportion changes. We used a more stringent expression cutoff (average ≥ 5 normalized counts across all samples) to reduce the transcriptional noise. Since it still may be possible that some marker genes are transcriptionally regulated under our experimental conditions, genes with reduced correlation (potentially regulated) to the majority of marker genes (assumed to reflect primarily cell type proportion change) were excluded from final estimates as described in⁷¹. Resulting MGP estimates were flagged if a high proportion of marker genes was removed in the previous step ($> 40\%$) and/or proportion of variance explained by the first principal component was low ($< 50\%$). Differences in expression of final marker gene sets were analyzed by linear mixed model in R project v3.6.0 using *lmerTest* package v3.1¹⁵⁴ on \log_2 transformed, DESeq2-normalized expression values. We used two-factor design (age, injury) with interaction as fixed effects and gene intercept as random effect. Significance was tested by Satterthwaite's method. Post-hoc t-tests were performed using *emmeans* package v1.3.5 and p-values were adjusted by Holm method.

To validate the first estimations, we employed CIBERSORT, a transcriptome deconvolution algorithm that uses gene expression matrix of individual cell types as a reference, and deconvolutes the cellular composition of mixed sample by linear support vector regression⁷². We used published single-cell RNA-Seq dataset of adult mouse cortex as a reference gene expression signature¹⁵⁵. From the normalized gene-expression matrix, we excluded all intermediate cells as defined by authors and used the remaining 1424 core cells assigned to the major CNS cell types. CIBERSORT was run in the relative mode with 1000 permutations, quantile normalization was disabled as recommended for RNA-Seq data and q-value cutoff was lowered to 0.15. The results were correlated to the corresponding cellular MGPs ([Figure S6](#)).

Weighted gene co-expression network analysis (WGCNA)

Standard WGCNA procedure was followed to create gene co-expression networks using *blockwiseModules* function from the *WGCNA* R package v1.68¹⁵⁶. Filtered DESeq2-normalized expression data were used to calculate Pearson correlation between all gene pairs. Soft-thresholding was then applied by raising correlation values to a power of 22 to amplify disparity between strong and weak correlations. The soft-thresholding power was chosen to achieve approximately scale-free network topology, as recommended for biological networks^{157,158}.

The resulting signed adjacency matrix was used to calculate topological overlap matrix (TO), which was then hierarchically clustered with (1-TO) as a distance measure. Genes were then assigned into co-expression modules by dynamic tree cutting algorithm requiring minimal module size of 20 genes. Modules with a distance between the module eigengenes (MEs) of less than 0.2 were merged. ME is the first principal component of the gene expression values within a module and is used to summarize the module's expression. Pearson correlation between each gene and ME was then calculated. This value is called module membership and represents how close a particular gene is to a module. Finally, each gene was assigned to a module for which it had the highest module membership.

Motif and transcription factor enrichment analysis

Cytoscape plugin iRegulon¹⁵⁹ with default parameters was used to search for over-represented motifs and their associated transcription factors 500 bp upstream of the transcription start site. All genes from a particular WGCNA module were used as an input. A transcription factor was considered a hit for a given module only if its gene belonged to the same module.

Protein-protein interaction network

Known interactions (minimal interaction score 0.4) between genes in the "more up MCAO18" DE gene set were downloaded from STRING database v10.5¹⁶⁰. Remaining unconnected genes from "more up MCAO18" gene set were then added to the network based on their correlation with any of the genes already present in the network requiring Pearson $r \geq 0.96$ (edges visualized with orange dotted line in Figure 5E). The resulting interaction network was then visualized and analyzed in Cytoscape v3.5.1. Spectral partition-based network clustering algorithm¹⁶¹ via Cytoscape ReactomeFI plugin v6.1.0¹⁶² was used for network clustering.

Custom gene set enrichment

Gene sets of interest were collected directly from relevant publications. For references and legends, see Table S1. R package *GeneOverlap* v1.20 was used to calculate the odds ratio (OR) and the significance of the overlap of the gene sets of interest with the Fisher's exact test.

High-throughput RT-qPCR

Samples were reverse transcribed in a reaction volume of 10 μ l containing: 5 μ l template (either 125 ng total tissue RNA or 100 sorted cells after direct lysis), 0.5 μ l spike-in RNA (Tataa Biocenter; $c = 0.1x$ for tissues or $0.01x$ for sorted cells), 0.5 μ l equimolar mixture of random hexamers with oligo(dT) ($c = 50 \mu$ M), 0.5 μ l dNTPs ($c = 10$ mM), 2 μ l 5 \times RT buffer, 0.5 μ l RNaseOUT, 0.5 μ l Maxima H- Reverse Transcriptase (all Thermo Fisher) and 0.5 μ l nuclease-free water. After the pre-incubation step at 65°C ($t = 5$ min), followed by the immediate cooling on ice, the main incubation was performed at 25°C ($t = 10$ min), 50°C ($t = 30$ min), 85°C ($t = 5$ min), after which the samples were immediately cooled on ice. cDNA from tissue samples was diluted 4x in nuclease-free water; sorted cell-cDNA was left undiluted. All cDNA samples were pre-amplified immediately after reverse transcription in 40 μ l total reaction volume containing 4 μ l cDNA, 20 μ l IQ Supermix buffer (Bio-Rad), 4 μ l primer mix of 96 assays ($c = 250$ nM each), and 12 μ l of nuclease-free water. Reactions were incubated at 95°C ($t = 3$ min) following by 18 cycles of 95°C ($t =$

20 s), 57°C ($t = 4$ min) and 72°C ($t = 20$ s). After thermal cycling, reactions were immediately cooled on ice and diluted in nuclease-free water (sorted cells 4x, tissue 50x). High-throughput qPCR was then performed on a 96.96 microfluidic platform BioMark (Fluidigm) as previously described¹⁶³. Cycling program consisted of activation at 95°C ($t = 3$ min), followed by 40 cycles of 95°C ($t = 5$ s), 60°C ($t = 15$ s) and 72°C ($t = 20$ s) and melting curve analysis.

RT-qPCR data analysis

Raw data were pre-processed with the Real-Time PCR analysis software v4.1.3 (Fluidigm); unspecific values were deleted based on melting-curve analysis. Further processing was done in GenEx v6.0.1 (MultiD Analyses AB): Cq value cutoff of 28 was applied; gDNA background was subtracted using ValidPrime¹⁶⁴ (Tataa Biocenter); data were normalized to the mean expression of 5 reference genes (*Actb*, *Gapdh*, *Ppia*, *Ywhaz*, *Tubb5*); outliers were deleted (within group Grubbs test, $p < 0.05$) and a gene was considered undetected for given group if more than 75% values per group were missing; technical replicates (RT and FACS) were averaged; if appropriate, missing data were imputed on a within-group basis and remaining missing data were replaced with $Cq_{max} + 2$ for tissue samples or $Cq_{max} + 0.5$ for sorted cells. Of note, the RT-qPCR and RNA-Seq data showed high correlation (Pearson $r \geq 0.942$; Figure S13).

Temporal expression of individual genes was first analyzed with two-way ANOVA in R project v3.6.0 using time-point and age as predictor variables, then differences between time-points were tested separately for each age by one-way ANOVA and post-hoc t-tests using *emmeans* package v1.3.5. P-values were adjusted with Benjamini-Hochberg method. Temporal expression of groups of cellular marker genes was first analyzed using linear mixed model in R project v3.6.0 (*lmerTest* package v3.1¹⁵⁴) with random gene intercept using two-factor design (time-point including control, age) with interaction, then differences between time-points were tested separately for each age. Significance was tested by Satterthwaite's method. Post-hoc pairwise t-tests were performed using *emmeans* package v1.3.5 and p-values were adjusted by Holm method. Temporal expression of IFN-I pathway was analyzed in similar steps using random slope and random intercept mixed model with 23 interferon-stimulated genes as response variables. IFN-I pathway expression in sorted cells was analyzed in the same way with two-factor design (age, injury) with interaction, using only detected ISGs per each cell type. Differential expression of individual genes (relative to age-matched control) in sorted cells was tested in GenEx v6.0.1 (MultiD Analyses AB) using ANOVA with Bonferroni's post-hoc test for selected pairwise comparisons and p-values were corrected using Benjamini-Hochberg method (Table S2).

Acknowledgements

This study was supported by: Czech Science Foundation (16-10214S and 19-02046S); Institutional support (RVO 86652036 and BIOCEV CZ.1.05/1.1.00/02.0109).

Conflict of interest statement

All authors declare that they have no conflict of interest.

Author contributions

P.An. prepared sequencing libraries, analyzed and interpreted the data, wrote the manuscript and prepared all figures. D.K., J.T. and J.K. performed animal surgeries, prepared cell suspensions and performed FACS experiments. D.Z. and E.R. performed RNA isolation and RT-qPCR experiments. P.Ab. performed raw read pre-processing, mapping and DESeq2 analysis. D.K. M.A. M.K. and L.V. commented on the manuscript. M.K. M.A. and L.V. conceived and supervised the study. All authors reviewed the manuscript.

Supplementary material

Supplementary material includes [Figures S1-S13](#); [Tables S1 and S2](#); Supplementary excel files containing gene expression matrix, DESeq2 results, gene membership in described gene sets and sets collected from the literature, and searchable excel file for visualization of user-defined genes.

Availability of data and materials

The dataset supporting the conclusions of this article is available in the Gene Expression Omnibus database under accession number [GSE137482](#). In addition, processed data are available as supplementary files.

References

- Giroud, M., Jacquin, A. & Béjot, Y. The worldwide landscape of stroke in the 21st century. *Lancet* **383**, 195–197 (2014).
- Chen, R.-L., Balami, J. S., Esiri, M. M., Chen, L.-K. & Buchan, A. M. Ischemic stroke in the elderly: an overview of evidence. *Nat. Rev. Neurol.* **6**, 256–265 (2010).
- Sturm, J. W. *et al.* Quality of Life After Stroke. *Stroke* **35**, 2340–2345 (2004).
- Saka, O., McGuire, A. & Wolfe, C. Cost of stroke in the United Kingdom. *Age Ageing* **38**, 27–32 (2008).
- Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics—2018 Update: A Report From the American Heart Association. *Circulation* **137**, (2018).
- Deb, P., Sharma, S. & Hassan, K. M. Pathophysiologic mechanisms of acute ischemic stroke: An overview with emphasis on therapeutic significance beyond thrombolysis. *Pathophysiology* **17**, 197–218 (2010).
- Puig, B., Brenna, S. & Magnus, T. Molecular Communication of a Dying Neuron in Stroke. *Int. J. Mol. Sci.* **19**, 2834 (2018).
- Majid, A. Neuroprotection in stroke: past, present, and future. *ISRN Neurol.* **2014**, 515716 (2014).
- Green, A. R. Pharmacological approaches to acute ischaemic stroke: reperfusion certainly, neuroprotection possibly. *Br. J. Pharmacol.* **153 Suppl 1**, S325-38 (2008).
- O'Collins, V. E. *et al.* 1,026 Experimental treatments in acute stroke. *Ann. Neurol.* **59**, 467–477 (2006).
- Roth, J. M. Recombinant tissue plasminogen activator for the treatment of acute ischemic stroke. *Proc. (Bayl. Univ. Med. Cent)*. **24**, 257–9 (2011).
- Liu, F., Yuan, R., Benashski, S. E. & McCullough, L. D. Changes in Experimental Stroke Outcome across the Life Span. *J. Cereb. Blood Flow Metab.* **29**, 792–802 (2009).
- Roy-O'Reilly, M. & McCullough, L. D. Age and Sex Are Critical Factors in Ischemic Stroke Pathology. *Endocrinology* **159**, 3120–3131 (2018).
- Worp, H. B., Haan, P., Morrema, E. & Kalkman, C. J. Methodological quality of animal studies on neuroprotection in focal cerebral ischaemia. *J. Neurol.* **252**, 1108–1114 (2005).
- Kammersgaard, L. P. *et al.* Short- and long-term prognosis for very old stroke patients. The Copenhagen Stroke Study. *Age Ageing* **33**, 149–154 (2004).
- Marini, C. *et al.* Burden of first-ever ischemic stroke in the oldest old: evidence from a population-based study. *Neurology* **62**, 77–81 (2004).
- Gargano, J. W. & Reeves, M. J. Sex Differences in Stroke Recovery and Stroke-Specific Quality of Life. *Stroke* **38**, 2541–2548 (2007).
- Di Carlo, A. *et al.* Sex Differences in the Clinical Presentation, Resource Use, and 3-Month Outcome of Acute Stroke in Europe. *Stroke* **34**, 1114–1119 (2003).
- Bushnell, C. D. *et al.* Sex differences in quality of life after ischemic stroke. *Neurology* **82**, 922–931 (2014).
- Haast, R. A., Gustafson, D. R. & Kiliaan, A. J. Sex Differences in Stroke. *J. Cereb. Blood Flow Metab.* **32**, 2100–2107 (2012).
- Rosen, C. L., Dinapoli, V. A., Nagamine, T. & Crocco, T. Influence of age on stroke outcome following transient focal ischemia. *J. Neurosurg.* **103**, 687–694 (2005).
- Shapira, S., Sapir, M., Wengier, A., Grauer, E. & Kadar, T. Aging has a complex effect on a rat model of ischemic stroke. *Brain Res.* **925**, 148–58 (2002).
- DiNapoli, V. A., Huber, J. D., Houser, K., Li, X. & Rosen, C. L. Early disruptions of the blood–brain barrier may contribute to exacerbated neuronal damage and prolonged functional recovery following stroke in aged rats. *Neurobiol. Aging* **29**, 753–764 (2008).
- Manwani, B. *et al.* Differential effects of aging and sex on stroke induced inflammation across the lifespan. *Exp. Neurol.* **249**, 120–131 (2013).
- Chisholm, N. C. & Sohrabji, F. Astrocytic response to cerebral ischemia is influenced by sex differences and impaired by aging. *Neurobiol. Dis.* **85**, 245–253 (2016).
- Sohrabji, F., Bake, S. & Lewis, D. K. Age-related changes in brain support cells: Implications for stroke severity. *Neurochem. Int.* **63**, 291–301 (2013).
- Soriano, M. A., Tessier, M., Certa, U. & Gill, R. Parallel Gene Expression Monitoring using Oligonucleotide Probe Arrays of Multiple Transcripts with an Animal Model of Focal Ischemia. *J. Cereb. Blood Flow Metab.* **20**, 1045–1055 (2000).
- Schmidt-Kastner, R. *et al.* DNA microarray analysis of cortical gene expression during early recirculation after focal brain ischemia in rat. *Brain Res. Mol. Brain Res.* **108**, 81–93 (2002).
- Lu, X.-C. M. *et al.* Microarray analysis of acute and delayed gene expression profile in rats after focal ischemic brain injury and reperfusion. *J. Neurosci. Res.* **77**, 843–857 (2004).
- Kim, J.-B. *et al.* Delayed genomic responses to transient middle cerebral artery occlusion in the rat. *J. Neurochem.* **89**, 1271–1282 (2004).
- Jin, K. *et al.* Microarray analysis of hippocampal gene expression in global cerebral ischemia. *Ann. Neurol.* **50**, 93–103 (2001).
- Ford, G., Xu, Z., Gates, A., Jiang, J. & Ford, B. D. Expression Analysis Systematic Explorer (EASE) analysis reveals differential gene expression in permanent and transient focal stroke rat models. *Brain Res.* **1071**, 226–236 (2006).
- Roth, A., Gill, R. & Certa, U. Temporal and spatial gene expression patterns after experimental stroke in a rat model and characterization of PC4, a potential regulator of transcription. *Mol. Cell. Neurosci.* **22**, 353–64 (2003).
- Mitsios, N. *et al.* A microarray study of gene and protein regulation in human and rat brain following middle cerebral artery occlusion. *BMC Neurosci.* **8**, 93 (2007).
- Dergunova, L. V. *et al.* Genome-wide transcriptome analysis using RNA-Seq reveals a large number of differentially expressed genes in a transient MCAO rat model. *BMC Genomics* **19**, 655 (2018).

36. Duan, X. *et al.* RNA Sequencing for Gene Expression Profiles in a Rat Model of Middle Cerebral Artery Occlusion. *Biomed Res. Int.* **2018**, 1–14 (2018).
37. Zhang, C. *et al.* Temporal Gene Expression Profiles after Focal Cerebral Ischemia in Mice. *Aging Dis.* **9**, 249–261 (2018).
38. VanGilder, R. L., Huber, J. D., Rosen, C. L. & Barr, T. L. The transcriptome of cerebral ischemia. *Brain Res. Bull.* **88**, 313–319 (2012).
39. Sieber, M. W. *et al.* Age-specific transcriptional response to stroke. *Neurobiol. Aging* **35**, 1744–1754 (2014).
40. Buga, A.-M. M. *et al.* The genomic response of the ipsilateral and contralateral cortex to stroke in aged rats. *J. Cell. Mol. Med.* **12**, 2731–2753 (2008).
41. Buga, A. M. *et al.* Transcriptomics of Post-Stroke Angiogenesis in the Aged Brain. *Front. Aging Neurosci.* **6**, 1–20 (2014).
42. Stroke Therapy Academic Industry Roundtable (STAIR). Recommendations for Standards Regarding Preclinical Neuroprotective and Restorative Drug Development. *Stroke* **30**, 2752–2758 (1999).
43. Hossmann, K.-A. Cerebral ischemia: Models, methods and outcomes. *Neuropharmacology* **55**, 257–270 (2008).
44. Kahle, M. P. & Bix, G. J. Successfully Climbing the “STAIRs”: Surmounting Failed Translation of Experimental Ischemic Stroke Treatments. *Stroke Res. Treat.* **2012**, 1–8 (2012).
45. Hossmann, K.-A. The Two Pathophysiologies of Focal Brain Ischemia: Implications for Translational Stroke Research. *J. Cereb. Blood Flow Metab.* **32**, 1310–1316 (2012).
46. McBride, D. W. & Zhang, J. H. Precision Stroke Animal Models: the Permanent MCAO Model Should Be the Primary Model, Not Transient MCAO. *Transl. Stroke Res.* **8**, 397–404 (2017).
47. Tsvigoulis, G., Katsanos, A. H. & Caso, V. Underrepresentation of women in stroke randomized controlled trials: inadvertent selection bias leading to suboptimal conclusions. *Ther. Adv. Neurol. Disord.* **10**, 241–244 (2017).
48. Sohrabji, F., Park, M. J. & Mahnke, A. H. Sex differences in stroke therapies. *J. Neurosci. Res.* **95**, 681–691 (2017).
49. Carcel, C. *et al.* Trends in recruitment of women and reporting of sex differences in large-scale published randomized controlled trials in stroke. *Int. J. Stroke* 174749301985129 (2019). doi:10.1177/1747493019851292
50. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–50 (2005).
51. Boisvert, M. M., Erikson, G. A., Shokhirev, M. N. & Allen, N. J. The Aging Astrocyte Transcriptome from Multiple Regions of the Mouse Brain. *Cell Rep.* **22**, 269–285 (2018).
52. Clarke, L. E. *et al.* Normal aging induces A1-like astrocyte reactivity. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E1896–E1905 (2018).
53. Holtman, I. R. *et al.* Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: a co-expression meta-analysis. *Acta Neuropathol. Commun.* **3**, 31 (2015).
54. Grabert, K. *et al.* Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nat. Neurosci.* **19**, 504–516 (2016).
55. Spitzer, S. O. *et al.* Oligodendrocyte Progenitor Cells Become Regionally Diverse and Heterogeneous with Age. *Neuron* **0**, (2019).
56. Emy, D. *et al.* Host microbiota constantly control maturation and function of microglia in the CNS. *Nat. Neurosci.* **18**, 965–977 (2015).
57. Srinivasan, K. *et al.* Untangling the brain’s neuroinflammatory and neurodegenerative transcriptional responses. *Nat. Commun.* **7**, 11295 (2016).
58. Wang, I.-M. *et al.* Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol. Syst. Biol.* **8**, 594 (2012).
59. Friedman, B. A. *et al.* Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer’s Disease Not Evident in Mouse Models. *Cell Rep.* **22**, 832–847 (2018).
60. Hammond, T. R. *et al.* Single-Cell RNA Sequencing of Microglia throughout the Mouse Lifespan and in the Injured Brain Reveals Complex Cell-State Changes. *Immunity* **50**, 253–271.e6 (2019).
61. Mathys, H. *et al.* Temporal Tracking of Microglia Activation in Neurodegeneration at Single-Cell Resolution. *Cell Rep.* **21**, 366–380 (2017).
62. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS One* **5**, e13984 (2010).
63. Mrdjen, D. *et al.* High-Dimensional Single-Cell Mapping of Central Nervous System Immune Cells Reveals Distinct Myeloid Subsets in Health, Aging, and Disease. *Immunity* **48**, 380–395.e6 (2018).
64. Zamanian, J. L. *et al.* Genomic Analysis of Reactive Astroglia. *J. Neurosci.* **32**, 6391–6410 (2012).
65. Badan, I. *et al.* Accelerated Glial Reactivity to Stroke in Aged Rats Correlates with Reduced Functional Recovery. *J. Cereb. Blood Flow Metab.* **23**, 845–854 (2003).
66. Gordon, M. N., Schreier, W. A., Ou, X., Holcomb, L. A. & Morgan, D. G. Exaggerated astrocyte reactivity after nigrostriatal deafferentation in the aged rat. *J. Comp. Neurol.* **388**, 106–119 (1997).
67. Popa-Wagner, A. *et al.* Accelerated infarct development, cytogenesis and apoptosis following transient cerebral ischemia in aged rats. *Acta Neuropathol.* **113**, 277–293 (2007).
68. Meng, H., Liu, T., Borjigin, J. & Wang, M. M. Ischemic stroke destabilizes circadian rhythms. *J. Circadian Rhythms* **6**, 9 (2008).
69. Sterr, A. *et al.* Post-stroke insomnia in community-dwelling patients with chronic motor stroke: Physiological evidence and implications for stroke care. *Sci. Rep.* **8**, 8409 (2018).
70. Duss, S. B. *et al.* The role of sleep in recovery following ischemic stroke: A review of human and animal data. *Neurobiol. Sleep Circadian Rhythm.* **2**, 94–105 (2017).
71. Mancarci, B. O. *et al.* Cross-Laboratory Analysis of Brain Cell Type Transcriptomes with Applications to Interpretation of Bulk Tissue Data. *eneuro* **4**, ENEURO.0212-17.2017 (2017).
72. Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
73. Hu, H., Gan, J. & Jonas, P. Fast-spiking, parvalbumin+ GABAergic interneurons: From cellular design to microcircuit function. *Science (80-.)*. **345**, 1255263–1255263 (2014).
74. Hattori, R., Kuchibhotla, K. V., Froemke, R. C. & Komiyama, T. Functions and dysfunctions of neocortical inhibitory neuron subtypes. *Nat. Neurosci.* **20**, 1199–1208 (2017).
75. Paul, A. *et al.* Transcriptional Architecture of Synaptic Communication Delineates GABAergic Neuron Identity. *Cell* **171**, 522–539.e20 (2017).
76. Rudy, B. & McBain, C. J. Kv3 channels: voltage-gated K⁺ channels designed for high-frequency repetitive firing. *Trends Neurosci.* **24**, 517–26 (2001).
77. Kaczmarek, L. K. & Zhang, Y. Kv3 Channels: Enablers of Rapid Firing, Neurotransmitter Release, and Neuronal Endurance. *Physiol. Rev.* **97**, 1431–1468 (2017).
78. Weiser, M. *et al.* The potassium channel subunit KV3.1b is localized to somatic and axonal membranes of specific populations of CNS neurons. *J. Neurosci.* **15**, 4298–314 (1995).
79. Ogiwara, I. *et al.* Nav1.1 Localizes to Axons of

- Parvalbumin-Positive Inhibitory Interneurons: A Circuit Basis for Epileptic Seizures in Mice Carrying an Scn1a Gene Mutation. *J. Neurosci.* **27**, 5903–5914 (2007).
80. Meylan, E. M. *et al.* Involvement of the agmatineric system in the depressive-like phenotype of the Crtc1 knockout mouse model of depression. *Transl. Psychiatry* **6**, e852 (2016).
81. Rossier, J. *et al.* Cortical fast-spiking parvalbumin interneurons enwrapped in the perineuronal net express the metalloproteinases Adamts8, Adamts15 and Nephylisin. *Mol. Psychiatry* **20**, 154–61 (2015).
82. Crow, M., Paul, A., Ballouz, S., Huang, Z. J. & Gillis, J. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat. Commun.* **9**, 884 (2018).
83. Mo, A. *et al.* Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. *Neuron* **86**, 1369–1384 (2015).
84. Sommeijer, J.-P. & Levelt, C. N. Synaptotagmin-2 Is a Reliable Marker for Parvalbumin Positive Inhibitory Boutons in the Mouse Visual Cortex. *PLoS One* **7**, e35323 (2012).
85. Pang, Z. P. *et al.* Synaptotagmin-2 Is Essential for Survival and Contributes to Ca²⁺ Triggering of Neurotransmitter Release in Central and Neuromuscular Synapses. *J. Neurosci.* **26**, 13493–13504 (2006).
86. Bouhours, B., Gjoni, E., Kochubey, O. & Schneggenburger, R. Synaptotagmin2 (Syt2) Drives Fast Release Redundantly with Syt1 at the Output Synapses of Parvalbumin-Expressing Inhibitory Neurons. *J. Neurosci.* **37**, 4604–4617 (2017).
87. Chen, C., Arai, I., Satterfield, R., Young, S. M. & Jonas, P. Synaptotagmin 2 Is the Fast Ca²⁺ Sensor at a Central Inhibitory Synapse. *Cell Rep.* **18**, 723–736 (2017).
88. Lucas, E. K. *et al.* PGC-1 Provides a Transcriptional Framework for Synchronous Neurotransmitter Release from Parvalbumin-Positive Interneurons. *J. Neurosci.* **34**, 14375–14387 (2014).
89. Bartley, A. F. *et al.* Interneuron Transcriptional Dysregulation Causes Frequency-Dependent Alterations in the Balance of Inhibition and Excitation in Hippocampus. *J. Neurosci.* **35**, 15276–90 (2015).
90. Zeisel, A. *et al.* Molecular Architecture of the Mouse Nervous System. *Cell* **174**, 999–1014.e22 (2018).
91. Tasic, B. *et al.* Shared and distinct transcriptomic cell types across neocortical areas. *Nature* **563**, 72–78 (2018).
92. Jordão, M. J. C. *et al.* Single-cell profiling identifies myeloid cell subsets with distinct fates during neuroinflammation. *Science* (80-.). **363**, eaat7554 (2019).
93. Han, X. *et al.* Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091–1107.e17 (2018).
94. Imperato, M. R., Cauchy, P., Obier, N. & Bonifer, C. The RUNX1–PU.1 axis in the control of hematopoiesis. *Int. J. Hematol.* **101**, 319–329 (2015).
95. Suzuki, E. *et al.* The transcription factor Fli-1 regulates monocyte, macrophage and dendritic cell development in mice. *Immunology* **139**, 318–327 (2013).
96. Kierdorf, K. & Prinz, M. Factors regulating microglia activation. *Front. Cell. Neurosci.* **7**, 44 (2013).
97. Wehrspaun, C. C., Haerty, W. & Ponting, C. P. Microglia recapitulate a hematopoietic master regulator network in the aging human frontal cortex. *Neurobiol. Aging* **36**, 2443.e9–2443.e20 (2015).
98. Owens, T., Khoroshii, R., Wlodarczyk, A. & Asgari, N. Interferons in the central nervous system: A few instruments play many tunes. *Glia* **62**, 339–355 (2014).
99. Mostafavi, S. *et al.* Parsing the Interferon Transcriptional Network and Its Disease Associations. *Cell* **164**, 564–578 (2016).
100. Zhang, M. *et al.* Type-I interferon signalling through IFNAR1 plays a deleterious role in the outcome after stroke. *Neurochem. Int.* **108**, 472–480 (2017).
101. Blank, T. & Prinz, M. Type I interferon pathway in CNS homeostasis and neurological disorders. *Glia* **65**, 1397–1406 (2017).
102. Takagi, Y., Harada, J., Chiarugi, A. & Moskowitz, M. A. STAT1 is Activated in Neurons after Ischemia and Contributes to Ischemic Brain Injury. *J. Cereb. Blood Flow Metab.* **22**, 1311–1318 (2002).
103. McDonough, A. *et al.* Ischemia/Reperfusion Induces Interferon-Stimulated Gene Expression in Microglia. *J. Neurosci.* **37**, 8292–8308 (2017).
104. Lécuyer, M.-A., Kebir, H. & Prat, A. Glial influences on BBB functions and molecular players in immune cell trafficking. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1862**, 472–482 (2016).
105. Cribbs, D. H. *et al.* Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study. *J. Neuroinflammation* **9**, 643 (2012).
106. Benayoun, B. A. *et al.* Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res.* (2019). doi:10.1101/gr.240093.118
107. Ori, A. *et al.* Integrated Transcriptome and Proteome Analyses Reveal Organ-Specific Proteome Deterioration in Old Rats. *Cell Syst.* **1**, 224–237 (2015).
108. Soreq, L. *et al.* Major Shifts in Glial Regional Identity Are a Transcriptional Hallmark of Human Brain Aging. *Cell Rep.* **18**, 557–570 (2017).
109. Davie, K. *et al.* A Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell* **174**, 982–998.e20 (2018).
110. Liddelow, S. A. *et al.* Neurotoxic reactive astrocytes are induced by activated microglia. *Nature* **541**, 481–487 (2017).
111. Miller, J. A., Woltjer, R. L., Goodenbour, J. M., Horvath, S. & Geschwind, D. H. Genes and pathways underlying regional and cell type changes in Alzheimer’s disease. *Genome Med.* **5**, 48 (2013).
112. Yu, Q. & He, Z. Comprehensive investigation of temporal and autism-associated cell type composition-dependent and independent gene expression changes in human brains. *Sci. Rep.* **7**, 1–12 (2017).
113. Peters, A. & Sethares, C. Oligodendrocytes, their Progenitors and other Neuroglial Cells in the Aging Primate Cerebral Cortex. *Cereb. Cortex* **14**, 995–1007 (2004).
114. Vostrikov, V. & Uranova, N. Age-related increase in the number of oligodendrocytes is dysregulated in schizophrenia and mood disorders. *Schizophr. Res. Treatment* **2011**, 174689 (2011).
115. Young, K. M. *et al.* Oligodendrocyte Dynamics in the Healthy Adult CNS: Evidence for Myelin Remodeling. *Neuron* **77**, 873–885 (2013).
116. del Pino, I. *et al.* Erbb4 Deletion from Fast-Spiking Interneurons Causes Schizophrenia-like Phenotypes. *Neuron* **79**, 1152–1168 (2013).
117. Lewis, D. A., Hashimoto, T. & Volk, D. W. Cortical inhibitory neurons and schizophrenia. *Nat. Rev. Neurosci.* **6**, 312–324 (2005).
118. Verret, L. *et al.* Inhibitory Interneuron Deficit Links Altered Network Activity and Cognitive Dysfunction in Alzheimer Model. *Cell* **149**, 708–721 (2012).
119. Zhou, Z. *et al.* Loss of Phenotype of Parvalbumin Interneurons in Rat Prefrontal Cortex Is Involved in Antidepressant- and Pro-psychotic-Like Behaviors Following Acute and Repeated Ketamine Administration. *Mol. Neurobiol.* **51**, 808–819 (2015).
120. Sauer, J.-F., Strüber, M. & Bartos, M. Impaired fast-spiking interneuron function in a genetic mouse model of depression. *Elife* **4**, (2015).
121. Donato, F., Rompani, S. B. & Caroni, P. Parvalbumin-expressing basket-cell network plasticity induced by experience regulates adult learning. *Nature* **504**, 272–276 (2013).
122. Sohal, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459**, 698–702 (2009).

123. Carmichael, S. T. Brain Excitability in Stroke. *Arch. Neurol.* **69**, 161 (2012).
124. Streckler, J.-K., Schmidt, A., Schäbitz, W.-R. & Minnerup, J. Neutrophil granulocytes in cerebral ischemia – Evolution from killers to key players. *Neurochem. Int.* **107**, 117–126 (2017).
125. Turner, R. J. & Sharp, F. R. Implications of MMP9 for Blood Brain Barrier Disruption and Hemorrhagic Transformation Following Ischemic Stroke. *Front. Cell. Neurosci.* **10**, 56 (2016).
126. Allen, C. L. & Bayraktutan, U. Oxidative Stress and Its Role in the Pathogenesis of Ischaemic Stroke. *Int. J. Stroke* **4**, 461–470 (2009).
127. Ritzel, R. M. *et al.* Aging alters the immunological response to ischemic stroke. *Acta Neuropathol.* **136**, 89–110 (2018).
128. Asano, S., Chantler, P. D. & Barr, T. L. Gene expression profiling in stroke: relevance of blood–brain interaction. *Curr. Opin. Pharmacol.* **26**, 80–86 (2016).
129. Prinz, M. & Knobeloch, K.-P. Type I Interferons as Ambiguous Modulators of Chronic Inflammation in the Central Nervous System. *Front. Immunol.* **3**, 67 (2012).
130. Goldmann, T. *et al.* USP18 lack in microglia causes destructive interferonopathy of the mouse brain. *EMBO J.* **34**, 1612–1629 (2015).
131. McDonough, A., Lee, R. V. & Weinstein, J. R. Microglial Interferon Signaling and White Matter. *Neurochem. Res.* **42**, 2625–2638 (2017).
132. Wang, R., Yang, B. & Zhang, D. Activation of interferon signaling pathways in spinal cord astrocytes from an ALS mouse model. *Glia* **59**, 946–958 (2011).
133. Taylor, J. M. *et al.* Type-1 interferon signaling mediates neuro-inflammatory events in models of Alzheimer’s disease. *Neurobiol. Aging* **35**, 1012–1023 (2014).
134. Sala Frigerio, C. *et al.* The Major Risk Factors for Alzheimer’s Disease: Age, Sex, and Genes Modulate the Microglia Response to A β Plaques. *Cell Rep.* **27**, 1293–1306.e6 (2019).
135. Roselli, F., Chandrasekar, A. & Morganti-Kossmann, M. C. Interferons in Traumatic Brain and Spinal Cord Injury: Current Evidence for Translational Application. *Front. Neurol.* **9**, 458 (2018).
136. Karve, I. P. *et al.* Ablation of Type-1 IFN Signaling in Hematopoietic Cells Confers Protection Following Traumatic Brain Injury. *eNeuro* **3**, (2016).
137. Abdullah, A. *et al.* STING-mediated type-I interferons contribute to the neuroinflammatory process and detrimental effects following traumatic brain injury. *J. Neuroinflammation* **15**, 323 (2018).
138. Chen, H.-Z. *et al.* A Critical Role for Interferon Regulatory Factor 9 in Cerebral Ischemic Stroke. (2014). doi:10.1523/JNEUROSCI.1545-14.2014
139. Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nat. Rev. Immunol.* **14**, 36 (2014).
140. de Weerd, N. A. *et al.* Structural basis of a unique interferon- β signaling axis mediated via the receptor IFNAR1. *Nat. Immunol.* **14**, 901–907 (2013).
141. Marsh, B. *et al.* Systemic lipopolysaccharide protects the brain from ischemic injury by reprogramming the response of the brain to stroke: a critical role for IRF3. *J. Neurosci.* **29**, 9839–49 (2009).
142. Hamner, M. A. *et al.* Ischemic Preconditioning in White Matter: Magnitude and Mechanism. *J. Neurosci.* **35**, 15599–611 (2015).
143. Jin, R., Yang, G. & Li, G. Inflammatory mechanisms in ischemic stroke: role of inflammatory cells. *J. Leukoc. Biol.* **87**, 779–89 (2010).
144. Savitz, S. I. Cell therapies: careful translation from animals to patients. *Stroke* **44**, S107-9 (2013).
145. Nolte, C. *et al.* GFAP promoter-controlled EGFP-expressing transgenic mice: A tool to visualize astrocytes and astrogliosis in living brain tissue. *Glia* **33**, 72–86 (2001).
146. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–20 (2014).
147. Kopylova, E., Noé, L. & Touzet, H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28**, 3211–3217 (2012).
148. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
149. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–9 (2015).
150. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
151. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
152. Arumugam, T. V *et al.* An atypical role for the myeloid receptor Mincle in central nervous system injury. *J. Cereb. Blood Flow Metab.* **37**, 2098–2111 (2017).
153. Bruttger, J. *et al.* Genetic Cell Ablation Reveals Clusters of Local Self-Renewing Microglia in the Mammalian Central Nervous System. *Immunity* **43**, 92–106 (2015).
154. Kuznetsova, A., Brockhoff, P. B. & Christensen, R. H. B. lmerTest Package: Tests in Linear Mixed Effects Models. *J. Stat. Softw.* **82**, 1–26 (2017).
155. Tasic, B. *et al.* Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat. Neurosci.* **19**, 335–346 (2016).
156. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
157. Barabási, A.-L. & Bonabeau, E. Scale-Free Networks. *Sci. Am.* **288**, 60–69 (2003).
158. Zhang, B. & Horvath, S. A General Framework for Weighted Gene Co-Expression Network Analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, Article17 (2005).
159. Janky, R. *et al.* iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput. Biol.* **10**, e1003731 (2014).
160. Szklarczyk, D. *et al.* STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–D452 (2015).
161. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl. Acad. Sci.* **103**, 8577–8582 (2006).
162. Wu, G., Feng, X. & Stein, L. A human functional protein interaction network and its application to cancer data analysis. *Genome Biol.* **11**, R53 (2010).
163. Rusnakova, V. *et al.* Heterogeneity of Astrocytes: From Development to Injury – Single Cell Gene Expression. *PLoS One* **8**, e69734 (2013).
164. Laurell, H. *et al.* Correction of RT–qPCR data for genomic DNA-derived signals with ValidPrime. *Nucleic Acids Res.* **40**, e51–e51 (2012).



PALACKÝ UNIVERSITY OLOMOUC
FACULTY OF SCIENCE
Laboratory of Growth Regulators

Peter Androvič

Methodology and application of quantitative microRNA analysis

Summary of the Ph.D. thesis

P1527 Biology
1501V019 Experimental Biology

Olomouc
2020

This Ph.D. thesis was carried out in the Laboratory of Gene Expression at the Institute of Biotechnology, Czech Academy of Sciences within the framework of internal Ph.D. Study of Experimental Biology, guaranteed by the Laboratory of Growth Regulators, Faculty of Science, Palacký University in Olomouc, between the years 2015-2020.

Ph.D. candidate: Ing. Peter Androvič

Supervisor: prof. Mikael Kubista, Ph.D. – Laboratory of Gene Expression, Institute of Biotechnology, Czech Academy of Sciences, Vestec, Czech Republic & Laboratory of Growth Regulators, Faculty of Science, Palacký University in Olomouc, Czech Republic

Consultant: Ing. Lukáš Valihrač, Ph.D. – Laboratory of Gene Expression, Institute of Biotechnology, Czech Academy of Sciences, Vestec, Czech Republic

Opponents: prof. RNDr. Ondřej Slabý, Ph.D. – Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic

RNDr. Dana Dlouhá Ph.D. – Institute for Clinical and Experimental Medicine (IKEM), Prague, Czech Republic

The evaluation of this Ph.D. thesis was written by Prof. Ing. Miroslav Strnad, CSc. DSc., Laboratory of Growth Regulators, Faculty of Science, Palacký University in Olomouc.

The oral defence will take place on **15.6.2020** before the Commission for the Ph.D. thesis of the Study Program Experimental Biology, **room.....**, Šlechtitelů 27, Olomouc – Holic.

The Ph.D. thesis and expert reviews will be available 14 days before the defence in the Study Department of Faculty of Science (Mgr. M. Karásková), Palacký University, 17. listopadu 12, Olomouc.

After the defense, the Ph.D. thesis will be stored in the Library of the Biological Departments of Faculty of Science, Palacký University, Šlechtitelů 27, Olomouc – Holic.

Prof. Ing. Miroslav Strnad, CSc. DSc.
Chairman of the Commission for the Ph.D. thesis,
Study Program Experimental Biology,
Faculty of Science, Palacký University in Olomouc

Table of Contents

1 Introduction.....	4
2 Aims and scope.....	5
3 Materials and methods	6
3.1 Samples	6
3.2 Primers and synthetic oligonucleotides	6
3.3 RNA isolation	6
3.4 RT-qPCR of miRNAs	7
3.5 Library preparation and sequencing	7
3.6 High-throughput RT-qPCR	7
3.7 Bioinformatic and statistical data analysis	8
4 Survey of results	10
4.1 Development of novel RT-qPCR-based method for highly accurate miRNA quantification	10
4.2 Development of quality control tool for circulating miRNA studies	10
4.3 Comprehensive performance comparison of small RNA-Seq library preparation methods from biofluids	11
4.4 Decoding the transcriptional response to ischemic stroke in young and aged mice.....	12
4.5 mRNA-miRNA regulatory networks in CNS injury (unpublished results)	14
5 Conclusion	15
6 References	16
7 List of publications.....	17
8 Súhrn (Summary, in Slovak).....	19

1 Introduction

According to the latest reports, the human genome encodes for approximately 19 000 protein-coding genes, which represents only 1.0 – 1.5% of 3 billion-plus base pairs it contains [1]. The rest was originally considered “junk” DNA. However, closer exploration of the genome revealed that in fact about 80% has biochemical role and is integral to the function of cells, particularly for the control of the gene activity [2]. In addition to various regulatory elements, the genome harbors at least 18 000 non-coding genes that produce over 47 000 non-coding RNAs including tRNAs, rRNAs, lncRNAs, snRNAs, piRNAs, circRNAs, siRNAs, miRNAs and many others, often with unknown function. One of the relatively well-characterized classes are microRNAs (miRNAs) – tiny, but all the more important regulators of gene expression. First miRNA was discovered by Victor Ambros and his co-workers Rosalind Lee and Rhonda Feinbaum in 1993 [3]. They found that *lin-4*, a gene known to control timing of *C. elegans* larval development by repression of the *lin-14* gene, does not code for a protein, but instead produces a pair of small RNAs. This discovery represented dramatic breakthrough in our understanding of the transcriptome. Since then, miRNAs have been shown to play important roles in modulation of an array of physiological and pathological processes ranging from embryonic development to neoplastic progression. This brought them significant attention as potential therapeutic targets. Another wave of excitement came with the discovery that miRNAs are released into extracellular environment and are stably present in the circulatory system and various body fluids. The realization that circulating miRNA levels change in response to pathophysiological processes meant that they might serve as promising and non-invasive clinical biomarkers to aid diagnosis, prognosis and monitoring of the response to treatment.

Whether the goal is to elucidate regulatory roles of the miRNAs, or find novel biomarkers, miRNA expression patterns provide essential information for these endeavors. Therefore, since the very beginning quantitative profiling of miRNA expression has played a pivotal role in the miRNA research. Several techniques for miRNA profiling are nowadays available for both targeted and global measurements, including microarrays, RT-qPCR and next generation sequencing methods. However, to obtain precise and reliable readouts of miRNA profiles is not straightforward because of many technical challenges associated with the workflow.

This thesis focuses on the development of novel methods and tools for improved miRNA quantitative workflows and on the assessment of small-RNA sequencing methods for better understanding of their impact on the resulting data. After establishment of the technical base, these methods are applied together with global mRNA expression profiling to dissect molecular and cellular response to acute central nervous system injuries including stroke and spinal cord injury in rodent models.

2 Aims and scope

The broader goal of this thesis was to develop and establish robust and reliable workflows for both, targeted and global miRNA expression profiling, including profiling of challenging samples such as biofluids. The ultimate goal was to utilize these workflows together with techniques for gene expression profiling to improve our understanding of molecular mechanisms of central nervous system (CNS) injury.

The specific aims of the work described in this thesis were:

- To develop, optimize and validate new method for miRNA quantification based on RT-qPCR that would allow precise and cost-effective quantification from various samples including animal and plant tissues, cells and biofluids.
- To develop easy-to-use tool that would allow convenient optimization and troubleshooting of the wet-lab workflow of quantitative miRNA studies as well as routine control of sample quality.
- To comprehensively evaluate all currently available methods for small RNA-Seq library preparation with focus on their performance with challenging samples such as biofluids to reveal their strengths, drawbacks, biases and impact on resulting data quality.
- To dissect global gene and miRNA expression changes after ischemic stroke and spinal cord injury (SCI) to reveal underlying molecular and cellular mechanisms.

3 Materials and methods

3.1 Samples

Animal samples were obtained from C57Black/6 or FVB mice, or from Wistar rats. Experimental stroke was induced by permanent MCAO on female mice, experimental SCI was induced by balloon compression lesion on male rats. For the preparation of rat serum, 1 ml of blood was collected from orbital plexus into 2 ml tubes (Eppendorf) using glass capillary. Blood was allowed to clot for 1 hour at room temperature and then centrifuged at 1000 g for 10 min. The clot was mechanically retracted from the tube wall before the centrifugation. Serum was transferred to another 2 ml tube and centrifuged a second time at 3000 g for 10 min. The supernatant was then transferred to cryovials (Biologix) and stored at -80°C until analysis. All procedures involving the use of laboratory animals were performed in accordance with appropriate regulations and efforts were made to minimize both the suffering and the number of animals used.

Human blood samples were obtained from healthy volunteers. Informed consent was obtained from all volunteers participating in the study. For the preparation of human serum, blood was collected into 8.5 ml BD Vacutainer SST II Advance tubes (Beckman Dickinson) and allowed to clot for at least 30 min before centrifugation at 1500 g for 10 min at room temperature. The serum was then transferred to 2 ml tubes (Eppendorf) and stored at -80°C . For the preparation of human plasma, blood was collected from four healthy volunteers into K_2EDTA BD Vacutainer tubes (Beckman Dickinson) and centrifuged within 30 min at 1500 g for 15 min at room temperature. The plasma fraction was aspirated and transferred to 2 ml tubes (Eppendorf) and centrifuged again for 15 min at 3000 g. The supernatant was transferred to new 2 ml tubes and stored at -80°C until analysis. Standardized human plasma sample was prepared by pooling high-quality RNA eluates from four healthy individuals.

3.2 Primers and synthetic oligonucleotides

Sequences of the miRNA oligonucleotides were obtained from the miRBase (www.mirbase.org). RNA oligonucleotides were synthesized and quantified by Integrated DNA Technologies. DNA primers were synthesized and quantified by Invitrogen. Precursor miRNAs were synthesized by *in vitro* transcription from corresponding PCR products using T7 RNA polymerase (New England Biolabs) according to the manufacturer's protocol. Secondary structure of the Two-tailed RT primers were predicted using the UNAFold web server (<http://unafold.rna.albany.edu/>). Spike-in miRNA sequences were screened *in silico* for homology against human, mouse and rat miRBase records.

3.3 RNA isolation

For RNA isolation from tissue, samples were homogenized using the Tissue-Lyser (Qiagen). Total RNA was extracted with TRI Reagent (Sigma-Aldrich) according to the manufacturer's protocol and treated with TURBO DNA-free kit (Thermo Fisher). RNA quantity and purity was assessed using NanoDrop 2000 spectrophotometer (Thermo Fisher) and RNA integrity was assessed using Fragment Analyzer (Agilent). For total RNA isolation from human plasma and human and rat serum miRNeasy Serum/Plasma Advanced Kit (Qiagen) was used according to the manufacturer's instructions. 1 μl of isolation spike-in mix containing synthetic cel-miR-54 ($1\text{e}+7$ copies/ μl), spike-A ($2\text{e}+5$ copies/ μl), spike-B ($4\text{e}+3$ copies/ μl) and, when appropriate 1 μl of GlycoBlue Coprecipitant (15 mg/mL) (Invitrogen), per sample was added at the lysis step. RNA was eluted into 20 μl nuclease-free water and stored at -80°C .

3.4 RT-qPCR of miRNAs

RT reactions were performed with the qScript flex cDNA kit (Quantabio) in a total reaction volume of 10 μ l. The reaction mixture contained either 10 ng of total RNA or synthetic miRNA template, 1 \times RT buffer, 0.05 μ M RT primer, 1 μ l GSP enhancer and 0.5 μ l RT enzyme. RT reactions were incubated in a 96-well plate in a Bio-Rad CFX 1000 thermocycler for 45 min at 25°C, 5 min at 85°C and then held at 4°C. Reactions using TaqMan miRNA assays (Thermo Fisher) and Quantabio qScript miRNA system (Quantabio) were performed according to the manufacturer's protocol except that the total reaction volume was scaled down to 10 μ l. Reactions using miQPCR method were performed as described in [4]. Quantitative PCR (qPCR) was performed in a total volume of 10 μ l. One reaction contained 1 \times SYBR Grandmaster Mix (Tataa Biocenter), forward and reverse primer (final concentration 0.4 μ M), and 2 μ l of diluted cDNA template. qPCR was performed in a total reaction volume of 10 μ l containing 1 \times SYBR Grandmaster Mix (TATAA Biocenter), 0.4 μ M forward and reverse primer and the cDNA product diluted at least 10 \times . Reactions were performed in duplicates and incubated in a 96- or 384-well plate in a CFX 96 or CFX 384 Real Time Detection System (Bio-Rad) at 95°C for 30 s, followed by 45 cycles of 95°C for 5 s and 60°C for 15 s. Reaction specificity was assessed by melting curve analysis immediately after the qPCR. qPCR with TaqMan miRNA assays and Quantabio qScript miRNA system were performed according to manufacturers' protocols in a total reaction volume of 10 μ l.

3.5 Library preparation and sequencing

Small RNA libraries were prepared in technical duplicates starting from 5 μ l of plasma RNA pool and 5 μ l of miRXplore Universal Reference (2e+6 copies/ μ l) with six commercial kits (from Lexogen, Norgen Biotek, Bioo Scientific, Takara, Qiagen and Somagenics) according to each manufacturer's protocol. Libraries were quantified on the Qubit 3 fluorometer (Thermo Fisher) and Fragment Analyzer (Agilent). Libraries generated by the same kit were pooled and run on 5% TBE-PAGE on Mini-PROTEAN tetra cell (BioRad). A region representing fragments with RNA inserts of length 22 nt \pm ~10 nt (i.e. fragments corresponding to the size of miRNAs) was excised from the gel and purified. All libraries were sequenced in one sequencing run on NextSeq 500 high-output (Illumina) with 85bp single-end reads. Small RNA libraries for tissue profiling were prepared from 100 ng total brain RNA with RealSeq kit (Somagenics) according to manufacturer's protocol. mRNA libraries were prepared from 400 ng total brain RNA with QuantSeq 3' Library Prep Kit FWD (Lexogen) according to manufacturer's protocol. 1 μ l of ERCC spike-in (c = 0.01 \times ; Thermo Fisher) per library was included. This library preparation method generates stranded libraries predominantly covering the 3' end of the transcript, thus producing gene-centric expression values.

3.6 High-throughput RT-qPCR

Samples were reverse transcribed in a reaction volume of 10 μ l containing: 5 μ l template (either 125 ng total tissue RNA or 100 sorted cells after direct lysis), 0.5 μ l spike-in RNA (Tataa Biocenter; c = 0.1 \times for tissues or 0.01 \times for sorted cells), 0.5 μ l equimolar mixture of random hexamers with oligo(dT) (c = 50 μ M), 0.5 μ l dNTPs (c = 10 mM), 2 μ l 5 \times RT buffer, 0.5 μ l RNaseOUT, 0.5 μ l Maxima H-Reverse Transcriptase (all Thermo Fisher) and 0.5 μ l nuclease-free water. After the pre-incubation step at 65°C (t = 5 min), followed by the immediate cooling on ice, the main incubation was performed at 25°C (t = 10 min), 50°C (t = 30 min), 85°C (t = 5 min), after which the samples were immediately cooled on ice. cDNA from tissue samples was diluted 4 \times in nuclease-free water; sorted cell-cDNA was left undiluted. All cDNA samples were pre-amplified immediately after RT in 40 μ l total reaction volume containing 4 μ l cDNA, 20 μ l IQ Supermix buffer (Bio-Rad), 4 μ l primer

mix of 96 assays ($c = 250$ nM each), and 12 μ l of nuclease-free water. Reactions were incubated at 95°C ($t = 3$ min) followed by 18 cycles of 95°C ($t = 20$ s), 57°C ($t = 4$ min) and 72°C ($t = 20$ s). After thermal cycling, reactions were immediately cooled on ice and diluted in nuclease-free water (sorted cells 4x, tissue 50x). High-throughput qPCR was then performed on a 96.96 microfluidic platform BioMark (Fluidigm) as previously described [5]. Cycling program consisted of activation at 95°C ($t = 3$ min), followed by 40 cycles of 95°C ($t = 5$ s), 60°C ($t = 15$ s) and 72°C ($t = 20$ s) and melting curve analysis (. V).

3.7 Bioinformatic and statistical data analysis

3.7.1 Differential gene expression analysis

Differential gene expressions from RNA-Seq data between desired groups of samples were analysed in R project using DESeq2 package.

3.7.2 Gene set enrichment analysis (GSEA)

GSEA was performed for pairwise differential expression comparisons. First, a gene score was calculated for every gene using DESeq2 output as $-\log_{10}(p_{adj})$ and assigned a positive or negative sign based on direction of regulation. Genes were ranked by their gene-scores and GSEA was run in a weighted pre-ranked mode with 1000 permutations. Gene sets were downloaded from http://download.baderlab.org/EM_Genesets/. Significantly overrepresented gene sets were visualized as a network using Enrichment Map. In the network, each node represents gene set and highly overlapping gene sets are connected with edges, resulting in a tight clustering of highly redundant gene sets. For functional annotation of discrete sets of genes, we used Cytoscape plugin ClueGO with all expressed genes (16048 genes) as the background set.

3.7.3 Cell-specific gene sets and cell type proportion estimation

Marker genes for major cell types specifically in the mouse cortex region were taken as an initial reference. In order to acquire marker genes with stable expression regardless of activation states, we have further removed the genes previously found to be differentially expressed under similar conditions in studies on purified cell types. DESeq2-normalized gene expression data and the cell-specific gene lists were used as an input into the marker-GeneProfile R package or the estimation of marker gene profiles (MGPs), which serve as a proxy for relative cell type proportion changes. Differences in expression of final marker gene sets were analyzed by linear mixed model in R project v3.6.0 using lmerTest package. To validate the first estimations, we employed CIBERSORT, a transcriptome deconvolution algorithm that uses gene expression matrix of individual cell types as a reference, and deconvolutes the cellular composition of mixed sample by linear support vector regression. We used published single-cell RNA-Seq dataset of adult mouse cortex as a reference gene expression signature.

3.7.4 Weighted gene co-expression network analysis (WGCNA)

Standard WGCNA procedure was followed to create gene co-expression networks using blockwiseModules function from the WGCNA R package. This analysis groups genes with highly correlated expression pattern across samples into modules.

3.7.5 Motif and transcription factor enrichment analysis

Cytoscape plugin iRegulon with default parameters was used to search for over-represented motifs and their associated transcription factors 500 bp upstream of the transcription start site. All genes from a particular WGCNA module were used as an input.

3.7.6 Protein-protein interaction network

Known interactions between genes in the desired gene sets were downloaded from STRING database v10.5. The resulting interaction network was then visualized and analyzed in Cytoscape. Spectral partition-based network clustering algorithm via Cytoscape ReactomeFI plugin was used for network clustering.

3.7.7 Custom gene set enrichment

Gene sets of interest were collected directly from relevant publications. R package GeneOverlap v1.20 was used to calculate the odds ratio (OR) and the significance of the overlap of the gene sets of interest with the Fisher's exact test.

3.7.8 RT-qPCR data analysis

Raw data were pre-processed with the Real-Time PCR analysis software v4.1.3 (Fluidigm); unspecific values were deleted based on melting-curve analysis. Further processing was done in GenEx v6.0.1 (MultiD Analyses AB): C_q value cutoff was applied; gDNA background was subtracted using ValidPrime (Tataa Biocenter); data were normalized to the mean expression of 5 reference genes (*Actb*, *Gapdh*, *Ppia*, *Ywhaz*, *Tubb5*); outliers were deleted (within group Grubbs test, $p < 0.05$) and a gene was considered undetected for given group if more than 75% values per group were missing; technical replicates (RT and FACS) were averaged; if appropriate, missing data were imputed on a within-group basis and remaining missing data were replaced with $C_{q_{max}} + 2$ for tissue samples or $C_{q_{max}} + 0.5$ for sorted cells.

Temporal expression of individual genes was first analyzed with two-way ANOVA in R project v3.6., then differences between time-points were tested separately for each age by one-way ANOVA and post-hoc t-tests using emmeans package v1.3.5. P-values were adjusted with Benjamini-Hochberg method. Temporal expression of groups of cellular marker genes was first analyzed using linear mixed model in R project v3.6.0 (lmerTest package), then differences between time-points were tested separately for each age. Significance was tested by Satterthwaite's method. Post-hoc pairwise t-tests were performed using emmeans package v1.3.5 and p-values were adjusted by Holm method. Temporal expression of IFN-I pathway was analyzed in similar steps using random slope and random intercept mixed model with 23 interferon-stimulated genes as response variables. IFN-I pathway expression in sorted cells was analyzed in the same way with two-factor design (age, injury) with interaction, using only detected ISGs per each cell type. Differential expression of individual genes (relative to age-matched control) in sorted cells was tested in GenEx v6.0.1 (MultiD Analyses AB) using ANOVA with Bonferroni's post-hoc test for selected pairwise comparisons and p-values were corrected using Benjamini-Hochberg method.

4 Survey of results

4.1 Development of novel RT-qPCR-based method for highly accurate miRNA quantification

Sensitive and specific quantification of miRNAs is challenging. Although several methods for RT-qPCR analysis of miRNAs have been developed, they suffer from one or more drawbacks and are typically available only commercially, and costly. We have therefore aimed to develop novel specific and cost-effective approach to quantify miRNA expression that would improve on the available methods.

We have designed a RT-qPCR system that utilizes specific structured primers for reverse transcription and SYBR-Green-based qPCR, which we named “Two-tailed RT-qPCR”. Briefly, RT primers containing two binding probes (“hemiprobcs”) complementary to the target miRNA joined by sequence forming hairpin structure were designed. We hypothesized that the introduction of 5' binding sequence would improve sensitivity and specificity as larger part of the miRNA sequence will be interrogated compared to previously available methods, which target only 3' region of the miRNA. Importance of the 5' hemiprobe was assessed on synthetic miRNA oligonucleotides with positive results. Next, parameters of the method, including repeatability, sensitivity and dynamic range, ability to discriminate between highly similar miRNAs, between mature and precursor miRNAs, and ability to capture isomiRs were assessed and compared against three other previously available methods, each of which employs different technical approach. These experiments revealed that Two-tailed RT-qPCR matches or outperforms other methods, while simultaneously being less costly. In addition, Two-tailed RT-qPCR was used to profile seven miRNA targets in various mouse tissues, both in singleplexed and multiplexed setting and compared against then-industry-standard TaqMan miRNA assays (Thermo Fisher), which revealed good agreement of relative quantification both between single and multiplexed setting as well as between the two methods.

In summary, we have developed a highly sensitive and exceedingly specific method called Two-tailed RT-qPCR, suitable for rapid and cost-effective miRNA profiling. At the same time, Two-tailed RT-qPCR reflects on the current state of miRNA field and confers several advantages over current RT-qPCR methods, including increased specificity and ability to capture the full isomiR profile.

For more details see:

Androvic, P., Valihrach, L., Elling, J., Sjoback, R., and Kubista, M. (2017). Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. Nucleic Acids Research 45, e144–e144

4.2 Development of quality control tool for circulating miRNA studies

Application of circulating miRNAs as clinical biomarkers is an exciting avenue of miRNA research. In our laboratory, we were interested to study circulating miRNA profiles from serum and plasma samples to identify potential biomarker candidates for acute spinal cord injury and its severity. However, the miRNA profiling workflow from liquid biopsy samples is even more challenging than from typical tissue or cellular samples, and proper quality control becomes even more critical. Because standard quantification and quality control tools are inappropriate for biofluid profiling, we have developed an RT-qPCR-based quality control tool for circulating miRNA studies. It is freely available and allows users to monitor quality of miRNA isolation, degree of inhibition, and erythrocyte contamination to ensure technical soundness of the obtained results.

Briefly, we have designed five synthetic spike-in miRNAs and eight two-tailed RT-qPCR assays targeting these synthetic spike-ins, and three endogenous miRNAs serving as controls for miRNA yield and indicators of hemolysis. We then demonstrated how the protocol can be utilized to optimize input volume of the sample to obtain the best yield and purity using human plasma, human serum and rat serum and identified optimal volumes for isolation of these biofluids with miRNeasy Serum/Plasma Advanced Kit (Qiagen). Next, we assessed the effect of carriers (specifically glycogen) on the isolation procedure and demonstrated that it improves yield and reproducibility of the repeated isolations. Next, we prepared hemolysis dilution series and constructed calibration curve to correlate hemolysis indicator based on ΔCq values (miR-23a – miR-451a) to the absorbance at 414 nm (wavelength indicative of hemolysis). This allowed us to establish reference ΔCq values that can be used to identify level of hemolysis in any new sample as long as the described workflow is used.

Taken together, we have developed Two-tailed RT-qPCR panel for quality control, monitoring of technical performance, and optimization of miRNA profiling experiments from biofluid samples. The detailed experimental protocol including guide to data interpretation, and sequences of the RNA oligonucleotides and RT-qPCR assays is provided to the community and will hopefully contribute to the increased quality and reliability of the results from circulating miRNA studies.

For more details see:

Androvic, P., Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., and Valihrach, L. (2019). Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. Scientific Reports 9, 4255.

Valihrach, L., Androvic, P., Kubista, M. (2019). Circulating miRNA Analysis for Cancer Diagnostics and Therapy. Molecular Aspects of Medicine, <https://doi.org/10.1016/j.mam.2019.10.002>.

4.3 Comprehensive performance comparison of small RNA-Seq library preparation methods from biofluids

During the exploratory phase of miRNA studies, it is desirable to interrogate expression of the whole miRNome in several samples at once. For this purpose, small RNA-Seq is becoming a leading technology. At the same time, library preparation methods are known to suffer from several biases, and new technical solutions including usage of randomized adapters, usage of poly(A)-tailing and template switching or usage of single-adaptor ligation and circularization were recently developed. In the past years, several studies compared different protocols for small RNA-seq. Although they provided valuable insights, they focused on the profiling of tissues and cells and typically covered only some of the available protocols. How various library preparation protocols perform with biofluid samples is not established. Therefore, we carried out comprehensive comparison of the six commercially available protocols for small RNA-Seq library preparation (Lexogen, Norgen Biotek, NextFlex, Qiagen, Takara, Somagenics) with focus on their technical biases and performance in biofluids. To our knowledge, this is the first study that covers all currently available approaches for small RNA library preparation.

Briefly, plasma samples were obtained from healthy volunteers and RNA was isolated using commercial kit. Quality of the isolation procedure was assessed using previously developed Two-tailed QC panel and RNA samples passing QC criteria were used to prepare standardized human plasma sample. Small RNA-Seq libraries were prepared with each kit in duplicates from human plasma as well as from synthetic mixture of 962 miRNAs in equimolar amount (miRXplore), which allowed us to assess technical biases of each method.

Mapping statistics for both plasma and miRXplore samples were examined, which revealed varying performance, particularly low mapping rate to miRNAs for the polyadenylation-based kit from Takara and surprisingly low mapping rate for all kits in human plasma. Next, ligation bias was assessed for each kit on the miRXplore samples as the fold-deviation from the expected number of reads for each miRNA. This analysis revealed that all kits suffer from substantial ligation bias, as percentage of miRNAs that can be considered unbiased ranged from 13.1% for Norgen kit to 50.7% for Takara kit. Notably, single-molecule ligation and circularization approach from Somagenics, that was recently claimed to be bias-free [6] showed only 21% unbiased miRNAs and was outperformed in this metric by three other kits. We also assessed the contribution of ligation bias vs PCR bias through UMIs that are incorporated in the Qiagen kit. We found that ligation is a dominant source of bias (ligation bias explained more than 75% variability for 518 out of 962 miRNAs), although for approximately quarter (227) of miRNAs, PCR is significant contributor to bias too and explains more than half of their variation. We then assessed how the sequencing reads are allocated to the detected spectrum of miRNAs for each kit and how sensitive for miRNA detection is each kit at various sequencing depths. We found that few most abundant miRNAs consumed majority of the reads in all kits and that the kits that detected most miRNAs at all sequencing depths in human plasma were Takara and NextFlex. Arguably, this is due to their lower rates of bias, which causes that reads are progressively allocated to the larger spectrum of miRNAs. To examine which kit quantified abundances in plasma samples most truthfully, we have performed absolute quantification of 19 miRNAs in our plasma samples by Two-tailed RT-qPCR and compared absolute quantities to abundances reported by each kit. Takara RNA-seq expression was closest to the true expression (Pearson $r = 0.94$), followed by Somagenics, Qiagen and NextFlex. Lexogen and Norgen data showed lower correlation (Pearson $r = 0.81$ and 0.73). In order to compare kits at the level of individual miRNAs and identify which miRNAs are most affected by kit-specific technical performance, we clustered kits based on miRNAs with highest differential expression between all kits. Such clustering of miRXplore samples showed that measured expression was similar in kits with similar technical procedures.

In summary, this study provides comprehensive overview of the performance of all currently available technical approaches for library preparation from biofluids and can serve as a guide for selection of optimal kit for each experiment. In addition, it contributes to our understanding of various technical biases, which is important for proper interpretation of the data and potential development of novel wet-lab protocols as well as *in silico* correction algorithms and data analysis pipelines.

For more details see:

Androvic, P., Benesova, S., Kubista, M., Valihrach, L. (2020). Performance comparison of small RNA-Seq library preparation methods for biofluid samples. Manuscript in preparation

4.4 Decoding the transcriptional response to ischemic stroke in young and aged mice

The previously described work covers the laboratory methodology for quantitative analysis of miRNA ranging from pre-analytical aspects to the establishment of methods for low and high-throughput measurements. In addition to this work, we reviewed the current state of circulating miRNA analysis.

In a biological setting, we apply miRNA profiling and complementary technologies (mRNA-Seq, single-cell RNA-Seq, high-throughput RT-qPCR) to study gene expression changes and gene regulatory networks underlying CNS injuries, such as spinal cord injury and ischemic stroke. For

similar biological questions, miRNA profiling alone provides only limited picture, considering that miRNAs have pleiotropic roles, and changes in their mRNA targets have to be inferred indirectly. Depending on the goal of biological experiments, direct mRNA profiling or integrative profiling of mRNA and miRNA expression from the same samples can provide more accurate picture of the gene expression changes. Therefore, we first focused on the mRNA analysis with the perspective of integrated analysis in the next study.

Ischemic stroke is one of the leading causes of mortality and major healthcare and economic burden. It is a well-recognized disease of aging, yet it is unclear how the age-dependent vulnerability occurs and what are the underlying mechanisms. To address these issues, we performed a comprehensive mRNA-Seq analysis of aging, ischemic stroke and their interaction. Briefly, we have modelled ischemic stroke in 3-month old and 18-month old mice and RNA-sequenced the brain of injured (at 3 days after stroke) and age-matched control mice. We have analyzed the gene expression changes with normal aging and after ischemic stroke on the level of genes, gene ontology (GO) terms, pathways and custom gene sets. These analyses revealed activation of glial subpopulations with normal aging and increased inflammatory environment after stroke in aged mice. In order to provide cell-specific context to the observed transcriptional changes, we assessed relative changes of the cell type proportions by computational deconvolution of the RNA-Seq data. This revealed increased infiltration of peripheral leukocytes and greater damage to Parvalbumin-positive interneurons in aged mice after stroke. To capture the full extent of expression trends from systems perspective, we complemented our results with weighted co-expression network analysis (WGCNA). WGCNA recapitulated results of cell type proportion estimates in an unsupervised manner and highlighted amplified activation of module of inflammatory and interferon-stimulated genes (ISGs).

It has been reported that activation of type-I interferon (IFN-I) signaling is detrimental to stroke outcome in young mice. To explore the IFN-I signaling in detail, we mapped our RNA-Seq data to the published IFN-I regulatory network, which revealed age-dependent activation of one of the signaling submodules and suggested that age-dependent amplification of ISG expression occurs through action of canonical regulators including *Stat1* and *Irf9*. To reveal temporal changes of interferon pathway, we analyzed expression of ISGs at several time-points after stroke by microfluidic RT-qPCR in both age groups. This revealed activation of IFN-I pathway around 1 day post-stroke, which prevailed at least until 14 days and was higher in aged mice. Because little was known about cell types that contribute to the IFN-I signaling post-stroke, we FACS-purified endothelial and three glial cell populations and analyzed their expression of ISGs. Our cell-specific analysis revealed that not only microglia, but also oligodendrocytes heavily induce IFN-I signaling following stroke and all cell types converge on higher ISG expression in aged mice, although their individual responses differ. In summary, this study provides detailed insights into transcriptional response to stroke in young and aged mice and may contribute to our understanding of the interplay between stroke pathology and aging.

For more details see:

Androvic, P., Kirdajova, D., Tureckova, J., Zucha, D., Rohlova, E., Abaffy, P., Kriska, J., Anderova, M., Kubista, M., and Valihrach, L. (2019). Decoding the transcriptional response to ischemic stroke in young and aged mouse brain. BioRxiv 769331.

4.5 mRNA-miRNA regulatory networks in CNS injury (unpublished results)

Several studies have implicated *in vitro* and *in vivo* effects of miRNAs after CNS injury [7, 8]. After previously exploring global mRNA changes at single time-point after experimental stroke, we focused on understanding of temporal alterations of mRNA-miRNA regulatory networks following CNS injury. Two models of CNS injuries have been employed: middle cerebral artery occlusion (MCAo) on mice to model ischemic stroke and spinal cord compression lesion on rat, to model spinal cord injury (SCI). Groups of injured and sham operated control animals have been sacrificed at different time points after injury: 3h, 7h, 24h, 3d and 7d. RNA from each sample was used for preparation of mRNA and small-RNA sequencing libraries, which were sequenced on Illumina NextSeq instrument. The detailed analysis is ongoing and here we report the initial results from the SCI model. Samples were clustered using PCA based on mRNA and miRNA expression separately. In both cases, naive controls cluster together with sham samples and injured samples clearly separate from controls indicating widespread changes in expression of both RNA modalities after injury. mRNA expression more clearly separated individual time-points, while miRNA expression separated injured samples into early (3h, 12h, 24h) and later (3d, 7d) time-points, indicating that overall post-transcriptional regulatory response is lacking behind mRNA changes and is lower in magnitude during early phase.

Next, we used WGCNA to cluster genes with similar expression profiles into modules. Several gene modules showing various patterns of differential expression with respect to control were identified. We then searched for significant enrichment of cell-specific markers in the modules and found that several of them are associated with single cell type, and likely reflect temporal changes of the particular cell type or cell-specific expression. We also annotated selected modules functionally by GO and pathway enrichment analysis. This revealed time-dependent changes in wound healing processes upon SCI. Positive regulation of wound healing was associated mainly with brown expression module with continuously increasing expression, while green-yellow module associated with negative regulation of wound healing was downregulated in early time-points and then its expression started to increase after 12h post injury.

Clustering of the top differently expressed miRNAs identified in total 135 miRNAs with distinctive profiles in each time-point. Similarly to mRNA expression, the miRNA expression variance is driven partly by cellular composition. We identified several cell-specific miRNAs corresponding to the distinct profiles including miR-142 (microglia), miR-124 and miR-129-2 (neurons), miR-144 and miR-451 (hematopoietic lineage), miR-221 and miR-27a (endothelial cells) or miR-92b (astrocytes). These initial results point to significant reorganization of both cellular composition and expression of mRNAs and miRNAs after SCI and highlight several interesting avenues for further exploration such as expression changes of wound-healing-associated genes.

As a next step, we aim to perform *in silico* integrative analysis of miRNA-mRNA expression to elucidate roles of miRNAs on their respective targets. However, the miRNA-mRNA correlative relationship may be confounded by the expression variation due to changing cellular composition. We are therefore developing procedures to correct bulk tissue gene expression as well as miRNA expression data for cellular composition before parsing them to integrative analysis algorithms in order to better reveal true targeting relationships. These analyses are subject to ongoing research.

5 Conclusion

This thesis deals with methods and workflows for targeted as well as global quantitative miRNA analysis and applies them together with global measurements of mRNA expression to understand pathophysiological responses after acute CNS injuries.

The most important outcomes of the work are summarized below:

- Novel RT-qPCR method called “Two-tailed RT-qPCR” for accurate measurement of miRNA expression in cells, tissues and biofluids was developed.
- Two-tailed RT-qPCR was utilized to develop quality control protocol allowing monitoring of technical workflow of miRNA expression studies, particularly studies of circulating miRNAs from biofluid samples.
- Methods for small RNA-Seq library preparation were comprehensively evaluated in a benchmarking study with focus on biofluid samples and understanding of technical biases distorting the small RNA-Seq data.
- Current state of the art of technical aspects of circulating miRNA analysis was reviewed.
- Global RNA-Seq analysis provided detailed insights into the impact of stroke, aging and their interaction on genome-wide expression profiles.
- Paired mRNA and miRNA profiling of neural tissue revealed temporal changes to the transcriptome and miRNome after SCI.
- Future detailed integrative analysis will reveal impact of post-transcriptional regulation of miRNAs on the mRNA targets after SCI and stroke.

In summary, this thesis provides novel tools for the field of miRNA analysis and contributes to the detailed understanding of the technical performance of small-RNA-Seq methods including various biases that hamper the accurate quantification. These results may serve for the miRNA research community to obtain reliable results from the quantitative studies and guide the choice of the small RNA-Seq platform. They can be utilized for the development of improved small RNA-Seq protocols and computational methods aimed at correction of small RNA-Seq data. In addition, global analyses of mRNA and miRNA expression reported in this thesis may contribute to our understanding of molecular and cellular mechanisms underlying CNS injuries such as stroke and spinal cord injury and open new avenues for the search for future therapeutic strategies.

6 References

1. Piovesan, A., Antonaros, F., Vitale, L., Strippoli, P., Pelleri, M. C., & Caracausi, M. (2019). Human protein-coding genes and gene feature statistics in 2019. *BMC Research Notes*, *12*(1), 315. <https://doi.org/10.1186/s13104-019-4343-8>
2. Dunham, I., Kundaje, A., Aldred, S. F., Collins, P. J., Davis, C. A., Doyle, F., ... Lochovsky, L. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, *489*(7414), 57–74. <https://doi.org/10.1038/nature11247>
3. Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*, *75*(5), 843–854. [https://doi.org/10.1016/0092-8674\(93\)90529-Y](https://doi.org/10.1016/0092-8674(93)90529-Y)
4. Benes, V., Collier, P., Kordes, C., Stolte, J., Rausch, T., Muckentaler, M. U., ... Castoldi, M. (2015). Identification of cytokine-induced modulation of microRNA expression and secretion as measured by a novel microRNA specific qPCR assay. *Scientific Reports*, *5*(1), 11590. <https://doi.org/10.1038/srep11590>
5. Rusnakova, V., Honsa, P., Dzamba, D., Ståhlberg, A., Kubista, M., & Anderova, M. (2013). Heterogeneity of Astrocytes: From Development to Injury – Single Cell Gene Expression. *PLoS ONE*, *8*(8), e69734. <https://doi.org/10.1371/journal.pone.0069734>
6. Barberán-Soler, S., Vo, J. M., Hogans, R. E., Dallas, A., Johnston, B. H., & Kazakov, S. A. (2018). Decreasing miRNA sequencing bias using a single adapter and circularization approach. *Genome Biology*, *19*(1), 105. <https://doi.org/10.1186/s13059-018-1488-z>
7. Shi, Z., Zhou, H., Lu, L., Li, X., Fu, Z., Liu, J., ... Feng, S. (2017). The roles of microRNAs in spinal cord injury. *International Journal of Neuroscience*, *127*(12), 1104–1115. <https://doi.org/10.1080/00207454.2017.1323208>
8. Saugstad, J. A. (2015). Non-Coding RNAs in Stroke and Neuroprotection. *Frontiers in Neurology*, *6*(MAR), 1–11. <https://doi.org/10.3389/fneur.2015.00050>

7 List of publications

Publications:

- I. **Androvc, P.***, Benesova, S.*, Kubista, M., Valihrach, L. (2020). Performance comparison of small RNA-Seq library preparation methods for biofluid samples. *Manuscript in preparation*. *These authors contributed equally
- II. **Androvc, P.**, Kirdajova, D., Tureckova, J., Zucha, D., Rohlova, E., Abaffy, P., Kriska, J., Anderova, M., Kubista, M., and Valihrach, L. (2019). Decoding the transcriptional response to ischemic stroke in young and aged mouse brain. BioRxiv 769331. *Online preprint. Manuscript is under review in an international peer-reviewed journal*.
- III. Valihrach, L., **Androvc, P.**, Kubista, M. (2019). Circulating miRNA Analysis for Cancer Diagnostics and Therapy. *Molecular Aspects of Medicine*, <https://doi.org/10.1016/j.mam.2019.10.002>.
- IV. Kolenicova, D., Tureckova, J., Pukajova, B., Harantova, L., Kriska, J., Kirdajova, D., Vorisek, I., Kamenicka, M., Valihrach, L., **Androvc, P.**, Kubista, M., Vargova L., and Anderova, M. (2019). High potassium exposure reveals the altered ability of astrocytes to regulate their volume in the aged hippocampus of GFAP/EGFP mice. *Neurobiology of Aging*. 86, 162–181
- V. Zucha, D., **Androvc, P.**, Kubista M., Valihrach, L. (2019). Performance comparison of reverse transcriptases for single-cell studies. *Clinical Chemistry* 66, 217-228
- VI. **Androvc, P.**, Romanyuk, N., Urdzikova-Machova, L., Rohlova, E., Kubista, M., and Valihrach, L. (2019). Two-tailed RT-qPCR panel for quality control of circulating microRNA studies. *Scientific Reports* 9, 4255.
- VII. Smieszek, A., Kornicka, K., Szłapka-Kosarzewska, J., **Androvc, P.**, Valihrach, L., Langerova, L., Rohlova, E., Kubista, M., and Marycz, K. (2019). Metformin Increases Proliferative Activity and Viability of Multipotent Stromal Stem Cells Isolated from Adipose Tissue Derived from Horses with Equine Metabolic Syndrome. *Cells* 8, 80.
- VIII. Valihrach, L., **Androvc, P.**, and Kubista, M. (2018). Platforms for Single-Cell Collection and Analysis. *International Journal of Molecular Sciences* 19, 807.
- IX. Valny, M., Honsa, P., Waloschkova, E., Matuskova, H., Kriska, J., Kirdajova, D., **Androvc, P.**, Valihrach, L., Kubista, M., and Anderova, M. (2018). A single-cell analysis reveals multiple roles of oligodendroglial lineage cells during post-ischemic regeneration. *Glia* 66, 1068–1081.
- X. **Androvc, P.**, Valihrach, L., Elling, J., Sjoback, R., and Kubista, M. (2017). Two-tailed RT-qPCR: a novel method for highly accurate miRNA quantification. *Nucleic Acids Research* 45, e144–e144
- XI. Honsa, P., Valny, M., Kriska, J., Matuskova, H., Harantova, L., Kirdajova, D., Valihrach, L., **Androvc, P.**, Kubista, M., and Anderova, M. (2016). Generation of reactive astrocytes from NG2 cells is regulated by sonic hedgehog. *Glia* 64, 1518–1531.

Conference presentations:

- I. Getting more out of bulk RNA-Seq: transcriptomic analysis of ischemic stroke in young and aged mice (poster) - *Single Cell Genomics, 2019, Djuronaset-Stockholm, Sweden*
- II. Integrated mRNA and microRNA transcriptional analysis reveals altered regulatory networks following spinal cord injury (poster) - *XIV European Meeting on Glial Cells in Health and Disease, 2019, Porto, Portugal*
- III. Getting more out of RNA-Seq data: transcriptomic analysis of ischemic stroke (invited talk) - *9th International Gene Quantification Event, 2019, Freising, Germany*
- IV. Diverse transcriptomic response after ischemic stroke in young and aged mice (poster) - *RNA Society Meeting, 2018, Berkeley, USA*
- V. Two-tailed RT-qPCR: a Novel Method for Highly Accurate miRNA Quantification (poster) - *Non-coding RNAs in Nervous System Development, Plasticity and Disease, 2017, Marburg, Germany*
- VI. Two-tailed RT-qPCR: a Novel Method for Highly Accurate miRNA Quantification (poster) - *8th International Gene Quantification Event, 2017, Freising, Germany*
- VII. New RT-qPCR system for miRNA analysis (selected for talk) - *RNA Club, 2016, Brno, Czech Republic*
- VIII. Highly Specific and Sensitive Method for miRNA Analysis (poster) - *International Symposium on Microgenomics, 2016, Paris, France*
- IX. Generation of reactive astrocytes from NG2 cells is regulated by sonic hedgehog (poster) - *Single Molecule & Single Cell Analysis, 2016, Madrid, Spain*

8 Súhrn (Summary, in Slovak)

MikroRNA (miRNA) sú krátke nekódujúce molekuly RNA, ktoré plnia dôležitú úlohu ako negatívne regulátory génovej expzie na posttranskripčnej úrovni. Deregulácia miRNA je implikovaná v mnohých fyziologických ako aj patologických procesoch a preto sú tieto molekuly stredobodom záujmu ako potenciálne terapeutiká. Navyše, objav extracelulárnych, cirkulujúcich miRNA v telesných tekutinách znamenal, že miRNA by mohli slúžiť ako nové, neinvazívne klinické biomarkery. Všetky tieto aplikácie si vyžadujú metódy pre presnú a citlivú kvantifikáciu hladín miRNA v rôznych vzorkách. Avšak spoľahlivá kvantifikácia miRNA je pomerne náročná, vzhľadom na ich vlastnosti ako je krátka dĺžka, ako aj kvôli ďalším technickým faktorom ako je variabilita izolácie a pod.

Táto práca sa zameriava na vývin nových metód a nástrojov pre vylepšenie kvantitatívnej analýzy miRNA. Cieľom práce bolo vyvinúť novú špecifickú a cenovo prijateľnú metódu pre kvantifikáciu miRNA, založenú na technológii RT-qPCR. Ďalej, vyvinúť protokol, ktorý by umožnil jednoduchú optimalizáciu postupu kvantitatívnej miRNA analýzy a tiež rutinné testovanie kvality vzoriek, najmä problematických vzoriek ako sú telesné tekutiny. Ďalším cieľom bolo komplexné porovnanie komerčne dostupných metód pre miRNA sekvenovanie, ktoré by umožnilo nájsť najvhodnejšiu metódu pre sekvenáciu z telesných tekutín a tiež lepšie pochopiť jednotlivé technické nedostatky, ktoré vplyvajú na kvalitu dát. Posledným cieľom bolo aplikovať takto získané poznatky a techniky, spolu s príbuznými technikami pre globálne profilovanie génovej expzie, v zvieracích modeloch poškodenia centrálného nervového systému (CNS) za účelom lepšieho porozumenia molekulárnych a bunkových zmien v poškodenom tkanive.

Najskôr sme navrhli novú RT-qPCR metódu pre kvantifikáciu miRNA zvanú „Two-tailed RT-qPCR“. Je založená na špecifických primeroch pre reverznú transkripciu. Tieto sú zložené z dvoch hybridizačných prôb („hemiprôb“), ktoré sú navzájom spojené DNA sekvenciou tvoriacou sekundárnu vlásenkovú štruktúru. Predpokladali sme, že pridanie druhej väzobnej hemipróby na 5' koniec miRNA zvýši špecificitu a senzitivitu oproti dostupným technológiám. Tento predpoklad sme otestovali s pozitívnym výsledkom na syntetických miRNA templátoch. Ďalej sme novú metódu po jej optimalizácii porovnali s tromi dostupnými RT-qPCR metódami. Každá z nich je založená na inom princípe. Vyhodnotenie parametrov ako senzitivita, špecificita, reproducibilita, dynamický rozsah či schopnosť zachytiť spektrum miRNA variánt ukázalo, že naša metóda funguje rovnako dobre, alebo lepšie ako predtým dostupné techniky a zároveň je cenovo výhodnejšia. Ďalej sme našu metódu použili na profilovanie niekoľkých miRNA v reálnych biologických vzorkách a porovnali výsledky s komerčnou technológiou, čo ukázalo dobrú zhodu tak v jednoduchom ako aj v „multiplex“ formáte.

Po tom čo sme vyvinuli nový RT-qPCR systém sme sa zamerali na vývin protokolu umožňujúceho monitorovanie kvality vzoriek a laboratórneho postupu miRNA štúdií. Konkrétnym zameraním bola kontrola kvality pre profilovanie telových tekutín ako plazma a sérum, ktoré sú obzvlášť problematické. Navrhli sme päť syntetických miRNA spike-in oligonukleotidov a osem Two-tailed RT-qPCR esejí, ktoré celiť tieto syntetické miRNA ako aj tri endogénne miRNA ciele. Celkovo exogénne spike-in spolu s endogénnymi cieľmi umožňujú kontrolovať výťažok miRNA izolácie, jej variabilitu, prítomnosť inhibítorov enzymatických reakcií, variabilitu RT-qPCR a tiež úroveň hemolýzy vo vzorkách. Pomocou takto vytvoreného panelu sme demonštrovali optimalizáciu

izolačného protokolu. Ďalej sme identifikovali najlepší vstupný objem ľudského a potkanieho séra a ľudskej plazmy do RNA izolácie pomocou komerčného kitu a demonštrovali sme užitočnosť prídania polyakrylamidu ako nosiča minimalizujúceho variabilitu izolácie. Panel pre kvalitu kontroly, ktorý sme vytvorili je spolu s detailným protokolom a návodom na interpretáciu dát voľne dostupný pre vedeckú komunitu.

Počas exploratívnej fázy miRNA štúdií je často žiadúce získať globálny profil všetkých miRNA prítomných vo vzorke. Za týmto účelom je čoraz viac využívanou metódou sekvenovanie malých RNA. Protokoly pre prípravu knižníc z malých RNA sú však známe tým, že sú zaťažené rôznymi technickými nedostatkami. V našej štúdii sme sa zamerali na porovnanie technického výkonu všetkých v súčasnosti dostupných komerčných kitov pre sekvenovanie malých RNA, konkrétne s cieľom otestovať ich výkon so vzorkami z telesných tekutín ako krvná plazma. Pripravili sme knižnice zo štandardizovanej vzorky ľudskej plazmy ako aj zo syntetického ekvimolárneho mixtu 962 miRNA, čo nám umožnilo vyhodnotiť jednak výkon kitov s reálnymi vzorkami plazmy a jednak spoľahlivo vyhodnotiť mieru jednotlivých technických chýb vznikajúcich počas procesu prípravy knižníc. Pre každý kit sme vyhodnotili rôzne parametre ako senzitivita, ligačná a amplifikačná chyba, presnosť atď. Naše výsledky poskytujú komplexný a detailný pohľad na technický výkon jednotlivých kitov a môžu slúžiť ako návod pre výber správnej metódy, pre lepšie pochopenie chýb a toho ako vplývajú na dáta, ako aj na vývin nových vylepšených protokolov.

Predošlá práca pokrýva technickú metodiku kvantitatívnej analýzy miRNA. Tieto techniky sme ďalej využili spolu s globálnym profilovaním génovej expície pomocou RNA sekvenovania na štúdium zmien expície po poškodení CNS, konkrétne po ichémii mozgu a po poranení miechy. Na myšách a potkaních modeloch sme indukovali jednotlivé poškodenia a výsledné sekvenáčne dáta sme analyzovali škálou komputačných techník ako diferenciálna expícia na úrovni génov, termínov génovej ontológie, signálnych dráh a vybraných génových setov a tiež dopĺňujúcimi technikami ako identifikácia ko-exprimovaných génových modulov a dekonvolúcia bunkových populácií. Naše analýzy poukázali na signifikantnú reorganizáciu expície mRNA ako aj miRNA a tiež na zmeny zastúpenia jednotlivých bunkových populácií. Identifikovali sme konkrétne signálne dráhy, ktoré môžu hrať rolu v progresii poranenia (napríklad interferón typu jeden) a poukázali na potenciálne regulátory. Nasledovať bude detailnejšia integrovaná analýza mRNA a miRNA expície vo viacerých časoch po poranení.

Záverom, táto práca prináša nové metódy a nástroje pre analýzu expície miRNA a prispieva k detailnému porozumeniu technických parametrov metód pre sekvenovanie miRNA. Tieto výsledky môžu slúžiť vedeckej komunitě na získanie spoľahlivých dát a slúžiť ako návod pre informovanú voľbu metódy pre miRNA sekvenovanie, či pre vývin nových metód. Okrem toho, výsledky globálneho profilovania mRNA a miRNA expície po poranení CNS môžu prispieť k lepšiemu porozumeniu mechanizmov poranenia nervového tkaniva.