TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Tierphysiologie und Immunologie

Assessment of extracellular vesicles, associated microRNAs, and of their utility as clinical biomarkers in liquid biopsies

Dominik Sebastian Buschmann

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Prof. Dr. Ulrich Kulozik
Prüfende der Dissertation:	1. apl. Prof. Dr. Michael W. Pfaffl
	2. apl. Prof. Dr. Gustav Schelling
	3. Prof. Dr. Martin Klingenspor

Die Dissertation wurde am 22.01.2019 bei der Technischen Universität München eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 30.05.2019 angenommen.

Table of contents

List of abbreviations

AGO2	Argonaute 2
ARE	AU-rich element
BNP	Brain natriuretic peptide
bp	Base pair
CABG	Coronary artery bypass graft
CAR-T	Chimeric antigen receptor T cell
cfDNA	Cell-free DNA
CSF	Cerebrospinal fluid
CTC	Circulating tumor cell
ctDNA	Circulating tumor DNA
CVD	Cardiovascular disease
DGE	Differential gene expression
DNA	Deoxyribonucleic acid
dPCR	Digital polymerase chain reaction
e.g.	Exempli gratia
ECM	Extracellular matrix
EGFR	Epidermal growth factor receptor
EpCAM	Epithelial cell adhesion molecule
ESCRT	Endosomal sorting complex required for transport
EV	Extracellular vesicle
EV-TRACK	Transparent Reporting and Centralizing Knowledge in Extracellular Vesicle Research
FDA	Food and Drug Administration
GR	Glucocorticoid receptor
HSP	Heat shock protein
ICU	Intensive care unit
ILV	Intraluminal vesicle
ISEV	International Society for Extracellular Vesicles
isomiR	miRNA isoform
lncRNA	Long non-coding RNA
MHC	Major histocompatibility complex

MIQE	Minimum Information for Publication of Quantitative Real-Time PCR Experiments
miRNA	microRNA
MISEV	Minimal Information for the Studies of EVs
mRNA	Messenger RNA
MV	Microvesicle
MVB	Multivesicular body
MVE	Multivesicular endosome
NGS	Next-Generation Sequencing
NR3C1	Nuclear receptor subfamily 3 group C member 1
nt	Nucleotide
NTA	Nanoparticle Tracking Analysis
P-body	Processing body
pre-miRNA	Precursor miRNA
pri-miRNA	Primary miRNA
QC	Quality control
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNP	Ribonucleoprotein complex
RT-qPCR	Reverse transcription quantitative polymerase chain reaction
SEC	Size-exclusion chromatography
SIRS	Systemic inflammatory response syndrome
Small RNA-Seq	Next-Generation small RNA Sequencing
SNARE	Soluble N-ethylmaleimide-sensitive-factor attachment receptor
SNP	Single-nucleotide polymorphism
SOP	Standard operating procedure
TEM	Transmission electron microscopy
TLR	Toll-like receptor
TNBC	Triple-Negative Breast Cancer
tRNA	Transfer RNA
UC	Differential ultracentrifugation
UTR	Untranslated region
WB	Western blot

Abstract

Extracellular vesicles (EVs) are secreted by all cells and play crucial roles in long-range and short-range intercellular communication. In addition to their involvement in physiological contexts, EVs are increasingly recognized for their role in disease pathogenesis, which makes them promising diagnostic targets in minimally invasive liquid biopsies. Amongst the biomolecules associated with EVs, non-coding RNAs such as microRNAs (miRNAs) are particularly attractive analytes because they can be easily isolated and amplified and might reflect pathological alterations in EV-secreting cells. The aim of this thesis was to assess the applicability and utility of exploiting EVs and their miRNA cargo as molecular biomarkers.

First, the technical basis for standardized and reproducible miRNA analyses via Next-Generation Sequencing (NGS) was established. The procedures involved in NGS sample preparation and sequencing itself are complex and prone to bias, which makes stringently optimized and standardized workflows crucial for experimental success. Next, several methods of EV isolation were compared in a cohort of sepsis patients and healthy volunteers in order to identify the most suitable approach for downstream experiments. Due to its high RNA yield, robust performance in sequencing experiments and flawless classification of patients and volunteers, EV isolation by precipitation proved to be particularly suited for our purposes.

In the next step, disease-associated alterations in vesicular miRNA profiles were assessed in several patient cohorts. Distinct miRNA dysregulation was detected in circulating EVs in a cohort of critically ill sepsis patients. These alterations were different from those in blood cells and total sera and carried specific diagnostic information. Interestingly, expression levels of several vesicular miRNAs correlated with patient survival, which points towards their potential applicability as biomarkers. Additionally, the prognostic utility of EV miRNAs was assessed in a cohort of cardiovascular disease patients scheduled for bypass surgery. Preoperative concentrations of several miRNAs significantly correlated with clinical variables that indicate cardiac instability during surgery, suggesting a potential advantage for the identification of high-risk patients.

Furthermore, initiatives to increase transparency and reproducibility of EV experiments, which are cornerstones of translating basic EV science into clinical applications, were established. While the developed guidelines provide directives and minimal requirements for EV experiments, the EV-TRACK platform evaluates the completeness of reporting crucial experimental parameters and offers a comprehensive knowledgebase on current EV studies.

Finally, experimental, logistical, and regulatory challenges for the utilization of EVs as clinical markers were assessed. Despite several anticipated obstacles, careful optimization of methods and technologies and standardization of analytical and pre-analytical steps will help to increase the reliability and reproducibility of experiments across all areas of EV science and thus facilitate the development of EV-based biomarkers.

Zusammenfassung

Extrazelluläre Vesikel (EVs) werden von lebenden Zellen in den extrazellulären Raum sekretiert und spielen eine elementare Rolle in der Kommunikation zwischen Zellen und Geweben. Neben ihrer eigentlichen physiologischen Bedeutung können die Konzentration und Zusammensetzung von EVs bei der Entstehung von Krankheiten verändert sein, was sie zu attraktiven Kandidaten für den Einsatz als Biomarker in minimal invasiven Flüssigbiopsien macht. Nicht-kodierende RNAs wie microRNAs (miRNAs) sind dabei von besonderer Bedeutung, da sie eine potenzielle Reflexion krankhafter zellulärer Veränderungen darstellen. Zielsetzung dieser Arbeit war es, die Anwendbarkeit und Nützlichkeit von EVs und den darin enthaltenen miRNAs als molekulare Biomarker abzuschätzen.

Zunächst wurden die technischen Voraussetzungen für eine standardisierte und reproduzierbare miRNA-Analyse per Next-Generation Sequencing (NGS) geschaffen. Die für die Probenvorbereitung und das NGS selbst benötigten Arbeitsabläufe sind komplex und können Artefakte produzieren, daher sind optimierte und standardisierte Verfahren essentiell für den experimentellen Erfolg. Im nächsten Schritt wurden in einer Stichprobe von Sepsis-Patienten und Gesunden verschiedene Methoden der EV-Isolation verglichen, um den bestmöglichen Ansatz für nachfolgende Experimente zu identifizieren. Aufgrund hoher RNA-Ausbeute, guter NGS-Performance und fehlerfreier Klassifikation von Patienten und Kontrollen erwies sich die EV-Isolation per Präzipitation als besonders geeignet für unsere Zwecke.

Anschließend wurden krankheitsbedingte Veränderungen im EV-miRNA-Profil verschiedener Patientenstichproben untersucht. In einer Kohorte kritisch kranker Sepsis-Patienten ließen sich spezifische Veränderungen im miRNA-Profil zirkulierender EVs nachweisen, die sich von denen in Blutzellen und Serum unterschieden. Dabei war von besonderem Interesse, dass die Expressionslevel bestimmter miRNAs mit der Überlebensrate der Patienten korrelierten, was auf eine potentielle Nutzbarkeit als Biomarker hinweist. Ferner wurde das prognostische Potenzial von EVs in einer Kohorte von Patienten mit koronarer Herzerkrankung, die sich einer Bypass-Operation unterziehen mussten, untersucht. Hier ließen sich die präoperativen Konzentrationen bestimmter miRNAs mit klinischen Variablen korrelieren, die auf eine Instabilität der Herzfunktion im Verlauf der Operation hindeuten.

Des Weiteren konnten Maßnahmen etabliert werden, um die Transparenz und Reproduzierbarkeit von EV-Experimenten zu verbessern. Abschließend wurden experimentelle, logistische und regulatorische Herausforderungen für die Translation akademischer EV-Forschung in klinisch nutzbare Biomarker evaluiert. Obwohl diverse Schwierigkeiten zu antizipieren sind, lassen sich Belastbarkeit und Reproduzierbarkeit der EV-Forschung durch Optimierung verwendeter Methoden und Technologien sowie Standardisierung aller analytischen und präanalytischen Verfahren erhöhen, was die Entwicklung EV-basierter klinischer Marker begünstigt.

1 Introduction

1.1 microRNAs: Powerful regulators of gene expression and cellular signaling

microRNAs (miRNAs) are short, non-coding RNAs that generally range from 18 to 22 nucleotides (nt) in length. After they were first discovered in 1993, early interest in miRNAs was limited, and it took them until 2001 to be formally recognized as an independent class of RNA [1]. In stark contrast, however, miRNA research has accelerated rapidly after unveiling their role in post-transcriptional regulation of gene expression via RNA interference and might now be at an all-time high.

In the nucleus of a mammalian cell, miRNAs are transcribed from genomic loci as primary miRNAs (pri-miRNAs) with several hundred nt in length by RNA polymerase II and III [2, 3]. After cropping the transcript to a stem-loop structure of about 70 nt by the type III ribonuclease Drosha, the resulting precursor miRNA (pre-miRNA) is exported into the cytosol by the energy-dependent Exportin-5 protein (Figure 1). The pre-miRNA's stem joins the complementary 3'- and 5'-arms, which will, upon further processing, generate the -3p and -5p variant of a particular miRNA, respectively. The next step in miRNA biogenesis includes trimming of the pre-miRNA's loop by Dicer, resulting in a duplex of about 22 nt in length. After being incorporated into the multi-component RNA-induced silencing complex (RISC), the duplex is unwound, and the guide strand is retained as a mature miRNA, while the complementary passenger strand is degraded. It was initially assumed that guide strand selection strand relies entirely on the thermodynamic stability of hydrogen bonds within the duplex, leading to the selection of the strand with lower stability at the 5'-end. Arm selection preference was therefore postulated to be an inherent feature of each pre-miRNA duplex, which should result in the dominant expression of either the -3p or -5p miRNA variant. We now know, however, that miRNA arm selection is a dynamic process, and that preferences for one of either arms change not only between species, but also between various tissues and even different pathophysiological tissue states [4]. The expression of a particular miRNA does therefore not solely depend on the expression of its precursor and the resulting duplex's stability but is biologically regulated according to a cell's demands at the time [5]. Consequently, the initial concept of a dominant arm has been adjusted, and it is widely recognized that both the -3p and -5p arm are expressed at comparable levels for many miRNAs [6].

In addition to this canonical pathway, several other routes of miRNA biogenesis have been discovered. For instance, some pre-miRNAs are cleaved by Argonaute 2 (AGO2) in a Dicer-independent manner before further processing by exonucleases [7]. The relative contributions

of different biogenesis pathways to the final pool of mature miRNAs are currently unknown and a subject of ongoing investigation.

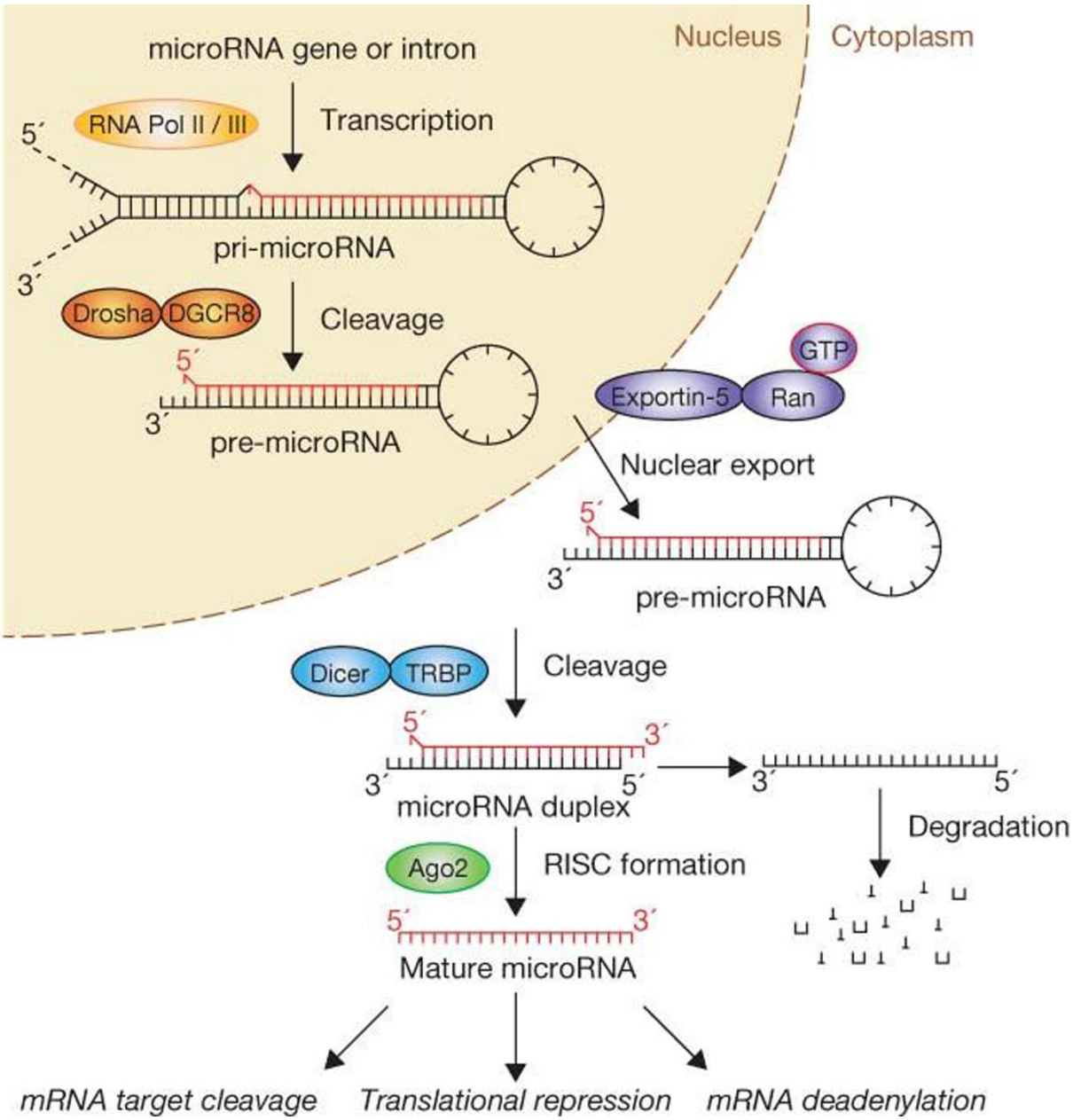

Figure 1. The canonical pathway of mammalian miRNA biogenesis. Primary miRNA transcripts are generated and processed in the nucleus before energy-dependent export of precursor miRNAs into the cytoplasm. Subsequent steps include production of miRNA duplexes by cleavage of the precursor's hairpin, loading into effector protein complexes, degradation of passenger strands and post-transcriptional repression of target mRNAs by mature miRNAs. RNA Pol: RNA polymerase; pri-microRNA: primary miRNA; pre-microRNA: precursor miRNA; AGO2: Argonaute 2; RISC: RNA-induced silencing complex; mRNA: messenger RNA. Figure reprinted from Winter et al. [8].

In animals, the canonical mode of action for mature miRNAs, regardless of their prior biogenesis, relies on binding to messenger RNA (mRNA) by Watson-Crick base pairing and

inhibiting their translation into proteins. Guided by RISC, mature miRNAs target complementary sequences in the 3'-untranslated region (UTR) or coding regions of mRNAs and negatively modulate their expression [9]. This process requires the concerted actions of several proteins, which are recruited to act as a multicomponent ribonucleoprotein complex (RNP). Crucial RNP components serve individual and coordinated functions such as binding the 5'-end of miRNAs (Argonaute proteins) and recruiting downstream effectors (GW182).

Efficiency and specificity of post-transcriptional repression rely heavily on a miRNA's seed region, which stretches from nucleotide two to nucleotide seven. While the 5'-end of miRNAs is anchored in a deep binding pocket of AGO2, the seed sequence is displayed on the protein's surface in a semi-helical conformation, thereby making it accessible for complementary mRNA sequences [10]. Upon attachment of miRNA RNPs, translation of target mRNAs is commonly repressed at the initiation step by impaired recognition of the 5'-methylguanosine cap, interference with translation initiation factors and inhibition of ribosome assembly [11-13]. Additionally, translation can also be repressed at post-initiation steps by mechanisms such as premature termination of protein synthesis or recruitment of proteolytic enzymes that degrade the nascent polypeptide chain [14].

Perfect complementarity between the seed sequence and its target usually leads to the enzymatic degradation of mammalian mRNAs. The mRNA decay machinery, recruited by miRNA RNPs, sequentially removes the 3'-Poly(A) tail and the 5'-cap, followed by exonucleolytic mRNA degradation [15]. Deadenylation and decapping were shown to be co-translational events, indicating that repression of translation might precede mRNA decay [16]. While miRNA-mediated repression seems to be initiated in the cytosol, repressed mRNAs accumulate in distinct, microscopically visible foci called processing bodies (P-bodies). As P-bodies were shown to be enriched in proteins of the decay machinery, it is likely that they are also the location of ultimate mRNA degradation [17]. Although a perfect seed match is the most effective way of mRNA repression, partial sequence matches between miRNA and mRNA can still stifle translation by blocking the assembly of essential translation machinery components and promoting premature termination of translation due to ribosome drop-off [18-20]. This canonical way of mRNA targeting has been expanded by more recent findings that describe an alternate mode of target recognition. Some protein-coding transcripts can bulge out a nucleotide, usually a guanosine, which allows base pairing with the entire seed sequence despite imperfect complementarity of primary miRNA and mRNA sequences. This "G-bulge" mechanism might account for a significant proportion of miRNA-mediated repression events and expands the repertoire of target sites for a given miRNA [21].

In addition to mediating canonical repression of protein synthesis, some miRNAs were found to also increase translation of target mRNAs. In 2007, Vasudevan et al. reported enhanced translation of specific mRNAs that was contingent on base pairing with miRNAs and recruitment of AGO2 [22]. The initially counterintuitive ability of miRNAs to stimulate protein synthesis was described to depend on physiological context and to oscillate during the cell cycle. Whereas the conventional downregulation of translation prevailed in proliferating cells, cell cycle arrest-inducing stress conditions such as nutrient deprivation increased translation of specific mRNA transcripts via miRNAs. Similarly, miR-10a, which regulates global protein synthesis by modifying the production of ribosomal proteins, was found to activate translation of specific targets usually repressed in starvation conditions [23]. In contrast to conventional 3'-UTR binding, miR-10a binds a particular sequence motif (5 TOP sequence) in the 5'-UTR of its targets, many of which are mRNAs for ribosomal proteins.

Additionally, AU-rich elements (ARE), decay sequences that target mRNAs for rapid cytosolic degradation, were found in the 5'-UTR of many protein-coding transcripts, particularly those of cytokines and proto-oncogenic transcription initiators. By competing with ARE-binding decay factors, some miRNAs can stabilize mRNAs and thus indirectly foster their translation [24]. In contrast to a direct activation of translation, this mode of action is more regulatory in nature, as it mediates relief of repression rather than activation itself [25]. Similar mechanisms were reported for miRNAs that decoy repressive proteins in a seed sequence-independent manner, which also leads to mRNA stabilization [26]. Importantly, the actions of a particular miRNA are not confined to be exclusively inhibitory or stimulatory: depending on cellular context, a single miRNA transcript represses some mRNAs while increasing the translation of others.

Regardless of the specific molecular interaction, transcriptional regulation by miRNAs is a highly sophisticated mechanism that appears to be essential for fundamental physiological processes such as cellular development and differentiation, metabolism and apoptosis [27, 28]. Dysregulation of miRNA expression and activity has been associated with a wide range of diseases including most types of cancer [29, 30]. While an individual miRNA putatively targets up to several hundreds of different mRNAs, a single mRNA is also regulated by many different miRNAs, leading to a complex network of regulatory interactions [31, 32]. Indeed, it was demonstrated that many mammalian mRNAs are highly conserved miRNA targets and that the majority of human protein-coding genes is targeted by miRNAs [33].

While canonical upregulation and downregulation of transcription are predicated on a miRNA's interaction with other nucleic acids, an additional mode of action, resembling hormonal signaling, was proposed in a 2018 publication by Fabbri [34]. Partly fueled by the finding that miRNAs can be detected in biofluids, where they mediate intercellular communication, it

suggested endocrine and paracrine activities of miRNAs. The existence of dedicated protein receptors, essential for hormone-like modes of action, has been documented for several miRNA transcripts. miR-21 and miR-29a were found to bind to and activate members of the Toll-like receptor (TLR) family in human and murine immune cells, leading to downstream signaling through the NF-κB pathway and increased secretion of proinflammatory cytokines [35, 36]. Similar findings were reported for miRNA let-7b, which fosters neurodegeneration by binding to TLR7 [37]. As the hormone-like action of miRNAs is a very novel field of investigation, many open questions remain to be elucidated, particularly which features of a miRNA determine receptor binding, the proportions of miRNA allocated to acting on mRNAs or protein receptors, respectively, and which additional non-coding RNA receptors are yet to be detected.

In addition to the various ways in which a miRNA can exert its biological function, sequence variants of miRNAs add another layer of complexity to the network of transcriptomic regulation. The advent of Next-Generation Sequencing (NGS) has increased both the throughput and granularity of transcriptomic analyses, allowing the inspection of transcripts at single-nucleotide resolution. Even though miRNA variants had been detected in a multitude of NGS data sets, they were initially thought to be sequencing artefacts. Due to the use of exogenous spike-in controls, we now know that detection rates of varied miRNA sequences far surpass the frequency of sequencing errors, and have come to appreciate the existence of genuine miRNA variants, so-called isomiRs [38]. Just as canonical miRNAs, isomiRs are loaded into RISC and associate with mRNAs in polysomes, indicating their significance in the regulation of the translation machinery.

isomiRs can display variations at the ends of a transcript or within internal nucleotides and oftentimes stem from imprecise cleavage by Drosha, Dicer and other proteins involved in miRNA biogenesis [39]. Although these enzymes have preferred cleavage sites in precursor sequences, positional shifts of exonucleolytic activity, which lead to templated miRNA variants, are commonly observed. Even canonical miRNAs, cut at dominant cleavage sites, can subsequently be modified by nibbling of exonucleases or addition of nucleotides by nucleotidyl transferases, resulting in shorter or longer non-templated isoforms, respectively. Finally, RNA editing and single-nucleotide polymorphisms (SNPs) also contribute to the multitude of isomiRs, albeit to a lesser extent [40, 41]. For a number of reasons, variations at the 3’-end are much more common than those at the 5’-end, which is buried in Argonaute’s binding pocket and thus less accessible to modifying enzymes [42]. Even though much remains to be learned about biogenesis and functions of isomiRs, we now know that many, if not most, miRNA loci generate several isoforms, and that isomiR profiles are dynamically regulated depending on the respective tissue and developmental stage. In fact, the canonical miRNA sequence

annotated in dedicated databases might not be the dominantly expressed variant in many instances [39].

The vast majority of isomiRs displays only minor modifications compared to the canonical miRNA, indicating a role in fine-tuning of gene expression rather than inducing major shifts. It has been demonstrated that modifying miRNA sequences impacts their stability (reducing or increasing their half-life), loading into Argonaute proteins (changing strand selection and preferential loading of isoforms into different Ago family members) and target repertoire. The latter is an obvious consequence of variation at the 5'-end: modifying the crucial first few nucleotides of a miRNA has a seed-shifting effect that directly changes the mRNA target repertoire. 3'-isomiRs, which are observed much more frequently, have a subtler effect on target selection, indicating that they drive similar biology as their canonical miRNAs with a common seed sequence [43]. An elegant model of isomiR function proposes the cooperative targeting of core biological processes by several isoforms, increasing the repressive pressure on target mRNAs while distributing the effects on off-target transcripts more broadly (Figure 2) [7]. By generating different isomiRs on an as-needed basis, gene expression can be precisely modified with increased specificity compared to regulation by canonical miRNAs alone.

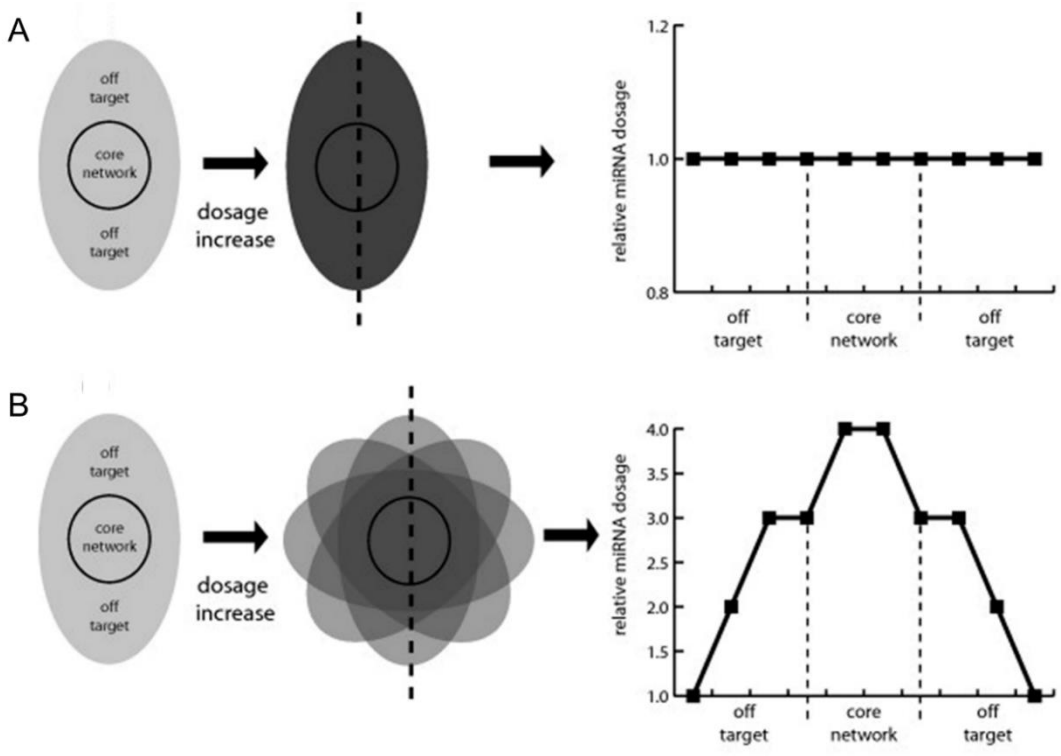

Figure 2. Cooperative repression of core targets by miRNAs and isomiRs. A: Increasing levels of a single miRNA affect core biological functions and off-targets alike. B: Cooperative action of a miRNA and its isomiRs with slightly different target repertoires increases repression of core biological functions while distributing off-target effects more broadly. Figure adapted from Cloonan et al. [7].

1.2 Liquid biopsy: A breakthrough method in molecular diagnostics?

In disease management, the ability to adequately treat patients depends on precise and timely diagnosis. Accurately identifying and characterizing disease states is crucial, as the majority of clinical decisions is based on analytical tests. Sampling cellular material in tissue biopsies is the gold standard for accessing disease-related information in many pathologies, particularly cancer. While undoubtedly useful, biopsies harbor several disadvantages. In addition to generally being invasive, costly and not free from risk, tissue sampling is not feasible for some maladies such as lung cancer [44]. Even for successful biopsies, subsequent analyses are cumbersome, and specimens must be evaluated by skilled pathologists.

Using material sampled from biofluids, most commonly serum or plasma, as a surrogate marker for tissue-derived information has been a highly anticipated development in modern medicine. These so-called liquid biopsies are minimally invasive, affordable and allow serial sampling, which is a crucial prerequisite for assessing treatment response. As most tissues throughout the body secrete biomolecular material into the circulation, sampling blood also allows remote surveillance of tissues inaccessible to needle biopsies. In oncology, a field historically struggling with the heterogeneity of tumors, liquid biopsies may provide significant diagnostic value, as they capture material from the entire tumor and potential metastases, as opposed to the specific section obtained from a tissue biopsy [45]. Potentially even lending themselves to preventive screening programs, liquid biopsies might help reach the holy grail of oncology: detecting and characterizing tumors in the early stages of development, when they are easier to treat and more susceptible to clinical interventions.

One of the first sample types to be utilized in liquid biopsies were circulating tumor cells (CTCs). Originally discovered in 1869, CTCs were intensely studied in the past few decades, as technological advances improved their detection, isolation and analysis. CTCs are generated by a multitude of tumors, but are rarely found in the circulation of healthy individuals or patients with non-malignant tumors, which lead to an initial euphoria for their use as diagnostic markers [46]. Even though CTCs reflect the mutational profile of the primary lesion, correlate with tumor burden and provide a higher predictive value than conventional imaging, their utilization has proven to be difficult. CTCs are extremely rare cells, with estimates of 1 – 10 CTCs per ml blood even for metastatic tumors. The same volume of blood contains billions of leukocytes, erythrocytes and other blood cells that generate an enormous level of background noise for CTC detection. Deciphering their genomic and transcriptomic profiles is highly informative, but despite continuous advances in both CTC isolation and sequencing technologies, gathering sufficient quantities of nucleic acids remains challenging [47].

Another type of material frequently assessed in liquid biopsies is cell-free DNA (cfDNA). It was first described in 1947 and can be detected in various biofluids including blood, urine and saliva [48]. cfDNA is known to be released from cells undergoing apoptosis and necrosis, but active secretion from living cells has also been proposed [49, 50]. Even though they are also detectable in biofluids from healthy individuals in low concentrations, cfDNA levels are much higher in disease states, particularly cancer [51]. In cancer patients, circulating tumor DNA (ctDNA) captures a quantitative and qualitative snapshot of both the primary tumor and potential metastases, which is of tremendous interest for diagnostic purposes. ctDNA has been detected in many types of cancer including breast cancer, colorectal cancer and lung cancer, and offers unique advantages that make it useful for oncology [52-54]. Levels of ctDNA were shown to correlate with tumor burden and thus allow the monitoring of disease progression. As the half-life of ctDNA was estimated at approximately two hours, much shorter than that of many protein biomarkers, it can be used to very precisely assess cancer progression and treatment success [55].

Even though ctDNA makes up only 1 – 40 % of all circulating DNA, it provides highly useful information, particularly on somatic genetic changes of the tumor cells it was shed from [56]. Screening actionable mutations in tumor driver genes such as epidermal growth factor receptor (EGFR), BRAF, KRAS and TP53 allows remote phenotyping of primary tumors and, importantly, prediction of treatment response. In fact, the first ctDNA-based liquid biopsy assay, commercialized by Roche to assess EGFR mutations in lung cancer patients, was approved by the American Food and Drug Administration (FDA) in 2016 [57]. Concordance between mutations detected in ctDNA and tissue biopsies is typically high, and ctDNA has been reported to detect more mutations than a single needle biopsy would [45]. It is therefore not only used to detect a disease state, but also to profile the respective tumor and monitor the dynamic evolution of its genome, generating valuable leads for treatment decisions and disease management. The assessment of druggable mutations in circulating tumor-derived DNA is particularly important for making decisions about endocrine treatments in hormone-sensitive tumors such as breast cancer [58, 59]. Although the clinical utilization of ctDNA benefited tremendously from recent advances in nucleic acid quantification technologies, it is hampered by minimal analyte concentrations and low signal-to-noise ratios in much the same way that CTCs are. Additionally, ctDNA fragments are usually rather short (< 200 base pairs (bp)), which impedes the detection of some genetic variations such as copy number variations and gene translocations [60].

While CTCs and ctDNA have been on the scene for decades, extracellular vesicles (EVs) just recently emerged as new and exciting players in liquid biopsies.

1.3 Extracellular vesicles and exosomes: A new paradigm in intercellular communication

"Extracellular vesicle" is an umbrella term for a broad variety of vesicles that are actively or passively released from cells. While our knowledge in the field is still evolving, EVs are generally categorized into three major classes: apoptotic bodies, microvesicles and exosomes. Despite sharing some similarities, these EV types differ significantly in their biogenesis, physiochemical properties, molecular composition and biological function (Table 1).

Table 1. Characteristic features of the three major classes of extracellular vesicles. mRNA: messenger RNA; ncRNA: non-coding RNA; ESCRT: endosomal sorting complex required for transport.

	Apoptotic bodies	**Microvesicles**	**Exosomes**
Size	0.5 – 5 µm	100 nm – 1.5 µm	30 – 120 nm
Biogenesis	Cellular blebbing	Pinched off at plasma membrane	Endosomal pathway
Release	Passive	Active	Active
Cargo	Nuclear fragments, organelles, cytosol	DNA, mRNA, ncRNA, soluble proteins, membrane receptors	DNA, mRNA, ncRNA, soluble proteins, membrane receptors
Function	Cellular decay	Intercellular communication	Intercellular communication
Markers	Phosphatidylserine exposed in outer leaflet	ARF6, flotillin-1, VAMP3, CD40, integrins	Tetraspanins, ESCRT components, Rab GTPases

Apoptotic bodies, generally the largest type of EV, are generated by cells undergoing programmed cell death. As apoptotic cells shrink, intracellular cascades degrade organelles, destroy mRNAs and ultimately fragment the cell's nucleus and cleave the DNA contained therein. During cellular disassembly, parts of the plasma membrane protrude outwards and pinch off, generating vesicles that are called apoptotic bodies or apoptotic blebs. These are subsequently incorporated by phagocytic cells, ensuring controlled degradation of cells without releasing their content into the extracellular space. Apoptotic bodies strongly differ in size, ranging from 0.5 µm to 5 µm in diameter. They contain cytoplasm, organelles and nuclear fragments as well as nucleic acids. As they are directly pinched off from plasma membranes, their membrane composition is similar to that of their parent cells. In contrast to healthy cells

and other classes of EVs, however, they display phosphatidylserine in the outer leaflet of their membrane. Initially thought to be rather inert byproducts of cellular degradation, apoptotic bodies have been shown to have biological functions such as horizontal DNA transfer [61].

In contrast to apoptotic bodies, microvesicles (MVs), sometimes also referred to as microparticles, are produced by living cells. They, too, are shed from the plasma membrane by outward budding and subsequent fission, which generates vesicles of 100 nm – 1.5 µm in diameter. Biogenesis of MVs takes place at designated membrane sites and involves specific lipids, membrane proteins and contractile cytoskeleton components. While composition and topology of MV membranes resemble the secreting cell's plasma membrane, they are additionally enriched in phosphatidylserine and lipid raft-like structures. Besides displaying surface receptors, adhesion molecules and other membrane-associated proteins, MVs also carry cargo in their lumenal space. MV-encapsulated biomolecules include enzymes, signaling molecules, cytokines, cytoskeletal components and nucleic acids. Whereas apoptotic bodies seem to enclose a random assortment of degraded cellular components, cargo packaging into MVs is a specific process [62]. Several studies have reported an enrichment of proteins related to adhesion and signal transduction, as well as mitochondrial, ribosomal and cytoskeletal proteins. While several key players in the recruitment of specific cargo to MVs have been identified, the underlying machinery is yet to be fully elucidated [63].

Nucleic acid cargo of MVs includes DNA and mRNA, as well as non-coding RNAs such as miRNAs. MV-encapsulated cargo is protected from degradation and biologically functional in recipient cells, representing a novel way of horizontal biomolecule transfer. Generally speaking, secreted MVs have the capacity to modify nearby extracellular matrix (ECM) as well as to physiologically impact proximal and distant cells. As the composition of MVs shed from different tissues varies, there is an ensuing heterogeneity of MV subpopulations with different biological functions. For instance, MVs were shown to mediate both pro-inflammatory and anti-inflammatory functions, depending on their cytokine cargo and the effects provoked in target cells [64, 65]. Additionally, MVs were implicated to play a role in coagulation, immunity and drug resistance [66-68]. MVs shed from tumor cells were shown to degrade ECM, induce angiogenesis and aid in cancer immune evasion [69].

While interest in MVs has steadily increased over the past decade, exosomes remain the most studied class of EVs. After their discovery in 1983, exosomes were initially regarded as the "garbage bin of the cell"; a system to remove waste, misfolded proteins and harmful substances by expulsion into the extracellular space [70, 71]. After the stunning 2007 discovery that they contain functional RNA molecules that can be transferred to recipient cells, where they induce phenotypic changes, however, exosome research increased exponentially [72].

Realizing their role in physiology and pathophysiology has brought about a paradigm shift in intercellular communication, and exosomes are now highly studied for both their biological functions and their potential biotechnological applications.

Compared to other EV types, exosomes stand out due to their molecular composition, physiochemical properties and biogenesis. While both apoptotic bodies and MV are pinched off at the plasma membrane, exosomes have an intracellular origin (Figure 3). They are generated in the endosomal trafficking pathway, the cell's major mechanism for internalizing and processing extracellular material [73]. During endocytosis, parts of the plasma membrane engulf extracellular fluid and bud inwards, eventually pinching off from the cell surface to create endosomes. Within these vesicles, the newly internalized cargo is sorted and some of the endocytosed membrane, including protein receptors, is recycled to the plasma membrane. During maturation from early endosome to late endosome, the vesicle's lumen is acidified in preparation for cargo degradation. Endosomal maturation is a highly dynamic process, in which parts of the vesicle's membrane can bud inwards to generate intraluminal vesicles (ILVs) within the endosome, which is now referred to as a multivesicular body (MVB) or multivesicular endosome (MVE). ILVs are enclosed by endosomal membranes and carry cytosolic material previously incorporated during inward budding. Typical MVBs range from 250 nm to 1 µm in diameter and contain up to several dozens of ILVs [74]. While most MVBs subsequently fuse with lysosomes, which leads to disintegration of the vesicle and degradation of its cargo, some MVBs escape this fate and instead fuse with the plasma membrane. This fusion process releases ILVs of 30 – 120 nm in diameter into the extracellular space, where they are now called exosomes. Exosome secretion is a highly conserved mechanism employed by virtually all mammalian cells, and exosomes have consequently been detected in a variety of biofluids including serum and plasma, breast milk, urine, saliva and cerebrospinal fluid (CSF) [75-77].

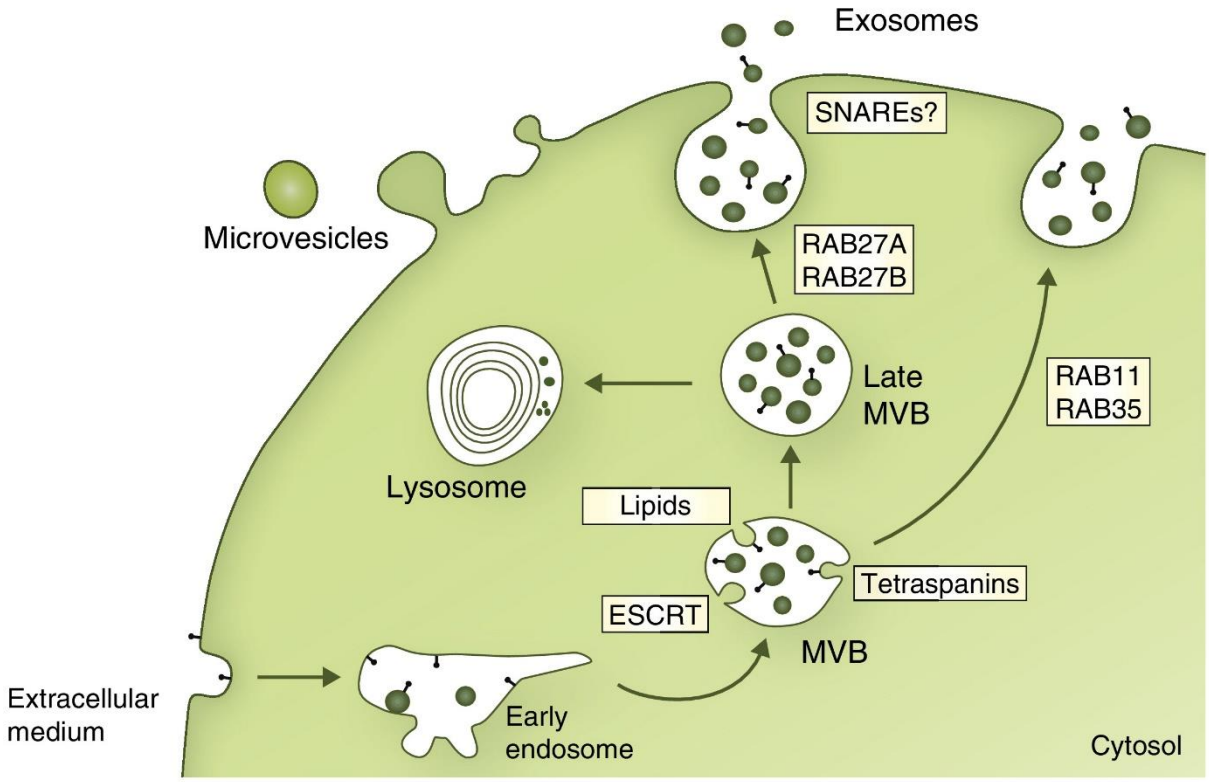

Figure 3. Schematic representation of the biogenesis of extracellular vesicles. While microvesicles bud off from the plasma membrane, exosomes are generated by inward budding of endosomal membranes, forming multivesicular bodies (MVB). During maturation, MVBs are either targeted to lysosomes for degradation or fuse with the plasma membrane to release enclosed exosomes into the extracellular space. ESCRT: endosomal sorting complex required for transport; SNARE: soluble N-ethylmaleimide-sensitive-factor attachment receptor. Figure reprinted from Kowal et al. [78].

Even more so than for MVs, cargo selection for exosomes is not a random event, but a specific process orchestrated by complex macromolecular mechanisms. The endosomal sorting complex required for transport (ESCRT) is the main machinery of MVB biogenesis and seems to be at least partly involved in sorting cargo into ILVs. Due to experiments in which ESCRT deletion did not completely abrogate exosome secretion, we now know that there are additional, ESCRT-independent mechanisms of MVB formation, potentially working competitively on the same MVB [79]. Many molecules have been associated with specific cargo sorting into ILVs, and mechanisms based on tetraspanins, posttranslational protein modification and lipid raft structures are discussed in the literature. A comprehensive understanding of MVB biogenesis and incorporation of various types of cargo into ILVs, however, remains elusive.

Due to their unique mode of biogenesis, exosomes carry marker proteins that might distinguish them from MVs [80]. Endosomal proteins involved in MVB biogenesis (Alix, TSG101) are enriched in exosomes, as are proteins related to membrane trafficking (Rab GTPases) and cell membrane-derived tetraspanins such as CD63, CD81 and CD9 (Figure 4). In addition to a

set of generic markers, exosomes also carry soluble proteins and surface receptors specific for their secreting cells. Depending on the parental cell, they might carry enzymes, cytokines and heat shock proteins as well as proteins involved in antigen presentation, signal transduction and cellular adhesion and targeting. Exosomal membranes are enriched in cholesterol, sphingomyelin, lipid raft-like structures and phospholipids carrying short saturated fatty acids, which render them highly stable and protect cargo from harsh conditions such as acidity, hypoxic environments and the presence of nucleases [81-83].

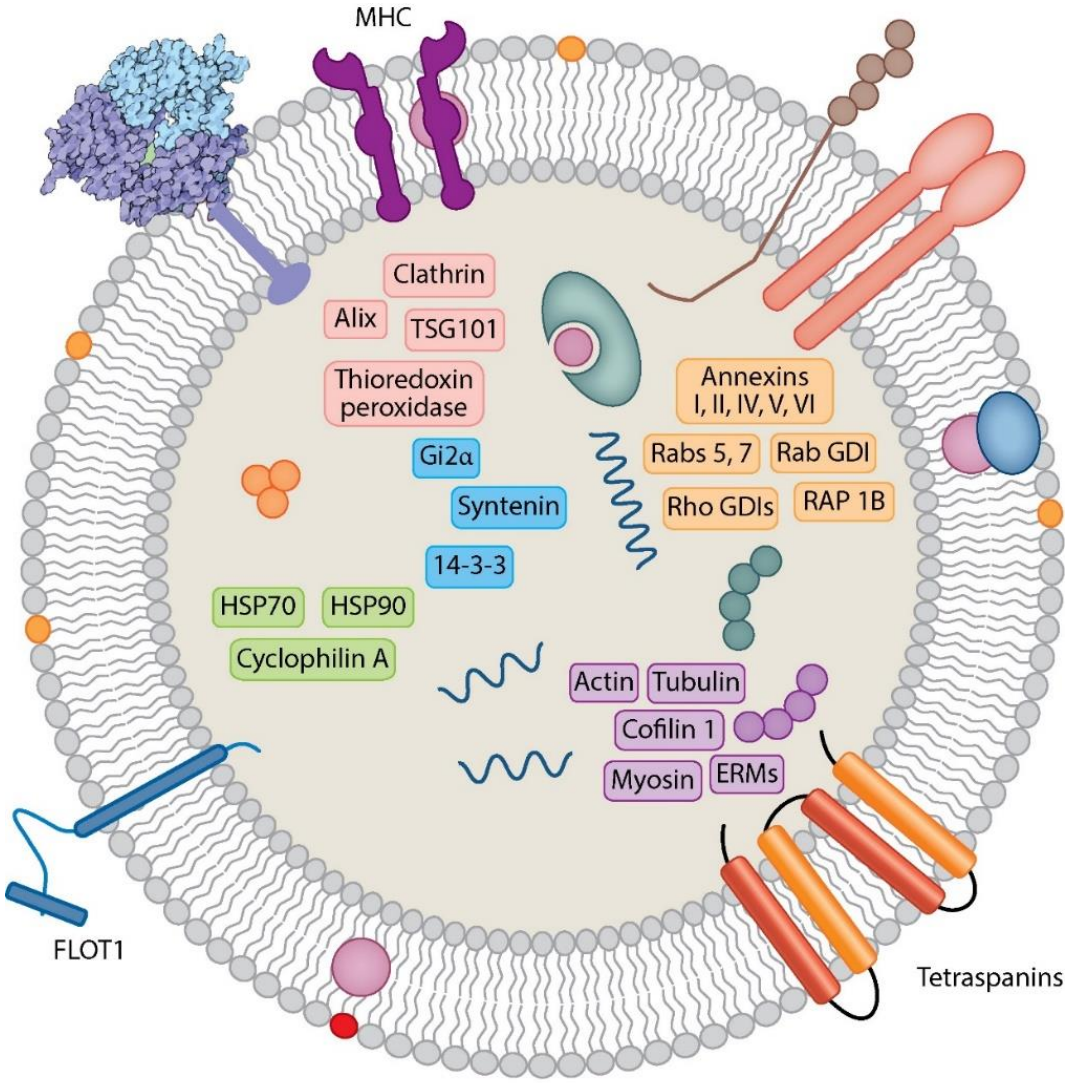

Figure 4. Schematic structure of a canonical exosome. Exosomal membranes carry integral and transmembrane proteins such as tetraspanins and major histocompatibility complexes (MHC). Soluble cargo enclosed by the phospholipid bilayer includes enzymes, cytoskeletal components and signaling factors. In addition to their protein cargo, exosomes also carry nucleic acids including DNA, mRNA and non-coding RNA. HSP: heat shock protein. Figure reprinted from Kourembanas et al. [84].

An important feature of exosomes is that they also carry nucleic acids, including small amounts of genomic and mitochondrial DNA, mRNA, but particularly non-coding RNAs such as transfer RNAs (tRNAs), miRNAs and long non-coding RNAs (lncRNAs). Early studies reporting that exosomal RNA is biologically active and can be translated to proteins (mRNA) or modulate gene expression (miRNA) in recipient cells have generated considerable interest in the selection and function of RNA in circulating exosomes. Exosomal RNA profiles are often reported to differ from those in their parental cells, hinting at selective cargo sorting [85]. Additionally, RNA profiles in exosomes from a given cell type differ depending on the physiological state of the cell. For instance, alterations in cargo were reported for exosomes secreted by hypoxic cells as well as cells undergoing inflammatory stress or oncogenic transformation [86-88]. Even though various mechanisms of how cytosolic RNA might be targeted to ILVs, ranging from specific sequence motifs to affinity-based interaction with MVB membranes, have been proposed, a comprehensive understanding of the underlying principles remains elusive [89-91].

In order to convey a biological function, exosomes must interact with target cells in one of several ways (Figure 5). The first and most direct way of interaction includes fusion of exosome membranes with plasma membranes, which releases vesicular cargo into the recipient cell's cytosol. Mechanistic studies on direct membrane fusion indicate a "rolling adhesion" process, during which exosomes initially attach to and roll across plasma membranes until reaching dedicated sites of internalization [92]. Even though little is known about its molecular mechanisms, the process was postulated to involve specific recognition of proteins on vesicular and cellular membranes, followed by protrusion of fusogenic proteins, lipid reorganization, a hemifusion transition state and, finally, full fusion of both membranes [93]. Secondly, intact exosomes can be internalized by several mechanisms including endocytosis, pinocytosis and macropinocytosis. Previous research demonstrated the involvement of both clathrin-dependent and -independent endocytosis, as well as lipid raft-mediated internalization [94-96].

Similar to competitive modes of exosome biogenesis, several mechanisms of internalization might simultaneously be at play. Alternatively, different subpopulations of exosomes might be taken up by distinct mechanisms, depending on the surface proteome of vesicles and recipient cells. Even though uptake kinetics differ between cell types, internalization was shown to be active and energy-dependent [97]. During internalization, exosomes enter recipient cells in endocytic vesicles, from which they can escape via back-fusion with the endosomal membrane. This "endosomal escape" releases exosomal cargo into the cytosol, enabling interaction with various cellular machineries [98].

In addition to targeting vesicles to specific tissues, surface proteins on exosomes also mediate signal transduction in a third way of interaction with recipient cells [99]. Exosomal surface ligands can bind to and activate plasma membrane receptors, which in turn initiates intracellular signaling cascades that lead to phenotypic changes by, e.g. altering the cellular transcriptome [100]. In its most extreme form, this mechanism can go as far as exosomes inducing apoptosis of immune cells via the ligands FasL and TRAIL displayed on their surface [101].

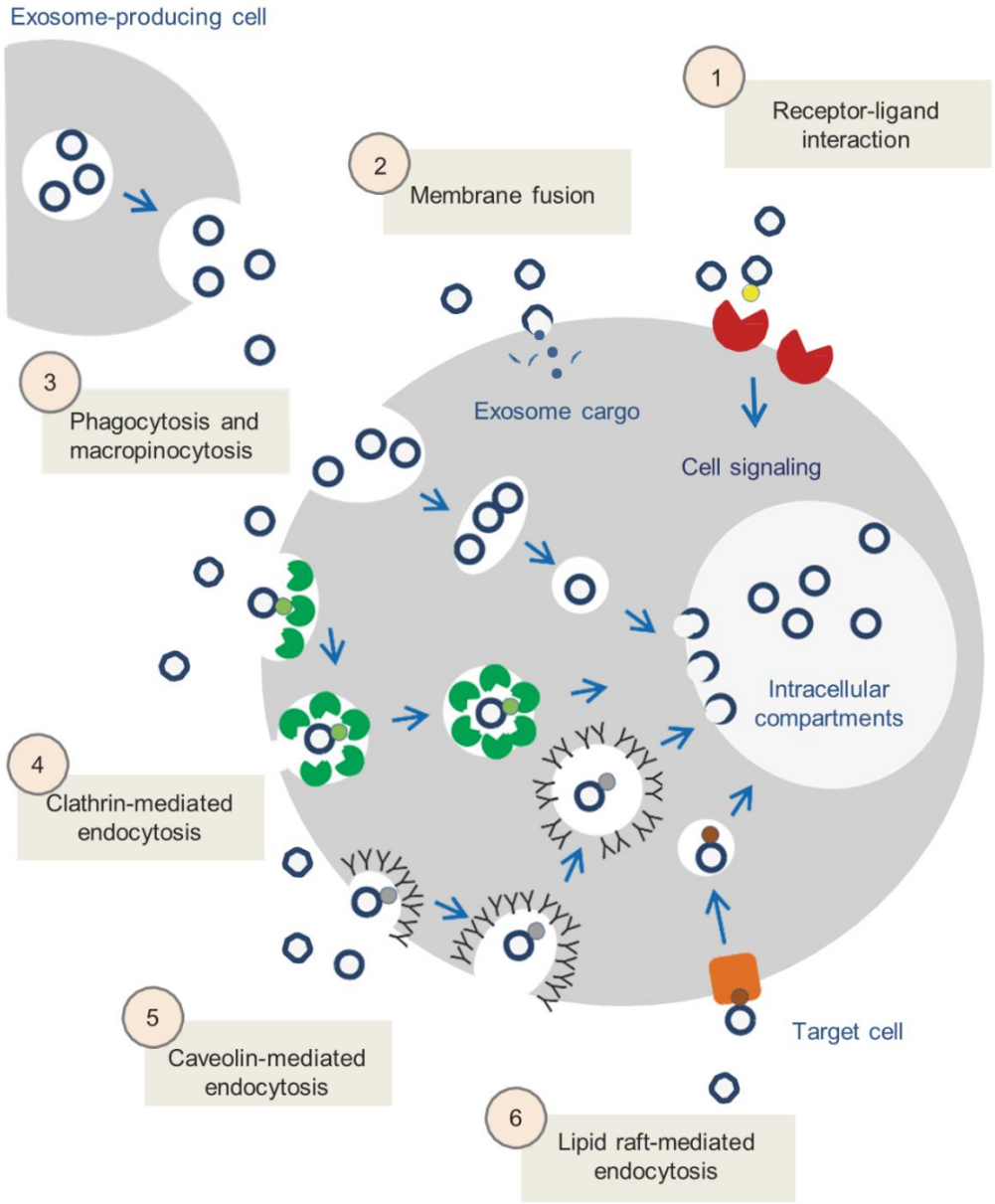

Figure 5. Schematic summary of modes of interaction for exosomes and recipient cells. Exosomes can initiate signaling cascades via cell surface receptors (1) or release their cargo into the cytosol by fusion with the recipient cell's plasma membrane (2). Alternatively, intact exosomes can be internalized by macropinocytosis (3) or endocytosis (4-6). Figure adapted from [102].

Initially regarded as a mere garbage disposal system, exosomes are now known to be involved in a myriad of biological functions ranging from immune modulation and antigen presentation to inflammation and tissue polarity [103-105]. Beside their role in physiology and homeostasis, pathological functions of exosomes have attracted particular interest. For instance, exosomes secreted from various tumor cells were shown to modulate ECM, suppress the anti-tumor immune response and foster the development of pre-metastatic niches [106]. Research on neurodegenerative diseases proved exosomes to propagate Alzheimer's Disease and Parkinson's Disease by spreading toxic amyloid-beta and α-synuclein, respectively [107, 108]. Additionally, exosomes were recently implicated in the pathogenesis of diabetes, cardiovascular disease (CVD) and infectious disease [109-111]. As they mirror both physiological and pathological conditions, circulating exosomes are intensely studied to understand molecular mechanisms of disease as well as to develop novel diagnostic, prognostic and predictive biomarkers.

1.4 Extracellular vesicles in liquid biopsies: Intercepting disease signaling

As the composition and concentration of EVs from infected, transformed and otherwise pathological cells differ from those of healthy tissues, EVs are a promising source of clinical biomarkers. Due to their protective shell, EVs harbor a stable and concentrated repertoire of biomolecules that lend themselves to quantification in downstream analytical assays. Indeed, nucleic acids including DNA, mRNA and miRNA as well as proteins and combinations thereof have been suggested as clinical markers in various types of cancer [112-114]. Similar to ctDNA, DNA in tumor EVs reflects the genomic makeup of secreting cells, thus allowing the remote screening of actionable mutations. Importantly, however, analysis of tumor EVs provides the additional benefit of deciphering the tumor's transcriptome, including aberrations such as fusion transcripts, in the same sample. By revealing malignant neoantigens, EVs might turn out to yield crucial clues for the development of personalized cancer immunotherapies, particularly those based on chimeric antigen receptor T cells (CAR-T) [115]. Indeed, combined analysis of ctDNA and EV RNA was recently demonstrated to increase detection rates of growth factor receptor mutations in lung cancer patients [116].

The ability to specifically enrich EVs from malignant cells represents additional benefits compared to other sample types used in liquid biopsies. Capturing EVs based on disease-specific surface proteins drastically reduces the background noise typically encountered when analyzing bulk populations of circulating vesicles and might provide a more concentrated source of biomarker candidates. For instance, tumor antigens were exclusively detected in

EVs isolated from patient sera using melanoma-specific antibodies as opposed to EVs from healthy tissues [117]. In a study on ovarian cancer, miRNA profiles in tumor EVs enriched by immunoaffinity to surface epithelial cell adhesion molecule (EpCAM) were suggested as a diagnostic marker in screening programs [118].

As the development of new drugs heavily relies on identifying patient populations that might benefit from a particular treatment and subsequently monitoring their response to it, the need for reliable biomarkers stretches beyond mere diagnostic purposes. In the age of precision medicine, developing effective personalized therapies is contingent on molecular disease profiling to guide clinical decisions. In the same vein, assessing a patient's response to therapy on a regular basis could indicate the potential need to change therapeutic approaches and thus improve favorable outcomes. Results from EV-based clinical trials are scare but encouraging, as demonstrated by a recent report on the utility of EVs to monitor therapy response in glioma patients [119].

Despite the undeniable interest in using EVs as clinical biomarkers, many questions remain to be answered before their potential can be fully realized. First, there is no consensus on which biofluid to sample for a given disease, which method of EV isolation is most suitable for various downstream assays, and how these might be affected by pre-analytical variables [96]. Second, given the stunning heterogeneity of EV classes and subtypes, it is unclear which EV fraction in a given biofluid is most informative of the disease state. For instance, a recent study on prostate cancer reported that larger tumor vesicles (oncosomes) carried significant amounts of high molecular weight DNA, which reflected tumor-specific genetic aberrations [120]. Exosomes from the same patients, on the other hand, contained only negligible amounts of DNA and might be a less suitable source of biomarkers. Third, different types of EV cargo, or combinations thereof, might harbor the most informative analytes for different diseases [121]. Fourth, the majority of promising biomarker candidates never make it to the clinic because they fail validation in larger cohorts, lack clinical utility or cannot be integrated into easily automated high-throughput analytical platforms [122, 123]. The success or failure of EVs as clinical biomarkers will ultimately be determined by how well they fare in facing these challenges.

1.5 Aim of the study

The goal of this study was to take a three-pronged approach to assess the utility of EVs and their miRNA cargo as disease biomarkers. Initially, a robust and reproducible workflow for the quantification of miRNAs via high-throughput Next-Generation small RNA Sequencing (small RNA-Seq) was to be established as the basis for evaluating disease-specific changes in

miRNA expression. Next, various methods for isolating EVs from human sera were to be compared in order to identify the most suitable approach for downstream small RNA-Seq experiments. Third, the utility of analyzing vesicular miRNA profiles for biomarker studies was to be examined *in vitro* and *in vivo* using the previously established techniques.

2 Methods

2.1 Establishing the technical foundations for Next-Generation Sequencing-based analysis of microRNAs in extracellular vesicles

To initially assess current methods and best practices for small RNA-Seq, a comprehensive literature review was performed. Special focus was put on miRNAs, which are arguably the most heavily researched type of small non-coding RNA. We aimed at establishing a robust and reproducible workflow, covering all aspects from pre-analytical variables to sequencing and data analysis, and identifying potential hurdles along the way (Figures 6 and 7). Additionally, measures to standardize experimental and analytical variables and thus enhance reproducibility and comparability of small RNA-Seq were highlighted in Buschmann et al. [124] (Appendix I). In line with the anticipated goal of analyzing miRNAs in EVs, a particular emphasis was put on handling and sequencing RNA from cell-free samples including material gathered in liquid biopsies.

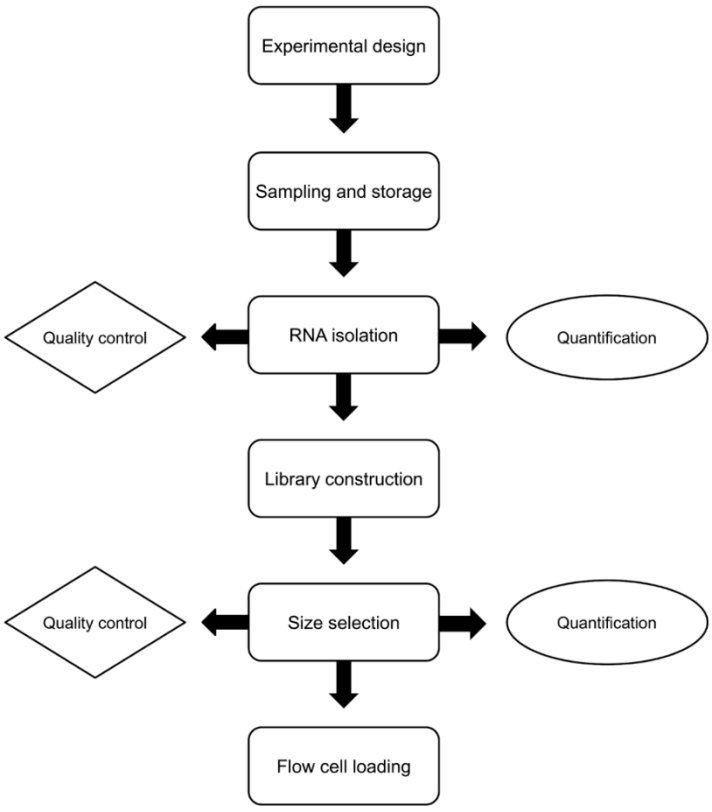

Figure 6. Schematic overview of small RNA-Seq experiments: Experimental design, pre-analytical procedures, library preparation and sequencing. Figure reprinted from Buschmann et al. [124] (Appendix I).

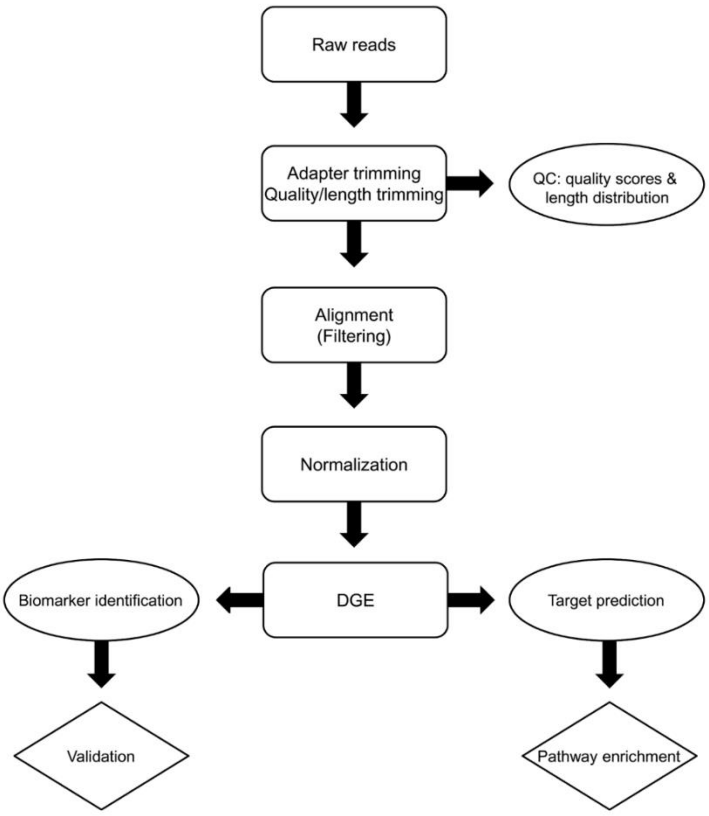

Figure 7. Schematic overview of small RNA-Seq experiments: Data pre-processing, alignment, normalization and biomarker-centric analysis. QC: quality control; DGE: differential gene expression analysis. Figure reprinted from Buschmann et al. [124] (Appendix I).

Next, several methods of isolating EVs from human sera were assessed regarding their suitability for downstream small RNA-Seq experiments. Specifically, approaches based on size-exclusion chromatography (SEC; qEV Columns, Izon Science, Oxford, UK; Exo-spin Midi Columns, Cell Guidance Systems, Cambridge, UK), precipitation (miRCURY Exosome Isolation Kit, Exiqon, Vedbaek, Denmark), membrane affinity (exoRNeasy Serum-Plasma Midi Kit, Qiagen, Hilden, Germany) and sedimentation (differential ultracentrifugation, Beckman Coulter, Brea, CA, USA) were compared in a cohort of sepsis patients (n=9) and healthy volunteers (n=10). EV isolation, small RNA-Seq and data analysis were carried out as published in Buschmann et al. [125] (Appendix II). Differential gene expression (DGE) analysis using DESeq2 [126] was performed to detect miRNAs dysregulated between patients and volunteers (Figure 8). Library sizes, miRNA profiles and relative mapping frequencies to several other classes of small non-coding RNA as well as results from DGE analysis were utilized to evaluate the suitability of each method for sequencing-based miRNA biomarker analysis. Additionally, EVs from each isolation method were analyzed by Nanoparticle Tracking Analysis (NTA; NanoSight LM10, Malvern Instruments, Malvern, UK) to assess method-specific differences in size and concentration. Purity, morphology and protein marker

composition of EVs from patients and volunteers were determined by transmission electron microscopy (TEM; Zeiss EM900, Carl Zeiss Microscopy, Jena, Germany) and Western blotting (WB; XCell SureLock Mini-Cell Electrophoresis System, Thermo Fisher Scientific, Waltham, MA, USA), respectively.

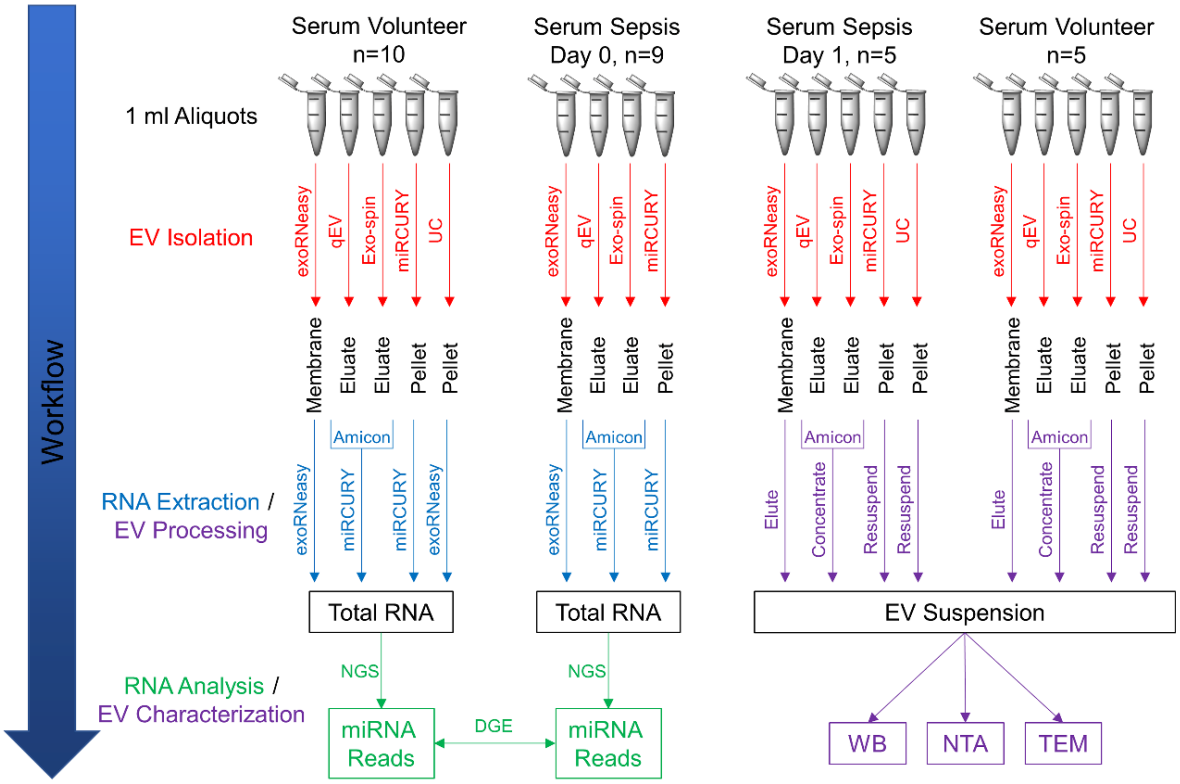

Figure 8. Schematic summary of EV isolation, RNA extraction and small RNA-Seq (left). EVs were isolated from patient sera sampled on the day of admission to the intensive care unit (day 0). In a subgroup of sepsis patients and volunteers, EVs were isolated and biologically characterized (right). For characterization experiments, patient EVs were isolated from sera sampled after 24 hours of hospitalization (day 1). EV: extracellular vesicle; UC: differential ultracentrifugation; NGS: small RNA-Seq; DGE: differential gene expression analysis; WB: Western blot; NTA: Nanoparticle Tracking Analysis; TEM: transmission electron microscopy. Figure reprinted from Buschmann et al. [125] (Appendix II).

Biomarker studies, particularly those focusing on critically ill patients, are frequently carried out using arterial or venous sera, but little is known about potential differences in concentration and composition of EVs isolated from these biofluids. Assessing the comparability of small RNA-Seq experiments based on arterial and venous EVs was therefore the next step. In a cohort of heart disease patients (n=20), paired sera were sampled from the radial artery and internal jugular vein of individual patients prior to coronary artery bypass graft (CABG) surgery. After precipitation-based isolation (miRCURY Exosome Isolation Kit, Exiqon, Vedbaek, Denmark) of EVs from these sera, associated miRNAs were profiled by small RNA-Seq as described in [127]. Potential differences in the abundance of miRNAs from arterial and venous

EVs were assessed via DGE analyses using DESeq2 and validated by reverse transcription quantitative polymerase chain reaction (RT-qPCR; miRCURY LNA miRNA PCR Kit, Exiqon, Vedbaek, Denmark). Additionally, EVs from both types of biofluid were biologically characterized to detect potential differences in size, concentration and composition.

2.2 Testing the utility of microRNAs in extracellular vesicles as biomarkers *in vitro*

Overexpression of the glucocorticoid receptor (GR) in Triple-Negative Breast Cancer (TNBC), which does not express receptors for estrogen, progesterone and human epidermal growth factor 2, is associated with therapy resistance and increased mortality. To assess the potential manifestation of this particularly oncogenic phenotype in secreted miRNAs, EVs shed from TNBC cells were analyzed in an *in vitro* model of GR overexpression. Three human TNBC cell lines (MDA-MB-231, MDA-MB-436, MDA-MB-468) were transfected with plasmids coding for the nuclear receptor subfamily 3 group C member 1 (*NR3C1*). The ensuing overexpression of GR and its target genes was validated by RT-qPCR (QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany), Sso Advanced Universal Supermix and SsoFast EvaGreen Supermix (Bio-Rad Laboratories, Munich, Germany)). Next, EVs were isolated (miRCURY Exosome Isolation Kit, Exiqon, Vedbaek, Denmark) from culture media of transfected and control cells and analyzed by small RNA-Seq as published in Buschmann, González et al. [128] (Appendix III). DGE analysis and validation of dysregulated miRNAs were carried out using DESeq2 and RT-qPCR (miScript II RT Kit and miScript SYBR-Green PCR Kit, Qiagen, Hilden, Germany), respectively. Additionally, cellular RNAs from transfected and control cells were analyzed in parallel to compare GR-dependent intracellular and extracellular changes in miRNA profiles.

2.3 Testing the utility of microRNAs in extracellular vesicles as biomarkers *in vivo*

Despite being a commonly performed medical procedure, open heart surgery carries substantial risks for adverse outcomes including postoperative organ failure and increased mortality. Early identification of patient populations at risk for perioperative cardiac instability might help to improve patient care and reduce adverse outcomes. The utility of EV miRNAs for patient stratification was assessed in a cohort of heart disease patients (n=19) and healthy volunteers (n=20) as described in [129]. EVs were isolated (miRCURY Exosome Isolation Kit,

Exiqon, Vedbaek, Denmark) from patient sera sampled prior to open heart surgery and from matched volunteers. EV-associated miRNAs were profiled by small RNA-Seq and compared in DGE analyses to detect miRNA species dysregulated between surgical patients and volunteers. Expression levels of these candidate miRNAs were then correlated to prospectively recorded perioperative clinical variables such as inflammation, intraoperative epinephrine dosing, serum lactate levels and duration of surgery.

In a separate study, the utility of cellular and extracellular miRNAs for disease detection was assessed in a cohort of patients in septic shock and matched healthy volunteers (n=7 each). Sepsis and septic shock represent the clinical manifestation of a massively derailed immune reaction to pathogens such as bacteria, fungi or viruses. Early disease detection, particularly the distinction between a septic state and a sterile inflammatory state, are critically important to administer appropriate clinical care and reduce mortality. Blood cells (PAXgene RNA Kit, PreAnalytiX, Hombrechtikon, Switzerland), total serum (S-Monovette, Sarstedt, Nürnbrecht, Germany) and serum-derived EVs (miRCURY Exosome Isolation Kit, Exiqon, Vedbaek, Denmark) were sampled from patients upon admission to the intensive care unit (ICU; day 0) and on the fourth day of treatment (day 4) if available. For each sample type, miRNAs were analyzed by small RNA-Seq and compared between patients and volunteers as published in [130] (Appendix IV). Dysregulated miRNAs in each sample type were further analyzed in a separate cohort of sepsis patients (n=9), septic shock patients (n=6) and matched volunteers (n=16) by RT-qPCR (miRCURY LNA miRNA PCR Kit, Exiqon, Vedbaek, Denmark). In addition to evaluating blood compartment-specific diagnostic information, samples were grouped according to disease severity and patient survival to detect miRNAs with expression levels correlated to progressive organ damage. Expression of miRNAs correlating with disease severity on day 0 was additionally evaluated by RT-qPCR in samples drawn on day 4 to detect potential markers for disease progression.

2.4 Enhancing experimental reproducibility and reporting in extracellular vesicle research

In order to assess current practices in experimental procedures and reporting thereof, an international consortium of researchers performed an extensive review of 1,226 manuscripts on EVs published between 2010 and 2015. Articles were distributed amongst participants and evaluated regarding a checklist of 115 parameters focusing on technical aspects of EV research; particularly pre-analytical variables and methods for EV isolation and characterization. Resulting data were used to create a crowdsourcing online

knowledgebase (http://evtrack.org) and analyzed to identify current practices in EV research. Additionally, the EV-TRACK (Transparent Reporting and Centralizing Knowledge in Extracellular Vesicle Research) platform was set up to assist EV researchers with designing and performing experiments, increase experimental reproducibility and enhance reporting as detailed in [131] (Figure 9). Briefly, researchers are encouraged to upload experiments and publications, which subsequently receive a numerical score (EV-METRIC) based on the completeness of reporting crucial experimental details regarding EV isolation and characterization. Non-public sections of the EV-TRACK platform were established to provide reviewers and editors with experimental details of unpublished manuscripts during the peer-review process. Additionally, a comprehensive search tool was implemented to allow researchers to browse the database for specific experimental details in published experiments.

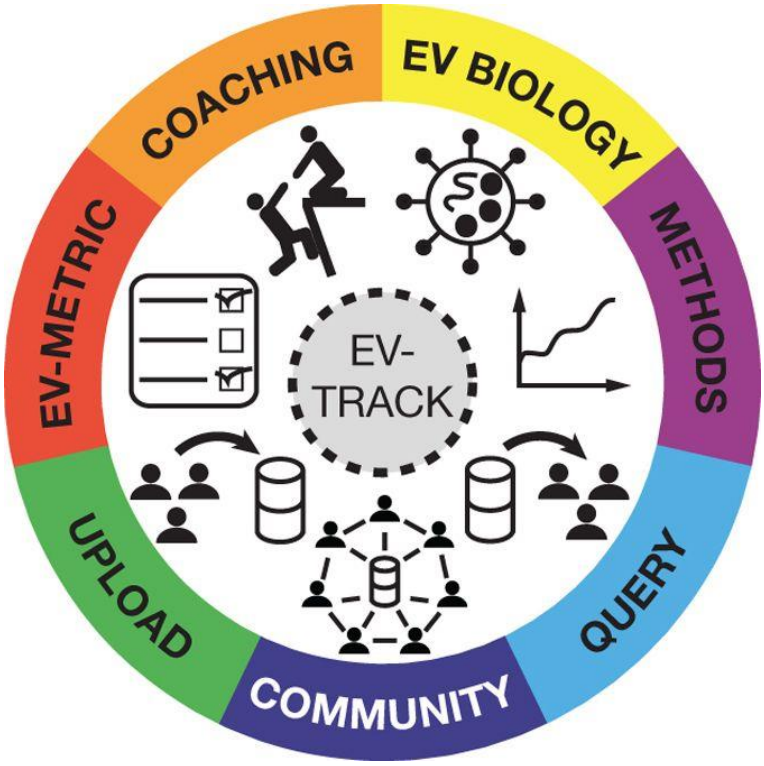

Figure 9. The seven enabling features of the EV-TRACK platform. EV-TRACK was established to improve methodology and reporting through community-driven coaching for researchers, experimental guidelines and objective scoring of published experiments. Uploading new experiments to EV-TRACK will ultimately expand the field's level of knowledge on EV biology, different subpopulations and suitable methods for their isolation and characterization. Figure adapted from [131].

In addition to the EV-TRACK initiative, a community-driven effort was made to improve research on EVs by providing specific recommendations for their isolation, characterization and functional analysis. Based on a previous initiative published in 2014 [132], current technological developments, novel biological insights and challenges in the rapidly evolving

EV field were assessed to assemble a comprehensive update of the recommended Minimal Information for the Studies of EVs (MISEV) as published in the MISEV2018 guidelines [133]. Along with updated guiding principles for separation, concentration and analysis of marker proteins, new recommendations regarding EV nomenclature and topology of analytes were made to adapt to the growing appreciation of EV heterogeneity and diverse subpopulations of vesicles with potentially distinct cargo and functionality.

2.5 Assessing current and future challenges for extracellular vesicle-based clinical biomarkers

Despite the potential utility of EV-based biomarkers, various obstacles need to be faced on the road from discovery research to the clinic. To highlight both the bright sides and challenges in the EV biomarker field, the International Society for Extracellular Vesicles (ISEV) held a workshop in Birmingham, UK, in 2017. Researchers from within the EV field discussed current achievements and problems, while participants not themselves working with EVs provided a general perspective on the challenges of moving biomarkers from bench to bedside. As published in [123], discussions were centered around biospecimen collection and pre-analytical procedures as well as challenges relating to EV heterogeneity and the detection of rare analytes, the need for more technically sophisticated analytical tools and how EVs can be integrated into routine clinical assays.

3 Results and Discussion

3.1 Small RNA-Seq is a promising but challenging technique for extracellular vesicle microRNA analysis

Current methods for small RNA-Seq were assessed in a comprehensive literature review as published in Buschmann et al. [124] (Appendix I). NGS-based techniques of transcriptomic analysis were found to provide exceptional advantages over previous technologies such as RT-qPCR. Specifically, its high throughput and sensitivity, ability to profile all transcripts without *a priori* sequence knowledge and single-nucleotide resolution have made small RNA-Seq invaluable for applications such as quantification of disease-related miRNAs and detection of novel transcripts [134, 135]. Indeed, miRNA-based biomarker signatures have been presented for a wide array of diseases including colorectal cancer, CVD and neurodegenerative disorders [136-138]. At the same time, various challenges and sources of technical bias were identified for small RNA-Seq experiments as demonstrated in Buschmann et al. [124] (Appendix I). Briefly, reliable and replicable experiments are contingent on sound study design and standardized pre-analytical procedures including sampling as well as RNA extraction, quality control and quantification. Circulating cell-free miRNAs and miRNAs in EVs present additional challenges due to yielding particularly low amounts of starting material. Furthermore, virtually all steps of library preparation protocols, ranging from the addition of adaptor and barcode sequences to PCR amplification and size selection of target fragments, were found to be prone to bias [139-141]. Even for technically sound sequencing experiments, strategies for normalization and data analysis need to be tailored for the respective study requirements, and results should be validated by orthogonal techniques such as RT-qPCR or digital PCR (dPCR). Crucial steps in the pre-analytical and analytical phase of small RNA-Seq experiments as well as corresponding recommendations to reduce bias and enhance reproducibility are summarized in Buschmann et al. [124] (Appendix I, Tables 1 and 2).

Recent reports claimed that biomedical science is facing an alarming reproducibility crisis and argue that a large proportion of research funding is wasted on studies that yield inaccurate or unreplicable findings [142, 143]. An analytical tool as complex and widely used as small RNA-Seq needs to be utilized correctly to not contribute to the multitude of invalid or inconclusive studies. In the context of biomarker research, reliable and reproducible results on disease-related alterations in miRNA expression can only be generated by rigorous standardization of pre-analytical factors and experimental workflows. Similar to the

recommendations made in Buschmann et al. [124] (Appendix I), previous guidelines such as the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) were implemented to improve experimental practices in PCR-based nucleic acid quantification [144]. Indeed, adoption of the frequently-cited MIQE guidelines was shown to improve reporting in the qPCR literature [145].

3.2 Precipitation is a highly suitable extracellular vesicle isolation method for downstream microRNA quantification

In a comparative study on sepsis patients and healthy volunteers, different commercial methods to isolate EVs from serum were assessed regarding their suitability for downstream analysis of EV-associated miRNAs by small RNA-Seq. EVs were isolated from patient and volunteer sera using methods based on SEC, precipitation, membrane affinity and sedimentation. As demonstrated in Buschmann et al. [125] (Appendix II), sequencing small RNA in kit-specific isolates generated libraries of vastly different size and composition for each method (Figure 10). Of the commercial approaches, precipitation and membrane affinity yielded the largest libraries in both sepsis patients and volunteers, clearly surpassing SEC-based methods. When assessing mapping frequencies to various classes of small non-coding RNA, precipitation distinctly outperformed its competitors with 27.56 % and 35.08 % of reads mapping to miRNAs for patients and volunteers, respectively.

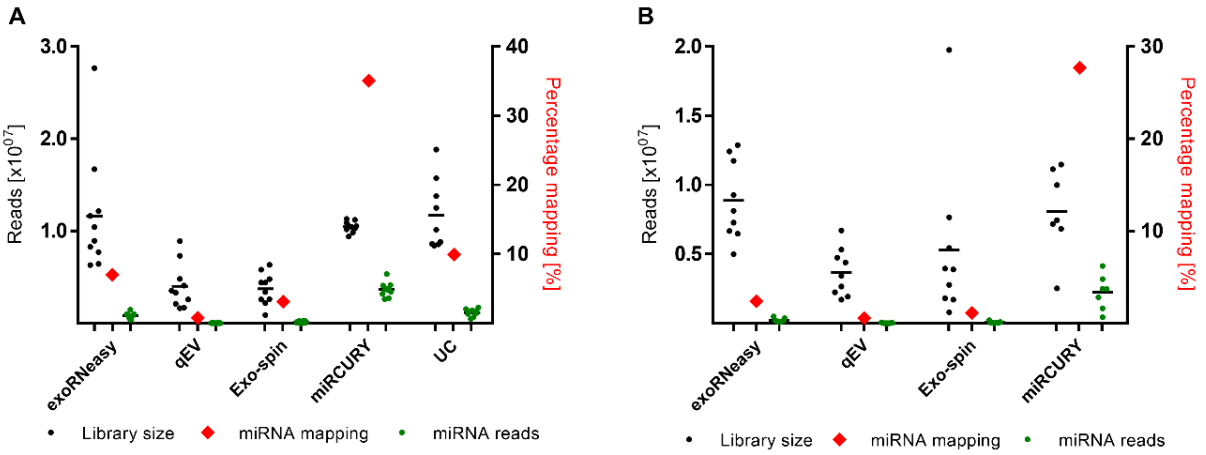

Figure 10. Library sizes and miRNA mapping in EVs isolated from volunteers (A) and sepsis patients (B). Mapping frequencies to miRNAs (red diamonds) are plotted against the right x-axes. EV isolation by precipitation (miRCURY) yielded the highest enrichment of miRNAs, while SEC-based methods (qEV, Exo-spin) resulted in smaller libraries and fewer miRNA reads. miRNA mapping percentages are mean for 10 volunteers and nine patients. UC: differential ultracentrifugation. Figure adapted from [125] (Appendix II)

In line with increased miRNA mapping rates, libraries from precipitation-based EV isolation also performed excellently in DGE analyses and accurately separated patients and volunteers in hierarchical clustering (Figure 11).

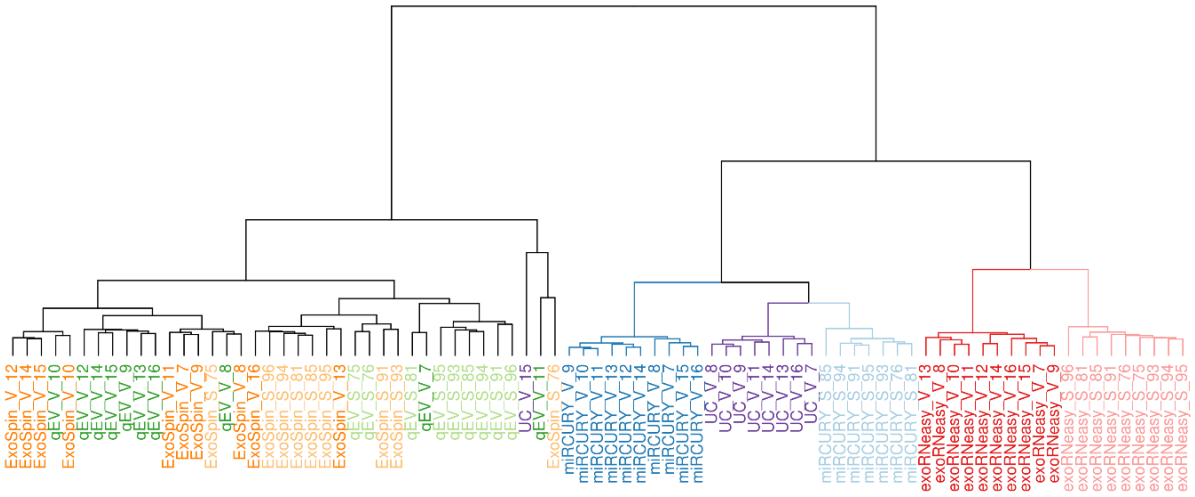

Figure 11. Hierarchical clustering of miRNAs in EVs isolated by commercial methods and UC. Primary clustering separated samples from precipitation (miRCURY), membrane affinity (exoRNeasy) and UC from SEC-based methods (qEV, Exo-spin). Volunteers (darker shades, V) and sepsis patients (lighter shades, S) were accurately distinguished by miRCURY and exoRNeasy, while miRNAs from qEV and Exo-spin isolation displayed noticeable heterogeneity and did not fully separate healthy and diseased individuals. UC: differential ultracentrifugation. Figure reprinted from [125] (Appendix II).

Biological characterization of isolated EVs revealed significant differences in particle size and concentration, with precipitation-derived samples featuring both the highest concentration and smallest diameter of EVs [125] (Appendix II, Figure 6). Immunoblot analysis of kit-specific isolates demonstrated an enrichment of common EV markers in preparations from SEC and membrane affinity but not from precipitation and sedimentation. As demonstrated by additional experiments using density gradient centrifugation, and in line with findings from other groups [146, 147], the latter methods co-isolated large amounts of serum albumin and other soluble proteins that diluted EV markers. These findings indicated a differential suitability of isolation methods for downstream analytical assays: isolates from SEC-based approaches were comparatively pure and low in non-EV proteins but yielded suboptimal results in small RNA-Seq and ensuing DGE analyses. EV preparations isolated by precipitation, on the other hand, were heavily contaminated with serum proteins but clearly outperformed SEC-based methods in sequencing and miRNA-based patient classification. These data suggested that a substantial proportion of miRNAs in precipitation-derived samples might not be associated with genuine EVs but rather co-precipitate in non-EV miRNA carriers such as lipoproteins and circulating Argonaute proteins. Indeed, previous studies demonstrated that significant amounts

of circulating miRNAs are associated with AGO2 [148] and that vesicle-free miRNAs are the dominant population in precipitation-derived EV isolates [149]. Similar findings were reported for urinary EVs, which indicates that despite being prone to co-isolate soluble proteins, precipitation is highly suitable for downstream small RNA-Seq as long as crude preparations are acceptable for the respective research question [150]. Additionally, impurities in EV preparations were shown to have little effect on subsequent miRNA quantification [151]. As biomarker applications generally aim at identifying circulating miRNA signatures that reliably detect diseases or separate patient populations rather than at attributing miRNAs to their specific carriers, precipitation was selected as an appropriate method to isolate serum EVs in further studies.

3.3 microRNA profiles in extracellular vesicles from arterial and venous sera can be compared for biomarker studies

As detailed above, reliable biomarker research relies on replicable studies and well-characterized biospecimens. Blood-derived biofluids such as serum and plasma are the most commonly used sample types in liquid biopsies and might be sampled from arterial or venous blood vessels, depending on the respective patient population. As a prerequisite to compare these sample types within a given study and across studies on either sample matrix, miRNA profiles in matched arterial and venous EVs were analyzed in a cohort of cardiac surgery patients (n=20). Analysis of resulting arterial and venous sequencing libraries revealed highly similar frequencies of reads mapped to miRNAs and other classes of small non-coding RNA [127]. As evidenced in unsupervised clustering, miRNA profiles in all samples overlapped significantly and tended to cluster by patient rather than by sampling site (Figure 12).

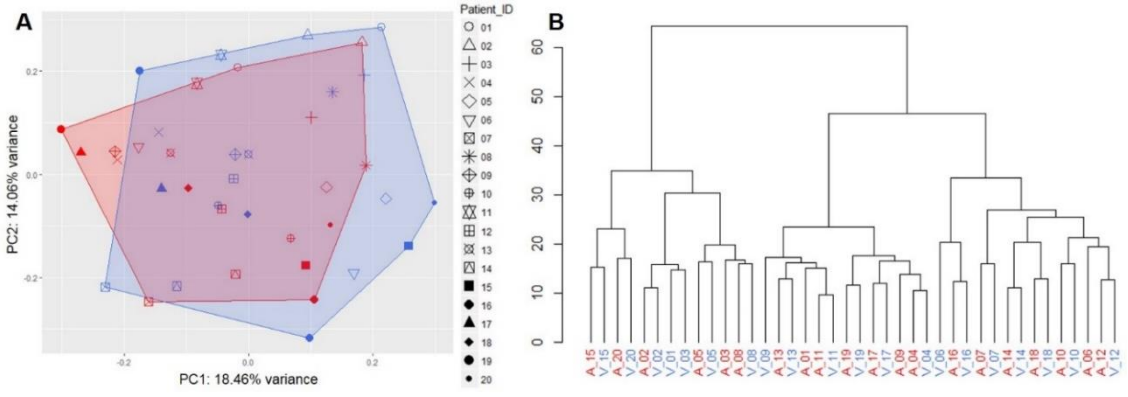

Figure 12. Principal component analysis (A) and hierarchical clustering (B) of miRNAs in EVs isolated from arterial (red, A) and venous (blue, V) sera. Individual miRNA profiles displayed significant overlap and did not reveal any systemic variation depending on sampling site. Figure reprinted from [127].

Additionally, EVs from venous and arterial sera were highly similar regarding particle size and concentration as well as morphology and marker protein profiles [127]. Even though a previous study on rats reported different miRNA profiles in total arterial and venous sera [152], differences in human serum-derived EVs were marginal. With the caveat of lung disease, which might skew arteriovenous balances due to altered interaction with arterial blood, it therefore seems feasible to sample arterial or venous blood and to compare studies utilizing either biofluid for biomarker studies on EV miRNAs.

3.4 Extracellular vesicles and their microRNA cargo are not a panacea in molecular diagnostics

Extracellular vesicles are heavily studied diagnostic candidates in various types of cancer. Dysregulated miRNA signatures in EVs have been proposed to indicate breast, colon and prostate cancer [153-155]. Beyond mere disease detection, they might also prove to be useful prognostic markers for patient survival and recurrence [156, 157]. Profiling DNA and RNA in EVs from pancreatic cancer patients was shown to detect actionable mutations that might guide treatment decisions in personalized medicine [115, 158]. Within the spectrum of breast cancers, variants that do not express receptors for estrogen, progesterone and human epidermal growth factor 2 are particularly difficult to treat due to their unresponsiveness to commonly used hormonal therapies. In these Triple-Negative Breast Cancers (TNBC), overexpression of the glucocorticoid receptor (GR) additionally correlates with therapy resistance and increased mortality. As GR impacts the expression of both protein-coding and non-coding RNAs, we assessed the potential extracellular reflection of GR overexpression by analyzing miRNA profiles in EVs shed from different TNBC cell lines *in vitro*. As published in Buschmann, González et al. [128] (Appendix III), artificially induced GR expression prompted only minor changes in secreted miRNAs. While EVs from individual cell lines were clearly distinguishable based on their miRNA composition, profiles in EVs from parental and transfected cells overlapped substantially. When assessing differentially regulated miRNAs between the two groups, statistical significance was not reached for any transcript. Additionally, GR overexpression induced only slight changes in intracellular miRNA profiles [128] (Appendix III, Figures 6 and 7). Even though transfected TNBC cells overexpressed functional GR that significantly upregulated downstream target genes, increased GR levels had little effect on intracellular and vesicular miRNA profiles. The glucocorticoid receptor mRNA itself is targeted by multiple miRNAs [159, 160] but little is known about the impact of GR signaling on miRNA expression [161]. Given our data, increased aggressiveness and therapy resistance in GR-overexpressing TNBC does not seem to be mediated by downstream

miRNA signaling. Additionally, quantifying miRNA levels in TNBC EVs is not informative about GR expression status. In a clinical situation, where phenotyping the hormone receptor repertoire of a tumor is crucial to devise a therapeutic strategy, analyzing EVs might not be an appropriate diagnostic tool. Despite their general utility as analytes in breast cancer liquid biopsy, circulating miRNAs do therefore not hold the answer to every clinical question.

3.5 Extracellular vesicles might help identify high-risk patients prior to cardiac surgery

The potential utility of EV miRNAs for risk stratification was assessed in a cohort of CVD patients undergoing open heart surgery (n=19). As detailed in [129], serum EVs were isolated prior to surgery, and miRNA profiles in EVs from patients and healthy volunteers (n=20) were analyzed by small RNA-Seq. Even though EVs from patients and volunteers did not differ in size, morphology and expression of marker proteins, levels of 29 miRNAs were significantly changed between the groups. In patient EVs, 15 miRNAs were downregulated with log2 fold changes between -1.03 and -1.76, while 14 miRNAs were upregulated with log2 fold changes between 1.02 and 2.34. Differentially regulated miRNAs were correlated with clinical variables recorded during and after surgery, revealing a subset of miRNAs that correlated significantly with intraoperative epinephrine dosing requirements ($p=0.008$), serum lactate levels ($p=0.036$) and decreased urine excretion ($p=0.031$) (Figure 13), which indicate perioperative cardiac instability and potential kidney damage, respectively. Patient demographics (e.g. age, body mass index) and clinical variables unrelated to organ dysfunction (e.g. duration of surgery, duration of postoperative ICU therapy, inflammation, intraoperative volume requirements) did not significantly correlate with miRNA expression.

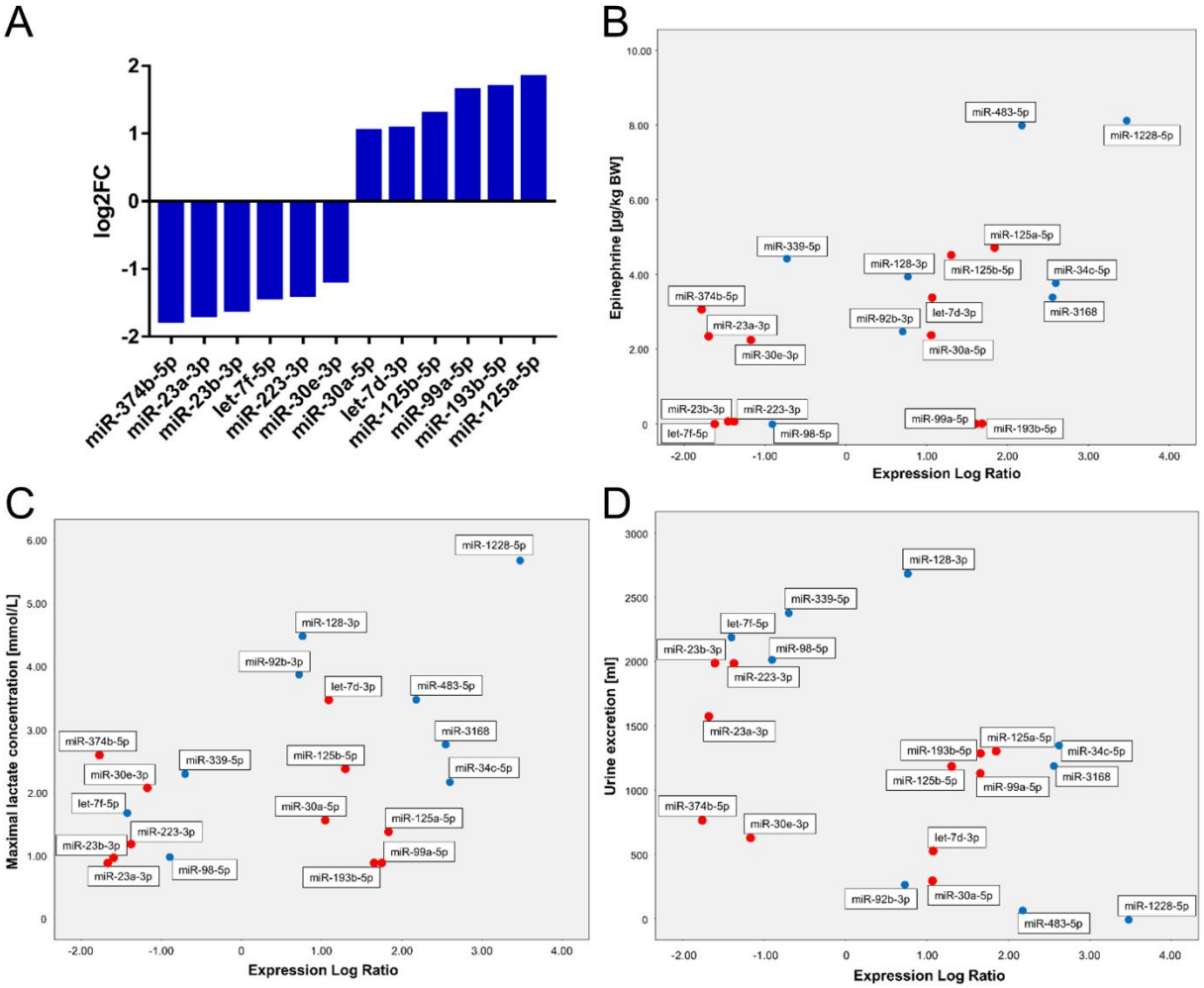

Figure 13. Differential expression of miRNAs that correlated with outcome-relevant variables (A). Positive log2 fold changes indicate miRNAs upregulated in patient EVs compared to volunteer EVs. Correlation of miRNA expression with intraoperative epinephrine requirements (B), serum lactate levels (C) and low urine excretion (D). Red dots indicate miRNAs with mean expression levels ≥ 50 reads and log2 fold changes ≥ |1|. Figure adapted from [129].

Cardiac failure during heart surgery is a rare but disastrous occurrence associated with a mortality of up to 50 % [162] and additional short- and long-term detrimental consequences including acute and chronic multiple organ failure [163]. Risk assessment prior to surgery is therefore essential to identify patients at increased risk for perioperative cardiac events. Once identified, appropriate risk mitigation measures such as close perioperative monitoring, careful selection of anesthetics, attendance of highly trained staff, use of specialized critical care facilities, advanced use of perioperative echocardiography, mechanical circulatory assist devices and customized pharmacologic management can be taken. As many risk factors such as age, comorbidities, hypertension, renal insufficiency and vascular disease stem from patient demographics and medical history, questioning and physical examination of patients prior to cardiac surgery is crucial [164]. The availability of circulating biomarkers for risk assessment, however, is scarce. Brain natriuretic peptide (BNP), a hormone secreted by stressed and

injured cardiomyocytes, was recently suggested for preoperative risk stratification in non-cardiac surgery [165]. In cardiac surgery, high postoperative levels of BNP were associated with longer hospitalization and early mortality [166]. A robust biomarker for identification of high-risk patients could mitigate adverse outcomes and enhance patient survival. While assessing miRNA profiles in EVs failed to reflect the GR expression status of TNBC, it might prove to be useful in preoperative patient stratification for cardiac surgery.

3.6 Extracellular vesicles carry specific diagnostic information in critically ill patients

The utility of EV miRNAs for disease detection was assessed in a cohort of septic shock patients and healthy volunteers. As published in [130] (Appendix IV), miRNA profiles in circulating EVs, total serum and blood cells were analyzed by small RNA-Seq and compared between patients and volunteers to identify differentially regulated transcripts. Across blood compartments, a total of 77 and 103 miRNAs was down- and upregulated in patients, respectively. The majority of regulated miRNAs was detected in blood cells with little overlap between compartments (Figure 14). Three miRNAs each were simultaneously down- and upregulated in EVs, serum and cells.

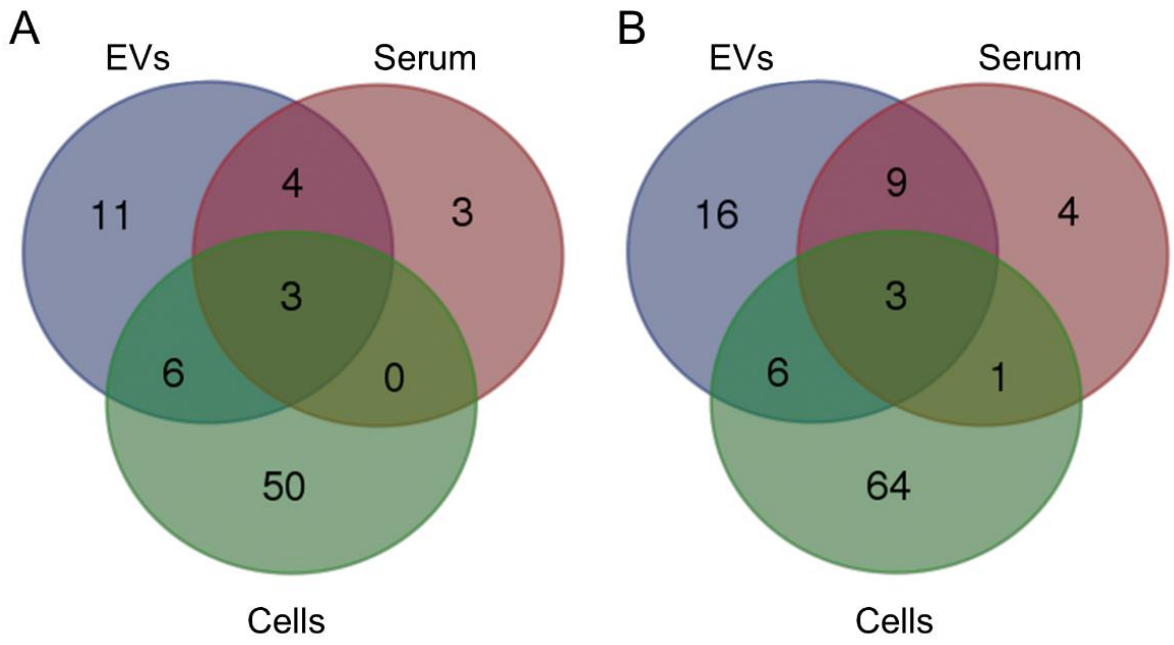

Figure 14. Compartment-specific changes in miRNA expression between septic shock patients and volunteers. In total, 77 miRNAs were downregulated in patients (A), whereas 103 miRNAs were upregulated (B). Despite some overlap between blood compartments, there was significant sample type-specific miRNA regulation.

In order to validate our findings from small RNA-Seq, a selection of 20 dysregulated miRNAs was analyzed by RT-qPCR in an independent cohort of sepsis patients (n=9), septic shock patients (n=6) and volunteers (n=16). In the validation study, nine miRNAs (EVs: two; serum: one; cells: six) were significantly dysregulated between volunteers and both sepsis patients and patients in septic shock upon admission to the ICU (Figure 15). Additionally, expression levels of several miRNAs were found to correlate with disease severity [130] (Appendix IV, Figure 4).

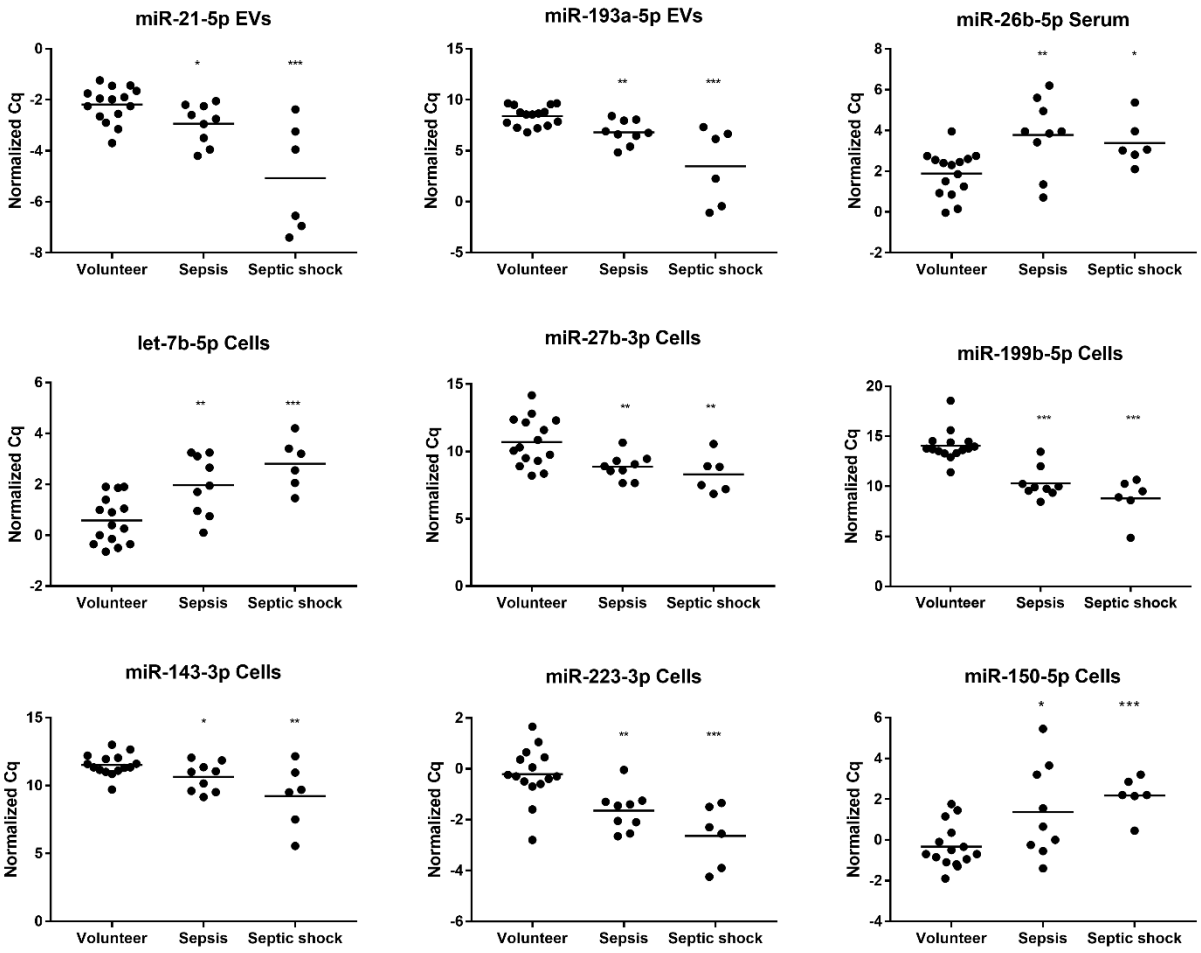

Figure 15. Expression levels of nine miRNAs in EVs, sera and blood cells were significantly increased or decreased in both sepsis patients and septic shock patients. Lower normalized Cq values indicate higher expression levels. *: $p<0.05$; **: $p<0.01$; ***: $p<0.001$.

In a separate analysis of RT-qPCR data, patients were grouped by outcome (survivors/non-survivors) regardless of disease severity upon hospitalization. Three miRNAs in extracellular samples, but not blood cells, were able to significantly distinguish between survivors and non-survivors (Figure 16).

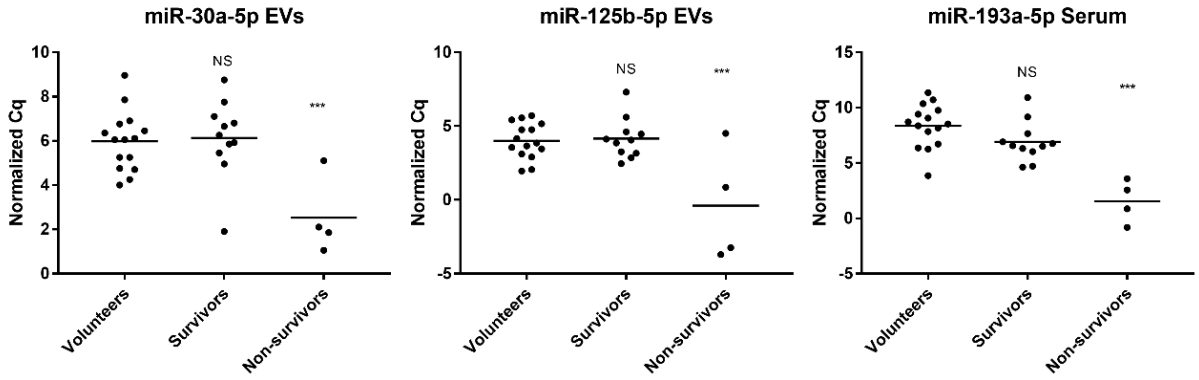

Figure 16. Survival prediction by extracellular miRNAs. Expression of three miRNAs in EVs (miR-30a-5p, miR-125b-5p) and sera (miR-193a-5p) sampled upon hospitalization correlated significantly with outcome in sepsis patients. Lower normalized Cq values indicate higher expression levels. NS: not significant; ***: p<0.001. Figure reprinted from [130] (Appendix IV).

Sepsis is a complex, life-threatening disease and a major cause of death in hospitals. Tragically, hospitalization rates for sepsis have continuously increased in the past decade, reaching approximately 970,000 annual admissions in the US alone [167]. In sepsis, the initial infectious insult is answered by a dysregulated inflammatory and oxidative host response, which can lead to sequential organ failure and death if it does not abate during treatment. Mortality in sepsis correlates with disease severity and ranges as high as 40 – 80 % for patients in septic shock [168]. The financial costs of sepsis management are disproportionally higher than for any other disease and increase for patients with delayed diagnosis and higher disease severity [167]. Delaying diagnosis and administration of appropriate treatment also accelerates disease progression and increases mortality [169, 170]. Despite being a frequently encountered condition in hospitals worldwide, sepsis poses significant diagnostic challenges. Due to its complex and only partially understood pathogenesis, heterogeneity of causative pathogens and lack of specific biomarkers, sepsis is oftentimes not immediately diagnosed upon hospitalization. Its pronounced inflammatory component further complicates the crucial distinction from non-infectious inflammatory conditions such as the systemic inflammatory response syndrome (SIRS).

Circulating EVs in sepsis are intensely studied for their role in pathogenesis [171] and as potential diagnostic and therapeutic agents [172]. Given the success of related efforts in oncology, it is no surprise that the utilization of EVs and their cargo as diagnostic biomarkers has gained traction in sepsis research. Similar to the results presented above, several studies reported dysregulated miRNA profiles in total plasma and circulating EVs sampled from sepsis patients [173, 174]. In our RT-qPCR data, the miRNA most significantly dysregulated in both sepsis and septic shock (miR-199b-5p) was detected in blood cells, which, in this case, might be the superior sample type for disease detection. A correlation with patient survival, on the

other hand, was only found for extracellular miRNAs (Figure 16), which underlines the compartment-specific miRNA signaling in sepsis. In a recent study on plasma exosomes from sepsis patients, Real et al. presented a signature of differentially regulated miRNAs that relate to cell cycle regulation and discriminate between survivors and non-survivors [174]. Interestingly, both EV miRNAs correlated with survival in our data (miR-30a-5p, miR-125b-5p) were also found to negatively regulate cell cycle and proliferation [175, 176].

As described in our study, disease-associated regulation of miRNAs is specific to individual blood compartments, and EVs might not be the most suitable sample type for all biomarker purposes. Analyzing miRNAs in blood cells could prove to be particularly useful for disease detection in sepsis, where the overshooting host response is bound to be reflected in circulating immune cells. Should EV miRNAs be selected as the sample type of choice for a given disease or clinical question, it is crucial to utilize appropriate methodology and standardize analytical assays. As detailed in 3.1 and 3.2, reliable small RNA-Seq studies are predicated on careful experimental design, stringently controlled laboratory workflows and consideration of the impact that different methods of EV isolation have on downstream miRNA quantification. To ensure specificity, miRNA candidates need to be validated in larger cohorts of sepsis patients and, optimally, patients with an inflammatory but non-infectious phenotype. Due to significant postoperative inflammation, patients undergoing CABG surgery (see 3.5) might be an appropriate control group to validate the specificity of potential sepsis markers for pathogen-associated inflammation.

3.7 Extracellular vesicle research is hampered by heterogeneity in experimental protocols and insufficient reporting

Current experimental practices and reporting thereof were assessed in 1,226 research articles published between 2010 and 2015. A total of 1,742 individual experiments was extracted from the articles and analyzed based on 115 parameters pertaining to sample type and pre-analytical variables as well as methods for EV isolation and characterization. As detailed in [131], our analyses detected 1,038 unique isolation protocols, with differential ultracentrifugation being used in 45 % of experiments. In addition to this multitude of isolation protocols, significant heterogeneity in reporting vital experiment details on experimental parameters and biochemical features of isolated vesicles was discovered. Regardless of sample type, many studies failed to provide specific information on the equipment and chemicals used in the experiments and did not extensively characterize EVs. In 17 % of experiments, EVs were not characterized at all, while characterization was limited to analysis

of proteins or particles in 29 % and 39 % of studies, respectively. Based on these findings, we developed the EV-METRIC, which is a reporting index that consolidates nine experimental parameters crucial for interpretation and reproducibility of experiments into a numerical score. The score ranges from 0 % to 100 % and reflects the completeness of reporting for EV isolation methods, protein analysis and particle analysis [131]. Across all sample types, reporting was found to be generally insufficient, with less than 6 % of experiments receiving an EV-METRIC above 50 % (Figure 17). Of all simple types included in the study, reporting was worst in experiments with serum EVs, as reflected by an average EV-METRIC of 11 %.

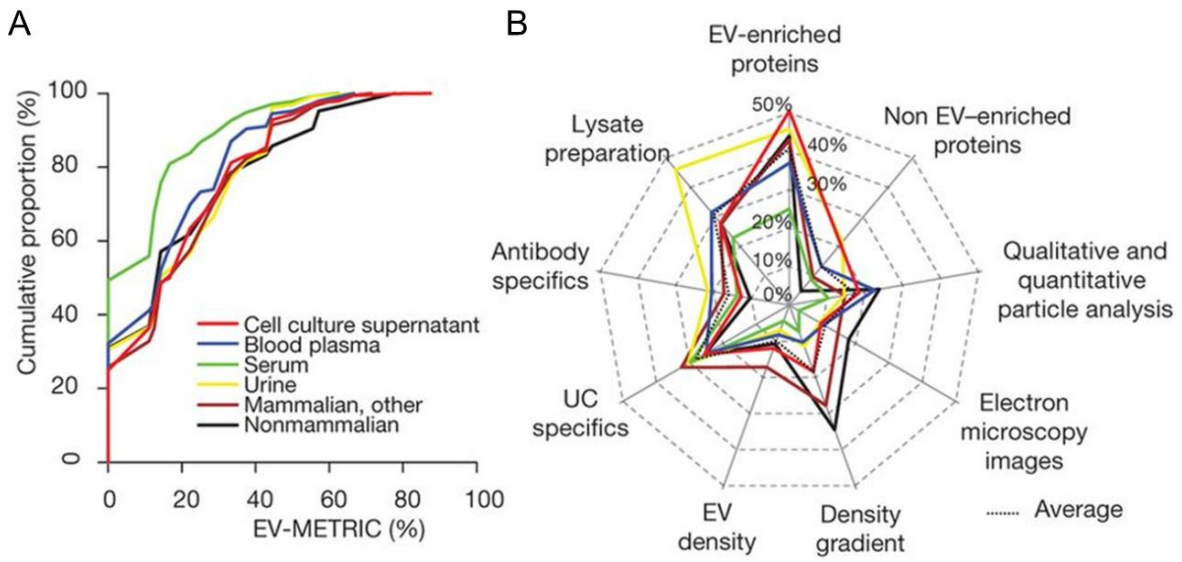

Figure 17. Cumulative EV-METRIC stratified by sample type (A). On average, experiments on serum-derived EVs obtained the lowest scores. Average adherence to each EV-METRIC parameter stratified by sample type (B). Characterization of EV-enriched proteins was more commonly reported than details on non EV-enriched proteins or antibodies used in the experiment. EV: extracellular vesicle; UC: differential ultracentrifugation. Figure adapted from [131].

As the EV-METRIC is intended as a tool to improve experimental rigor and reproducibility rather than a recommendation on which experiments to perform, higher scores could oftentimes be achieved by merely providing extensive experimental details. Indeed, 81 % of experiments analyzed in this study could have obtained a better score by improved reporting without performing any additional experiments.

EV research is a new and rapidly evolving area of science. Biogenesis, composition, functional activity and biotechnological applications of EVs from various biofluids are studied by many groups around the globe, which advances our understanding of EV biology and potential clinical uses. As demonstrated in EV-TRACK [131], the accelerated interest in EVs has brought about a vast heterogeneity of protocols and techniques. Both inappropriate methodology and

insufficient reporting hamper the interpretation of individual studies and the comparability of results across experiments. By making researcher aware of crucial experimental details and improving reporting in articles, the introduction of EV-TRACK and the EV-METRIC might elevate all areas of EV research regardless of sample type and scientific question. Even though average EV-METRIC scores improved over the past few years (Figure 18), it is too early to tell if EV-TRACK will have a positive impact on EV science. Still, the fact that EV-TRACK has now been cited in well over 100 articles from various branches of EV research indicates that there is at least an increased awareness throughout the community.

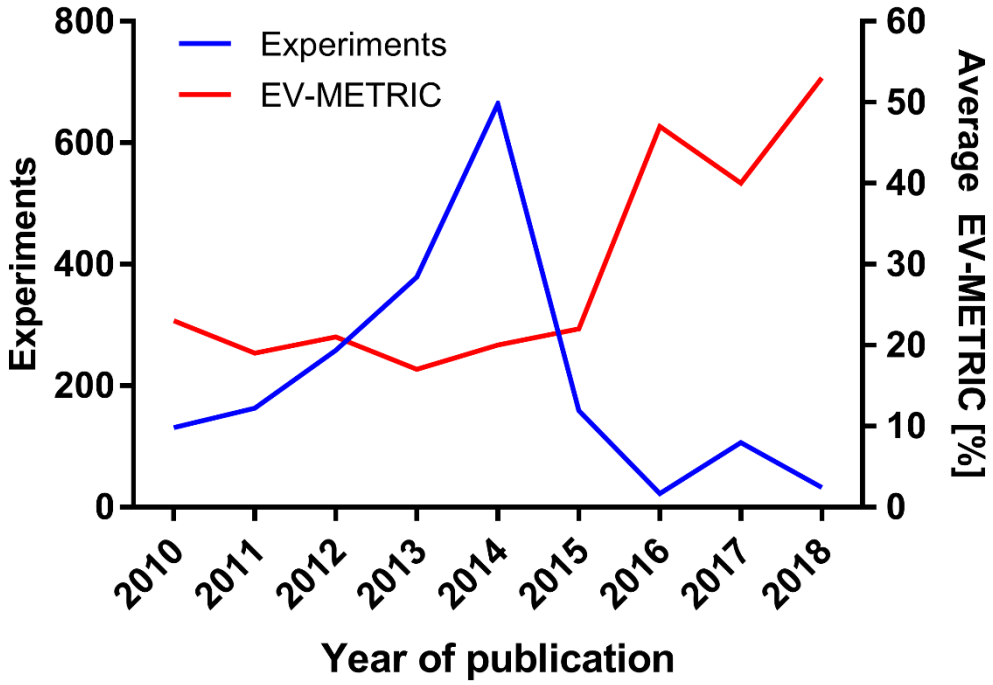

Figure 18. Number of experiments added to the EV-TRACK knowledgebase (blue, left x-axis) and average EV-METRIC (red, right x-axis) over the past nine years. While fewer articles were added after the initial release of EV-TRACK, rising EV-METRIC scores indicate improved reporting in publications. Data extracted from http://evtrack.org, accessed in January of 2019.

3.8 How to improve extracellular vesicle research: Guidelines for experiments and reporting

The initial guidelines on Minimal Information for the Studies of EVs (MISEV) were published in 2014 with the goal of sensitizing the EV community to crucial experimental and reporting requirements and providing recommendations for reliable experiments [132]. As the field evolved rapidly, the guidelines were now revisited and adapted to recent technological

advances and the increase in biological knowledge accumulated after their first release. As published in MISEV2018 [133], recommendations on EV isolation and characterization were updated and supplemented by new chapters on nomenclature and topology of analytes. Prompted by the growing recognition of many different types and subpopulations of EVs, the exosome-centric view of MISEV2014 was amended, and suggestions for markers to classify specific EV types were removed. Still, MISEV2018 provides actionable references for isolation and quantification of EVs as well as characterization of associated biomolecules and the study of EVs in functional assays. Furthermore, the updated guidelines specifically endorse submission of experimental details to EV-TRACK (discussed above in 3.7) and deposition of EV profiling data to appropriate public repositories. Major aspects of MISEV2018 were summarized in a checklist that allows researchers to quickly assess how well their experiments comply with the updated guidelines (Table 2).

Table 2. Quick-reference checklist summarizing key aspects of EV research discussed in the MISEV2018 guidelines. Table adapted from [133].

Section of MISEV2018 guidelines	Subsection
Nomenclature	Generic or specific description of EV populations
Collection and pre-processing of specimens	Tissue culture medium
	Biofluids or tissues
	Storage and recovery
EV separation and concentration	Experimental details of isolation method
EV characterization	Quantification
	Characterization of bulk EVs
	Characterization of single EVs
Functional studies	Quantitative assessment of EV-specific activity
Reporting	Data submission to EV-TRACK and relevant repositories

Results of a survey released to the EV community in 2016 revealed that the vast majority of researchers agreed on the importance of minimal requirements in experimental procedures and reporting thereof [177]. Similarly, most participants felt a need for both the continued revision of guidelines and community participation in updating these recommendations. MISEV2018 is therefore based on involvement of the ISEV community and reflects wide-ranging consensus wherever possible. As many participants perceived the guidelines in MISEV2014 to be too restrictive, MISEV2018 provides detailed explanations of mandatory and

optional recommendations as well as suggestions for experiments in which only limited adherence to the guidelines can be achieved.

Human biofluids and clinical applications are still areas of significant interest in EV science, but many groups also work on non-mammalian EVs and their role in various scientific contexts. While some of the recommendations made in MISEV2018, particularly those on protein markers, relate to EVs from specific species, the general principles can be applied to all experiments regardless of organism and research focus. In addition to these underlying minimal requirements, comprehensive guidelines for various biofluids [178] and EV-associated analytes [179] were recently published elsewhere.

While the EV-TRACK initiative (discussed above in 3.7) aims at improving reporting and building a comprehensive knowledgebase for EV studies, MISEV2018 provides updated guidelines for essential experimental parameters. In conjunction, both are intended to advance the field by improving reliability, reproducibility and our understanding of EV biology. Given that MISEV2014 was widely cited in articles from various areas of EV research and that publications citing the guidelines achieved significantly higher EV-METRIC scores [177], it is likely that the updated recommendations in MISEV2018 will have a positive impact on the field.

3.9 The promise of extracellular vesicles as clinical biomarkers: Opportunities and challenges

To assess current and future challenges for EV-based clinical biomarkers, a workshop involving researchers from within the community as well as participants with a strong background in biomarker development was held in 2017. Participants presented their research and compiled accomplishments and pitfalls in round table discussions. As published in [123], there was an overall agreement on the potential usefulness of EVs as diagnostic and prognostic biomarkers. Utilizing EVs for mutational profiling in cancer, remote detection of liver injury, and distinction between benign prostate hyperplasia and prostate cancer were but a few of the promising presented applications. Several contributions centered on miRNAs as EV-associated analytes with diagnostic value and emphasized the importance of high-throughput technologies such as small RNA-Seq (discussed above in 3.1) to establish biomarker signatures.

On the other hand, considerable challenges in moving EV-based biomarkers from academic research to a marketable product were identified (Figure 19). Special emphasis was put on the importance of well-characterized and properly handled biospecimens, which constitute the

very basis for all subsequent experiments. Unfortunately, the accessibility of suitable patient samples is often limited, and current biobanking practices might not be optimal for EV analysis. As discussed above in 3.2 and 3.7, participants also agreed that sampling, pre-analytical factors and methods of EV isolation impact results in downstream analytical assays, and that there is an urgent need for stringently standardized workflows.

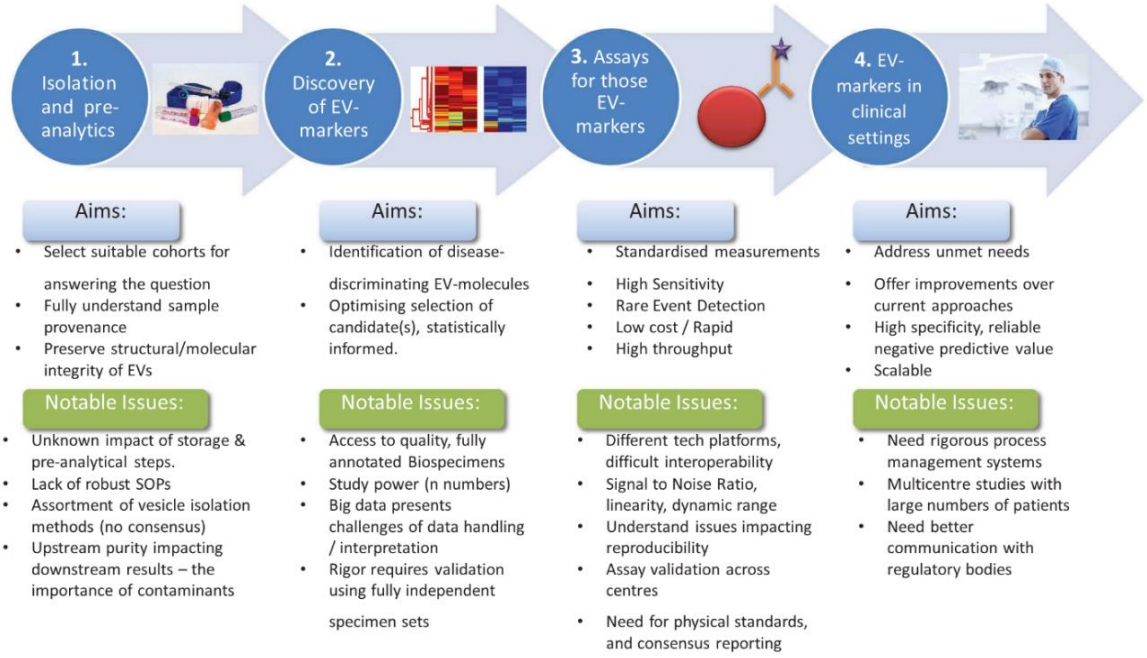

Figure 19. Topics of discussion at the ISEV workshop on EVs as disease biomarkers, which covered a wide range of aspects in the process of moving EV biomarkers from bench to bedside. Crucial goals and anticipated challenges as well as approaches to overcome them were compiled for each step. Figure reprinted from [123].

Additional challenges compiled in the workshop related to EV isolation from limited sample volumes, the absence of suitable reference standards, and the detection limits of current analytical devices. While novel technologies are developed concurrently with our evolving understanding of EV biology, participants reported having to modify established platforms according to their needs until more suitable solutions are presented. Accordingly, closer collaboration of academic researchers and biotech companies would likely accelerate the development of specialized equipment to study bulk and even single EVs.

In the same vein, collaborations between research groups, sharing of standard operating procedures (SOPs), and a common understanding of anticipated obstacles were stated to be necessary to overcome the current compartmentalization of the EV biomarker field. Taking advantage of platforms such as EV-TRACK and MISEV (discussed above in 3.7 and 3.8, respectively) was emphatically endorsed in order to increase transparency and reproducibility,

which have been a particular challenge in the biomarker field [180]. Major efforts to develop novel biomarkers, particularly in oncology, have prompted exciting academic findings, many of which could not be validated or translated into viable products. Despite initial enthusiasm, a comparably small number of markers is regularly used in medicine, and most seemingly promising candidates ended up never being approved for clinical application [181]. The reasons for this failure are manifold. A large proportion of studies on early miRNA biomarkers is underpowered and carried out using small sample sizes, increasing rates of both false positive and false negative findings and additionally inflating observed effect sizes [182]. Indeed, the effect sizes in many highly cited biomarker publications were shown to be larger than those in subsequent validation studies, which indicates that studies reporting extreme changes might receive more attention regardless of their validity [183]. Outsized effect sizes and unreliable conclusions gathered in underpowered studies hamper the selection and validation of biomarker candidates, as does a strong publication bias in this field. The tendency to selectively publish significant and exciting results at the expense of negative ones, as documented for cancer biomarkers [184] and CVD biomarkers [185], is bound to add to the overall unreliability in the early biomarker literature.

In addition to these general challenges, there are additional obstacles to be overcome for miRNA-based biomarkers. Circulating miRNAs are usually detected at low concentrations, and specialized techniques might have to be implemented for their isolation and quantification. Many demographic and lifestyle factors such as age [186], gender [187], diet [188] and activity levels [189] impact miRNA expression, resulting in high interindividual variability. This patient-to-patient variability, as well as small study populations and amply documented methodological challenges including the choice of sample type, profiling platform and appropriate normalization strategy, contribute to the heterogeneity of results from miRNA biomarker studies [190-192]. Consequently, reports of seemingly disease-specific miRNAs oftentimes failed to be reproduced even when the same disease was studied in highly similar experiments [193]. Furthermore, changes in several frequently studied circulating miRNAs were found to be associated with a number of unrelated diseases [194], prompting the notion that circulating miRNAs might be nonspecific indicators of a general pathological state rather than a reflection of altered expression in diseased tissues [195]. Compared to other transcriptomic biomarkers, disease-associated changes in miRNA expression are also generally small, and their functional relevance is hard to interpret due to the complexity of miRNA-mediated transcriptional regulation [196].

4 Conclusions

The historical development of disease diagnosis reflects a fascinating progression towards ever-increasing magnification and granularity. While early medicine was restricted to symptom-based patient evaluation, advances in biology and technology soon moved the diagnostic focus to specific organs and tissues, populations of cells and, eventually, single cells. Facilitated by the advent of precise high-throughput profiling technologies, the exploration of even smaller structures, such as EVs, is progressively taking on greater significance for establishing molecular biomarkers, which are a fundamental requirement for precision medicine.

As our understanding of EV biology increased rapidly over the past decade, so did the interest in exploiting them for biotechnological applications. EVs are now studied for their utility as vaccines, therapeutics and diagnostic targets. Several unique features, including their ubiquity in easily accessible biofluids, stability in circulation, and compositional reflection of secreting cells, moved EVs to the forefront of clinical biomarker research. Accordingly, the analysis of EVs in liquid biopsies is commonly thought of as a breakthrough technique in molecular diagnostics. As vesicular miRNAs are stable, can be easily detected and amplified, and might reflect specific disease-induced alterations in malignant cells, they are amongst the most promising analytes associated with EV-based liquid biopsies. Yet, both the study of EVs and their translation into approved clinical markers are fraught with obstacles.

The track record of molecular biomarkers is historically poor, with many promising candidates failing due to low sensitivity and specificity or irreproducibility in follow-up experiments. As these limitations became apparent, individual markers were increasingly superseded by biomarker signatures. The information gathered from a single marker is limited and might not be able to sufficiently capture the complexity of a disease. Facilitated by advances in profiling platforms and computer science, disease-related changes in multivariate data sets can be explained by the combined contribution of a set of markers. These biomarker signatures are often more robust and allow a more accurate disease detection.

For EV-based biomarkers, additional challenges are to be expected on the road from bench to bedside. Our incomplete understanding of EV biogenesis, secretion, biodistribution and uptake complicates the selection of appropriate sample types, EV populations and analytes as well as the association of disease-related molecules with EVs. As the specificity of altered miRNA expression and secretion for individual diseases seems to be low, panels of several miRNAs or combined analyses of miRNAs and other EV-associated analytes such as proteins or lipids might be more suitable to establish robust biomarker signatures. Furthermore, diagnostic

markers need to be validated in prospective studies, which are generally laborious, time-consuming, and costly [197]. Serial prospective sampling will be a particular challenge for EV-based biomarker candidates because current practices for biobanking might not be optimally suited for subsequent EV analysis, and it is not clear to which extent various pre-analytical variables and storage parameters impact results in downstream assays [198].

Nevertheless, there might be a bright future for EVs as clinical markers. As our knowledge of basic EV biology evolves, crucial issues such as sampling the most suitable biofluid to diagnose a given disease and identifying the most informative EV populations in said biofluid will be resolved. Some of the challenges for utilizing circulating miRNAs as biomarkers might be circumvented by only sampling expedient EV types or even EVs secreted by a particular tissue, once appropriately specific surface markers are identified.

This work and many others aimed at characterizing altered vesicular miRNA profiles and correlating changes to the presence of disease, or even different levels of severity. Although this seems to be a promising approach, additional classes of EV-associated biomolecules should not be neglected. Potential diagnostic utility has been ascribed to vesicular proteins [113], mRNAs [199], and lipids [200], and it is currently not clear which of these will be the most suitable analyte in different disease scenarios. Virtually all types of molecules associated with EVs, particularly RNA, DNA, proteins and metabolites, lend themselves to analysis by high-throughput technologies. As many diseases are a manifestation of complex aberrations in different layers of cellular biology, it is likely that they cannot be accurately captured by a single biomarker. Combining genomic, transcriptomic, lipidomic and metabolomic profiles to establish integrative biomarker signatures will help to detect complex diseases more sensitively and specifically [201]. Algorithm-driven data integration will also uncover relationships and synergies between single markers that were not initially obvious. Additionally, the combination of multi-omics data with clinical parameters might further improve the performance of marker signatures.

Regardless of the EV populations and associated molecules to be analyzed, standardization of pre-analytical variation and careful optimization of procedures for EV isolation and characterization are crucial to realize their potential as clinical biomarkers. If adhered to, the experimental guidelines and reporting criteria presented in MISEV2018 and EV-TRACK, respectively, will enhance transparency and reproducibility of EV studies. As biomarker development is increasingly data-driven, critical evaluation of frequently used profiling technologies such as RNA-seq [202] and RT-qPCR [203] is equally important.

At this time, the initial enthusiasm for EVs as biomarkers appears to be dampened by the growing appreciation of various experimental, analytical and regulatory obstacles. If the

necessary measures to standardize experiments, avoid biases, and validate findings are undertaken, however, the potential utility of EVs as clinical markers is immense.

5 References

1. Lee, R.C. and V. Ambros, *An extensive class of small RNAs in Caenorhabditis elegans.* Science, 2001. **294**(5543): p. 862-4.
2. Hammond, S.M., *Dicing and slicing: the core machinery of the RNA interference pathway.* FEBS Lett, 2005. **579**(26): p. 5822-9.
3. Borchert, G.M., W. Lanier, and B.L. Davidson, *RNA polymerase III transcribes human microRNAs.* Nat Struct Mol Biol, 2006. **13**(12): p. 1097-101.
4. Li, S.C., et al., *miRNA arm selection and isomiR distribution in gastric cancer.* BMC Genomics, 2012. **13 Suppl 1**: p. S13.
5. Guo, L., et al., *Selected isomiR expression profiles via arm switching?* Gene, 2014. **533**(1): p. 149-55.
6. Guo, L. and F. Chen, *A challenge for miRNA: multiple isomiRs in miRNAomics.* Gene, 2014. **544**(1): p. 1–7.
7. Cloonan, N., et al., *MicroRNAs and their isomiRs function cooperatively to target common biological pathways.* Genome biology, 2011. **12**(12): p. R126.
8. Winter, J., et al., *Many roads to maturity: microRNA biogenesis pathways and their regulation.* Nat Cell Biol, 2009. **11**(3): p. 228-34.
9. Bartel, D.P., *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. **116**(2): p. 281-97.
10. Fabian, M.R., N. Sonenberg, and W. Filipowicz, *Regulation of mRNA translation and stability by microRNAs.* Annu Rev Biochem, 2010. **79**: p. 351-79.
11. Mathonnet, G., et al., *MicroRNA inhibition of translation initiation in vitro by targeting the cap-binding complex eIF4F.* Science, 2007. **317**(5845): p. 1764-7.
12. Kiriakidou, M., et al., *An mRNA m7G cap binding-like motif within human Ago2 represses translation.* Cell, 2007. **129**(6): p. 1141-51.
13. Eulalio, A., E. Huntzinger, and E. Izaurralde, *GW182 interaction with Argonaute is essential for miRNA-mediated translational repression and mRNA decay.* Nat Struct Mol Biol, 2008. **15**(4): p. 346-53.
14. Nottrott, S., M.J. Simard, and J.D. Richter, *Human let-7a miRNA blocks protein production on actively translating polyribosomes.* Nat Struct Mol Biol, 2006. **13**(12): p. 1108-14.
15. Coller, J. and R. Parker, *Eukaryotic mRNA decapping.* Annu Rev Biochem, 2004. **73**: p. 861-90.
16. Fabian, M.R., et al., *Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation.* Mol Cell, 2009. **35**(6): p. 868-80.
17. Parker, R. and U. Sheth, *P bodies and the control of mRNA translation and degradation.* Mol Cell, 2007. **25**(5): p. 635-46.
18. Chen, S. and G. Gao, *MicroRNAs recruit eIF4E2 to repress translation of target mRNAs.* Protein Cell, 2017. **8**(10): p. 750-761.
19. Fukaya, T., H.O. Iwakawa, and Y. Tomari, *MicroRNAs block assembly of eIF4F translation initiation complex in Drosophila.* Mol Cell, 2014. **56**(1): p. 67-78.
20. Petersen, C.P., et al., *Short RNAs repress translation after initiation in mammalian cells.* Mol Cell, 2006. **21**(4): p. 533-42.
21. Chi, S.W., G.J. Hannon, and R.B. Darnell, *An alternative mode of microRNA target recognition.* Nat Struct Mol Biol, 2012. **19**(3): p. 321-7.
22. Vasudevan, S., Y. Tong, and J.A. Steitz, *Switching from repression to activation: microRNAs can up-regulate translation.* Science, 2007. **318**(5858): p. 1931-4.
23. Orom, U.A., F.C. Nielsen, and A.H. Lund, *MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation.* Mol Cell, 2008. **30**(4): p. 460-71.

24. Ma, F., et al., *MicroRNA-466l upregulates IL-10 expression in TLR-triggered macrophages by antagonizing RNA-binding protein tristetraprolin-mediated IL-10 mRNA degradation.* J Immunol, 2010. **184**(11): p. 6053-9.
25. Vasudevan, S., *Posttranscriptional upregulation by microRNAs.* Wiley Interdiscip Rev RNA, 2012. **3**(3): p. 311-30.
26. Eiring, A.M., et al., *miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts.* Cell, 2010. **140**(5): p. 652-65.
27. Mattick, J.S. and I.V. Makunin, *Small regulatory RNAs in mammals.* Hum Mol Genet, 2005. **14 Spec No 1**: p. R121-32.
28. Li, M., et al., *MicroRNAs: control and loss of control in human physiology and disease.* World J Surg, 2009. **33**(4): p. 667-84.
29. Ardekani, A.M. and M.M. Naeini, *The Role of MicroRNAs in Human Diseases.* Avicenna J Med Biotechnol, 2010. **2**(4): p. 161-79.
30. McManus, M.T., *MicroRNAs and cancer.* Semin Cancer Biol, 2003. **13**(4): p. 253-8.
31. Lewis, B.P., C.B. Burge, and D.P. Bartel, *Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets.* Cell, 2005. **120**(1): p. 15-20.
32. Bartel, D.P., *MicroRNAs: target recognition and regulatory functions.* Cell, 2009. **136**(2): p. 215-33.
33. Friedman, R.C., et al., *Most mammalian mRNAs are conserved targets of microRNAs.* Genome Res, 2009. **19**(1): p. 92-105.
34. Fabbri, M., *MicroRNAs and miRceptors: a new mechanism of action for intercellular communication.* Philos Trans R Soc Lond B Biol Sci, 2018. **373**(1737).
35. Fabbri, M., et al., *MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response.* Proc Natl Acad Sci U S A, 2012. **109**(31): p. E2110-6.
36. Ranganathan, P., et al., *Serum miR-29a Is Upregulated in Acute Graft-versus-Host Disease and Activates Dendritic Cells through TLR Binding.* J Immunol, 2017. **198**(6): p. 2500-2512.
37. Lehmann, S.M., et al., *An unconventional role for miRNA: let-7 activates Toll-like receptor 7 and causes neurodegeneration.* Nat Neurosci, 2012. **15**(6): p. 827-35.
38. Wyman, S.K., et al., *Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity.* Genome Res, 2011. **21**(9): p. 1450-61.
39. Neilsen, C.T., G.J. Goodall, and C.P. Bracken, *IsomiRs--the overlooked repertoire in the dynamic microRNAome.* Trends Genet, 2012. **28**(11): p. 544-9.
40. Kawahara, Y., et al., *Redirection of silencing targets by adenosine-to-inosine editing of miRNAs.* Science, 2007. **315**(5815): p. 1137-40.
41. Chen, K. and N. Rajewsky, *Natural selection on human microRNA binding sites inferred from SNP data.* Nat Genet, 2006. **38**(12): p. 1452-6.
42. Wu, C.W., et al., *A Comprehensive Approach to Sequence-oriented IsomiR annotation (CASMIR): demonstration with IsomiR profiling in colorectal neoplasia.* BMC Genomics, 2018. **19**(1): p. 401.
43. Burroughs, A.M., et al., *A comprehensive survey of 3' animal miRNA modification events and a possible role for 3' adenylation in modulating miRNA targeting effectiveness.* Genome Res, 2010. **20**(10): p. 1398-410.
44. Rolfo, C., et al., *Liquid Biopsy for Advanced Non-Small Cell Lung Cancer (NSCLC): A Statement Paper from the IASLC.* J Thorac Oncol, 2018. **13**(9): p. 1248-1268.
45. Lebofsky, R., et al., *Circulating tumor DNA as a non-invasive substitute to metastasis biopsy for tumor genotyping and personalized medicine in a prospective trial across all tumor types.* Mol Oncol, 2015. **9**(4): p. 783-90.
46. Lianidou, E.S., A. Strati, and A. Markou, *Circulating tumor cells as promising novel biomarkers in solid cancers.* Crit Rev Clin Lab Sci, 2014. **51**(3): p. 160-71.
47. Zhu, Z., et al., *Progress and challenges of sequencing and analyzing circulating tumor cells.* Cell Biol Toxicol, 2018. **34**(5): p. 405-415.

48. Chu, D. and B.H. Park, *Liquid biopsy: unlocking the potentials of cell-free DNA.* Virchows Arch, 2017. **471**(2): p. 147-154.
49. Choi, J.J., C.F. Reich, 3rd, and D.S. Pisetsky, *The role of macrophages in the in vitro generation of extracellular DNA from apoptotic and necrotic cells.* Immunology, 2005. **115**(1): p. 55-62.
50. Stroun, M., et al., *About the possible origin and mechanism of circulating DNA apoptosis and active DNA release.* Clin Chim Acta, 2001. **313**(1-2): p. 139-42.
51. Spindler, K.L., et al., *Cell-free DNA in healthy individuals, noncancerous disease and strong prognostic value in colorectal cancer.* Int J Cancer, 2014. **135**(12): p. 2984-91.
52. Bidard, F.C., C. Proudhon, and J.Y. Pierga, *Circulating tumor cells in breast cancer.* Mol Oncol, 2016. **10**(3): p. 418-30.
53. Hardingham, J.E., et al., *Detection and Clinical Significance of Circulating Tumor Cells in Colorectal Cancer--20 Years of Progress.* Mol Med, 2015. **21 Suppl 1**: p. S25-31.
54. Tong, B., et al., *Prognostic significance of circulating tumor cells in non-small cell lung cancer patients undergoing chemotherapy.* Oncotarget, 2017. **8**(49): p. 86615-86624.
55. Diehl, F., et al., *Circulating mutant DNA to assess tumor dynamics.* Nat Med, 2008. **14**(9): p. 985-90.
56. Neumann, M.H.D., et al., *ctDNA and CTCs in Liquid Biopsy - Current Status and Where We Need to Progress.* Comput Struct Biotechnol J, 2018. **16**: p. 190-195.
57. U.S. Food and Drug Administration. *cobas EGFR Mutation Test v2*. 2016 [cited 10.01.2019; Available from: https://www.fda.gov/drugs/informationondrugs/approveddrugs/ucm504540.htm.
58. De Mattos-Arruda, L. and C. Caldas, *Cell-free circulating tumour DNA as a liquid biopsy in breast cancer.* Mol Oncol, 2016. **10**(3): p. 464-74.
59. Angus, L., et al., *ESR1 mutations: Moving towards guiding treatment decision-making in metastatic breast cancer patients.* Cancer Treat Rev, 2017. **52**: p. 33-40.
60. Cheung, A.H., C. Chow, and K.F. To, *Latest development of liquid biopsy.* J Thorac Dis, 2018. **10**(Suppl 14): p. S1645-S1651.
61. Bergsmedh, A., et al., *Horizontal transfer of oncogenes by uptake of apoptotic bodies.* Proc Natl Acad Sci U S A, 2001. **98**(11): p. 6407-11.
62. Muralidharan-Chari, V., et al., *ARF6-regulated shedding of tumor cell-derived plasma membrane microvesicles.* Curr Biol, 2009. **19**(22): p. 1875-85.
63. Tricarico, C., J. Clancy, and C. D'Souza-Schorey, *Biology and biogenesis of shed microvesicles.* Small GTPases, 2017. **8**(4): p. 220-232.
64. Distler, J.H., et al., *Microparticles as regulators of inflammation: novel players of cellular crosstalk in the rheumatic diseases.* Arthritis Rheum, 2005. **52**(11): p. 3337-48.
65. Gasser, O. and J.A. Schifferli, *Activated polymorphonuclear neutrophils disseminate anti-inflammatory microparticles by ectocytosis.* Blood, 2004. **104**(8): p. 2543-8.
66. Timar, C.I., et al., *Antibacterial effect of microvesicles released from human neutrophilic granulocytes.* Blood, 2013. **121**(3): p. 510-8.
67. Oehmcke, S., et al., *A novel role for pro-coagulant microvesicles in the early host defense against streptococcus pyogenes.* PLoS Pathog, 2013. **9**(8): p. e1003529.
68. Jaiswal, R., et al., *Microparticles shed from multidrug resistant breast cancer cells provide a parallel survival pathway through immune evasion.* BMC Cancer, 2017. **17**(1): p. 104.
69. Muralidharan-Chari, V., et al., *Microvesicles: mediators of extracellular communication during cancer progression.* J Cell Sci, 2010. **123**(Pt 10): p. 1603-11.
70. Harding, C., J. Heuser, and P. Stahl, *Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes.* J Cell Biol, 1983. **97**(2): p. 329-39.

71. Pan, B.T. and R.M. Johnstone, *Fate of the transferrin receptor during maturation of sheep reticulocytes in vitro: selective externalization of the receptor.* Cell, 1983. **33**(3): p. 967-78.
72. Valadi, H., et al., *Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells.* Nat Cell Biol, 2007. **9**(6): p. 654-9.
73. Greening, D.W. and R.J. Simpson, *Understanding extracellular vesicle diversity - current status.* Expert Rev Proteomics, 2018. **15**(11): p. 887-910.
74. Von Bartheld, C.S. and A.L. Altick, *Multivesicular bodies in neurons: distribution, protein content, and trafficking functions.* Prog Neurobiol, 2011. **93**(3): p. 313-40.
75. Lasser, C., et al., *Human saliva, plasma and breast milk exosomes contain RNA: uptake by macrophages.* J Transl Med, 2011. **9**: p. 9.
76. Street, J.M., et al., *Identification and proteomic profiling of exosomes in human cerebrospinal fluid.* J Transl Med, 2012. **10**: p. 5.
77. Raj, D.A., et al., *A multiplex quantitative proteomics strategy for protein biomarker studies in urinary exosomes.* Kidney Int, 2012. **81**(12): p. 1263-72.
78. Kowal, J., M. Tkach, and C. Thery, *Biogenesis and secretion of exosomes.* Curr Opin Cell Biol, 2014. **29**: p. 116-25.
79. Edgar, J.R., E.R. Eden, and C.E. Futter, *Hrs- and CD63-dependent competing mechanisms make different sized endosomal intraluminal vesicles.* Traffic, 2014. **15**(2): p. 197-211.
80. Tkach, M., J. Kowal, and C. Thery, *Why the need and how to approach the functional diversity of extracellular vesicles.* Philos Trans R Soc Lond B Biol Sci, 2018. **373**(1737).
81. Ban, J.J., et al., *Low pH increases the yield of exosome isolation.* Biochem Biophys Res Commun, 2015. **461**(1): p. 76-9.
82. King, H.W., M.Z. Michael, and J.M. Gleadle, *Hypoxic enhancement of exosome release by breast cancer cells.* BMC Cancer, 2012. **12**: p. 421.
83. Liao, Y., et al., *Human milk exosomes and their microRNAs survive digestion in vitro and are taken up by human intestinal cells.* Mol Nutr Food Res, 2017. **61**(11).
84. Kourembanas, S., *Exosomes: vehicles of intercellular signaling, biomarkers, and vectors of cell therapy.* Annu Rev Physiol, 2015. **77**: p. 13-27.
85. Guduric-Fuchs, J., et al., *Selective extracellular vesicle-mediated export of an overlapping set of microRNAs from multiple cell types.* BMC Genomics, 2012. **13**: p. 357.
86. Kucharzewska, P., et al., *Exosomes reflect the hypoxic status of glioma cells and mediate hypoxia-dependent activation of vascular cells during tumor development.* Proc Natl Acad Sci U S A, 2013. **110**(18): p. 7312-7.
87. de Jong, O.G., et al., *Cellular stress conditions are reflected in the protein and RNA content of endothelial cell-derived exosomes.* J Extracell Vesicles, 2012. **1**.
88. Lobb, R.J., et al., *Oncogenic transformation of lung cells results in distinct exosome protein profile similar to the cell of origin.* Proteomics, 2017. **17**(23-24).
89. Batagov, A.O., V.A. Kuznetsov, and I.V. Kurochkin, *Identification of nucleotide patterns enriched in secreted RNAs as putative cis-acting elements targeting them to exosome nano-vesicles.* BMC Genomics, 2011. **12 Suppl 3**: p. S18.
90. Villarroya-Beltri, C., et al., *Sumoylated hnRNPA2B1 controls the sorting of miRNAs into exosomes through binding to specific motifs.* Nat Commun, 2013. **4**: p. 2980.
91. Janas, T., et al., *Mechanisms of RNA loading into exosomes.* FEBS Lett, 2015. **589**(13): p. 1391-8.
92. Meldolesi, J., *Exosomes and Ectosomes in Intercellular Communication.* Curr Biol, 2018. **28**(8): p. R435-R444.
93. Prada, I. and J. Meldolesi, *Binding and Fusion of Extracellular Vesicles to the Plasma Membrane of Their Cell Targets.* Int J Mol Sci, 2016. **17**(8).

94. Costa Verdera, H., et al., *Cellular uptake of extracellular vesicles is mediated by clathrin-independent endocytosis and macropinocytosis.* J Control Release, 2017. **266**: p. 100-108.
95. Tian, T., et al., *Exosome uptake through clathrin-mediated endocytosis and macropinocytosis and mediating miR-21 delivery.* J Biol Chem, 2014. **289**(32): p. 22258-67.
96. Xu, R., et al., *Extracellular vesicles in cancer - implications for future improvements in cancer care.* Nat Rev Clin Oncol, 2018. **15**(10): p. 617-638.
97. Horibe, S., et al., *Mechanism of recipient cell-dependent differences in exosome uptake.* BMC Cancer, 2018. **18**(1): p. 47.
98. Subra, C., et al., *Exosomes account for vesicle-mediated transcellular transport of activatable phospholipases and prostaglandins.* J Lipid Res, 2010. **51**(8): p. 2105-20.
99. Mathivanan, S., H. Ji, and R.J. Simpson, *Exosomes: extracellular organelles important in intercellular communication.* J Proteomics, 2010. **73**(10): p. 1907-20.
100. Mulcahy, L.A., R.C. Pink, and D.R. Carter, *Routes and mechanisms of extracellular vesicle uptake.* J Extracell Vesicles, 2014. **3**.
101. Stenqvist, A.C., et al., *Exosomes secreted by human placenta carry functional Fas ligand and TRAIL molecules and convey apoptosis in activated immune cells, suggesting exosome-mediated immune privilege of the fetus.* J Immunol, 2013. **191**(11): p. 5515-23.
102. Arenaccio, C. and M. Federico, *The Multifaceted Functions of Exosomes in Health and Disease: An Overview.* Adv Exp Med Biol, 2017. **998**: p. 3-19.
103. Thery, C., M. Ostrowski, and E. Segura, *Membrane vesicles as conveyors of immune responses.* Nat Rev Immunol, 2009. **9**(8): p. 581-93.
104. McDonald, M.K., et al., *Functional significance of macrophage-derived exosomes in inflammation and pain.* Pain, 2014. **155**(8): p. 1527-39.
105. Lakkaraju, A. and E. Rodriguez-Boulan, *Itinerant exosomes: emerging roles in cell and tissue polarity.* Trends Cell Biol, 2008. **18**(5): p. 199-209.
106. Becker, A., et al., *Extracellular Vesicles in Cancer: Cell-to-Cell Mediators of Metastasis.* Cancer Cell, 2016. **30**(6): p. 836-848.
107. Sardar Sinha, M., et al., *Alzheimer's disease pathology propagation by exosomes containing toxic amyloid-beta oligomers.* Acta Neuropathol, 2018. **136**(1): p. 41-56.
108. Wu, X., T. Zheng, and B. Zhang, *Exosomes in Parkinson's Disease.* Neurosci Bull, 2017. **33**(3): p. 331-338.
109. Kim, A., A.S. Shah, and T. Nakamura, *Extracellular Vesicles: A Potential Novel Regulator of Obesity and Its Associated Complications.* Children (Basel), 2018. **5**(11).
110. Zamani, P., et al., *The therapeutic and diagnostic role of exosomes in cardiovascular diseases.* Trends Cardiovasc Med, 2018.
111. Zhang, W., et al., *Exosomes in Pathogen Infections: A Bridge to Deliver Molecules and Link Functions.* Front Immunol, 2018. **9**: p. 90.
112. Thakur, B.K., et al., *Double-stranded DNA in exosomes: a novel biomarker in cancer detection.* Cell Res, 2014. **24**(6): p. 766-9.
113. Melo, S.A., et al., *Glypican-1 identifies cancer exosomes and detects early pancreatic cancer.* Nature, 2015. **523**(7559): p. 177-82.
114. Madhavan, B., et al., *Combined evaluation of a panel of protein and miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and specificity.* Int J Cancer, 2015. **136**(11): p. 2616-27.
115. San Lucas, F.A., et al., *Minimally invasive genomic and transcriptomic profiling of visceral cancers by next-generation sequencing of circulating exosomes.* Ann Oncol, 2016. **27**(4): p. 635-41.
116. Krug, A.K., et al., *Improved EGFR mutation detection using combined exosomal RNA and circulating tumor DNA in NSCLC patient plasma.* Ann Oncol, 2018. **29**(3): p. 700-706.

117. Sharma, P., et al., *Immunoaffinity-based isolation of melanoma cell-derived exosomes from plasma of patients with melanoma.* J Extracell Vesicles, 2018. **7**(1): p. 1435138.
118. Taylor, D.D. and C. Gercel-Taylor, *MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer.* Gynecol Oncol, 2008. **110**(1): p. 13-21.
119. Muller, L., et al., *Exosomes isolated from plasma of glioma patients enrolled in a vaccination trial reflect antitumor immune activity and might predict survival.* Oncoimmunology, 2015. **4**(6): p. e1008347.
120. Vagner, T., et al., *Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma.* J Extracell Vesicles, 2018. **7**(1): p. 1505403.
121. Whiteside, T.L., *The potential of tumor-derived exosomes for noninvasive cancer monitoring.* Expert Rev Mol Diagn, 2015. **15**(10): p. 1293-310.
122. Kern, S.E., *Why your new cancer biomarker may never work: recurrent patterns and remarkable diversity in biomarker failures.* Cancer Res, 2012. **72**(23): p. 6097-101.
123. Clayton, A., et al., *Summary of the ISEV workshop on extracellular vesicles as disease biomarkers, held in Birmingham, UK, during December 2017.* Journal of Extracellular Vesicles, 2018. **7**(1): p. 1473707.
124. 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, 2016. **44**(13): p. 5995-6018.
125. Buschmann, D., et al., *Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing.* J Extracell Vesicles, 2018. **7**(1): p. 1481321.
126. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
127. Hermann, S., et al., *Transcriptomic profiling of miRNAs in extracellular vesicles from matched arterial and venous serum.* Journal of Extracellular Vesicles, under review, 2018.
128. Buschmann, D., et al., *Glucocorticoid receptor overexpression slightly shifts microRNA expression patterns in triple-negative breast cancer.* Int J Oncol, 2018. **52**(6): p. 1765-1776.
129. Buschmann, D., et al. *Can vesicular microRNAs predict negative perioperative outcomes in cardiac surgery?* in *Poster presented at The 7th Annual Meeting of the International Society for Extracellular Vesicles*. 2018. Barcelona.
130. Reithmair, M., et al., *Cellular and extracellular miRNAs are blood-compartment-specific diagnostic targets in sepsis.* J Cell Mol Med, 2017. **21**(10): p. 2403-2411.
131. EV-Track Consortium, et al., *EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research.* Nat Methods, 2017. **14**(3): p. 228-232.
132. Lotvall, J., et al., *Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles.* J Extracell Vesicles, 2014. **3**: p. 26913.
133. Théry, C., et al., *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, 2018. **7**(1): p. 1535750.
134. Coenen-Stass, A.M.L., et al., *Evaluation of methodologies for microRNA biomarker detection by next generation sequencing.* RNA Biol, 2018. **15**(8): p. 1133-1145.
135. Wake, C., et al., *Novel microRNA discovery using small RNA sequencing in post-mortem human brain.* BMC Genomics, 2016. **17**(1): p. 776.
136. Huang, S., et al., *microRNA biomarkers in colorectal cancer liver metastasis.* J Cancer, 2018. **9**(21): p. 3867-3873.
137. Zhou, S.S., et al., *miRNAS in cardiovascular diseases: potential biomarkers, therapeutic targets and challenges.* Acta Pharmacol Sin, 2018. **39**(7): p. 1073-1084.

138. Viswambharan, V., et al., *miRNAs as biomarkers of neurodegenerative disorders.* Biomark Med, 2017. **11**(2): p. 151-167.
139. Alon, S., et al., *Barcoding bias in high-throughput multiplex sequencing of miRNA.* Genome Res, 2011. **21**(9): p. 1506-11.
140. Head, S.R., et al., *Library construction for next-generation sequencing: overviews and challenges.* Biotechniques, 2014. **56**(2): p. 61-4, 66, 68, passim.
141. Locati, M.D., et al., *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, 2015. **43**(14): p. e89.
142. Bustin, S.A., *The reproducibility of biomedical research: Sleepers awake!* Biomol Detect Quantif, 2014. **2**: p. 35-42.
143. Baker, M., *1,500 scientists lift the lid on reproducibility.* Nature, 2016. **533**(7604): p. 452-4.
144. Bustin, S.A., et al., *The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments.* Clin Chem, 2009. **55**(4): p. 611-22.
145. Bustin, S.A., et al., *The need for transparency and good practices in the qPCR literature.* Nat Methods, 2013. **10**(11): p. 1063-7.
146. Lobb, R.J., et al., *Optimized exosome isolation protocol for cell culture supernatant and human plasma.* J Extracell Vesicles, 2015. **4**: p. 27031.
147. Caradec, J., et al., *Reproducibility and efficiency of serum-derived exosome extraction methods.* Clin Biochem, 2014. **47**(13-14): p. 1286-92.
148. Arroyo, J.D., et al., *Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma.* Proc Natl Acad Sci U S A, 2011. **108**(12): p. 5003-8.
149. Karttunen, J., et al., *Precipitation-based extracellular vesicle isolation from rat plasma co-precipitate vesicle-free microRNAs.* J Extracell Vesicles, 2019. **8**(1): p. 1555410.
150. Alvarez, M.L., et al., *Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers.* Kidney Int, 2012. **82**(9): p. 1024-32.
151. Quek, C., et al., *Defining the purity of exosomes required for diagnostic profiling of small RNA suitable for biomarker discovery.* RNA Biol, 2017. **14**(2): p. 245-258.
152. Xu, W., et al., *Transcriptome analysis reveals non-identical microRNA profiles between arterial and venous plasma.* Oncotarget, 2017. **8**(17): p. 28471-28480.
153. Eichelser, C., et al., *Increased serum levels of circulating exosomal microRNA-373 in receptor-negative breast cancer patients.* Oncotarget, 2014. **5**(20): p. 9650-63.
154. Ogata-Kawata, H., et al., *Circulating exosomal microRNAs as biomarkers of colon cancer.* PLoS One, 2014. **9**(4): p. e92921.
155. Bryant, R.J., et al., *Changes in circulating microRNA levels associated with prostate cancer.* Br J Cancer, 2012. **106**(4): p. 768-74.
156. Liu, Q., et al., *Circulating exosomal microRNAs as prognostic biomarkers for non-small-cell lung cancer.* Oncotarget, 2017. **8**(8): p. 13048-13058.
157. Liu, C., et al., *Serum exosomal miR-4772-3p is a predictor of tumor recurrence in stage II and III colon cancer.* Oncotarget, 2016. **7**(46): p. 76250-76260.
158. Kahlert, C., et al., *Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer.* J Biol Chem, 2014. **289**(7): p. 3869-75.
159. Clayton, S.A., et al., *The role of microRNAs in glucocorticoid action.* J Biol Chem, 2018. **293**(6): p. 1865-1874.
160. Tessel, M.A., et al., *Role for microRNAs in regulating glucocorticoid response and resistance in multiple myeloma.* Horm Cancer, 2011. **2**(3): p. 182-9.
161. Wang, H., et al., *The effects of microRNAs on glucocorticoid responsiveness.* J Cancer Res Clin Oncol, 2017. **143**(6): p. 1005-1011.
162. Henes, J. and P. Rosenberger, *Systolic heart failure: diagnosis and therapy.* Curr Opin Anaesthesiol, 2016. **29**(1): p. 55-60.

163. Palant, C.E., R.L. Amdur, and L.S. Chawla, *Long-term consequences of acute kidney injury in the perioperative setting.* Curr Opin Anaesthesiol, 2017. **30**(1): p. 100-104.
164. Cornelissen, H. and J.E. Arrowsmith, *Preoperative assessment for cardiac surgery.* Continuing Education in Anaesthesia Critical Care & Pain, 2006. **6**(3): p. 109-113.
165. Clerico, A., M. Emdin, and C. Passino, *Cardiac biomarkers and risk assessment in patients undergoing major non-cardiac surgery: time to revise the guidelines?* Clin Chem Lab Med, 2014. **52**(7): p. 959-63.
166. Attaran, S., et al., *Brain natriuretic peptide a predictive marker in cardiac surgery.* Interact Cardiovasc Thorac Surg, 2009. **9**(4): p. 662-6.
167. Paoli, C.J., et al., *Epidemiology and Costs of Sepsis in the United States-An Analysis Based on Timing of Diagnosis and Severity Level.* Crit Care Med, 2018. **46**(12): p. 1889-1897.
168. Martin, G.S., *Sepsis, severe sepsis and septic shock: changes in incidence, pathogens and outcomes.* Expert Rev Anti Infect Ther, 2012. **10**(6): p. 701-6.
169. Whiles, B.B., A.S. Deis, and S.Q. Simpson, *Increased Time to Initial Antimicrobial Administration Is Associated With Progression to Septic Shock in Severe Sepsis Patients.* Crit Care Med, 2017. **45**(4): p. 623-629.
170. Liu, V.X., et al., *The Timing of Early Antibiotics and Hospital Mortality in Sepsis.* Am J Respir Crit Care Med, 2017. **196**(7): p. 856-863.
171. Terrasini, N. and V. Lionetti, *Exosomes in Critical Illness.* Crit Care Med, 2017. **45**(6): p. 1054-1060.
172. Wu, J., Y. Wang, and L. Li, *Functional significance of exosomes applied in sepsis: A novel approach to therapy.* Biochim Biophys Acta Mol Basis Dis, 2017. **1863**(1): p. 292-297.
173. Goodwin, A.J., et al., *Plasma levels of microRNA are altered with the development of shock in human sepsis: an observational study.* Crit Care, 2015. **19**: p. 440.
174. Real, J.M., et al., *Exosomes from patients with septic shock convey miRNAs related to inflammation and cell cycle regulation: new signaling pathways in sepsis?* Crit Care, 2018. **22**(1): p. 68.
175. Xie, M., et al., *MicroRNA-30a regulates cell proliferation and tumor growth of colorectal cancer by targeting CD73.* BMC Cancer, 2017. **17**(1): p. 305.
176. Fan, Y.X., et al., *MicroRNA-125b inhibits cell proliferation and induces cell apoptosis in esophageal squamous cell carcinoma by targeting BMF.* Oncol Rep, 2018. **40**(1): p. 61-72.
177. Witwer, K.W., et al., *Updating the MISEV minimal requirements for extracellular vesicle studies: building bridges to reproducibility.* J Extracell Vesicles, 2017. **6**(1): p. 1396823.
178. Witwer, K.W., et al., *Standardization of sample collection, isolation and analysis methods in extracellular vesicle research.* J Extracell Vesicles, 2013. **2**.
179. Mateescu, B., et al., *Obstacles and opportunities in the functional analysis of extracellular vesicle RNA - an ISEV position paper.* J Extracell Vesicles, 2017. **6**(1): p. 1286095.
180. Scherer, A., *Reproducibility in biomarker research and clinical development: a global challenge.* Biomark Med, 2017. **11**(4): p. 309-312.
181. Diamandis, E.P., *Cancer biomarkers: can we turn recent failures into success?* J Natl Cancer Inst, 2010. **102**(19): p. 1462-7.
182. Kok, M.G.M., et al., *Small sample sizes in high-throughput miRNA screens: A common pitfall for the identification of miRNA biomarkers.* Biomol Detect Quantif, 2018. **15**: p. 1-5.
183. Ioannidis, J.P. and O.A. Panagiotou, *Comparison of effect sizes associated with biomarkers reported in highly cited individual articles and in subsequent meta-analyses.* JAMA, 2011. **305**(21): p. 2200-10.

184. Kyzas, P.A., D. Denaxa-Kyza, and J.P. Ioannidis, *Almost all articles on cancer prognostic markers report statistically significant results.* Eur J Cancer, 2007. **43**(17): p. 2559-79.
185. Tzoulaki, I., et al., *Bias in associations of emerging biomarkers with cardiovascular disease.* JAMA Intern Med, 2013. **173**(8): p. 664-71.
186. Huan, T., et al., *Age-associated microRNA expression in human peripheral blood is associated with all-cause mortality and age-related traits.* Aging Cell, 2018. **17**(1).
187. Guo, L., et al., *miRNA and mRNA expression analysis reveals potential sex-biased miRNA expression.* Sci Rep, 2017. **7**: p. 39812.
188. Ross, S.A. and C.D. Davis, *MicroRNA, nutrition, and cancer prevention.* Adv Nutr, 2011. **2**(6): p. 472-85.
189. Dufresne, S., et al., *A Review of Physical Activity and Circulating miRNA Expression: Implications in Cancer Risk and Progression.* Cancer Epidemiol Biomarkers Prev, 2018. **27**(1): p. 11-24.
190. Moldovan, L., et al., *Methodological challenges in utilizing miRNAs as circulating biomarkers.* J Cell Mol Med, 2014. **18**(3): p. 371-90.
191. Kirschner, M.B., N. van Zandwijk, and G. Reid, *Cell-free microRNAs: potential biomarkers in need of standardized reporting.* Front Genet, 2013. **4**: p. 56.
192. Pritchard, C.C., H.H. Cheng, and M. Tewari, *MicroRNA profiling: approaches and considerations.* Nat Rev Genet, 2012. **13**(5): p. 358-69.
193. Leidner, R.S., L. Li, and C.L. Thompson, *Dampening enthusiasm for circulating microRNA in breast cancer.* PLoS One, 2013. **8**(3): p. e57841.
194. Haider, B.A., et al., *A critical evaluation of microRNA biomarkers in non-neoplastic disease.* PLoS One, 2014. **9**(2): p. e89565.
195. Witwer, K.W., *Circulating microRNA biomarker studies: pitfalls and potential solutions.* Clin Chem, 2015. **61**(1): p. 56-63.
196. Witwer, K.W. and M.K. Halushka, *Toward the promise of microRNAs - Enhancing reproducibility and rigor in microRNA research.* RNA Biol, 2016. **13**(11): p. 1103-1116.
197. Pesch, B., et al., *Biomarker research with prospective study designs for the early detection of cancer.* Biochim Biophys Acta, 2014. **1844**(5): p. 874-83.
198. Mora, E.M., S. Alvarez-Cubela, and E. Oltra, *Biobanking of Exosomes in the Era of Precision Medicine: Are We There Yet?* Int J Mol Sci, 2015. **17**(1).
199. Skog, J., et al., *Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers.* Nat Cell Biol, 2008. **10**(12): p. 1470-6.
200. Skotland, T., et al., *Molecular lipid species in urinary exosomes as potential prostate cancer biomarkers.* Eur J Cancer, 2017. **70**: p. 122-132.
201. Gallo Cantafio, M.E., et al., *From Single Level Analysis to Multi-Omics Integrative Approaches: A Powerful Strategy towards the Precision Oncology.* High Throughput, 2018. **7**(4).
202. Xu, J., et al., *Comprehensive Assessments of RNA-seq by the SEQC Consortium: FDA-Led Efforts Advance Precision Medicine.* Pharmaceutics, 2016. **8**(1).
203. Mestdagh, P., et al., *Evaluation of quantitative miRNA expression platforms in the microRNA quality control (miRQC) study.* Nat Methods, 2014. **11**(8): p. 809-15.

Acknowledgements

I would like to extend my gratitude to everybody who supported me during my time as a PhD student at the Division of Animal Physiology and Immunology and the Institute of Human Genetics.

First of all, I would like to thank Prof. Dr. Michael Pfaffl for his excellent supervision, his scientific input, his people management skills, and for always supporting me on this journey. The combination of having ample scientific freedom and receiving guidance whenever needed resulted in a productive and enjoyable work environment. I am very thankful for everything you did and continue to do for us and will always cherish my years at the institute.

I would also like to thank Prof. Dr. Dietmar Zehn for providing me with the opportunity to carry out my research at the Division of Animal Physiology and Immunology and for his sovereign leadership in times of change.

I am grateful for having had the opportunity to work with Prof. Dr. Ortrud Steinlein and Dr. Marlene Reithmair at the Institute of Human Genetics. Our continued collaboration enriched my scientific development and created many interesting projects.

Thank you to Prof. Dr. Gustav Schelling for our excellent collaboration and for kindly agreeing to serve on my thesis committee. You always came up with fascinating ideas and supplemented my academic work with real-life clinical insights, for which I am thankful.

To my EV colleagues at the institute: I appreciate what a great team we have become and would like to thank you for your support and for the fun times we spent in the lab and on our trips. Special thanks to Benedikt Kirchner, with whom I had many fruitful discussions and from whom I learned a lot over the years.

Thank you to all colleagues and staff at the Division of Animal Physiology and Immunology. The atmosphere you created made it easy to always enjoy my work, be it in the lab or in the office, and I do not take this for granted.

I would also like to thank Dr. Melanie Spornraft and the GZW team for their advice and swift help with all organizational and formal matters.

Last but not least, I am very grateful to my friends and family for your encouragement, for sticking with me on this journey and for always being there when I needed you. Thank you.

List of scientific communications

Peer-reviewed publications

Dominik Buschmann, Anna Haberberger, Benedikt Kirchner, Melanie Spornraft, Irmgard Riedmaier, Gustav Schelling, Michael W. Pfaffl

Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow

Nucleic Acids Research, July 27, 2016, DOI: 10.1093/nar/gkw545

EV-TRACK Consortium

EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research

Nature Methods, February 28, 2017, DOI: 10.1038/nmeth.4185

Marlene Reithmair, Dominik Buschmann, Melanie Märte, Benedikt Kirchner, Daniel Hagl, Ines Kaufmann, Martina Pfob, Alexander Chouker, Ortrud K. Steinlein, Michael W. Pfaffl, Gustav Schelling

Cellular and extracellular miRNAs are blood-compartment-specific diagnostic targets in sepsis

Journal of Cellular and Molecular Medicine, April 6, 2017, DOI: 10.1111/jcmm.13162

Dominik Buschmann, Ricardo González, Benedikt Kirchner, Claudia Mazzone, Michael W. Pfaffl, Gustav Schelling, Ortrud Steinlein, Marlene Reithmair

Glucocorticoid receptor overexpression slightly shifts microRNA expression patterns in triple-negative breast cancer

International Journal of Oncology, March 27, 2018, DOI: 10.3892/ijo.2018.4336

Aled Clayton, Dominik Buschmann, J. Brian Byrd, David R. F. Carter, Lesley Cheng, Carolyn Compton, George Daaboul, Andrew Devitt, Juan Manuel Falcon-Perez, Chris Gardiner, Dakota Gustafson, Paul Harrison, Clemens Helmbrecht, An Hendrix, Andrew Hill, Andrew Hoffman, Jennifer C. Jones, Raghu Kalluri, Ji Yoon Kang, Benedikt Kirchner, Cecilia Lässer, Charlotte Lawson, Metka Lenassi, Carina Levin, Alicia Llorente, Elena S. Martens-Uzunova, Andreas Möller, Luca Musante, Takahiro Ochiya, Ryan C Pink, Hidetoshi Tahara, Marca H. M. Wauben, Jason P. Webber, Joshua A. Welsh, Kenneth W. Witwer, Hang Yin, Rienk Nieuwland

Summary of the ISEV workshop on extracellular vesicles as disease biomarkers, held in Birmingham, UK, during December 2017

Journal of Extracellular Vesicles, May 17, 2018, DOI: 10.1080/20013078.2018.1473707

Dominik Buschmann, Benedikt Kirchner, Stefanie Hermann, Melanie Märte, Christine Wurmser, Florian Brandes, Stefan Kotschote, Michael Bonin, Ortrud K. Steinlein, Michael W. Pfaffl, Gustav Schelling, Marlene Reithmair

Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing

Journal of Extracellular Vesicles, June 4, 2018, DOI: 10.1080/20013078.2018.1481321

Clotilde Théry, Kenneth W Witwer et al.

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, November 23, 2018, DOI: 10.1080/20013078.2018.1535750

Stefanie Hermann, Dominik Buschmann, Benedikt Kirchner, Melanie Märte, Florian Brandes, Stefan Kotschote, Michael Bonin, Marlene Reithmair, Gustav Schelling, Michael W. Pfaffl

Transcriptomic profiling of miRNAs in extracellular vesicles from matched arterial and venous serum

Journal of Extracellular Vesicles, under review

Poster presentations

Dominik Buschmann, Marlene Eggert, Melanie Märte, Martina Pfob, Alexander Chouker, Michael W. Pfaffl, Gustav Schelling

The volatile anesthetic agent sevoflurane influences cancer-related miRNAs in circulating extracellular nanovesicles

The 5th Annual ISEV Meeting, May 2016, Rotterdam, The Netherlands

Benedikt Kirchner, Dominik Buschmann, Vijay Paul, Michael W. Pfaffl

Exosomal isoforms of microRNAs as a novel group of tissue-specific biomarkers

The 5th Annual ISEV Meeting, May 2016, Rotterdam, The Netherlands

Marlene Reithmair, Dominik Buschmann, Melanie Märte, Benedikt Kirchner, Daniel Hagl, Ines Kaufmann, Alexander Chouker, Ortrud Steinlein, Michael W. Pfaffl, Gustav Schelling

Characterization of exosomal miRNA profiles in patients with sepsis and septic shock

The 6th Annual ISEV Meeting, May 2017, Toronto, Canada

Benedikt Kirchner, Dominik Buschmann, Stefan Kotschote, Michael Bonin, Marlene Reithmair, Gustav Schelling, Michael W. Pfaffl

Influence of commercially available exosomal isolation kits on holistic small RNA expression profiles of serum in healthy and critically ill individuals

The 6th Annual ISEV Meeting, May 2017, Toronto, Canada

Dominik Buschmann, Marlene Reithmair, Benedikt Kirchner, Stefanie Hermann, Melanie Märte, Florian Brandes, Ortrud Steinlein, Michael W. Pfaffl, Gustav Schelling

Can vesicular microRNAs predict negative perioperative outcomes in cardiac surgery?

The 7th Annual ISEV Meeting, May 2018, Barcelona, Spain

Dominik Buschmann, Benedikt Kirchner, Stefanie Hermann, Stefan Kotschote, Michael Bonin, Marlene Reithmair, Gustav Schelling, Michael W. Pfaffl

Quantitative comparison of EV isolation methods for miRNA-based biomarker detection in critically ill patients

The 2018 Gordon Research Conference on Diagnostic and Therapeutic Applications of Extracellular Vesicles, August 2018, Newry, ME, USA

Book chapters

Dominik Buschmann, Benedikt Kirchner, Veronika Mussack, Michael W. Pfaffl

Nucleic Acids: RNA Identification and Quantification Via Next-Generation Sequencing

Encyclopedia of Analytical Science (Third Edition), Elsevier, 2019 (accepted)

Michael W. Pfaffl, Dominik Buschmann, Veronika Mussack, Benedikt Kirchner, Bajram Berisha

Nucleic Acids: RNA Identification and Quantification via RT-qPCR

Encyclopedia of Analytical Science (Third Edition), Elsevier, 2019 (accepted)

Veronika Mussack, Stefanie Hermann, Dominik Buschmann, Benedikt Kirchner, Michael W. Pfaffl

MIQE-compliant validation of microRNA biomarker signatures established by small RNA sequencing

Quantitative Real-Time PCR (Second Edition), Methods in Molecular Biology, Springer Nature, 2019 (accepted)

Oral presentations

Dominik Buschmann

Liquid biopsies, biomarker signatures, and beyond - why standardization of small RNA-Seq matters

The 8th international Gene Quantification Event, April 2017, Freising, Germany

Appendix

Appendix I:

Dominik Buschmann, Anna Haberberger, Benedikt Kirchner, Melanie Spornraft, Irmgard Riedmaier, Gustav Schelling, Michael W. Pfaffl

Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow

Nucleic Acids Research, July 27, 2016, DOI: 10.1093/nar/gkw545

Appendix II:

Dominik Buschmann, Benedikt Kirchner, Stefanie Hermann, Melanie Märte, Christine Wurmser, Florian Brandes, Stefan Kotschote, Michael Bonin, Ortrud K. Steinlein, Michael W. Pfaffl, Gustav Schelling, Marlene Reithmair

Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing

Journal of Extracellular Vesicles, June 4, 2018, DOI: 10.1080/20013078.2018.1481321

Appendix III:

Dominik Buschmann, Ricardo González, Benedikt Kirchner, Claudia Mazzone, Michael W. Pfaffl, Gustav Schelling, Ortrud Steinlein, Marlene Reithmair

Glucocorticoid receptor overexpression slightly shifts microRNA expression patterns in triple-negative breast cancer

International Journal of Oncology, March 27, 2018, DOI: 10.3892/ijo.2018.4336

Appendix IV:

Marlene Reithmair, Dominik Buschmann, Melanie Märte, Benedikt Kirchner, Daniel Hagl, Ines Kaufmann, Martina Pfob, Alexander Chouker, Ortrud K. Steinlein, Michael W. Pfaffl, Gustav Schelling

Cellular and extracellular miRNAs are blood-compartment-specific diagnostic targets in sepsis

Journal of Cellular and Molecular Medicine, April 6, 2017, DOI: 10.1111/jcmm.13162

Appendix I

Substantial contributions by Dominik Buschmann:

- Conception of the manuscript
- Acquisition and analysis of data
- Drafting of figures
- Writing of the manuscript

Dominik Buschmann

Michael W. Pfaffl

Nucleic Acids Research Advance Access published June 17, 2016

doi: 10.1093/nar/gkw545

SURVEY AND SUMMARY

Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow

Dominik Buschmann[1,2], Anna Haberberger[1], Benedikt Kirchner[1], Melanie Spornraft[1], Irmgard Riedmaier[3,4], Gustav Schelling[5] and Michael W. Pfaffl[1,*]

[1]Department of Animal Physiology and Immunology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Weihenstephaner Berg 3, 85354 Freising, Germany, [2]Institute of Human Genetics, University Hospital, Ludwig-Maximilians-University Munich, Goethestraße 29, 80336 München, Germany, [3]Department of Physiology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Weihenstephaner Berg 3, 85354 Freising, Germany, [4]Eurofins Medigenomix Forensik GmbH, Anzinger Straße 7a, 85560 Ebersberg, Germany and [5]Department of Anesthesiology, University Hospital, Ludwig-Maximilians-University Munich, Marchioninistraße 15, 81377 München, Germany

Received April 12, 2016; Revised May 31, 2016; Accepted June 3, 2016

ABSTRACT

Small RNA-Seq has emerged as a powerful tool in transcriptomics, gene expression profiling and biomarker discovery. Sequencing cell-free nucleic acids, particularly microRNA (miRNA), from liquid biopsies additionally provides exciting possibilities for molecular diagnostics, and might help establish disease-specific biomarker signatures. The complexity of the small RNA-Seq workflow, however, bears challenges and biases that researchers need to be aware of in order to generate high-quality data. Rigorous standardization and extensive validation are required to guarantee reliability, reproducibility and comparability of research findings. Hypotheses based on flawed experimental conditions can be inconsistent and even misleading. Comparable to the well-established MIQE guidelines for qPCR experiments, this work aims at establishing guidelines for experimental design and pre-analytical sample processing, standardization of library preparation and sequencing reactions, as well as facilitating data analysis. We highlight bottlenecks in small RNA-Seq experiments, point out the importance of stringent quality control and validation, and provide a primer for differential expression analysis and biomarker discovery. Following our recommendations will encourage better sequencing practice, increase experimental transparency and lead to more reproducible small RNA-Seq results. This will ultimately enhance the validity of biomarker signatures, and allow reliable and robust clinical predictions.

INTRODUCTION TO BIOMARKERS AND LIQUID BIOPSIES

The importance of biomarkers in molecular diagnostics is undisputed. A valid biomarker should be able to reveal a specific biological trait or a measurable change, which is directly associated with a change in the physiological condition of an organism. At the molecular and cellular levels, analysis of gene expression changes is the first step of exploration for any regulatory activity. Activating early response genes is a very dynamic process, allowing the organism to rapidly adapt to external or internal stimuli (1,2). Thus, gene expression profiling is the technique of choice to discover and identify transcriptional biomarkers that describe these changes affecting cells, tissues or the entire organism (3,4). Accessing this molecular information via biomarkers in tiny biopsies is a common procedure for many malignancies, but sampling tissues can be costly, painful and potentially impose additional risks on the patient (5). The readout of transcriptional biomarker signatures from minimally invasive sampling methods is therefore highly valued (6). Sampling patient biofluids, such as blood, urine, sweat, saliva or milk in liquid biopsies is currently being thought of

*To whom correspondence should be addressed. Tel: +49 8161 713511; Fax: +49 8161 713539; Email: michael.pfaffl@wzw.tum.de

as a crucial next step in biomarker research and molecular or clinical diagnostics (7).

The existence of extracellular DNA has been acknowledged for decades, and finds applications ranging from oncology to prenatal diagnostics (8,9). In 2005, the first study indicating the importance of microRNAs (miRNAs) in tumor diagnosis and monitoring was published (10). Ever since, the dysregulation of miRNAs in diseased tissues has gained significant prominence and expanded to an interest in extracellular miRNA as reflections of the malignant or dysfunctional alterations. The easy accessibility by blood sampling and remarkable stability of circulating miRNAs make them promising candidates in biomarker discovery. Numerous diseases and disorders, such as tumors, cardiovascular diseases, multiple sclerosis and liver injury have now been associated with altered extracellular miRNA profiles (11). Still, levels of circulating miRNA are presumably non-specific, and few overlapping reports of studies on the same disease have been published, possibly due to technical or methodological inconsistencies (12). Furthermore, miRNA levels seem to be associated with a wide range of conditions and outcomes in cancer research (13). It has therefore been hypothesized that changes in the profile of circulating miRNAs indicate a general state of disease or inflammation and rather derive from a non-specific response to the disease than the malady itself (14).

To date, gene expression profiling is the approach of choice for detecting diagnostic and prognostic biomarkers, or predicting drug safety. Reverse transcriptase quantitative real-time polymerase chain reaction (RT-qPCR) is considered the gold standard for exact and valid gene expression measurements, either for mRNA or small RNA specimens (15). More recently, digital PCR has emerged as a powerful and sensitive technique for absolute quantification of DNA molecules without the need for external calibration curves. Since RNA is converted into cDNA with varying efficiency, however, its applicability for RNA quantification is limited mostly by the reverse transcription (RT) reaction, which might lead to a skewed representation of initial RNA (16). Nowadays, the discovery and identification of potential new transcriptional biomarkers by RNA sequencing (RNA-Seq) is the holistic state of the art technique. The evaluation and validation of miRNA biomarkers by small RNA-Seq is now routinely being adopted for the identification of physiological or dysregulated miRNAs. Nevertheless, the subsequent validation of identified biomarker signatures by RT-qPCR is mandatory (17–19). But there is a lack of consensus regarding optimal methodologies or technologies for miRNA detection in liquid biopsies, their subsequent quantification and standardization strategies when different sequencing technologies or platforms, and library preparation chemistries are used.

Goal of this review

In this review we present a standardization procedure to discover and validate new biomarkers from liquid biopsies with focus on the entire small RNA-Seq workflow - from experimental design, sample stabilization, RNA extraction and quality control to library preparation, next generation RNA sequencing and all steps of small RNA-Seq data analysis, including validation and interpretation (Figure 1 and 2). Our goal is to point out the importance of experimental standardization and validation (20). The review will explain why and where problems in the small RNA-Seq workflow arise, discuss the real bottlenecks, and how one can resolve or at least circumvent them. We want to improve the quality of small RNA-Seq results by optimizing and standardizing the entire quantification procedure to receive better and more reproducible results. As a broader goal, the outcome of this expression profiling should result in valid biomarker signatures in order to make better predictions in molecular diagnostics. The review should follow the 'general MIQE and dMIQE idea' as published earlier, describing optimization strategies in the qPCR and dPCR workflow (21,22). Following our recommendations will encourage better experimental sequencing practice, lead to more reproducible results, and hence allow unequivocal interpretation of small RNA-Seq results. In summary, the outcome of miRNA analysis in liquid biopsies should be more reliable and valid for future predictions.

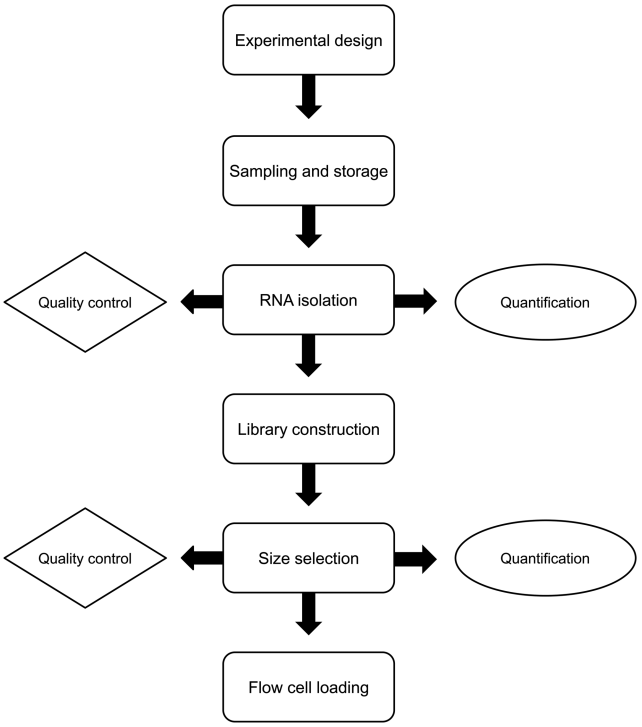

Figure 1. An overview of the small RNA library preparation workflow.

PRE-NGS AND PRE-PCR - THE SAMPLING BIAS

Experimental design and replication

The first step in planning a small RNA expression experiment is to set up a meaningful experimental design, including a reasonable number of replicates on the biological as well as technical level. Both biological and technical replicates have their place in biomarker discovery using RNA sequencing experiments. Biological replicates are crucial to correct for endogenous variability between experimental groups in order to ultimately draw generalized bi-

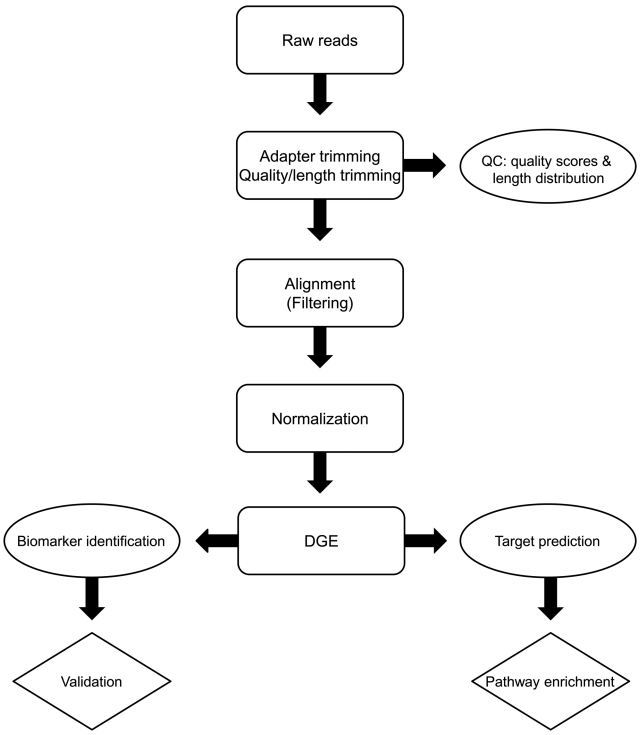

Figure 2. An overview of the small RNA-Seq data analysis workflow.

ological conclusions, whereas technical replicates can help assess the man-made bias introduced by the entire experimental setup and sequencing process itself.

Replicates in reality mean either biological replicates, representing the number of real individuals per experimental group, or technical replicates, repeated measurements of a biological sample with the goal of reducing technical noise. Technical replicates can be further subdivided and introduced either on the level of extraction, RT reactions per sample, sequencing depth, or the number of technical replicates in the sequencing step. Regarding biological sample size, one has to consider inter-individual genetic variation within the studied population. Within the human population, genetic variation is elevated in contrast to highly standardized and inbred animal models. This high variation in human populations as such is based on various factors. Human study groups can be standardized by age, weight or sex, but never by genetic background or lifestyle habits, which might have a remarkable impact on gene expression (nutrition e.g. coffee, alcohol, nicotine consumption, daily rhythms, sleep, stress and more). Regarding domesticated animals such as cattle or pigs, genetic variation is intermediate due to controlled reproduction with a limited number of male semen donors. Laboratory animals, including mice, rats or insects show very low genetic variation within one highly standardized and inbred animal strain. Genetic variation in cell-culture is dependent on the kind of cells used. Primary cell cultures from distinct, genetically different donors show species-specific biological variance, while the largely used permanent cell lines derived from one clone or one individual are genetically identical, and show no biological variance at all.

Different researchers already dealt with the question whether a higher number of biological replicates or a higher sequencing depth leads to better outcomes in RNA-Seq experiments. Increasing sequencing depth results in a higher number of reads, and thereby increases statistical power for the detection of differential gene expression (23). Hart *et al*. and Liu *et al*. concluded that a sequencing depth of 10 million (10M) reads is sufficient for mRNA expression analysis, and that increasing sequencing depth over 10M reads does not improve statistical power significantly. Both publications, however, stated that increasing the number of real biological samples significantly enhances statistical power of the experiment (24,25). Therefore, a higher number of biological samples is preferable over deeper sequencing. Previous reports further suggested including at least three biological replicates per group, depending on the inherent biological variation (26). For experiments involving samples with higher variability, such as human biofluids or specimens from diseased patients, even more replicates might be needed to correctly assess differential gene expression without detecting false-negative reads from biological noise. When biological variability is low, increasing replicates renders statistical power to the experiment. It has been shown that increasing the number of biological replicates in RNA-Seq experiments from two to five facilitates the detection of differential gene expression, but extensive biological replication to improve statistical power is still not utilized in most experiments (27). A recent publication on RNA-Seq additionally reported that experiments entailing only three biological replicates severely lack power to detect the majority of differentially expressed genes, and are only suited to identify transcripts with major fold changes (28). Increased replication markedly improved the correct assessment of differential expression. The authors suggested including at least 12 replicates in order to detect more than 90% of all truly differentially expressed genes.

Technical replicates are useful to characterize the technical variation of an experiment. In general, variability between technical replicates derives from the random sampling nature of sequencing and matches a Poisson distribution (29). Even though it can therefore be accounted for in downstream statistical analysis, some genes are known to deviate from Poisson sampling and thus falsely increase intra-group variability (30).

Another point that has to be considered is whether technical replicates are generally mandatory in gene expression analysis. Liu *et al*. stated that RNA-Seq shows a high reproducibility, concluding that technical replicates are not necessary (25). Due to fairly high library preparation and sequencing costs, technical replicates in RNA-Seq are mostly not realized. Comparing this with replicates in RT-qPCR analysis, Tichopad *et al*. investigated the effect of replicates on different levels of the RT-qPCR quantification workflow (31). The authors stated that replicates in qPCR are not essential, because inhibiting molecules should have been removed before qPCR takes place. The bias introduced by RT was multiple times higher compared to qPCR, wherefore RT replicates are reasonable and necessary (32).

To summarize, inter-individual or biological variation seems to have the highest impact, therefore replicates in biological samples are advisable. The authors recommend in-

troducing replicates early in the quantification workflow by including as many biological replicates as possible (31).

Tissue and RNA sampling and storage

When working with cell-culture, RNA sampling and later storage is unproblematic due to working directly in the molecular biology laboratory in a clean and non-contaminated environment. In contrast, if one collects diagnostic samples in the field, the sampling process cannot be performed under clean and safe laboratory conditions. Therefore, optimal tissue preservation, and thus total RNA preservation and stabilization, are essential points in the experimental workflow. Widely used methods include snap freezing tissues in liquid nitrogen, formalin fixation or storing tissues in RNAlater (Life Technologies), a solution that preserves tissue RNA from degradation, 'freezes' the RNA profile and allows storage of conserved tissue for several hours or days at room temperature.

In clinical research, tissue conservation is routinely performed by formalin fixation and paraffin embedding. Those formalin-fixed paraffin-embedded (FFPE) samples are advantageous for a number of downstream applications. But RNA analysis in FFPE samples is problematic because RNA is cross-linked and partly degraded. RNA extracted from FFPE is thereby of lesser quality for further expression studies (33,34). Due to the growing field of RNA expression analysis in clinical samples, the generation of biobanks for non-fixed frozen tissue is coming into focus (35). Therefore, snap freezing in liquid nitrogen or storage in RNA-preserving agents are the preferred methods for conservation of intact RNA. Publications dealing with the influence of both methods on RNA integrity concluded that high quality RNA can be extracted from tissues conserved in both ways (35,36). The method of tissue fixation has to be planned in detail for each individual experiment according to the given preconditions, which for instance means whether working with liquid nitrogen is generally possible.

RNA quality and RNA integrity

Good RNA quality and high RNA integrity are of great importance in any quantitative gene expression measurement. Degradation of RNA by RNAses, freezing and thawing, UV-light or heat leads to RNA fragmentation. Any RNA degradation influences the results obtained by quantitative downstream applications (37). Several methods to measure RNA integrity and quality exist. Most methods are based on high resolution agarose gel electrophoresis, monitoring the intensity of the major ribosomal 18S and 28S bands. High RNA quality is indicated by a 28S:18S ratio around 2.0 (37). In the past, results of RNA degradation relied on vague human interpretation of the agarose gel image. Nowadays, there are fully automated methods allowing digital interpretation and automatic estimation of the RNA integrity results. With those systems, minor amounts of RNA are labeled with an intercalating dye, and RNA is separated according to its molecular weight using capillary electrophoresis in a microfluidic device. By measuring laser-induced fluorescence detection, the retention time of RNA molecules is displayed in an electropherogram. Applying digital data analysis software, the 18S and 28S ribosomal intensity peaks in an electropherogram are automatically analyzed by a specific algorithm, and a numerical RNA quality score is calculated, whereby a score of 10 indicates intact RNA and a score of 1 completely degraded RNA. It should be mentioned, however, that the concept of RNA Integrity Number (RIN) values is optimized for total RNA profiles from higher eukaryotes, which inherently limits its applicability for studies on other species. Since RIN calculation is majorly based on ribosomal RNA subunit peaks, researchers working with samples differing from the prototypical mammalian RNA need to pay close attention to potentially shifted ribosomal bands. Integrity analysis of plant RNA is further complicated by the presence of additional chloroplast-derived ribosomal RNA that could be recognized as a degradation product and thus falsely lead to lower RIN values. Still, the assessment of RNA quality by measuring RIN has been successfully applied to a variety of non-mammalian organisms such as plants and bacteria (38–40). Although the importance of RNA integrity on downstream applications is well established, even excellent RIN values do not guarantee experimental success since they are unable to report the potential presence of contaminants that might inhibit further RNA processing.

An alternative way of determining transcript integrity is the so-called 3′/5′assay which is based on the quantification of mRNAs at the 3′-end and at the 5′-end. The ratio of the two fractions indicates the mRNA degradation status of the sample (41). The assay, however, is more labor-intensive and has another weakness due to unbalanced RT efficiency at the 3′-end and at the 5′-end.

There are various publications confirming the importance of high RNA quality for mRNA expression profiling studies using microarray and RT-qPCR assays (42–44). For RNA-Seq experiments, high quality RNA is of great importance as well. Degraded RNA leads to decreased quality of RNA-Seq data (45). Particularly the 3′ bias observed in degraded RNA has been shown to have an impact on the quality of RNA-Seq experiments (46). Feng *et al*. developed an algorithm that calculates an RNA quality parameter—the mRIN number— for each sample by quantifying the 3′ bias of read coverage for each measured gene (46).

Small RNAs include the highly prominent miRNAs, which are proven to show higher stability compared to longer RNAs, in particular mRNAs. Due to their short length they are less susceptible to RNA degradation by RNAses (47). The impact of RNA quality on small RNA-Seq has not been evaluated up to now, but it is well known that a high level of RNA degradation in a sample leads to a seemingly increased percentage of small RNAs due to degradation products. It is therefore likely that with decreasing RNA quality, short fragments are included in the sequencing library more frequently, and could thereby lead to a higher number of ambiguous hits after data mapping. The impact of RNA quality on miRNA quantification by SYBR green-based RT-qPCR was shown previously: decreasing RNA quality/integrity is correlated with an increasing Cq value (47).

Circulating RNA and microvesicles

Circulating RNAs are the preferred target in liquid biopsies, and are therefore highly accessed in molecular diagnostics. The RNA, mainly small RNA, found in cell-free blood plasma/serum is either packaged in microvesicles (e.g. exosomes, apoptotic bodies), associated to lipoproteins such as HDL (high-density lipoprotein) particles, or bound by stabilizing proteins (48). Circulating miRNAs are partly bound to proteins such as Argonaute 2 and lipoproteins, which contributes to their enhanced stability (49,50). A seminal paper published in 2007 reported functional miRNA encapsulated in extracellular vesicles (mainly exosomes) secreted by human and murine mast cell lines (51). Soon thereafter, additional reports described the applicability of extracellular vesicular miRNA as biomarkers in blood (52,53). The term circulating miRNA thus has to be used with caution, since it does not state whether the RNA is bound or encapsulated. Circulating or microvesicle-derived RNAs have already shown to be promising diagnostic biomarker for various diseases such as cardiovascular diseases or different kinds of cancer (54).

The composition of circulating vesicles reflects the physiological and pathological status of a patient, and is therefore of considerable diagnostic interest (55). Extracellular vesicles act as a protective shield and delivery vehicle for RNA, and are a treasure trove of easily accessible biological information. Both vesicular RNA and protein were shown to be potential targets for biomarker research (56). Even though considerable advances have been made in the field of extracellular vesicles, there is still no universal consensus on vesicle nomenclature (57). Despite inconsistent terminology, many researchers consider exosomes, the smallest class of extracellular vesicles, as a newly discovered and important mediator in intercellular communication. Since most circulating miRNAs derive from blood or endothelial cells and the contribution of diseased cells is arguably low, exosomes might provide a sampling fraction enriched in tissue-specific biomolecules (14).

There are numerous protocols and commercially available kits for the isolation of extracellular vesicles and extraction of circulating RNAs, in majority from human blood. Principles for isolating vesicles from biofluids include, among others, ultracentrifugation, precipitation, size exclusion chromatography, ultrafiltration, immunopurification and microfluidic approaches (58–61). While differential ultracentrifugation in conjunction with density gradient centrifugation is still considered the gold standard in vesicle isolation and generally yields preparations of high purity, it is labor-intensive, time-consuming and requires substantial sample material, rendering it unsuitable for many clinical and diagnostic applications. Choosing an appropriate isolation method for the particular study has been a topic of extensive debate, and multiple investigations have provided insights into the suitability of respective methods (62–65). Even though most methods were found to be able to isolate extracellular vesicles from various biofluids, yield and purity often differ substantially. Similarly, isolation methods also impact downstream applications: profiles of mRNA (66), miRNA (67) and vesicular protein (68) were shown to vary depending on the respective isolation. Generating pure isolates is complicated by both the complexity of biofluids and the tremendous heterogeneity of extracellular vesicles that even within a particular size range present various subpopulations with different molecular constitution (69,70). Although time-consuming, density gradient centrifugation is highly efficient in removing contaminating proteins and protein complexes, leading to reasonably pure vesicle preparations (62). Polymer-based precipitation methods, on the other hand, require less hands-on time, but suffer from co-isolating non-vesicular contaminants and residual precipitation reagents that can interfere with downstream processing and reduce the vesicle's biological activity (71). Recently, size exclusion chromatography has emerged as a less tedious alternative able to generate vesicles of purity comparable to density gradient-based methods, albeit with low throughput and yield (61,71). Excellent in-depth comparisons of methods for isolating extracellular vesicles from various biofluids can be found elsewhere (63,64,72). Regardless of the particular isolation approach, extraction of RNAs from liquid biopsies is well established. Measuring their concentration is nevertheless challenging due to low concentrations in biofluids. New advances in both sequencing and vesicle research, including careful optimization and standardization of techniques and protocols, will certainly foster progress toward highly specific biomarker signatures.

Blood sampling

In molecular diagnostics, blood is the primary and most important matrix for RNA expression analysis. In humans, minimally invasive sampling is of great advantage, hence blood is the matrix of choice for so-called liquid biopsies. Different and highly standardized methods and kit systems are available for the extraction of high quality RNA from blood, including total circulating RNA and microvesicular RNA. Which sampling system is applicable depends on the particular sample type (whole blood or only a cellular fraction, e.g. white blood cells, red blood cells or platelets), or whether cell-free circulating RNAs of interest are obtained from plasma or serum. For conservation of whole blood for RNA expression analysis, integrated systems for RNA degradation protection and freezing of the current RNA profile are available; namely the PAXgene System (PreAnalytix) and the Tempus System (Life Technologies). Both allow storage of whole blood samples at room temperature for several days or frozen for months without losing RNA quality. For both systems, dedicated kits are commercially available for extraction of RNA longer than 200 nt, or extraction of total RNA including small RNAs (<20 nt). Häntzsch *et al*. and Nikula *et al*. compared the two conservation systems and concluded that both result in high quality RNA samples (73,74). LeukoLock (Life Technologies) allows the extraction of leukocyte RNAs. Within this system, leukocytes are collected in a filter, and RNA is fixed using RNAlater which allows storage and extraction of high quality RNA from white blood cells (75–77).

Quantification of minimal amounts of RNA

The quantification of minimal amounts of total RNA from biopsies or microvesicle isolates is challenging. The de-

tection limit of conventional photometric RNA quantification methods is around 2 ng/μl (78). Due to diminished specificity in the lower concentration range, absorption and therefore quantification is mostly unspecific, because DNA contaminations cannot be distinguished from RNA. Fluorescence-based quantification methods use a fluorescent dye that specifically intercalates or associates with RNA, enabling precise quantification down to as little as 1 pg/μl (78). This method is based on conversion of the fluorescence signal of an unknown sample to a standard curve created from samples with known concentration. The Bioanalyzer 2100 small RNA assay (Agilent Technologies) also allows quantification of small RNAs, especially miRNAs. As mentioned above, this method is only valid in samples of high RNA quality and reasonable RNA quantity. It might result in false positive signals due to contamination of measured small RNA by RNA degradation products with ongoing RNA degradation (47). Due to very low concentrations in RNA samples extracted from plasma or microvesicles, fluorescence-based methods are preferable for small RNA-Seq studies.

How to improve RNA extraction

The extraction of extracellular small RNAs from serum, plasma or other biofluids such as urine or saliva is challenging due to low RNA concentrations. Using carriers to increase RNA output is helpful, whereby glycogen, yeast tRNA, or MS2 phage RNA are widely used. Due to potential interference of biological carrier RNAs with downstream applications, glycogen is the carrier of choice. The use of glycogen increases total RNA yield using most commercially available small RNA extraction kits (79,80). When establishing an extraction method, spiking starting material with known quantities of artificial or exogenous ribonucleotides, so-called spike-in controls, and quantifying their recovery is an easy way to assess the efficacy and reproducibility of the respective approach. Spike-in controls for miRNA extraction are, for example, artificial short RNAs in the length range of miRNAs or miRNA extracts from other species, such as *Caenorhabditis elegans*. Indeed, Burgos *et al*. (81) optimized RNA extraction from human cerebrospinal fluid by measuring the recovery of three previously spiked-in *C. elegans* miRNAs and found significant variation between commercially available kits, and even within technical replicates (81). It is recommended to add spike-in controls directly to the extraction buffer instead of adding it to the plasma or serum sample due to the presence of RNAses in biological samples, which might lead to degradation of the spike-in miRNA (79,80). Spike-in controls can be easily quantified by RT-qPCR in order to determine extraction recovery rate, and appropriately normalize resulting expression data (79). Furthermore, such spike-in controls are also useful to test the efficiency of the RT reaction step or to control for qPCR inhibitors.

LIBRARY PREPARATION - THE RT AND LIBRARY PREPARATION BIAS

The biases based in library preparation

Ultra-high-throughput sequencing allows global sequence profiling of the small RNA transcriptome. To this end, transcriptional targets need to be converted into sequencing libraries, entailing molecular modifications to make targets suitable for the small RNA-Seq chemistry. This pre-sequencing library preparation, however, introduces technical bias into the fine-tuned transcriptional screening and *de novo* discovery of transcripts (82).

In this chapter, we examine critical steps in preparing sequencing libraries from total RNA, and highlight the challenge of creating them in high quality. For the implementation of Next-Generation Sequencing (NGS) of small RNAs, the main task is to convert native small RNAs into sequencable molecules while minimizing technical bias. Preparing small RNA for expression profiling requires multiple enzymatic manipulation steps. These typically include sequential adaptor ligations to both ends of small RNAs, RT, and PCR-based amplification. The 3′-adaptor ligation introduces primer binding sites for first strand cDNA synthesis. The PCR step specifically enriches functional small RNAs with adaptors on both ends, and permits multiplexing through introducing unique barcodes to each sample. Ultimately, a size selection step ensures that only fragments pertaining to small RNAs are included in the final library. In the interest of comparing datasets generated in multiple RNA-Seq experiment with minimal distortion, the problem of pre-sequencing bias needs to be addressed according to the idea of the widely accepted MIQE guidelines (21). Previously, published experimental data showed that using identical starting RNA led to entirely different results concerning small RNA expression ratios due to the implementation of different library preparation strategies (83). Surprisingly, the choice of sequencing platform contributed little to the reported differences (Spearman's $\rho = 0.79$–0.95). Library replicates to test for reproducibility yielded comparable results ($\rho = 0.84$–0.99), indicating that data distortion was likely caused by differences inherent to cDNA construction protocols.

Bias resulting from low RNA input

Besides the quality of extracted total RNA (as discussed above), RNA quantity available for the particular experiment is crucial for successfully generating high-quality sequencing libraries. Various sample types such as plasma, serum or urine contain limited concentrations of small RNA due to lack of cellular material, which complicates library preparation. However, several efficient and sensitive methods for preparing libraries from sparse input material address this problem (84,85). Generally, it is recommended to use RNAs of similar quality and quantity for each sample within an experiment (54,86). Additionally, the capture efficiency of small RNAs from cell-free samples might be limited: Kim *et al*. reported that biological samples with low RNA concentration lack GC poor or highly structured miRNAs when extracting with the phenol/guanidine isothiocyanate reagent Trizol (Thermo Fisher Scientific) (87).

They hypothesized that small RNAs base pair with longer RNA species acting as carrier molecules, and thus compensate their limited capacity to precipitate in RNA extraction. Small RNAs with low GC content and stable secondary structures might interact with carriers less efficiently, reducing their representation in RNA preparations. For samples with low total RNA content, such as a small number of cells or biofluid specimens, the availability of longer RNAs that serve as carriers might be limiting the efficient recovery of this specific fraction of small RNAs. In order to minimize this bias, they recommended to avoid Trizol extractions, or to only compare samples with similar concentrations of total RNA. It was additionally suggested to stabilize RNA–RNA interactions by adding $MgCl_2$ in an attempt to equalize the extraction efficiency of all small RNA species.

The challenge of adapter and barcode ligation

Since the ligation step introduces the largest bias in RNA-Seq results, several studies investigated the effect of ligating 5′- and 3′-adapter or barcodes (88–93). Hafner *et al*. concluded that ligation efficiency depends on the sequence and secondary and tertiary self-structure of miRNAs and/or miRNA/adapter products (94). To reduce ligation bias, many researchers suggest using randomized adaptor pools containing various adapter sequences adjacent to the ligation junction (89,91,93,95). A recently published follow-up paper, however, observed that it is not necessary to design the randomized region near the ligation junction (96). Instead, this might complicate identification of the end of a miRNA sequence with an unknown sequence directly attached to it. Furthermore, the authors found out that miRNAs prefer to ligate to adapters with which they can form a particular structure, whereas the primary sequence is not the main contributor to ligation bias. Even better results can be achieved when the 5′- and 3′-adapter have complementary regions. The only commercial kit employing a similar strategy is the new NEXTflex Small RNA-Sequencing Kit (Bioo Scientific). It uses randomized sequences at the ligation site in massive concentrations to present small RNAs their optimal adapter. According to recent work by Baran-Gale *et al*., the NEXTflex protocol has shown a great reduction in bias and the best differential expression correlation to RT-qPCR (97).

Barcodes are very short distinct sequences which can be introduced in the sequence of interest to enable distinction of multiple samples at the same time and in the same lane of a flow cell. To enable multiplexing, a variety of barcode sets are commercially available (e.g. Illumina TruSeq Small RNA Library Preparation Kit: 48 unique indexes, New England Biolabs NEBNext® Multiplex Small RNA Library Prep Set for Illumina: 24 unique indexes, Bioo Scientific NEXTflex™ Illumina Small RNA-Sequencing Kit v3: 48 unique indexes). Depending on the library preparation kits used, barcodes can be introduced at three points in the library preparation: (i) during adapter ligation (94), (ii) during RT (89) or (iii) during PCR (98). Beside the fact that barcoding is a very useful tool, it causes technical bias by influencing the ligation efficacy, RT efficiency and PCR amplification (92,98). The above findings about the strong impact of base compositions in the core adapter sequence prove that it is crucially important to include barcodes only during RT or later in PCR (89,92,96). When carefully designing the library preparation strategy, it is therefore highly recommended to avoid barcode sequences near primer annealing sites, and to include barcodes only downstream of ligation reactions. It is, however, well described that multiple-template PCR amplification can result in sequence-dependent amplification bias due to template differences (18,89,99). In order to measure the PCR amplification bias resulting from barcodes, Van Nieuwerburgh *et al*. designed a new strategy named post-amplification ligation-mediated (PALM) barcoding, where the ligation of barcodes occurs after PCR without further purification of the library. No bias was observed when comparing PALM with Illumina's TrueSeq miRNA protocol, which introduces barcodes during the PCR step (98).

RNA modifications lead to ligation and RT bias

A simultaneous library construction for all small RNA species is challenging because of their different modified ends. Small RNAs possess different 5′- and 3′-modifications depending on their classes (e.g. miRNA or piRNA) and species origins (e.g. mammals, insects, or plants). While miRNAs in mammals carry a 2′-OH-modification at the 3′-end, many mammalian piRNAs or plant-derived miRNAs feature a 2′-O-methyl group on the ribose at the 3′-end (100,101). This may influence the efficiency of enzymes involved in ligation and cDNA synthesis. To minimize bias, it is important to notice that polyadenylation-based libraries are less suited for 2′-O-methylated RNAs. RNA tailing with poly(A) or poly(C) is significantly less efficient for modified 3′-ends, which might conceivably lead to the underrepresentation or even absence of some RNA species in cDNA libraries (82).

In ligation-based libraries, the ligation efficiency of RNAs with 2′-O-methyl groups can be significantly improved by a longer incubation time, reduced temperature, and the use of T4 RNA Ligase2 instead of T4 RNA Ligase1 (102,103).

Choosing appropriate enzymes for the RT step can also tone down the bias because of their known sensitivity to 2′-O-methyl groups. It is recommended to use avian myeloblastosis virus RTase or murine leukemia virus RTase to prevent favoring the transcription of some RNAs over others (103).

PCR amplification bias in library preparation

The efficiency of PCR amplification depends on the base composition of different types of templates, type of polymerase, PCR buffer composition, and potential presence of any inhibitory substances (104). It is well known that a varying GC-content is associated with unequal PCR amplification efficiencies and leads to template-specific preferences (105–108). To avoid that RNAs with high GC-content remain under-represented, one can perform an optimized PCR program with an extended initial denaturation time of 3 min and subsequent melt cycles of 80 s (109). Furthermore, choosing an appropriate polymerase will not only minimize GC-bias, but also narrow the length dis-

Table 1. Crucial steps and recommendations for small RNA sampling and library preparation

Step	To consider	Recommendation
Experimental design and replication	Type and number of samples Outcome of interest Variance within samples	Employ sufficient replication for question at hand Favor biological replicates over technical ones
Sequencing depth	Outcome of interest Replication	For a rough snapshot of gene expression or analysis of high-level transcripts, lower coverage is sufficient Sequencing depth needs to be increased for analysis of rare transcripts
Sampling and storage	Sampling environment Sample type Embedding/fixation Freezing/storage	Keep sampling conditions as clean as possible Choose an appropriate sampling system for the particular sample type Use agents to preserve and stabilize RNA Freeze samples as quickly as possible and store at appropriate temperature
RNA extraction	Quantity of input material Type of extraction kit Use of a carrier	Carefully optimize the method of extraction for the particular type and quantity of starting material Carrier material might be considered to increase small RNA yield
Total RNA	Expected yield and quantification system Quality of extracted RNA	Opt for fluorescence-based quantification of extracted RNA Check RNA quality and integrity by capillary electrophoresis
Addition of adapter	Type of RNA (e.g. miRNA, piRNA) modified ends	Be aware of ligation biases For small RNAs with modified 3′-ends avoid poly(A) or poly(C)-based approaches or modify protocol accordingly
Reverse transcription	Type of enzyme Introduction of barcodes	Choose appropriate enzyme for given experimental conditions Introduce barcodes during PCR
PCR amplification	Necessity Type of enzyme Number of cycles	Choose pre-amplification strategy based on the quantity of starting material Opt for high fidelity polymerases with low error rates Perform as few PCR cycles as possible
Size selection	Appropriate size range Precision of selection system	Select for cDNA fragments that reflect the size of the RNA of interest High-resolution gel electrophoresis to effectively separate small RNA species
Library purity and quantification	Contamination with adapter dimers Accurate quantification for precise flow cell loading	Assess library purity by capillary electrophoresis Quantify library by fluorimetric assays or qPCR/dPCR
Quality control	Quality and purity of samples at each step of the workflow	Control for sample quality throughout workflow: purity and integrity of initial sample, extracted RNA, cDNA library before and after size selection

tribution of generated PCR products. Several PCR polymerases such as Kapa HiFi (Kapa Biosystems) or AccuPrime Taq DNA Polymerase High Fidelity (Life Technologies) are recommended because of their ability to amplify difficult templates with higher efficiency and lower error rates (109,110). It was furthermore demonstrated that it is of high importance to select a suitable polymerase/buffer system, which can significantly reduce the PCR-mediated bias. In an attempt to optimally amplify DNA sequencing libraries, Dabney and Meyer tested 10 commercially available DNA polymerase/buffer systems and recommended the Herculase II Fusion enzyme as the best performer (107). Generally, it is recommended to use as few PCR cycles as possible for library amplification, and to compare only technical or biological replicates with the identical number of PCR cycles, since PCR noise accumulates with higher cycle number (110).

Library preparation of samples with limited starting material is challenging: researchers have to make a compromise between introducing PCR bias and not detecting lowly expressed transcripts that might not have been sufficiently amplified. Okino *et al.* recently presented a highly multiplexed pre-amplification approach that massively increases the abundance of target genes while keeping amplification bias at bay (111). Since gene expression patterns were maintained throughout up to 14 PCR cycles, analysis of pre-amplified samples yielded similar results to samples not undergoing pre-amplification. Gene expression profiling studies on low input samples might greatly benefit from such a distortion-free enrichment strategy. Recently, more sophisticated library preparation strategies to avoid PCR bias altogether were developed for both bulk and single cell analyses (112,113). By introducing unique molecular identifiers (UMI), researchers are able to detect absolute numbers of DNA or RNA molecules, since each nucleic acid in the

Table 2. Crucial steps and recommendations for small RNA-Seq data analysis

Step	To consider	Recommended tools or algorithms
Data pre-processing	Trimming adapters Removing short reads	Btrim, FASTX-Toolkit
Quality control	Library size and read distribution across samples Per base/sequence Phred score Read length distribution Assess degradation Check for over-represented sequences	Btrim, FASTX-Toolkit, FaQCs
Read alignment (Filtering)	Reference database or genome Annotation Mismatch rate Handling of multi-reads	Bowtie, BWA, HTSEQ, SAMtools, SOAP2
Normalization	Library sizes and sequencing depth Batch effects Read distribution Replication level	DESeq2, EdgeR, svaseq
DGE analysis	Data distribution Replication level False discovery rate	DESeq2, EdgeR, SAMSeq, voom limma
Target prediction of miRNAs / siRNAs	*Insilico* prediction or experimental validation Canonical and non-canonical target regulation	miRanda, miRTarBase, TarBase
Biomarker identification	Sensitivity Specificity Classification rate	DESeq2, Simca-Q, Numerous R packages: base, pcaMethods, Mixomics

starting material is tagged with a unique sequence during RT. After sequencing and mapping, UMI are counted to infer absolute copy numbers without including PCR duplicates in the analysis. Even though UMI-based library preparation has only been applied to mRNA sequencing so far, similar approaches might also be developed for small RNA-Seq in the future.

Gel size selection

The fragmentation of DNA by acoustic shearing, sonication or enzymatic digestion to attain the desired target length of 100–500 bp fragments is not necessary for sequencing small RNAs, which are usually considered to be shorter than 200 nt (110). For miRNA sequencing, fragment sizes of adaptor–transcript complexes and adaptor dimers hardly differ in size. An accurate and reproducible size selection procedure is therefore a crucial element in small RNA library generation. To assess size selection bias, Locati *et al*. used a synthetic spike-in set of 11 oligoribonucleotides ranging from 10 to 70 nt that was added to each biological sample at the beginning of library preparation (114). Monitoring library preparation for size range biases minimized technical variability between samples and experiments even when allocating as little as 1–2 % of all sequenced reads to the spike-ins. Potential biases introduced by purification of individual size-selected products can be reduced by pooling barcoded samples before gel or bead purification.

Since small RNA library preparation products are usually only 20–30 bp longer than adapter dimers, it is strongly recommended to opt for an electrophoresis-based size selection (110). High-resolution matrices such as MetaPhor™ Agarose (Lonza Group Ltd.) or UltraPure™ Agarose-1000 (Thermo Fisher Scientific) are often employed due to their enhanced separation of small fragments. To avoid sizing variation between samples, gel purification should ideally be carried out in a single lane of a high resolution agarose gel. When working with a limited starting quantity of RNA, such as from liquid biopsies or a small number of cells, however, cDNA libraries might have to be spread across multiple lanes. Based on our expertise, we recommend freshly preparing all solutions for each gel electrophoresis to obtain maximal reproducibility and optimal selective properties. Electrophoresis conditions (e.g. percentage of the respective agarose, buffer, voltage, run time, and ambient temperature) should be carefully optimized for each experimental setup. Improper casting and handling of gels might lead to skewed lanes or distorted cDNA bands, thus hampering precise size selection. Additionally, extracting the desired product while avoiding contaminations with adapter dimers can be challenging due to their similar sizes. Bands might be cut from the gel using scalpel blades or dedicated gel cutting tips. DNA gels are traditionally stained with ethidium bromide and subsequently visualized by UV transilluminators. It should be noted, however, that short-wavelength UV light damages DNA and leads to reduced functionality in downstream applications (115). Although the susceptibility to UV damage depends on the DNA's length, even short fragments of <200 bp are affected (116). For size selection of sequencing libraries, it is therefore preferable to use transilluminators that generate light with longer wavelengths and lower energy, or to opt for visualization techniques based on visible blue or green light which do not cause photodamage to DNA samples (117,118). In order not to lose precious sample material, size-selected libraries should always be handled in dedicated tubes with reduced nucleic acid binding capacity.

Precision of size selection and purity of resulting libraries are closely tied together, and thus have to be examined carefully. Contaminations can lead to competitive sequencing of adaptor dimers or fragments of degraded RNA, which reduces the proportion of miRNA reads. Rigorous quality control checkpoints and size selection steps are therefore

crucial. In order to assess length distribution and potential contaminations, it is recommended to use high sensitivity capillary gel electrophoresis assays. The size profile of final library preparation products is dictated by the initial small RNA's size distribution extended with respective sequencing adapters.

Library quantification and flow cell loading

Methods of quantitating final cDNA libraries are still highly debated in the field, and have a significant impact on the sequencing experiment since precise loading of flow cells is crucial for optimal cluster densities. Overloading results in overlapping clusters, reduced quality of reads, and ultimately diminishes the data output of the experiment (119). Low numbers of clusters, or underclustering, on the other hand, yields high-quality data, but a less-than-ideal output. Impurities in sequencing libraries not only skew library quantitation, but also affect cluster generation: shorter fragments such as adapter dimers cluster more efficiently and thus restrict clustering of target RNAs. Capillary gel electrophoresis is a useful tool to assess library integrity, insert size and contaminations, but detects both amplifiable and non-amplifiable molecules (120). Spectrophotometrical methods of nucleic acid quantification are not sensitive enough to precisely quantitate cDNA libraries, and suffer from also measuring single-stranded DNA and free nucleotides. Fluorometric assays such as PicoGreen (Thermo Fisher Scientific) or Qubit (Thermo Fisher Scientific) are more applicable due to increased sensitivity, and specifically quantify double-stranded DNA. Another common approach is quantifying cDNA libraries via qPCR with primers designed to adaptor sequences. Since only functional molecules are captured in the analysis, qPCR and its derivatives seem to precisely predict actual cluster densities (121). Increasingly sensitive methods of library quantification allow for both less input material and fewer PCR cycles, which in turn facilitates sequencing of limited samples and reduces distortion of the initial sequence distribution. Although more costly than other methods, calibration-free absolute quantification of cDNA libraries by digital PCR was found to be a highly accurate tool for quantification of amplifiable molecules in sequencing libraries (122,123).

Loading precision can also be increased by using artificial or exogenous spike-ins. Adding known quantities of a synthetic sequence to samples and quantifying their read count allows for additional control of sequencing parameters. Additionally, technical biases and sequencing errors can be assessed by correlating the amount of spiked-in RNA to read counts mapping to those standards. Fahlgren *et al*. spiked sequencing libraries with three synthetic 21-nt sequences, and found a linear correlation between spike-in concentration and mapped spike-in reads that reached saturation at 10 pmol spike-in per 100 μg of total RNA (124). Another publication using poly-A-tailed mRNA-mimetic standards reported a linear correlation spanning six orders of magnitude while suggesting that the detection of standards is robust to the endogenous complexity of RNA samples (125). As for the analysis of target transcripts, the recovery of standard reads was limited by sequence abundance and sequencing depth, both of which increased spike-in detection.

Critical steps in small RNA-Seq experimental design, sampling and library preparation as well as recommendations by the authors are summarized in Table 1.

SEQUENCING - THE SEQUENCING BIAS

Introduction to sequencing bias

While researchers used to increase sequencing depth rather than introduce additional biological replicates, the ever-subsiding costs of sequencing assays nowadays allow for more replication (126). This, in turn, increases specificity and sensitivity of NGS experiments, and helps correct for biases that cannot be mitigated by bioinformatics methods, such as batch or library preparation effects. Merely increasing sequencing depth in order to improve the specificity of experiments might seem a straightforward strategy, but in reality does not help alleviate sequencing-specific errors (126). Even though a major cause of bias lies in the library preparation of small RNA samples, the sequencing reaction itself can also lead to substantial errors in NGS data. A great number of factors pertaining to the sequencing reaction have to be considered when conceptualizing RNA-Seq experiments. Regardless of the particular experimental question, fundamental aspects such as randomization, replication and blocking need to be properly addressed (127). The most basic decisions relate to choosing a particular sequencing platform and type of flow cell, and designing an experiment that tailors the sequencing chemistry specifically to the question at hand. Additionally, insufficient replication, unsatisfactory sequencing depth and PCR errors are known to increase bias in sequencing data. It is also important to notice that batch effects may result from different kits, reagents, chips, platforms, instruments, handling by different technicians, and day-to-day variations. Batch effects may even occur between different lanes on an Illumina flow cell, or between sequencing runs (110,128). In light of Illumina's dominance in the NGS market, most types of bias discussed in this review are focused on this particular sequencing chemistry.

Batch, lane- and flow cell effects

A major concern in all experiments is detaching biological from technical variation since confounding both makes it impossible to interpret changes in data. For RNA-Seq experiments, it was shown that library preparation introduces the largest bias. This so-called batch effect is an often underestimated problem in high-throughput techniques. As shown above, variations in cDNA preparation from a singular biological source can arise from laboratory conditions, varying quality and reagent lots, skills of the particular operator, changes in personnel, or more subtle factors such as laboratory temperature or ozon levels (128). Quality and quantity of input material, primer concentration, size selection and number of PCR cycles are only a few of many critical parameters of an RNA-Seq protocol that can lead to profound batch effects. A recent article even reported that the composition of small RNA sequencing libraries is more heavily influenced by RNA extraction than by library preparation itself (129). Confounding batch effects with the question of interest, e.g. preparing sequenc-

ing libraries of all treated and control samples on different days or by different operators, can skew the data and directly lead to false biological conclusions. While shown to be of less impact than batch effects, there are also lane and flow cell effects that need to be taken into consideration when designing RNA-Seq experiments (29). These effects pertain to technical variations arising after the cDNA library is loaded onto the sequencer. Marioni *et al*. reported a high replicability in Illumina sequencing data with only a small percentage of genes featuring a systematic difference between different lanes of a flow cell (30). Ross *et al*., on the other hand, found substantial variation between separate flow cells, but not between lanes within a flow cell (130). It should be noted, though, that intra- and inter-assay variation was shown to be less prominent than variation between sequencing platforms.

Multiplexing

The ability to multiplex—adding specific barcodes to separate samples and sequencing them on the same lane of a flow cell—nowadays allows researchers to mitigate lane effects and create more effective experimental designs. Auer *et al*. proposed creating 'balanced blocks' by subjecting all samples to the same experimental conditions, including library preparation and sequencing (i.e. equal proportions of all samples are loaded onto all lanes of the flow cell) (131). For more sophisticated and larger experiments, it is advisable to spread library preparation batches, sequencing lanes and flow cells across all biological groups and replicates to minimize technical variability. Multiplexing and pooling samples as early as possible is advantageous since they can then be processed through the library preparation workflow together, which further alleviates batch effects. Multiplexing also helps to reduce sampling bias when loading the cDNA library onto the flow cell. Loading entails a large dilution step since only a fraction of the cDNA pool is used for cluster generation. An uneven distribution of molecules results in skewed library representation on the flow cell, and thus profoundly alters data output (132). Multiplexing and pooling all samples tones down sequencing errors by reducing sampling bias to only one dilution step.

Paired-end versus single-end sequencing

Paired-end sequencing is a powerful innovation in transcriptomics, yielding more information on transcripts at the same sequencing depth (29). While useful for detection of alternative splice variants and chimeric transcripts, paired-end sequencing usually offers no advantage in small RNA-Seq. Since inserts are short, most experiments do not exceed 50 cycles of sequencing even for small RNA discovery applications. Illumina in fact suggests lowering cycle numbers to 18–36 for miRNA expression profiling studies. Even for profiling of protein-coding genes, 50-bp single-end reads were previously recommended in the literature (26).

Sequencing depth

Since the amount of binding sites on a flow cell is a finite resource, the number of samples in a sequencing run and the sequencing depth are intimately connected. While depth usually refers to the number of reads contributing to an assembly, the respective coverage depends on the abundance of the transcript of interest. For high-level transcripts, even a lower depth might be sufficient to analyze differential gene expression, whereas low-level transcripts require much higher sequencing depths to yield sufficient coverage. Since small RNA copy numbers span a wide range of expression, higher depth is usually required to accurately capture less abundant transcripts. When designing RNA-Seq experiments, sequencing depth has to be tailored to the outcome of interest: a rough snapshot of gene expression requires far lower coverage than the analysis of rare transcripts. For miRNA discovery, Illumina nowadays recommends at least 10M mapped reads. Metpally *et al*. found that while increasing sequencing depth facilitates the detection of new miRNAs, even a moderate depth of only 1.5M mapped reads reliably represents the miRNA distribution in the sample (133). For a given sample type, increasing sequencing depth seems to positively correlate with increasing the proportion of mapped reads. Previous RNA-Seq studies stated that increasing sequencing depth reduces errors in differential gene expression experiments with the caveat of diminishing returns at a certain level of coverage (134). For mRNA-Seq experiments, a stable detection of transcripts seems to be reached at coverage of about 30× with greater coverage only yielding marginal error reduction rates (23). These guidelines could also be applied to small RNA-Seq studies. Since the percentage of initial reads mapping to known miRNAs varies across sample types and library preparation batches, it might be advisable to run a small pilot study in order to determine how many mapped reads are appropriate for the particular biological problem, and how much coverage is needed to generate those reads (133). This ultimately also determines how many samples can be multiplexed on each flow cell of the main experiment. The decision as to whether increase sequencing depth or include more samples depends on the outcome of interest, and is oftentimes limited by the given research budget.

Systematic PCR error

While careful experimental design, library preparation, and loading of the flow cell support bias reduction, the sequencing reaction itself bears additional risk for skewing NGS data. PCR errors induce bias not only during library preparation, but also affect cluster generation and sequencing by synthesis chemistry. Even in the days of high-fidelity DNA polymerases, false incorporation of nucleotides cannot be prevented completely, resulting in DNA strands deviating from the original template. Sequence errors during cluster generation are particularly detrimental since erroneous molecules are exponentially amplified and impair base calling during the subsequent sequencing reaction, ultimately resulting in poor read quality. Growing mixed clusters from more than one template molecule results in a heterogeneous colony of PCR products, and thus an inconclusive fluorescence signal during imaging (135). While amplification efficiency is a significant cause of bias in library preparation, differences in template-specific amplification during cluster generation do not majorly skew read count results since

only the fluorescence intensity of the respective cluster is affected.

Polymerase errors also occur during the sequencing reaction itself. Phasing, the lagging behind of a strand that failed to incorporate a base, hampers base-calling since a more heterogeneous fluorescence signal of the cluster is recorded in each imaging cycle. The enzyme can also erroneously insert multiple bases, which is referred to as prephasing (136). Both of these problems are independent of the template DNA sequence, and lead to an increased frequency of base-calling errors toward the end of a read since more and more noise from preceding and ensuing cycles is introduced. Imaging is further impeded by cross-talk, the partial overlap of emission spectra of the four dyes used in Illumina sequencing technology. This additional noise factor seems to be cycle-dependent and also increases error rates in later cycles (137). Further factors contributing to sequence-independent base-calling errors are dead fluorophores and uneven signal intensities across each tile of the flow cell (138,139). Base-calling algorithms need to be aware of and account for these biases. After signal detection and error correction, the base with the highest intensity is chosen. Remaining uncertainties about called bases are then expressed in quality metrics such as the widely adopted Phred score (140). Originally published in 1998, Phred employs log-transformed error probabilities to generate ASCII-encoded quality scores for each nucleobase. According to the algorithm $q = -10 \times \log_{10}(p)$ where p is the probability of an incorrect base-call, high quality scores equal low error probabilities and the ubiquitous benchmark of Q30 reads corresponds to an error probability of 0.001. The better a base-caller works, the higher the accuracy of sequencing, which ultimately reduces coverage requirements.

Sequence-specific PCR errors in Illumina sequencing

In addition to the abovementioned systematic errors, there are also several sequence-dependent biases in sequencing by synthesis. It is well-known that miscalls on the Illumina platform occur more frequently in GC-rich regions and increase in later cycles (141). Sequence-related biases resulting in failed single-nucleotide elongation might be induced by altered substrate preference of the DNA polymerase or specific inhibition of the enzyme. Indeed, Nakamura *et al.* identified sequence-specific dephasing triggered by GGC sequences to be a consistent bias in Illumina datasets (142). Another cause of sequence-specific errors in Illumina sequencing, albeit potentially of less relevance for small RNA-Seq, are secondary structures of the flow cell-bound single-stranded DNA (ssDNA). According to Nakamura *et al.*, ssDNA folding induced by inverted repeats contributes to polymerase inhibition, while Stein *et al.* illustrated how secondary structures can facilitate or hinder priming during Illumina bridge amplification (142,143). Sequence-induced errors are not only detrimental for applications such as SNP detection or transcriptome assembly, but can also interfere with small RNA-Seq due to the close homology of miRNAs.

Platform-specific error profiles

Previous publications about NGS error rates reported that a majority of miscalled bases is not associated with insufficient coverage, but rather stems from systematic biases in the respective sequencing chemistry (144). It is well known that single base substitutions are the dominant error in Illumina data, while pyrosequencing and ion semiconductor sequencing are more prone to insertions and deletions (indels) (145). In a recent comparison of common platforms, Illumina MiSeq sequencing was shown to produce the highest quality data with a substitution rate of 0.1/100 bases and an indel rate of <0.001/100 bases (146). The frequency of indels was markedly higher when using the Life Technologies Ion Torrent Personal Genome Machine (PGM) and Roche 454 GS Junior systems, featuring 1.5/100 bases and 0.38/100 bases, respectively. Another publication on Illumina sequencing reported error rates as low as 0.3% and an increased frequency of A>C conversion (141). Since the early days of high-throughput sequencing, significant improvements in sequencing chemistry and software have markedly lowered error rates in Illumina data and led to more robust performance. Still, certain error patters characteristic for the technology and independent of the input sequence still pertain to newer generations of sequencers (147). Error rates were shown to be reproducible and predictable across multiple samples in a recent publication on cellular barcoding (148). While indels are fairly rare in Illumina data, they can account for up to two thirds of all errors in 454 pyrosequencing (149). Both Ion Torrent and 454 are known to struggle with homopolymer stretches that oftentimes induce frameshifts. In 454 sequencing, homopolymer errors are more frequent in A and T rich regions and increase with longer sequences of identical bases, while Illumina errors are more randomly distributed (150).

DATA ANALYSIS - THE DATA ANALYSIS BIAS

Small RNA data analysis

Having successfully avoided any pitfalls and biases during experimental setup, library preparation and sequencing, scientists are challenged by processing the frequently huge amounts of sequence data, and extracting meaningful and reliable information from millions and millions of reads. Although digital datasets provide the opportunity to test and validate a seemingly endless array of analyses without spending more than time and computational resources, beginners in the field are often overwhelmed and deterred by the multitude of offered software tools and pipelines. Since a complete discussion of all possible analyses would go beyond the scope of this review, the following part will be centered on the currently prevalent aim of most small RNA-Seq experiments: the detection and comparison of small RNA (mainly miRNA) expression profiles in differently treated samples. In addition, we will focus on 'free to use' software tools or R packages (151) that, while sometimes lacking in user friendliness, are readily available to anyone. Even though most of the software provides comprehensive manuals and tutorials, scientists not already familiar with command line tools may want to try a more intuitively usable software suite, in particular Galaxy (152–

154) or eRNA (155), which implement many of the tools discussed here in a user-friendly graphical interface or invest in commercially distributed programs such as CLC Genomics Workbench (Qiagen), Ingenuity Pathway Analysis (Qiagen) or Genomatix Genome Analyzer (Genomatix). Unfortunately, due to the complexity of varying genomes, small RNA species, data bases and constant updates and improvements of existing software tools, a uniformly valid and standardized analysis approach for all datasets has yet to be established. The fact that most extensive evaluations of methods are carried out on sequencing runs of longer RNAs, and do not take into account the special nature of small RNA datasets further complicates this. The following chapter will highlight all major sources of bias or unwanted variation that need to be addressed and reported to nonetheless guarantee reproducibility and comparability between experimental setups or computational pipelines.

The starting point for all explorations is a fastq file comprising all read sequences with their associated quality scores, indicating the probability of a wrong base call for any given nucleotide. Small RNA data analysis can be generally divided into four individual parts of equal importance: **data preprocessing,** including quality control and adapter trimming, the **alignment** of reads to the respective reference genome or small RNA database, **normalization** of mapped reads, and **differential expression analysis** between samples. A summarizing overview of critical steps and recommended tools for small RNA-Seq data analysis is provided in Table 2.

Data preprocessing

As discussed previously, sequencing errors accumulate with read length, and quality of sequencing data drastically affects downstream analysis (141). Furthermore, sizes of many small RNA transcripts such as miRNAs (∼22 nt) and piRNAs (∼31 nt) (156) fall short of usual sequencing lengths (∼36–50 nt), and resulting reads inevitably incorporate 3′-end adapter sequences from library preparation. To facilitate correct alignments, small RNA read data must therefore be trimmed of adapter artifacts. Complementarily, a significant reduction in false positive alignments to multiple genomic locations can be achieved by filtering for sequences with inadequate lengths (157,158). Removal of these reads with less than 16–18 nt, representing almost exclusively degraded RNA or adapter dimers from library preparation, can also crucially save computational time and associated costs. With the adapter sequences supplied by library preparation kit manufacturers, this can be achieved by a number of programs including Btrim (159), the fastx_clipper tool from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), cutadapt (160) or FaQCs (161). Although current library preparation and sequencing protocols, in conjunction with small read lengths after adapter trimming, do a good job of minimizing sequencing errors, low quality datasets can still occur and will struggle finding accurate alignments. While there are algorithms such as Quake (162) or ALLPATHS-LG (163) that try to correct unreliable base callings by superimposing the most frequent, similar patterns on them, the intrinsically non-uniform sequence abundances found in small RNA-Seq (164) prohibit their application. Low quality reads can nonetheless be mitigated in part by removing bases with low Phred scores from reads up to a minimum length (∼18 nt) or, less preferably, by filtering them out completely (165). Popular quality trimming algorithms implement either some variation of a running sum of the quality scores from 3′- to 5′-end looking for a minimal (Cutadapt), or a moving window that determines the longest continuous stretch of nucleotides above the threshold and trims the rest (Btrim, fastq_quality_trimmer from FASTX Toolkit, FaQCs, SolexaQA (166)). Prior to alignment, filtered and adapter- as well as quality-trimmed reads should then be evaluated in terms of quality scores and typical length distribution of reads. Remaining reads should be free of low quality sequences indicating sequencing errors (quality score <20), and read lengths should show a distinct peak for the targeted small RNA species (e.g. 21–23 nt for miRNA, 30–32 nt for piRNA). An absence of these typical read lengths can originate from a multitude of causes, including incorrect small RNA isolation, inaccurate size selection during library preparation, as well as degradation during, for instance, storage of samples. A fairly uniform increase in read numbers from longer to shorter reads is further proof of low RNA integrity. Additionally, read data can be examined for over-represented sequences potentially deriving from amplification bias during library preparation or contamination with longer RNAs, especially rRNA. k-mer distribution can be assessed by, inter alia, FAQCs or FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Readers interested in benchmarking performances (computation time, memory consumption, possibility of multi-threading etc.) or further quality control checks can find short overviews of existing software tools in (165), (167) or (161).

Small RNA read alignment

To extract meaning from the carefully preprocessed data, reads must be mapped to their respective reference and matched with an appropriate annotation. Almost all existing tools start this process by creating an index for either the reads or the reference, which can then be used to find the corresponding sequence or genomic position. Using these indices allows alignment tools to quickly reduce the number of potential locations on the reference by a first heuristic match of reads, followed by a thorough local alignment for each possible match to evaluate the correct alignment. Without this inexact first pass, alignment of millions of nucleotides would take prohibitively long and overtax all but the most sophisticated computational clusters. Common indexing algorithms include hash tables based on principles used by the well-known BLAST aligner (168), or suffix/prefix tries based on Burrows-Wheeler Transform (169). While hash table based aligners have fewer problems identifying even complicated mismatches between read and reference, the computational requirements to do so escalate quickly. Burrows-Wheeler Transform aligners, on the other hand, are extremely fast and efficient in mapping closely matching read-reference pairs, but slow down significantly when challenged with complex misalignments. In general, there is no single 'best' software tool, and the individual per-

formance varies, among other things, with the error rate or genome type of the particular dataset, as well as the allowed mismatch rate (158), although reference indexing tends to outperform read indexing. Frequently used aligning software for small RNA-Seq include Bowtie (170), BWA (171), or SOAP2 (172), but an evaluation of mapping sensitivity and specificity based on an actual dataset is strongly recommended. Readers interested in benchmarking performances (indexing time, mapping throughput, mapping sensitivity etc.), as well as software-specific algorithm variations such as spaced seeding, q-gram filters, and FM-indices can find short overviews of existing software tools in (173), (174) or (158). Researchers with exceptionally large datasets or facing limiting time constraints could benefit from exploring the possibilities offered by multiple processors in high-end graphic cards (e.g. BarraCUDA (175) or SOAP3-dp (176)), or high-performance computing clusters (e.g. MICA (177)).

Classic read alignment strategies include mapping to a reference genome or a specific small RNA database such as mirBase (178,179) or Rfam (180). While reference genomes enable researchers to get the most comprehensive view of their data, allocating reads to all small RNA classes, as well as potential degraded mRNAs and rRNAs, their annotations often lack the extensiveness found in specific small RNA databases, especially in the case of less researched organisms. Additionally, alignment to a genome can lead to problems with reads that map to multiple genomic locations (multireads). Reads without unique genomic locations are mostly caused by sequencing errors or repetitive sequences, but can also originate from genes with multiple genuine copies in the genome (e.g. hsa-let-7a), and incorrect handling of them can lead to a severe bias (181–183). On the other hand, mapping to a reference genome allows for further characterization of unannotated sequences on the basis of their location or accumulation (e.g. novel miRNA prediction). Alignment to a specific small RNA database, however, has its own pros and cons, mostly stemming from a vastly downscaled mapping reference. Most noticeably, alignment is significantly faster and has a considerably reduced memory footprint. Even though multireads are extremely improbable to occur, the likelihood of false positive mappings of reads from non-targeted small RNAs is increased manifold due to the absence of their sequences in the reference. A more conservative mapping with less mismatches is as crucial in avoiding false positive mappings as is filtering for non-targeted small RNA classes (184). Further complicating this is the existence of functionally relevant isoforms such as isoMirs that often differ substantially from their canonical sequence, but have to be taken into account when determining mismatch thresholds and, ultimately, differential expression (185–187). By comparing reads directly to specific sequences, researchers can also take advantage of homologous datasets from well-explored organisms due to the strong conservation of seed sequences between most small RNA classes in different species (e.g miRNAs or piRNAs (188)). After deciding on a mapping strategy, the final step in alignment is matching the database sequence or genomic position to its corresponding small RNA and counting all reads related to the same feature. With annotations available for all major sequenced genomes, these countlists can be easily generated using HTSEQ (189) or R packages such as IRanges, GenomicFeatures (190) or, in the case of an alignment against a specific sequence database, with e.g. SAMtools (191).

Normalization strategies

Although small RNA-Seq features distinctively less noise and technical bias compared to former holistic screening methods such as microarrays (192), it still generates systematic variation that needs to be addressed prior to differential expression analysis. Unwanted differences between libraries commonly occur in size (sequencing depth) (193) as well as within libraries in GC-content (194) or as batch effects (128). Variation introduced by different gene lengths (195), as is frequently encountered in sequencings of longer RNAs, has a negligible effect. Since usual sequencing lengths cover the whole transcript and fragmentation is not necessary during library prep, the still popular Reads-per-Kilobase-per-Million-mapped-reads (193) is therefore not suited for small RNA-Seq. Overall, the general importance of normalization and its impact on differential expression was clearly shown by Bullard *et al*. in 2010 (196). Special attention has to be paid to experimental setups such as degradation studies, where read distributions differ fundamentally from the underlying assumptions of most methods. Most of the currently established and preferred normalization strategies evolve around a global scaling factor per sample to adjust read counts with. Widespread normalization methods include: (i) library size or total mapped reads, where individual read counts are first divided by their respective library size and then multiplied by the arithmetic mean of all library sizes or counts of total mapped reads, respectively. Since individual read counts are not only directly related to sequencing depth, but also dependent on their relative expression compared to all other small RNA expression levels in a sample, this normalization should be avoided (196). (ii) Upper quartile of reads, where transcripts with zero counts across all samples are filtered from the dataset, and a scaling factor is derived for each sample from the 75th quartile of the remaining reads (196). (iii) Quantile, where the distribution of each gene is assumed to be identical, and read counts are adjusted according to a reference obtained from the median of each quantile across all samples (197). (iv) Trimmed Mean of M-values, where a weighted trimmed mean of log expression ratios is calculated for each sample compared to a reference sample. Working under the assumption that expression of most genes will not be significantly altered in the experiment, these means should be close to 1, and a scaling factor is derived from this difference, and finally adjusted by the mean of the normalized libraries (198). (v) Median of expression ratios from geometric means, where a pseudoreference is first created by computing the geometric mean of all genes across samples, and then the ratio for each count to its respective mean is determined. The scaling factor is finally obtained from the median of all ratios for each sample. Similar to (iv), median normalization also assumes most genes to not be differentially expressed (199,200). (vi) Artificial spike-in standards, where reads are quantified using a standard curve derived from a set of pre-determined small RNAs independent of the samples (125). (vii) Surrogate variable analysis, which is

specifically targeted on batch effects, and helps identifying genomic data affected by artifacts. It adjusts read counts by estimating these artifacts with the help of singular vectors of the specific subset of the data (201).

Although the variety in experimental and genomic set ups so far makes it impossible to universally recommend a single normalization strategy, recent evaluations of these methods have found Median normalizing of expression ratios from geometric means to work favorably with various kinds of datasets (202,203). Additionally, Zyprich-Walczak *et al*. proposed a step-by-step workflow to determine the most appropriate normalization method for a specific dataset in terms of bias, variance, sensitivity, specificity and prediction errors to avoid data distortion by using the wrong normalization (203).

Differential expression analysis

In comparison to normalization strategies that were mostly extending existing methods for microarrays, the distinctly different data type of NGS made the development of new algorithms for differential expression analysis imperative. While microarray data consists of continuous intensities coupled with a high background, NGS read counts give discrete measurements for each gene, and should not, unlike microarray intensities, be modeled on a normal distribution. Although early RNA-Seq reported a good fit to a Poisson distribution for single sample sequencings and technical replicates (30,196), studies with biological replicates are extremely likely to show variances greater than the mean for many genes (204). This so-called overdispersion makes analyses working under the Poisson assumption prone to high false-positive rates due to an underestimation of sampling error. One way to overcome this is an extension of the Poisson model with a quasi-likelihood approach, where each gene is tested individually for overdispersion (Two-Stage-Poisson-Model (205)). Another way to account for biological variability is the negative binomial distribution, which adds the dispersion to the mean as a second parameter (206). Correct estimation of gene-wise dispersion factors is crucial, but unfortunately also hampered by the still prevalent low number of sample in most RNA-Seq studies. To obtain more accurate dispersion factors, analysis tools share information across all genes in the dataset by, among other things, a weighted likelihood approach toward the common dispersion (edgeR (207)) or by modeling the observed mean-variance relationship for all genes via regression (DESeq (199,200)). Differential expression can then be tested by either exact tests (edgeR, DESeq) or empirical Bayesian frameworks (EBSeq (208), baySeq (209)). Apart from these distribution assumptions, differential expression can also be assessed by non-parametric approaches based for instance on Wilcoxon rank statistics and resampling strategies (SAMSeq (210)), or by comparing the absolute and relative expression differences between and within experimental conditions (NOISeq (211)). A major drawback of these methods is their relatively low power and specificity in experiments with low sample numbers. In addition, robust methods established for microarrays (limma (212,213)) can be made applicable through transformation of discrete read count data (voom (214)). Irrelevant of the employed algorithm, all tools will produce a list of significantly regulated genes that should be treated with caution. Due to the large number of tests, the false discovery rate should be controlled for all results to avoid accumulation of type-1-errors (215). Additionally, the ratio of expression signal to experimental noise should be monitored for lowly expressed genes by assessing the biological relevance of the fold change, as well as absolute read count values.

More so than any other tools, software for differential expression is subject to frequent updates, which can alter their behavior dramatically and new algorithms are published continually. Even though comparisons of software performances on small RNA-Seq data are scarce, a number of independent and extensive evaluations for mRNAs based on either synthetic data with clearly defined properties (216), or on biological datasets with validated gene expressions (217,218) have been made recently. While it was shown that statistical power of almost all methods is heavily dependent on the number of samples per condition and less on sequencing depth, the variability of expression changes in biological datasets affects each analysis tool differently. Outliers, 'ON/OFF' expression changes, where a gene is detected in only one condition, and lopsided expression patterns, where upregulations drastically outweigh downregulations or *vice versa*, influence specificity (false positive rate) and sensitivity (false negative rate) of each method unequally. Nonetheless, some methods appear to capture the true expression status of small RNAs better than others. Most independent evaluations seem to agree that calling differential expression with SAMSeq works well for datasets with sufficient sample sizes of 10 or more. For smaller datasets, edgeR and especially the more conservative DESeq (or DESeq2) are found to be the methods of choice. On top of that, the voom + limma method was reported to generally perform well for different datasets (216). Additionally, a recent publication on RNA-Seq showed that most of the frequently used tools correctly assess differential gene expression when sufficient biological replication is employed (28). For a low number of replicates, edgeR outperformed its competitors, while DESeq exceled in experiments with more than 12 replicates, suggesting that data analysis tools need to fit the respective experimental setup. Efforts with mixed results have also been made to weigh differential expression results of various methods and combine them to an optimized consensus bypassing the individual flaws of each algorithm (219). Considering all this, choosing the optimal tool for differential expression analysis is still strongly dependent on the individual dataset, highlighting once again the fact that researchers need to thoroughly acquaint themselves with the details and specifics of their individual setup and data distribution before starting any analyses.

BIOMARKER IDENTIFICATION AND VALIDATION

After biomarker candidates have been identified in the differential expression analysis, these markers have to be statistically validated. Since univariate analyses, like most differential expression tests, treat each biomarker (i.e. small RNA) as independent, they are unable to capture the complete reality of highly multivariate (variables >> observa-

tions) and correlated datasets such as NGS read counts. By taking the synergies, antagonisms and redundancy inherent in each NGS dataset into consideration, multivariate analyses can reach much higher discriminative power and separate noise from signal (19,220). In reality, there will most likely be no single valid transcriptional biomarker for the physiological situation of interest. In most cases, only a set of multiple biomarkers can ensure the high sensitivity, specificity and reliability needed for diagnostic and prognostic analyses. Appropriately dealing with these data to retrieve the desired outcome of a stable and valid biomarker signature is, however, not trivial.

The most promising approach is to first screen read counts for general trends or potential outliers in an unsupervised manner (no classification information is given to the algorithm), and subsequently assess the discriminative power of potential biomarker candidates (221). These analyses generate clusters of similarities, specifically similar gene expression patterns in the case of RNA-Seq, by using methods for dimension reduction combined with pattern recognition technologies and visualize them in two- or three-dimensional graphs (222). Similar to differential expression profiling, read count lists need to be preprocessed. Input data for any cluster or classification analysis can either be normalized read counts, as described previously, or ratios thereof, and in addition should be transformed to address their skewed distribution. A simple shifted log transformation ($\log_2$ (n + 0.5)) to make the data conform to normality is most commonly used, but more sophisticated alternatives such as regularized log transformation (rlog, (200)) and variance stabilizing transformation (vst, (223)) might be better suited for small RNA-Seq data (both algorithms are implemented in DESeq2, (200)). Cluster algorithms are implemented, for instance, in the base distribution of R, as well as more comprehensive packages such as pcaMethods (224) and the excellent mixOmics (225), or the commercially available Simca-Q software (Umetrics).

Widely accepted unsupervised multivariate analyses include clustering analyses such as hierarchical clustering (HCA), partitioning methods such as k-means and self-organizing maps (SOM), as well as projections on latent variables such as the powerful principal component analysis (PCA). In agglomerative HCA, samples (or genes) start as single entity clusters and are then joined step-by-step based on a similarity measure and a linkage function, defining inter-cluster distances. For log-transformed data, it was shown that Euclidian distances and Pearson correlation perform well as distance measures, while complete linkage (or Ward's method) strictly surpass single or average linkage functions (226). The result and graphical output of HCA is a tree dendrogram emphasizing the distances between the individual samples (or genes) with rising node lengths and clusters can be obtained by, among others things, cutting at fixed heights (227,228). Combining HCA of samples and genes with a two-dimensional color-coded description of the whole experimental matrix creates a heatmap, which allows for easy detection of similarities and dissimilarities in a read count list. Although HCA is the still the most common clustering algorithm, it is in most cases outperformed by partitioning methods such as k-means and SOM (226,229). Both work by subdividing the dataset into a predetermined number of unhierarchical subsets based on randomly chosen centroids. In k-means, samples (or genes) are iteratively assigned to the closest centroid with each iteration replacing the former centroid by the average of each entity in its cluster until all samples (or genes) are set. In SOM, the centroids are linked by a grid structure, and with each iteration the closest centroid, as well as its neighbors, is moved toward a randomly chosen sample (or gene). By gradually shrinking the radius of each adjacent centroid, this will result in a grid of clusters comprising all samples (or genes) with related expression patterns. Since both k-means and SOM start with randomly placed centroids and the optimal number of cluster is usually not apparent, these algorithms should be rerun with random seeds and different numbers of clusters to obtain a stable classification.

Even more information on potential biomarkers can be obtained by PCA, which converts a multidimensional dataset into a lower number of variables called principal components (PCs) (228,230). The read count data is thus decomposed in a score matrix describing small RNA genes, a loadings matrix describing the samples, and a residual matrix expressing deviations between the original variables and the projections. PCs are calculated ranked with the first PC accounting for the greatest variance in the dataset and subsequent PCs comprising the respective maximum residual variance. Since PCs are computed orthogonally to each other, they each describe independent sources of information, and with decreasing variance explained by later PCs, they can be used to separate systematic effects, explained by the molecular biomarker set, from random expression noise (227). Variance derived from experimental study design is expected to be systematic, while confounding variance is expected to be small and random and can therefore be found in later PCs. The advantage of PCA in comparison to clustering and partitioning methods is obvious, since it allows a much clearer recognition and more precise differentiation of the experimental groups. In PCA, the commonalities (or differences) in gene expression pattern are clearly visualized by the symbol interspaces in at least two dimensions (228,231). By plotting scores and loadings plots side by side and looking at their corresponding positioning, it is also possible to identify which small RNA genes are responsible for the separations of samples. Potential biomarkers can be assessed by their contribution plots, and outliers can be detected by either Hotelling's T^2 or by their residual standard deviation (distance to model, DModX) (232).

All of the unsupervised methods mentioned above generate groupings of samples (or genes) with similar expression patterns. While this allows for easy detection of outliers and inconsistencies in experimental setup, it does not necessarily mean that resulting clusters will reflect the desired classification of samples or genes. An underlying treatment effect can sometimes be veiled by other dominating effects, be they intentional (different cell types, time points etc.) or not (batch effects). By incorporating information on experimental setup, researchers are able to filter out genes inducing the greatest separation between treatment groups or, in other words, potential biomarkers. Although a number of supervised classifications algorithms exist, it was shown that the widely used partial least squares projection to latent structures (PLS) and its modifications such as PLS discrim-

inant analysis (PLS-DA, (233)), sparse PLS-DA (sPLS-DA, (234)) or orthogonal PLS (OPLS, (235)) are well suited for dimension reduction and discrimination (233,236).

PLS is related to linear discriminant analyses (LDA), and is a regression extension of PCA that shares many characteristics with it. By adding a second matrix containing the responses or dependent variables to the read count matrix, PLS attempts to find latent variables (LV) that predict the responses from gene expression profiles and describe the common structure of both matrices. LVs are calculated hierarchically similar to PCs, but LVs maximize covariance instead of variance. In PLS-DA, the response matrix is replaced by an optimized dummy matrix containing only 0 and 1 for every respective class, and the resulting projection model therefore focuses on maximum discrimination between classes in the responses rather than 'optimal class modeling' (221). Biomarkers can then be evaluated by a number of variable selection methods including variable importance in projection (equivalent to a contribution plot in PCA) or target projection with selectivity ratio test (237), and by drawing a consensus between differentially expressed genes and multivariate analyses.

The biological functionality of detected small RNA biomarkers, mainly based on miRNAs, can be further verified in functional experimental tests using miRNA overexpression, knockdown or even knockout experiments. Various tools and software packages are available for the *in silico* functional analysis of miRNAs. For *in silico* target prediction, we recommend the TargetScan package (http://www.targetscan.org/) (238,239) or miRanda (http://www.microrna.org/) (240,241). For analyzing the inverse relation of expressed miRNAs and mRNAs in conjunction with target predictions, we recommend using a Lasso regression model (242,243). If an integrative analysis of miRNAs and their target genes is of interest, the miRNA–mRNA relations can be tested on the basis of regression analysis, and further processed by testing for enrichment in gene ontology terms or KEGG pathways (http://www.genome.jp/kegg/pathway.html), amongst others (244,245). In addition, several all-in-one software packages such as CLC Genomics Workbench (Qiagen), Ingenuity Pathway Analysis (Qiagen) or Genomatix Genome Analyzer (Genomatix) are available to allow a relatively easy, graphic user interface (GUI)-based *in silico* functional analysis of miRNAs. Applying Genomatix Pathway System (GEPS) or Ingenuity Pathway Analysis facilitates the creation and extension of miRNA networks based on information extracted from public and proprietary databases and co-citations in the literature.

Conclusion - where are the real bottlenecks?

Today, liquid biopsies and the small RNA biomarker signatures they may inclose are considered the promising new generation of transcriptional biomarkers. The RNA is easily accessible, often by non-invasive procedures, physiologically stable and protected by microvesicles or associated proteins. Due to its chemical nature, it can be rapidly amplified and quantified using RT and PCR-related methods. Small RNA-based biomarker signatures can therefore be detected at low concentrations and early disease stages, and the discovery workflow can be further optimized and standardized. This sustains the idea of the MIQE and dMIQE guidelines previously published by an international consortium (headed by SA Bustin and JF Huggett) in the field of qPCR and dPCR (21,22).

Thoroughly and accurately following our recommendations by optimizing and standardizing the small RNA-Seq workflow will result in reproducible data and, subsequently, reliable hypotheses. The digital and holistic nature of the small RNA-Seq approach provides vast transcriptional data that is highly informative in terms of both quality and quantity (246). The subsequent complex, comparative and multivariate data analysis can result in valid biomarker signatures. The technological developments in the entire workflow (from sampling to multivariate data analysis) are very dynamic, and will continue to improve in the future. While proven standards and optimized methodologies to identify promising biomarkers in liquid biopsies are still lacking, the optimization and validation process will continue to develop.

Where are the real bottlenecks in small RNA-Seq analysis of liquid biopsies? The most significant factor leading to success is probably the number of variables and conditions being tested, and the number of real biological replicates used for sequencing. What appears to be specific in the particular biological samples analyzed by small RNA-Seq may not necessarily be reflected in a larger group, or even in the entire population. Therefore, the more individuals tested, and the more conditions or variables being evaluated, the better the outcome of the prediction and the validity of the discovered biomarker signature will be (247,248).

No step in the workflow is free of bias, but some are more prone to produce noise in the resulting data. Due to financial reasons, researchers still employ too few biological replicates. Only biological replicates can explain any biological difference, while technical replicates are limited to only report the technical noise researchers introduce. In our opinion, the largest noise impact is introduced by RNA extraction and the complex library preparation, which can be performed in various ways, but always highly depends on enzyme efficiency. Depending on the respective library preparation chemistry, numerous individual barcodes are used. These not only cause technical bias, but also affect RT efficiency and PCR amplification.

In general, it is recommended to perform as few PCR cycles as possible for pre-amplification, and to only compare replicates with the identical number of cycles. The sequencing or clonal amplification as such is not a major source of variation, since error rates of polymerases are acceptably low, sequencing chemistry exhibits high purity and the hardware operates very precisely and reproducibly. A further big challenge is the off-instrument data analysis, which requires the majority of manpower and time in the quantification workflow. We should put major focus on alignment, normalization and differential expression analysis, since these are the most critical steps. Biases introduced at earlier stages can in part be corrected and compensated by an appropriate normalization strategy.

As a final and essential step after small RNA-Seq, we recommend additional validation of the identified transcriptional biomarker signatures. This confirmation should be

carried out using established and highly standardized methods such as RT in combination with real-time PCR or digital PCR. The consistency and correctness of the discovered transcriptional biomarker signature in the liquid biopsy can only be assumed after data verification and demonstration of a statistically validated correlation between small RNA-Seq and RT-qPCR or dPCR.

ACKNOWLEDGEMENT

This work was supported by the Vereinigung zur Förderung der Milchwissenschaftlichen Forschung an der Technischen Universität München e.V.

FUNDING

Funding for open access charge: Internal institutional budget.
Conflict of interest statement. None declared.

REFERENCES

1. Riedmaier,I., Pfaffl,M.W. and Meyer,H.H. (2012) The physiological way: monitoring RNA expression changes as new approach to combat illegal growth promoter application. *Drug Test Anal.*, **4**(Suppl. 1), 70–74.
2. Pfaffl,M.W. (2013) Transcriptional biomarkers. *Methods*, **59**, 1–2.
3. Sewall,C.H., Bell,D.A., Clark,G.C., Tritscher,A.M., Tully,D.B., Vanden Heuvel,J. and Lucier,G.W. (1995) Induced gene transcription: implications for biomarkers. *Clin. Chem.*, **41**, 1829–1834.
4. Pfaffl,M.W. (2015) Guest editor's introduction for BDQ special issue: 'Advanced Molecular Diagnostics for Biomarker Discovery'. *Biomol. Detect. Quantif.*, **5**, 1–2.
5. Karachaliou,N., Mayo-de-Las-Casas,C., Molina-Vila,M.A. and Rosell,R. (2015) Real-time liquid biopsies become a reality in cancer treatment. *Ann. Transl. Med.*, **3**, doi:10.3978/j.issn.2305-5839.2015.01.16.
6. Buder,A., Tomuta,C. and Filipits,M. (2016) The potential of liquid biopsies. *Curr. Opin. Oncol.*, **28**, 130–134.
7. Properzi,F., Logozzi,M. and Fais,S. (2013) Exosomes: the future of biomarkers in medicine. *Biomark. Med.*, **7**, 769–778.
8. Fleischhacker,M. and Schmidt,B. (2007) Circulating nucleic acids (CNAs) and cancer–a survey. *Biochim. Biophys. Acta*, **1775**, 181–232.
9. Pinzani,P., Salvianti,F., Pazzagli,M. and Orlando,C. (2010) Circulating nucleic acids in cancer and pregnancy. *Methods*, **50**, 302–307.
10. Calin,G.A., Ferracin,M., Cimmino,A., Di Leva,G., Shimizu,M., Wojcik,S.E., Iorio,M.V., Visone,R., Sever,N.I., Fabbri,M. *et al.* (2005) A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N. Engl. J. Med.*, **353**, 1793–1801.
11. Reid,G., Kirschner,M.B. and van Zandwijk,N. (2011) Circulating microRNAs: association with disease and potential use as biomarkers. *Crit. Rev. Oncol. Hematol.*, **80**, 193–208.
12. Kirschner,M.B., van Zandwijk,N. and Reid,G. (2013) Cell-free microRNAs: potential biomarkers in need of standardized reporting. *Front. Genet.*, **4**, doi:10.3389/fgene.2013.00056.
13. Ono,S., Lam,S., Nagahara,M. and Hoon,D.S. (2015) Circulating microRNA biomarkers as liquid biopsy for cancer patients: pros and cons of current assays. *J. Clin. Med.*, **4**, 1890–1907.
14. Witwer,K.W. (2015) Circulating microRNA biomarker studies: pitfalls and potential solutions. *Clin. Chem.*, **61**, 56–63.
15. Hayes,C.J. and Dalton,T.M. (2015) Microfluidic droplet-based PCR instrumentation for high-throughput gene expression profiling and biomarker discovery. *Biomol. Detect. Quantif.*, **4**, 22–32.
16. Sanders,R., Mason,D.J., Foy,C.A. and Huggett,J.F. (2013) Evaluation of digital PCR for absolute RNA quantification. *PLoS One*, **8**, e75296.
17. Meyer,S.U., Kaiser,S., Wagner,C., Thirion,C. and Pfaffl,M.W. (2012) Profound effect of profiling platform and normalization strategy on detection of differentially expressed microRNAs–a comparative study. *PLoS One*, **7**, e38946.
18. Meyer,S.U., Pfaffl,M.W. and Ulbrich,S.E. (2010) Normalization strategies for microRNA profiling experiments: a 'normal' way to a hidden layer of complexity? *Biotechnol. Lett.*, **32**, 1777–1788.
19. Spornraft,M., Kirchner,B., Pfaffl,M.W. and Irmgard,R. (2015) The potential of circulating extracellular small RNAs (smexRNA) in veterinary diagnostics—Identifying biomarker signatures by multivariate data analysis. *Biomol. Detect. Quantif.*, **5**, 15–22.
20. Bustin,S.A. (2014) The reproducibility of biomedical research: sleepers awake! *Biomol. Detect. Quantif.*, **2**, 35–42.
21. Bustin,S.A., Benes,V., Garson,J.A., Hellemans,J., Huggett,J., Kubista,M., Mueller,R., Nolan,T., Pfaffl,M.W., Shipley,G.L. *et al.* (2009) The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.*, **55**, 611–622.
22. Huggett,J.F., Foy,C.A., Benes,V., Emslie,K., Garson,J.A., Haynes,R., Hellemans,J., Kubista,M., Mueller,R.D., Nolan,T. *et al.* (2013) The digital MIQE guidelines: minimum information for publication of quantitative digital PCR experiments. *Clin Chem*, **59**, 892–902.
23. Tarazona,S., Garcia-Alcalde,F., Dopazo,J., Ferrer,A. and Conesa,A. (2011) Differential expression in RNA-seq: a matter of depth. *Genome Res.*, **21**, 2213–2223.
24. Hart,S.N., Therneau,T.M., Zhang,Y., Poland,G.A. and Kocher,J.-P. (2013) Calculating sample size estimates for RNA sequencing data. *J. Comput. Biol.*, **20**, 970–978.
25. Liu,Y., Zhou,J. and White,K.P. (2013) RNA-seq differential expression studies: more sequence or more replication? *Bioinformatics*, **30**, 301–304.
26. Williams,A.G., Thomas,S., Wyman,S.K. and Holloway,A.K. (2014) RNA-seq data: challenges in and recommendations for experimental design and analysis. *Curr. Protoc. Hum. Genet.*, **83**, 11–20.
27. Kvam,V.M., Liu,P. and Si,Y. (2012) A comparison of statistical methods for detecting differentially expressed genes from RNA-seq data. *Am. J. Bot.*, **99**, 248–256.
28. Schurch,N.J., Schofield,P., Gierliński,M., Cole,C., Sherstnev,A., Singh,V., Wrobel,N., Gharbi,K., Simpson,G.G., Owen-Hughes,T. *et al.* (2016) How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? *RNA*, **22**, 839–851.
29. Fang,Z. and Cui,X. (2011) Design and validation issues in RNA-seq experiments. *Brief. Bioinform.*, **12**, 280–287.
30. Marioni,J.C., Mason,C.E., Mane,S.M., Stephens,M. and Gilad,Y. (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res.*, **18**, 1509–1517.
31. Tichopad,A., Kitchen,R., Riedmaier,I., Becker,C., Stahlberg,A. and Kubista,M. (2009) Design and optimization of reverse-transcription quantitative PCR experiments. *Clin. Chem.*, **55**, 1816–1823.
32. Stahlberg,A., Kubista,M. and Pfaffl,M. (2004) Comparison of reverse transcriptases in gene expression analysis. *Clin. Chem.*, **50**, 1678–1680.
33. Scicchitano,M.S., Dalmas,D.A., Bertiaux,M.A., Anderson,S.M., Turner,L.R., Thomas,R.A., Mirable,R. and Boyce,R.W. (2006) Preliminary comparison of quantity, quality, and microarray performance of RNA extracted from formalin-fixed, paraffin-embedded, and unfixed frozen tissue samples. *J. Histochem. Cytochem.*, **54**, 1229–1237.
34. Li,J., Smyth,P., Cahill,S., Denning,K., Flavin,R., Aherne,S., Pirotta,M., Guenther,S.M., O'Leary,J.J. and Sheils,O. (2008) Improved RNA quality and TaqMan® Pre-amplification method (PreAmp) to enhance expression analysis from formalin fixed paraffin embedded (FFPE) materials. *BMC Biotechnol.*, **8**, doi:10.1186/1472-6750-8-10.
35. Micke,P., Ohshima,M., Tahmasebpoor,S., Ren,Z.P., Ostman,A., Ponten,F. and Botling,J. (2006) Biobanking of fresh frozen tissue: RNA is stable in nonfixed surgical specimens. *Lab. Invest.*, **86**, 202–211.
36. Dekairelle,A.F., Van der Vorst,S., Tombal,B. and Gala,J.L. (2007) Preservation of RNA for functional analysis of separated alleles in

yeast: comparison of snap-frozen and RNALater solid tissue storage methods. *Clin. Chem. Lab. Med.*, **45**, 1283–1287.
37. Schroeder,A., Mueller,O., Stocker,S., Salowsky,R., Leiber,M., Gassmann,M., Lightfoot,S., Menzel,W., Granzow,M. and Ragg,T. (2006) The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol. Biol.*, **7**, doi:10.1186/1471-2199-7-3.
38. Jahn,C.E., Charkowski,A.O. and Willis,D.K. (2008) Evaluation of isolation methods and RNA integrity for bacterial RNA quantitation. *J. Microbiol. Methods*, **75**, 318–324.
39. Die,J.V. and Roman,B. (2012) RNA quality assessment: a view from plant qPCR studies. *J. Exp. Bot.*, **63**, 6069–6077.
40. Heera,R., Sivachandran,P., Chinni,S.V., Mason,J., Croft,L., Ravichandran,M. and Yin,L.S. (2015) Efficient extraction of small and large RNAs in bacteria for excellent total RNA sequencing and comprehensive transcriptome analysis. *BMC Res. Notes*, **8**, doi:10.1186/s13104-015-1726-3.
41. Die,J.V., Obrero,Á., González-Verdejo,C.I. and Román,B. (2011) Characterization of the 3′:5′ ratio for reliable determination of RNA quality. *Anal. Biochem.*, **419**, 336–338.
42. Copois,V., Bibeau,F., Bascoul-Mollevi,C., Salvetat,N., Chalbos,P., Bareil,C., Candeil,L., Fraslon,C., Conseiller,E., Granci,V. *et al.* (2007) Impact of RNA degradation on gene expression profiles: assessment of different methods to reliably determine RNA quality. *J. Biotechnol.*, **127**, 549–559.
43. Fleige,S. and Pfaffl,M.W. (2006) RNA integrity and the effect on the real-time qRT-PCR performance. *Mol. Aspects Med.*, **27**, 126–139.
44. Fleige,S., Walf,V., Huch,S., Prgomet,C., Sehm,J. and Pfaffl,M.W. (2006) Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time RT-PCR. *Biotechnol. Lett.*, **28**, 1601–1613.
45. Gallego Romero,I., Pai,A.A., Tung,J. and Gilad,Y. (2014) RNA-seq: impact of RNA degradation on transcript quantification. *BMC Biol.*, **12**, doi:10.1186/1741-7007-12-42.
46. Feng,H., Zhang,X. and Zhang,C. (2015) mRIN for direct assessment of genome-wide and gene-specific mRNA integrity from large-scale RNA-sequencing data. *Nat. Commun.*, **6**, 7816–7826.
47. Becker,C., Hammerle-Fickinger,A., Riedmaier,I. and Pfaffl,M.W. (2010) mRNA and microRNA quality control for RT-qPCR analysis. *Methods*, **50**, 237–243.
48. Etheridge,A., Gomes,C.P.C., Pereira,R.W., Galas,D. and Wang,K. (2013) The complexity, function and applications of RNA in circulation. *Front. Genet.*, **4**, doi:10.3389/fgene.2013.00115.
49. Vickers,K.C., Palmisano,B.T., Shoucri,B.M., Shamburek,R.D. and Remaley,A.T. (2011) MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nat. Cell Biol.*, **13**, 423–433.
50. Arroyo,J.D., Chevillet,J.R., Kroh,E.M., Ruf,I.K., Pritchard,C.C., Gibson,D.F., Mitchell,P.S., Bennett,C.F., Pogosova-Agadjanyan,E.L., Stirewalt,D.L. *et al.* (2011) Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 5003–5008.
51. Valadi,H., Ekstrom,K., Bossios,A., Sjostrand,M., Lee,J.J. and Lotvall,J.O. (2007) Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.*, **9**, 654–659.
52. Taylor,D.D. and Gercel-Taylor,C. (2008) MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.*, **110**, 13–21.
53. Chim,S.S., Shing,T.K., Hung,E.C., Leung,T.Y., Lau,T.K., Chiu,R.W. and Lo,Y.M. (2008) Detection and characterization of placental microRNAs in maternal plasma. *Clin. Chem.*, **54**, 482–490.
54. Spornraft,M., Kirchner,B., Haase,B., Benes,V., Pfaffl,M.W. and Riedmaier,I. (2014) Optimization of extraction of circulating RNAs from plasma—enabling small RNA sequencing. *PLoS One*, **9**, e107259.
55. Yanez-Mo,M., Siljander,P.R., Andreu,Z., Zavec,A.B., Borras,F.E., Buzas,E.I., Buzas,K., Casal,E., Cappello,F., Carvalho,J. *et al.* (2015) Biological properties of extracellular vesicles and their physiological functions. *J. Extracell. Vesicles*, **4**, doi:10.3402/jev.v4.27066.
56. Yokoi,A., Yoshioka,Y. and Ochiya,T. (2015) Towards the realization of clinical extracellular vesicle diagnostics: challenges and opportunities. *Expert Rev. Mol. Diagn.*, **12**, 1555–1566.
57. Gould,S.J. and Raposo,G. (2013) As we wait: coping with an imperfect nomenclature for extracellular vesicles. *J. Extracell. Vesicles*, **2**, doi:10.3402/jev.v2i0.20389.
58. Lobb,R.J., Becker,M., Wen,S.W., Wong,C.S., Wiegmans,A.P., Leimgruber,A. and Moller,A. (2015) Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J. Extracell. Vesicles*, **4**, doi:10.3402/jev.v4.27031.
59. He,M., Crow,J., Roth,M., Zeng,Y. and Godwin,A.K. (2014) Integrated immunoisolation and protein analysis of circulating exosomes using microfluidic technology. *Lab. Chip*, **14**, 3773–3780.
60. Welton,J.L., Webber,J.P., Botos,L.A., Jones,M. and Clayton,A. (2015) Ready-made chromatography columns for extracellular vesicle isolation from plasma. *J. Extracell. Vesicles*, **4**, doi:10.3402/jev.v4.27269.
61. Baranyai,T., Herczeg,K., Onodi,Z., Voszka,I., Modos,K., Marton,N., Nagy,G., Mager,I., Wood,M.J., El Andaloussi,S. *et al.* (2015) Isolation of exosomes from blood plasma: qualitative and quantitative comparison of ultracentrifugation and size exclusion chromatography methods. *PLoS One*, **10**, e0145686.
62. Greening,D.W., Xu,R., Ji,H., Tauro,B.J. and Simpson,R.J. (2015) A protocol for exosome isolation and characterization: evaluation of ultracentrifugation, density-gradient separation, and immunoaffinity capture methods. *Methods Mol. Biol.*, **1295**, 179–209.
63. Kalra,H., Drummen,G.P. and Mathivanan,S. (2016) Focus on extracellular vesicles: introducing the next small big thing. *Int. J. Mol. Sci.*, **17**, doi:10.3390/ijms17020170.
64. Szatanek,R., Baran,J., Siedlar,M. and Baj-Krzyworzeka,M. (2015) Isolation of extracellular vesicles: determining the correct approach (Review). *Int. J. Mol. Med.*, **36**, 11–17.
65. Zeringer,E., Barta,T., Li,M. and Vlassov,A.V. (2015) Strategies for isolation of exosomes. *Cold Spring Harb. Protoc.*, **2015**, 319–323.
66. Van Deun,J., Mestdagh,P., Sormunen,R., Cocquyt,V., Vermaelen,K., Vandesompele,J., Bracke,M., De Wever,O. and Hendrix,A. (2014) The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. *J. Extracell. Vesicles*, **3**, doi:10.3402/jev.v3.24858.
67. Rekker,K., Saare,M., Roost,A.M., Kubo,A.L., Zarovni,N., Chiesi,A., Salumets,A. and Peters,M. (2014) Comparison of serum exosome isolation methods for microRNA profiling. *Clin. Biochem.*, **47**, 135–138.
68. Royo,F., Zuniga-Garcia,P., Sanchez-Mosquera,P., Egia,A., Perez,A., Loizaga,A., Arceo,R., Lacasa,I., Rabade,A., Arrieta,E. *et al.* (2016) Different EV enrichment methods suitable for clinical settings yield different subpopulations of urinary extracellular vesicles from human samples. *J. Extracell. Vesicles*, **5**, doi:10.3402/jev.v5.29497.
69. Kowal,J., Arras,G., Colombo,M., Jouve,M., Morath,J.P., Primdal-Bengtson,B., Dingli,F., Loew,D., Tkach,M. and Thery,C. (2016) Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, E968–E977.
70. Willms,E., Johansson,H.J., Mager,I., Lee,Y., Blomberg,K.E., Sadik,M., Alaarg,A., Smith,C.I., Lehtio,J., El Andaloussi,S. *et al.* (2016) Cells release subpopulations of exosomes with distinct molecular and biological properties. *Sci. Rep.*, **6**, doi:10.1038/srep22519.
71. Paolini,L., Zendrini,A., Noto,G.D., Busatto,S., Lottini,E., Radeghieri,A., Dossi,A., Caneschi,A., Ricotta,D. and Bergese,P. (2016) Residual matrix from different separation techniques impacts exosome biological activity. *Sci. Rep.*, **6**, doi:10.1038/srep23550.
72. Abramowicz,A., Widlak,P. and Pietrowska,M. (2016) Proteomic analysis of exosomal cargo: the challenge of high purity vesicle isolation. *Mol. Biosyst.*, **12**, 1407–1419.
73. Müller,R., Häntzsch,M., Tolios,A., Beutner,F., Nagel,D., Thiery,J., Teupser,D. and Holdt,L.M. (2014) Comparison of whole blood RNA preservation tubes and novel generation RNA extraction kits for analysis of mRNA and MiRNA profiles. *PLoS One*, **9**, e113298.
74. Nikula,T., Mykkänen,J., Simell,O. and Lahesmaa,R. (2013) Genome-wide comparison of two RNA-stabilizing reagents for transcriptional profiling of peripheral blood. *Transl. Res.*, **161**, 181–188.
75. Hammerle-Fickinger,A., Riedmaier,I., Becker,C., Meyer,H.H., Pfaffl,M.W. and Ulbrich,S.E. (2010) Validation of extraction

methods for total RNA and miRNA from bovine blood prior to quantitative gene expression analyses. *Biotechnol. Lett.*, **32**, 35–44.
76. Jiang,Z., Uboh,C.E., Chen,J. and Soma,L.R. (2013) Isolation of RNA from equine peripheral blood cells: comparison of methods. *Springerplus*, **2**, 478–484.
77. Bayatti,N., Cooper-Knock,J., Bury,J.J., Wyles,M., Heath,P.R., Kirby,J. and Shaw,P.J. (2014) Comparison of blood RNA extraction methods used for gene expression profiling in amyotrophic lateral sclerosis. *PLoS One*, **9**, e87508.
78. Wieczorek,D., Delauriere,L., Schagat,T. and Promega Corporation. (2012) Methods of RNA Quality Assessment. http://www.promega.de/resources/pubhub/methods-of-rna-quality-assessment/.
79. McAlexander,M.A., Phillips,M.J. and Witwer,K.W. (2013) Comparison of methods for miRNA extraction from plasma and quantitative recovery of RNA from cerebrospinal fluid. *Front. Genet.*, **4**, 1–8.
80. Li,Y. and Kowdley,K.V. (2012) Method for microRNA isolation from clinical serum samples. *Anal. Biochem.*, **431**, 69–75.
81. Burgos,K.L., Javaherian,A., Bomprezzi,R., Ghaffari,L., Rhodes,S., Courtright,A., Tembe,W., Kim,S., Metpally,R. and Van Keuren-Jensen,K. (2013) Identification of extracellular miRNA in human cerebrospinal fluid by next-generation sequencing. *RNA*, **19**, 712–722.
82. Raabe,C.A., Tang,T.H., Brosius,J. and Rozhdestvensky,T.S. (2014) Biases in small RNA deep sequencing data. *Nucleic Acids Res.*, **42**, 1414–1426.
83. Linsen,S.E., de Wit,E., Janssens,G., Heater,S., Chapman,L., Parkin,R.K., Fritz,B., Wyman,S.K., de Bruijn,E., Voest,E.E. *et al.* (2009) Limitations and possibilities of small RNA digital gene expression profiling. *Nat. Methods*, **6**, 474–476.
84. Sterling,C.H., Veksler-Lublinsky,I. and Ambros,V. (2015) An efficient and sensitive method for preparing cDNA libraries from scarce biological samples. *Nucleic Acids Res.*, **43**, e1.
85. Blomquist,T., Crawford,E.L., Yeo,J., Zhang,X. and Willey,J.C. (2015) Control for stochastic sampling variation and qualitative sequencing error in next generation sequencing. *Biomol. Detect. Quantif.*, **5**, 30–37.
86. van Dijk,E.L., Jaszczyszyn,Y. and Thermes,C. (2014) Library preparation methods for next-generation sequencing: tone down the bias. *Exp. Cell Res.*, **322**, 12–20.
87. Kim,Y.K., Yeo,J., Kim,B., Ha,M. and Kim,V.N. (2012) Short structured RNAs with low GC content are selectively lost during extraction from a small number of cells. *Mol. Cell*, **46**, 893–895.
88. Jackson,T.J., Spriggs,R.V., Burgoyne,N.J., Jones,C. and Willis,A.E. (2014) Evaluating bias-reducing protocols for RNA sequencing library preparation. *BMC Genomics*, **15**, doi:10.1186/1471-2164-15-569.
89. 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.
90. Zhuang,F., Fuchs,R.T. and Robb,G.B. (2012) Small RNA expression profiling by high-throughput sequencing: implications of enzymatic manipulation. *J. Nucleic Acids*, **2012**, doi:10.1155/2012/360358.
91. Sun,G., Wu,X., Wang,J., Li,H., Li,X., Gao,H., Rossi,J. and Yen,Y. (2011) A bias-reducing strategy in profiling small RNAs using Solexa. *RNA*, **17**, 2256–2262.
92. Alon,S., Vigneault,F., Eminaga,S., Christodoulou,D.C., Seidman,J.G., Church,G.M. and Eisenberg,E. (2011) Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Res.*, **21**, 1506–1511.
93. Jayaprakash,A.D., Jabado,O., Brown,B.D. and Sachidanandam,R. (2011) Identification and remediation of biases in the activity of RNA ligases in small-RNA deep sequencing. *Nucleic Acids Res.*, **39**, e141.
94. Hafner,M., Renwick,N., Brown,M., Mihailovic,A., Holoch,D., Lin,C., Pena,J.T., Nusbaum,J.D., Morozov,P., Ludwig,J. *et al.* (2011) RNA-ligase-dependent biases in miRNA representation in deep-sequenced small RNA cDNA libraries. *RNA*, **17**, 1697–1712.
95. Sorefan,K., Pais,H., Hall,A.E., Kozomara,A., Griffiths-Jones,S., Moulton,V. and Dalmay,T. (2012) Reducing ligation bias of small RNAs in libraries for next generation sequencing. *Silence*, **3**, doi:10.1186/1758-907X-3-4.
96. Fuchs,R.T., Sun,Z., Zhuang,F. and Robb,G.B. (2015) Bias in ligation-based small RNA sequencing library construction is determined by adaptor and RNA structure. *PLoS One*, **10**, e0126049.
97. Baran-Gale,J., Kurtz,C.L., Erdos,M.R., Sison,C., Young,A., Fannin,E.E., Chines,P.S. and 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**, doi:10.3389/fgene.2015.00352.
98. Van Nieuwerburgh,F., Soetaert,S., Podshivalova,K., Ay-Lin Wang,E., Schaffer,L., Deforce,D., Salomon,D.R., Head,S.R. and Ordoukhanian,P. (2011) Quantitative bias in Illumina TruSeq and a novel post amplification barcoding strategy for multiplexed DNA and small RNA deep sequencing. *PLoS One*, **6**, e26969.
99. Lopez-Barragan,M.J., Quinones,M., Cui,K., Lemieux,J., Zhao,K. and Su,X.Z. (2011) Effect of PCR extension temperature on high-throughput sequencing. *Mol. Biochem. Parasitol.*, **176**, 64–67.
100. Kirino,Y. and Mourelatos,Z. (2007) Mouse Piwi-interacting RNAs are 2′-O-methylated at their 3′ termini. *Nat. Struct. Mol. Biol.*, **14**, 347–348.
101. Ohara,T., Sakaguchi,Y., Suzuki,T., Ueda,H. and Miyauchi,K. (2007) The 3′ termini of mouse Piwi-interacting RNAs are 2′-O-methylated. *Nat. Struct. Mol. Biol.*, **14**, 349–350.
102. Viollet,S., Fuchs,R.T., Munafo,D.B., Zhuang,F. and Robb,G.B. (2011) T4 RNA ligase 2 truncated active site mutants: improved tools for RNA analysis. *BMC Biotechnol.*, **11**, doi:10.1186/1472-6750-11-72.
103. 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.
104. Svec,D., Tichopad,A., Novosadova,V., Pfaffl,M.W. and Kubista,M. (2015) How good is a PCR efficiency estimate: recommendations for precise and robust qPCR efficiency assessments. *Biomol. Detect. Quantif.*, **3**, 9–16.
105. Orpana,A.K., Ho,T.H. and Stenman,J. (2012) Multiple heat pulses during PCR extension enabling amplification of GC-rich sequences and reducing amplification bias. *Anal. Chem.*, **84**, 2081–2087.
106. Sendler,E., Johnson,G.D. and Krawetz,S.A. (2011) Local and global factors affecting RNA sequencing analysis. *Anal. Biochem.*, **419**, 317–322.
107. Dabney,J. and Meyer,M. (2012) Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques*, **52**, 87–94.
108. Mamedov,T.G., Pienaar,E., Whitney,S.E., TerMaat,J.R., Carvill,G., Goliath,R., Subramanian,A. and Viljoen,H.J. (2008) A fundamental study of the PCR amplification of GC-rich DNA templates. *Comput. Biol. Chem.*, **32**, 452–457.
109. Aird,D., Ross,M.G., Chen,W.S., Danielsson,M., Fennell,T., Russ,C., Jaffe,D.B., Nusbaum,C. and Gnirke,A. (2011) Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. *Genome Biol.*, **12**, doi:10.1186/gb-2011-12-2-r18.
110. Head,S.R., Komori,H.K., LaMere,S.A., Whisenant,T., Van Nieuwerburgh,F., Salomon,D.R. and Ordoukhanian,P. (2014) Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*, **56**, 61–68.
111. Okino,S.T., Kong,M., Sarras,H. and Wang,Y. (2016) Evaluation of bias associated with high-multiplex, target-specific pre-amplification. *Biomol. Detect. Quantif.*, **6**, 13–21.
112. Kivioja,T., Vaharautio,A., Karlsson,K., Bonke,M., Enge,M., Linnarsson,S. and Taipale,J. (2012) Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods*, **9**, 72–74.
113. Islam,S., Zeisel,A., Joost,S., La Manno,G., Zajac,P., Kasper,M., Lonnerberg,P. and Linnarsson,S. (2014) Quantitative single-cell RNA-seq with unique molecular identifiers. *Nat. Methods*, **11**, 163–166.
114. 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. *et al.* (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**, e89.

115. Grundemann,D. and Schomig,E. (1996) Protection of DNA during preparative agarose gel electrophoresis against damage induced by ultraviolet light. *Biotechniques*, **21**, 898–903.
116. Cariello,N.F., Keohavong,P., Sanderson,B.J. and Thilly,W.G. (1988) DNA damage produced by ethidium bromide staining and exposure to ultraviolet light. *Nucleic Acids Res.*, **16**, 4157–4157.
117. Alba,F.J., Bermudez,A. and Daban,J.R. (2001) Green-light transilluminator for the detection without photodamage of proteins and DNA labeled with different fluorescent dyes. *Electrophoresis*, **22**, 399–403.
118. Seville,M. (2001) A whole new way of looking at things: the use of Dark Reader technology to detect fluorophors. *Electrophoresis*, **22**, 814–828.
119. Quail,M.A., Kozarewa,I., Smith,F., Scally,A., Stephens,P.J., Durbin,R., Swerdlow,H. and Turner,D.J. (2008) A large genome center's improvements to the Illumina sequencing system. *Nat. Methods*, **5**, 1005–1010.
120. Buehler,B., Hogrefe,H.H., Scott,G., Ravi,H., Pabon-Pena,C., O'Brien,S., Formosa,R. and Happe,S. (2010) Rapid quantification of DNA libraries for next-generation sequencing. *Methods*, **50**, S15–S18.
121. Laurie,M.T., Bertout,J.A., Taylor,S.D., Burton,J.N., Shendure,J.A. and Bielas,J.H. (2013) Simultaneous digital quantification and fluorescence-based size characterization of massively parallel sequencing libraries. *Biotechniques*, **55**, 61–67.
122. Robin,J.D., Ludlow,A.T., LaRanger,R., Wright,W.E. and Shay,J.W. (2016) Comparison of DNA quantification methods for next generation sequencing. *Sci. Rep.*, **6**, doi:10.1038/srep24067.
123. White,R.A. 3rd, Blainey,P.C., Fan,H.C. and Quake,S.R. (2009) Digital PCR provides sensitive and absolute calibration for high throughput sequencing. *BMC Genomics*, **10**, doi:10.1186/1471-2164-10-116.
124. Fahlgren,N., Sullivan,C.M., Kasschau,K.D., Chapman,E.J., Cumbie,J.S., Montgomery,T.A., Gilbert,S.D., Dasenko,M., Backman,T.W., Givan,S.A. *et al.* (2009) Computational and analytical framework for small RNA profiling by high-throughput sequencing. *RNA*, **15**, 992–1002.
125. Jiang,L., Schlesinger,F., Davis,C.A., Zhang,Y., Li,R., Salit,M., Gingeras,T.R. and Oliver,B. (2011) Synthetic spike-in standards for RNA-seq experiments. *Genome Res.*, **21**, 1543–1551.
126. Robasky,K., Lewis,N.E. and Church,G.M. (2014) The role of replicates for error mitigation in next-generation sequencing. *Nat. Rev. Genet.*, **15**, 56–62.
127. Fisher,R.A. (1935) *The Design of Experiments*. Oliver and Boyde, London.
128. Leek,J.T., Scharpf,R.B., Bravo,H.C., Simcha,D., Langmead,B., Johnson,W.E., Geman,D., Baggerly,K. and Irizarry,R.A. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.*, **11**, 733–739.
129. t Hoen,P.A., Friedlander,M.R., Almlof,J., Sammeth,M., Pulyakhina,I., Anvar,S.Y., Laros,J.F., Buermans,H.P., Karlberg,O., Brannvall,M. *et al.* (2013) Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nat. Biotechnol.*, **31**, 1015–1022.
130. Ross,M.G., Russ,C., Costello,M., Hollinger,A., Lennon,N.J., Hegarty,R., Nusbaum,C. and Jaffe,D.B. (2013) Characterizing and measuring bias in sequence data. *Genome Biol.*, **14**, doi:10.1186/gb-2013-14-5-r51.
131. Auer,P.L. and Doerge,R.W. (2010) Statistical design and analysis of RNA sequencing data. *Genetics*, **185**, 405–416.
132. McIntyre,L.M., Lopiano,K.K., Morse,A.M., Amin,V., Oberg,A.L., Young,L.J. and Nuzhdin,S.V. (2011) RNA-seq: technical variability and sampling. *BMC Genomics*, **12**, doi:10.1186/1471-2164-12-293.
133. Metpally,R.P., Nasser,S., Malenica,I., Courtright,A., Carlson,E., Ghaffari,L., Villa,S., Tembe,W. and Van Keuren-Jensen,K. (2013) Comparison of analysis tools for miRNA high throughput sequencing using nerve crush as a model. *Front. Genet.*, **4**, doi:10.3389/fgene.2013.00020.
134. Fonseca,N.A., Marioni,J. and Brazma,A. (2014) RNA-Seq gene profiling–a systematic empirical comparison. *PLoS One*, **9**, e107026.
135. Kircher,M., Sawyer,S. and Meyer,M. (2012) Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res.*, **40**, e3.
136. Ledergerber,C. and Dessimoz,C. (2011) Base-calling for next-generation sequencing platforms. *Brief. Bioinform.*, **12**, 489–497.
137. Erlich,Y., Mitra,P.P., delaBastide,M., McCombie,W.R. and Hannon,G.J. (2008) Alta-Cyclic: a self-optimizing base caller for next-generation sequencing. *Nat. Methods*, **5**, 679–682.
138. Rougemont,J., Amzallag,A., Iseli,C., Farinelli,L., Xenarios,I. and Naef,F. (2008) Probabilistic base calling of Solexa sequencing data. *BMC Bioinformatics*, **9**, doi:10.1186/1471-2105-9-431.
139. Fuller,C.W., Middendorf,L.R., Benner,S.A., Church,G.M., Harris,T., Huang,X., Jovanovich,S.B., Nelson,J.R., Schloss,J.A., Schwartz,D.C. *et al.* (2009) The challenges of sequencing by synthesis. *Nat. Biotechnol.*, **27**, 1013–1023.
140. Ewing,B. and Green,P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
141. Dohm,J.C., Lottaz,C., Borodina,T. and Himmelbauer,H. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic Acids Res.*, **36**, e105.
142. Nakamura,K., Oshima,T., Morimoto,T., Ikeda,S., Yoshikawa,H., Shiwa,Y., Ishikawa,S., Linak,M.C., Hirai,A., Takahashi,H. *et al.* (2011) Sequence-specific error profile of Illumina sequencers. *Nucleic Acids Res.*, **39**, e90.
143. Stein,A., Takasuka,T.E. and Collings,C.K. (2010) Are nucleosome positions in vivo primarily determined by histone-DNA sequence preferences? *Nucleic Acids Res.*, **38**, 709–719.
144. Harismendy,O., Ng,P.C., Strausberg,R.L., Wang,X., Stockwell,T.B., Beeson,K.Y., Schork,N.J., Murray,S.S., Topol,E.J., Levy,S. *et al.* (2009) Evaluation of next generation sequencing platforms for population targeted sequencing studies. *Genome Biol.*, **10**, doi:10.1186/gb-2009-10-3-r32.
145. Yang,X., Chockalingam,S.P. and Aluru,S. (2013) A survey of error-correction methods for next-generation sequencing. *Brief. Bioinform.*, **14**, 56–66.
146. Loman,N.J., Misra,R.V., Dallman,T.J., Constantinidou,C., Gharbia,S.E., Wain,J. and Pallen,M.J. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat. Biotechnol.*, **30**, 434–439.
147. Schirmer,M., D'Amore,R., Ijaz,U.Z., Hall,N. and Quince,C. (2016) Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. *BMC Bioinformatics*, **17**, doi:10.1186/s12859-016-0976-y.
148. Beltman,J.B., Urbanus,J., Velds,A., van Rooij,N., Rohr,J.C., Naik,S.H. and Schumacher,T.N. (2016) Reproducibility of Illumina platform deep sequencing errors allows accurate determination of DNA barcodes in cells. *BMC Bioinformatics*, **17**, doi:10.1186/s12859-016-0999-4.
149. Huse,S.M., Huber,J.A., Morrison,H.G., Sogin,M.L. and Welch,D.M. (2007) Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.*, **8**, doi:10.1186/gb-2007-8-7-r143.
150. Luo,C., Tsementzi,D., Kyrpides,N., Read,T. and Konstantinidis,K.T. (2012) Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS One*, **7**, e30087.
151. Team, R.C. (2014) *R Foundation for Statistical Computing*. Vienna.
152. Blankenberg,D., Kuster,G.V., Coraor,N., Ananda,G., Lazarus,R., Mangan,M., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.*, doi:10.1002/0471142727.mb1910s89.
153. Giardine,B., Riemer,C., Hardison,R.C., Burhans,R., Elnitski,L., Shah,P., Zhang,Y., Blankenberg,D., Albert,I., Taylor,J. *et al.* (2005) Galaxy: a platform for interactive large-scale genome analysis. *Genome Res.*, **15**, 1451–1455.
154. Goecks,J., Nekrutenko,A. and Taylor,J. (2010) Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.*, **11**, doi:10.1101/gr.4086505.
155. Yuan,T., Huang,X., Dittmar,R.L., Du,M., Kohli,M., Boardman,L., Thibodeau,S.N. and Wang,L. (2014) eRNA: a graphic user interface-based tool optimized for large data analysis from high-throughput RNA sequencing. *BMC Genomics*, **15**, doi:10.1186/1471-2164-15-176.
156. Kim,V.N., Han,J. and Siomi,M.C. (2009) Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.*, **10**, 126–139.

157. Yang,J., Ding,X., Sun,X., Tsang,S.-Y. and Xue,H. (2015) SAMSVM: A tool for misalignment filtration of SAM-format sequences with support vector machine. *J. Bioinform. Comput. Biol.*, **13**, doi:10.1142/s0219720015500250.
158. Hatem,A., Bozdağ,D., Toland,A.E. and Çatalyürek,Ü.V. (2013) Benchmarking short sequence mapping tools. *BMC Bioinformatics*, **14**, doi:10.1186/1471-2105-14-184.
159. Kong,Y. (2011) Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics*, **98**, 152–153.
160. Martin,M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, **17**, 10–12.
161. Lo,C.-C. and Chain,P.S.G. (2014) Rapid evaluation and quality control of next generation sequencing data with FaQCs. *BMC Bioinformatics*, **15**, doi:10.1186/s12859-014-0366-2.
162. Kelley,D.R., Schatz,M.C. and Salzberg,S.L. (2010) Quake: quality-aware detection and correction of sequencing errors. *Genome Biol.*, **11**, doi:10.1186/gb-2010-11-11-r116.
163. Gnerre,S., Maccallum,I., Przybylski,D., Ribeiro,F.J., Burton,J.N., Walker,B.J., Sharpe,T., Hall,G., Shea,T.P., Sykes,S. *et al.* (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 1513–1518.
164. Robertson,G., Schein,J., Chiu,R., Corbett,R., Field,M., Jackman,S.D., Mungall,K., Lee,S., Okada,H.M., Qian,J.Q. *et al.* (2010) De novo assembly and analysis of RNA-seq data. *Nat. Methods*, **7**, 909–912.
165. Del Fabbro,C., Scalabrin,S., Morgante,M. and Giorgi,F.M. (2013) An extensive evaluation of read trimming effects on Illumina NGS data analysis. *PLoS One*, **8**, e85024.
166. Cox,M.P., Peterson,D.A. and Biggs,P.J. (2010) SolexaQA: at-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, **11**, doi:10.1186/1471-2105-11-485.
167. Chen,C., Khaleel,S.S., Huang,H. and Wu,C.H. (2014) Software for pre-processing Illumina next-generation sequencing short read sequences. *Source Code Biol. Med.*, **9**, doi:10.1186/1751-0473-9-8.
168. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
169. Burrows,M. and Wheeler,D.J. (1994) *Technical Report*. Systems Research Center, Paöp Alto.
170. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, doi:10.1186/gb-2009-10-3-r25.
171. Li,H. and Durbin,R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**, 1754–1760.
172. Li,R., Yu,C., Li,Y., Lam,T.-W., Yiu,S.-M., Kristiansen,K. and Wang,J. (2009) SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics*, **25**, 1966–1967.
173. Li,H. and Homer,N. (2010) A survey of sequence alignment algorithms for next-generation sequencing. *Brief. Bioinform.*, **11**, 473–483.
174. Bao,S., Jiang,R., Kwan,W., Wang,B., Ma,X. and Song,Y.-Q. (2011) Evaluation of next-generation sequencing software in mapping and assembly. *J. Hum. Genet.*, **56**, 406–414.
175. Klus,P., Lam,S., Lyberg,D., Cheung,M., Pullan,G., McFarlane,I., Yeo,G.S.H. and Lam,B.Y.H. (2012) BarraCUDA - a fast short read sequence aligner using graphics processing units. *BMC Res. Notes*, **5**, doi:10.1186/1756-0500-5-27.
176. Luo,R., Wong,T., Zhu,J., Liu,C.-M., Zhu,X., Wu,E., Lee,L.-K., Lin,H., Zhu,W., Cheung,D.W. *et al.* (2013) SOAP3-dp: fast, accurate and sensitive GPU-based short read aligner. *PLoS One*, **8**, e65632.
177. Luo,R., Cheung,J., Wu,E., Wang,H., Chan,S.-H., Law,W.-C., He,G., Yu,C., Liu,C.-M., Zhou,D. *et al.* (2015) MICA: a fast short-read aligner that takes full advantage of many integrated core architecture (MIC). *BMC Bioinformatics*, **16**, doi:10.1186/1471-2105-16-s7-s10.
178. Griffiths-Jones,S. (2004) The microRNA registry. *Nucleic Acids Res.*, **32**, D109–D111.
179. Kozomara,A. and Griffiths-Jones,S. (2014) miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.*, **42**, D68–D73.
180. Nawrocki,E.P., Burge,S.W., Bateman,A., Daub,J., Eberhardt,R.Y., Eddy,S.R., Floden,E.W., Gardner,P.P., Jones,T.A., Tate,J. *et al.* (2015) Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.*, **43**, D130–D137.
181. Lipson,D., Speed,T.P. and Taub,M. (2010) Methods for allocating ambiguous short-reads. *Commun. Inform. Syst.*, **10**, 69–82.
182. Li,B., Ruotti,V., Stewart,R.M., Thomson,J.A. and Dewey,C.N. (2010) RNA-Seq gene expression estimation with read mapping uncertainty. *Bioinformatics*, **26**, 493–500.
183. Ji,Y., Xu,Y., Zhang,Q., Tsui,K.-W., Yuan,Y., Norris,C., Liang,S. and Liang,H. (2011) BM-map: Bayesian mapping of multireads for next-generation sequencing data. *Biometrics*, **67**, 1215–1224.
184. Motameny,S., Wolters,S., Nürnberg,P. and Schumacher,B. (2010) Next generation sequencing of miRNAs—strategies, resources and methods. *Genes*, **1**, 70–84.
185. Guo,L. and Chen,F. (2014) A challenge for miRNA: multiple isomiRs in miRNAomics. *Gene*, **544**, 1–7.
186. 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**, doi:10.1186/gb-2011-12-12-r126.
187. Neilsen,C.T., Goodall,G.J. and Bracken,C.P. (2012) IsomiRs–the overlooked repertoire in the dynamic microRNAome. *Trends Genet.*, **28**, 544–549.
188. Sai Lakshmi,S. and Agrawal,S. (2008) piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Res.*, **36**, D173–D177.
189. Anders,S., Pyl,P.T. and Huber,W. (2015) HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics*, **31**, 166–169.
190. Lawrence,M., Huber,W., Pagès,H., Aboyoun,P., Carlson,M., Gentleman,R., Morgan,M.T. and Carey,V.J. (2013) Software for computing and annotating genomic ranges. *PLoS Comput. Biol.*, **9**, e1003118.
191. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G. and Durbin,R. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
192. Zhao,S., Fung-Leung,W.-P., Bittner,A., Ngo,K. and Liu,X. (2014) Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One*, **9**, e78644.
193. Mortazavi,A., Williams,B.A., McCue,K., Schaeffer,L. and Wold,B. (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods*, **5**, 621–628.
194. Risso,D., Schwartz,K., Sherlock,G. and Dudoit,S. (2011) GC-content normalization for RNA-Seq data. *BMC Bioinformatics*, **12**, doi:10.1186/1471-2105-12-480.
195. Oshlack,A. and Wakefield,M.J. (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biol. Direct*, **4**, doi:10.1186/1745-6150-4-14.
196. Bullard,J.H., Purdom,E., Hansen,K.D. and Dudoit,S. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics*, **11**, doi:10.1186/1471-2105-11-94.
197. Bolstad,B.M., Irizarry,R.A., Astrand,M. and Speed,T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.
198. Robinson,M.D. and Oshlack,A. (2010) A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.*, **11**, doi:10.1186/gb-2010-11-3-r25.
199. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data. *Genome Biol.*, **11**, doi:10.1186/gb-2010-11-10-r106.
200. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.*, **15**, doi:10.1186/s13059-014-0550-8.
201. Leek,J.T. (2014) svaseq: removing batch effects and other unwanted noise from sequencing data. *Nucleic Acids Res.*, **42**, e161.
202. Dillies,M.-A., Rau,A., Aubert,J., Hennequet-Antier,C., Jeanmougin,M., Servant,N., Keime,C., Marot,G., Castel,D., Estelle,J. *et al.* (2013) A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Brief. Bioinform.*, **14**, 671–683.

203. Zyprych-Walczak,J., Szabelska,A., Handschuh,L., Górczak,K., Klamecka,K., Figlerowicz,M. and Siatkowski,I. (2015) The impact of normalization methods on RNA-Seq data analysis. *Biomed Res. Int.*, **2015**, doi:10.1155/2015/621690.
204. Oberg,A.L., Bot,B.M., Grill,D.E., Poland,G.A. and Therneau,T.M. (2012) Technical and biological variance structure in mRNA-Seq data: life in the real world. *BMC Genomics*, **13**, doi:10.1186/1471-2164-13-304.
205. Auer,P.L. and Doerge,R.W. (2011) A two-stage poisson model for testing RNA-Seq data. *Stat. Appl. Genet. Mol. Biol.*, **10**, doi:10.2202/1544-6115.1627.
206. Di,Y., Schafer,D.W., Cumbie,J.S. and Chang,J.H. (2011) The NBP negative binomial model for assessing differential gene expression from RNA-Seq. *Stat. Appl. Genet. Mol. Biol.*, **10**, doi:10.2202/1544-6115.1637.
207. Robinson,M.D., McCarthy,D.J. and Smyth,G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, **26**, 139–140.
208. Leng,N., Dawson,J.A., Thomson,J.A., Ruotti,V., Rissman,A.I., Smits,B.M.G., Haag,J.D., Gould,M.N., Stewart,R.M. and Kendziorski,C. (2013) EBSeq: an empirical Bayes hierarchical model for inference in RNA-seq experiments. *Bioinformatics*, **29**, 1035–1043.
209. Hardcastle,T.J. and Kelly,K.A. (2010) baySeq: empirical Bayesian methods for identifying differential expression in sequence count data. *BMC Bioinformatics*, **11**, doi:10.1186/1471-2105-11-422.
210. Li,J. and Tibshirani,R. (2013) Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data. *Stat. Methods Med. Res.*, **22**, 519–536.
211. Tarazona,S., Furió-Tarí,P., Turrà,D., Di Pietro,A., Nueda,M.J., Ferrer,A. and Conesa,A. (2015) Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.*, **43**, e140.
212. Smyth,G.K. (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, doi:10.2202/1544-6115.1027.
213. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
214. Law,C.W., Chen,Y., Shi,W. and Smyth,G.K. (2014) voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.*, **15**, doi:10.1186/gb-2014-15-2-r29.
215. Benjamini,Y. and Hochberg,Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B*, **57**, 289–300.
216. Soneson,C. and Delorenzi,M. (2013) A comparison of methods for differential expression analysis of RNA-seq data. *BMC Bioinformatics*, **14**, doi:10.1186/1471-2105-14-91.
217. Rapaport,F., Khanin,R., Liang,Y., Pirun,M., Krek,A., Zumbo,P., Mason,C.E., Socci,N.D. and Betel,D. (2013) Comprehensive evaluation of differential gene expression analysis methods for RNA-seq data. *Genome Biol.*, **14**, doi:10.1186/gb-2013-14-9-r95.
218. Seyednasrollah,F., Laiho,A. and Elo,L.L. (2015) Comparison of software packages for detecting differential expression in RNA-seq studies. *Brief. Bioinformatics*, **16**, 59–70.
219. Moulos,P. and Hatzis,P. (2015) Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns. *Nucleic Acids Res.*, **43**, e25.
220. Marcello Manfredi,E.R. (2013) Biomarkers discovery through multivariate statistical methods: a review of recently developed methods and applications in proteomics. *J. Proteomics Bioinform.*, **s3**, doi:10.4172/jpb.S3-003.
221. Eriksson,L., Antti,H., Gottfries,J., Holmes,E., Johansson,E., Lindgren,F., Long,I., Lundstedt,T., Trygg,J. and Wold,S. (2004) Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Anal. Bioanal. Chem.*, **380**, 419–429.
222. Herrero,J., Valencia,A. and Dopazo,J. (2001) A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics*, **17**, 126–136.
223. Tibshirani,R. (1988) Estimating transformations for regression via additivity and variance stabilization. *J. Am. Stat. Assoc.*, **83**, 394–405.
224. Stacklies,W., Redestig,H., Scholz,M., Walther,D. and Selbig,J. (2007) pcaMethods–a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, **23**, 1164–1167.
225. Le Cao,K.A., Gonzalez,I. and Dejean,S. (2009) integrOmics: an R package to unravel relationships between two omics datasets. *Bioinformatics*, **25**, 2855–2856.
226. Gibbons,F.D. and Roth,F.P. (2002) Judging the quality of gene expression-based clustering methods using gene annotation. *Genome Res.*, **12**, 1574–1581.
227. Bergkvist,A., Rusnakova,V., Sindelka,R., Garda,J.M., Sjogreen,B., Lindh,D., Forootan,A. and Kubista,M. (2010) Gene expression profiling–clusters of possibilities. *Methods*, **50**, 323–335.
228. Beyene,J., Tritchler,D., Bull,S.B., Cartier,K.C., Jonasdottir,G., Kraja,A.T., Li,N., Nock,N.L., Parkhomenko,E., Rao,J.S. *et al.* (2007) Multivariate analysis of complex gene expression and clinical phenotypes with genetic marker data. *Genet. Epidemiol.*, **31**(Suppl. 1), S103–S109.
229. Costa,I.G., Carvalho,F.d.A.T.d. and Souto,M.C.P.d. (2004) Comparative analysis of clustering methods for gene expression time course data. *Genet. Mol. Biol.*, **27**, 623–631.
230. Kubista,M., Andrade,J.M., Bengtsson,M., Forootan,A., Jonak,J., Lind,K., Sindelka,R., Sjoback,R., Sjogreen,B., Strombom,L. *et al.* (2006) The real-time polymerase chain reaction. *Mol. Aspects Med.*, **27**, 95–125.
231. Lee,G., Rodriguez,C. and Madabhushi,A. (2008) Investigating the efficacy of nonlinear dimensionality reduction schemes in classifying gene and protein expression studies. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **5**, 368–384.
232. Shaffer,R.E. (2002) Multi- and megavariate data analysis. Principles and applications, I. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold, Umetrics Academy, Umeå, 2001, ISBN 91-973730-1-X, 533pp. *J. Chemom.*, **16**, 261–262.
233. Barker,M. and Rayens,W. (2003) Partial least squares for discrimination. *J. Chemom.*, **17**, 166–173.
234. Le Cao,K.A., Boitard,S. and Besse,P. (2011) Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*, **12**, doi:10.1186/1471-2105-12-253.
235. Trygg,J. and Wold,S. (2002) Orthogonal projections to latent structures (O-PLS). *J. Chemom.*, **16**, 119–128.
236. Liu,Y. and Rayens,W. (2007) PLS and dimension reduction for classification. *Comput. Stat.*, **22**, 189–208.
237. Rajalahti,T., Arneberg,R., Kroksveen,A.C., Berle,M., Myhr,K.-M. and Kvalheim,O.M. (2009) Discriminating variable test and selectivity ratio plot: quantitative tools for interpretation and variable (biomarker) selection in complex spectral or chromatographic profiles. *Anal. Chem.*, **81**, 2581–2590.
238. Lewis,B.P., Burge,C.B. and Bartel,D.P. (2005) Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*, **120**, 15–20.
239. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
240. Betel,D., Wilson,M., Gabow,A., Marks,D.S. and Sander,C. (2008) The microRNA.org resource: targets and expression. *Nucleic Acids Res.*, **36**, D149–D153.
241. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, doi:10.1186/gb-2003-5-1-r1.
242. Lu,Y., Zhou,Y., Qu,W., Deng,M. and Zhang,C. (2011) A Lasso regression model for the construction of microRNA-target regulatory networks. *Bioinformatics*, **27**, 2406–2413.
243. Meyer,S.U., Stoecker,K., Sass,S., Theis,F.J. and Pfaffl,M.W. (2014) Posttranscriptional regulatory networks: from expression profiling to integrative analysis of mRNA and microRNA data. *Methods Mol. Biol.*, **1160**, 165–188.
244. 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. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
245. Meyer,S.U., Sass,S., Mueller,N.S., Krebs,S., Bauersachs,S., Kaiser,S., Blum,H., Thirion,C., Krause,S., Theis,F.J. *et al.* (2015) Integrative analysis of microRNA and mRNA data reveals an orchestrated function of microRNAs in skeletal myocyte

differentiation in response to TNF-alpha or IGF1. *PLoS One*, **10**, e0135284.
246. McCormick,K.P., Willmann,M.R. and Meyers,B.C. (2011) Experimental design, preprocessing, normalization and differential expression analysis of small RNA sequencing experiments. *Silence*, **2**, doi:10.1186/1758-907X-2-2.
247. Hruz,T., Laule,O., Szabo,G., Wessendorp,F., Bleuler,S., Oertle,L., Widmayer,P., Gruissem,W. and Zimmermann,P. (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinformatics*, **2008**, doi:10.1155/2008/420747.
248. Zimmermann,P., Laule,O., Schmitz,J., Hruz,T., Bleuler,S. and Gruissem,W. (2008) Genevestigator transcriptome meta-analysis and biomarker search using rice and barley gene expression databases. *Mol. Plant*, **1**, 851–857.

Appendix II

Substantial contributions by Dominik Buschmann:

- Conception and execution of experiments
- Curation, analysis and interpretation of data
- Drafting of figures
- Writing of the manuscript

Dominik Buschmann D. Buschmann

Michael W. Pfaffl Michael Pfaffl

JOURNAL OF EXTRACELLULAR VESICLES
2018, VOL. 7, 1481321
https://doi.org/10.1080/20013078.2018.1481321

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing

Dominik Buschmann [a,b], Benedikt Kirchner [b,c], Stefanie Hermann [b], Melanie Märte [d], Christine Wurmser [e], Florian Brandes [d], Stefan Kotschote[f], Michael Bonin[f], Ortrud K. Steinlein [a], Michael W. Pfaffl [b], Gustav Schelling [d] and Marlene Reithmair [a]

[a]Institute of Human Genetics, University Hospital, LMU Munich, Munich, Germany; [b]Division of Animal Physiology and Immunology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany; [c]Dr. von Hauner Children's Hospital, LMU Munich, Munich, Germany; [d]Department of Anesthesiology, University Hospital, LMU Munich, Munich, Germany; [e]Chair of Animal Breeding, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Freising, Germany; [f]IMGM Laboratories GmbH, Planegg, Germany

ABSTRACT

Extracellular vesicles (EVs) are intercellular communicators with key functions in physiological and pathological processes and have recently garnered interest because of their diagnostic and therapeutic potential. The past decade has brought about the development and commercialization of a wide array of methods to isolate EVs from serum. Which subpopulations of EVs are captured strongly depends on the isolation method, which in turn determines how suitable resulting samples are for various downstream applications. To help clinicians and scientists choose the most appropriate approach for their experiments, isolation methods need to be comparatively characterized. Few attempts have been made to comprehensively analyse vesicular microRNAs (miRNAs) in patient biofluids for biomarker studies. To address this discrepancy, we set out to benchmark the performance of several isolation principles for serum EVs in healthy individuals and critically ill patients. Here, we compared five different methods of EV isolation in combination with two RNA extraction methods regarding their suitability for biomarker discovery-focused miRNA sequencing as well as biological characteristics of captured vesicles. Our findings reveal striking method-specific differences in both the properties of isolated vesicles and the ability of associated miRNAs to serve in biomarker research. While isolation by precipitation and membrane affinity was highly suitable for miRNA-based biomarker discovery, methods based on size-exclusion chromatography failed to separate patients from healthy volunteers. Isolated vesicles differed in size, quantity, purity and composition, indicating that each method captured distinctive populations of EVs as well as additional contaminants. Even though the focus of this work was on transcriptomic profiling of EV-miRNAs, our insights also apply to additional areas of research. We provide guidance for navigating the multitude of EV isolation methods available today and help researchers and clinicians make an informed choice about which strategy to use for experiments involving critically ill patients.

ARTICLE HISTORY
Received 12 February 2018
Accepted 18 May 2018

KEYWORDS
Extracellular vesicle; exosome isolation; miRNA; small RNA sequencing; next-generation sequencing; sepsis; biomarker; precipitation; ultracentrifugation

Introduction

A multitude of isolation methods for extracellular vesicles (EVs) has been developed and commercialized in the last decade. Many methods claim rapid, reliable and highly efficient isolation from serum, yet there is no consensus on each method's suitability for scientific and clinical applications. Comparative data on methods for isolating EVs from patient biofluids are scarce, despite clear interest in utilizing EVs and their miRNA cargo for biomarker studies. Further, few attempts have been made to comprehensively analyse vesicular miRNAs in biofluid samples from critically ill patients, a population highly relevant to many clinical situations. This work compares five different methods of EV isolation and their suitability for miRNA-based biomarker discovery. We isolated serum EVs from sepsis patients and healthy volunteers, sequenced their small RNA cargo and performed differential miRNA expression analysis. Additional experiments assessed method-specific differences in vesicle composition and morphology. Our data reveal that precipitation and membrane affinity are highly suitable for both small RNA-Seq and patient classification based on cell-free miRNAs. Comparative evaluation demonstrates that miRNA yield correlates with robust separation of sepsis patients and healthy individuals, while vesicle purity seems less

CONTACT Dominik Buschmann dominik.buschmann@wzw.tum.de Institute of Human Genetics, University Hospital, LMU Munich, Germany
Supplementary data can be accessed here

relevant for RNA-based biomarker applications. Differences in size, quantity and composition of isolated vesicles indicate that each method captures distinctive, but partially overlapping EV populations accompanied by varying degrees of contamination with non-EV material.

EVs are intercellular communicators with key functions in physiological and pathological processes and have recently garnered significant interest as potential diagnostic and therapeutic agents. Rapidly increasing research in this field is accompanied by the demand for reproducible, time-efficient and economic isolation methods. A recent survey conducted by Gardiner et al. revealed that although ultracentrifugation (UC) remains the most commonly used isolation method, other approaches have gained preference when starting volume is limited [1]. Capturing EVs from blood-based biofluids such as serum and plasma is of particular interest for clinical applications. As a consequence, manufacturers offer a wide array of commercial isolation kits. These rely on principles ranging from filtration, precipitation and sedimentation to size-exclusion chromatography (SEC) and immunocapture.

One of the most important aspects of EV research is analysing their nucleic acid cargo, particularly small RNAs. These are commonly quantified by RT-qPCR or, increasingly, comprehensive transcriptomic profiling by next-generation sequencing (NGS, small RNA-Seq). Applications of EV transcriptomics range from basic research to biomarker discovery and drug development, making use of EVs as an easily accessible, enriched sampling fraction [2,3]. Inferring credible information from the transcriptome relies on precise quantification of target RNA, which in turn requires samples of high quality and integrity [4]. Additionally, methods for RNA extraction itself influence downstream analyses by yielding non-identical, kit-specific isolates [5]. This holds true particularly for extracellular RNA, which bears additional challenges such as low concentrations, diminished RNA integrity and high variability between individuals. Indeed, recent publications have highlighted the impact of cell-free RNA extraction strategies on small RNA-Seq, reporting quantitative and qualitative differences in resulting sequencing libraries [6,7].

Similarly, the impact of EV isolation strategies on RNA quantification assays has been demonstrated for cell culture supernatant [8], urine [9,10], milk [11] and serum [12]. Depending on the respective isolation principle, different populations of EVs with varying degrees of contamination seem to be isolated, resulting in only partially overlapping RNA profiles. Being able to detect specific RNA patterns in bulk populations of blood-derived EVs is challenging due to the vesicular secretome's complexity. Although most EVs in blood are secreted by erythrocytes, platelets and endothelial cells, various other tissues also secrete vesicles into the circulation, further complicating analysis [13,14]. Multiple classes of EVs are secreted from even one specific cell type, each carrying its individual RNA signature [15]. Beyond RNA profiles, kit-specific isolates also differ in EV composition, size, concentration, purity and functionality [16–20]. Selecting appropriate isolation methods is therefore a critical step in all areas of EV research.

There are excellent publications comparing different strategies of isolating EVs from human serum for RNA analyses. Rekker et al., Andreu et al. and Crossland et al. relied on RT-qPCR to profile vesicular miRNAs, comparing isolation based on UC, precipitation and filtration [12,21,22]. Helwa et al. isolated EVs from different starting volumes by precipitation and UC and quantified associated miRNAs by droplet digital PCR [23]. Analysing EV miRNAs using PCR-based assays is an important and well-established approach supported by excellent protocols and methods [24–26].

However, as NGS has become an increasingly popular downstream application to study miRNAs in EVs, it is crucial to define the EV isolation method most suitable for this particular technique. Several previous publications reported the feasibility and utility of sequencing small RNA in EVs isolated from serum, plasma, urine, and cell culture supernatant [2,27–30]. Small RNA-Seq experiments often focus on valuable applications such as liquid biopsy-based diagnostics and, consequently, clinical samples. Screening potential isolation methods should therefore include samples from healthy individuals as well as diseased patients, who often display severe anomalies in blood parameters. These matrix effects could conceivably interfere with EV isolation and hamper the transfer of methodologies from healthy to diseased subjects. Prime examples of critically ill patients are individuals suffering from sepsis and septic shock. This complex, life-threatening disease comes along with various clinical complications such as multiple organ failure, dysregulated coagulation and altered blood lipid profiles [31,32]. Findings derived from comparing EV isolation strategies for healthy donors, however, might not be readily transferred to such challenging samples. We therefore believe it is important to verify each method's applicability in samples relevant for the respective clinical situation.

The objective of the current study was to compare several methods of isolating EVs from healthy and septic sera and to identify the one most suitable for biomarker-focused small RNA-Seq in this population

of critically ill patients. Routine biomarker applications call for time-efficient, simple and streamlined procedures, ideally provided to clinical laboratories as one-box solutions. We did therefore not screen all potential combinations of EV isolation and RNA extraction methods but focused on either recommended RNA kits by the same manufacturer or combinations commonly used in the EV field. Additionally, isolates from each method were comparatively characterized in order to assess method-specific differences in captured EV populations and potential contaminating material.

Material and methods

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Medical Faculty of the University of Munich (protocol #551-14). Written informed consent and approval of a patient's legal representative was obtained when the patient lacked capacity to give informed consent for participation in the study. The study was carried out in accordance with approved guidelines, and all study samples were anonymized during analysis. Written informed consent for publication of blinded individual person's data was obtained from each participant or the patient's legal representative.

Patient recruitment

Four patients with sepsis and five patients in septic shock were included in the study and sex-matched to 10 healthy volunteers (Supplemental Table 1). Patients included in the study were >18 years of age and within 24 h of admission to the intensive care unit (ICU). Exclusion criteria were pregnancy, immunosuppression, leukopenia, haematological malignancies or the initiation of palliative care. Healthy volunteers were recruited from hospital personnel and by advertisement. Only volunteers with a Charlson Comorbidity Index [33] of ≤1 were included.

Sample collection

Blood was drawn from 20G catheters within the radial artery of sepsis patients on the day of admission to the ICU (day 0) and 24 h later (day 1). Healthy volunteers were sampled by venipuncture using 20G needles. In order to prevent haemolysis, aspiration was performed slowly and evenly for both procedures. Blood was collected in 9 ml serum tubes (S-Monovette, Sarstedt AG&Co) and centrifuged at 3400 *g* for 10 min at room temperature (RT) within 10 min of sampling. Resulting serum was aliquoted and stored at −80 °C.

Isolation of extracellular vesicles

EVs were isolated from serum using four commercially available isolation kits as well as differential UC (Table 1). One millilitre serum from each patient and volunteer was used as starting material for all isolation methods. EV isolation was performed as detailed below, following manufacturer's recommendations for pre-clearing of serum and subsequent steps. For all commercial isolation methods, we sequenced vesicular RNA from both patients and volunteers. Small RNA-Seq was performed for all samples except in the case of UC-derived EVs where it was only performed for healthy volunteers, but not for sepsis patients, as serum availability was limited. Serum EVs from day 0 were used for RNA extraction and small RNA-Seq.

Five septic shock patients and five matched volunteers from our small RNA-Seq cohort were selected for additional biological characterization of EVs. In these supplemental experiments, we isolated EVs from 1 ml serum sampled on day 1 of intensive care therapy in patients with sepsis. These day 1 EVs were isolated as described below, concentrated to 50 µl using Amicon Ultra-4 30 kDa NMWL spin filters (Merck Millipore) and split into separate aliquots for protein analysis and particle characterization, respectively. A schematic diagram that summarizes all steps of the EV isolation and characterization workflow is provided in Figure 1.

Precipitation

EVs were precipitated from 1 ml serum using the miRCURY Exosome Isolation Kit (Exiqon) according to the manufacturer's instructions. For RNA extraction, EV pellets were lysed with the provided miRCURY biofluid

Table 1. EV isolation methods and RNA extraction kits utilized in this study.

Principle of EV isolation	Method	RNA extraction kit
Precipitation	miRCURY Exosome Isolation Kit (Exiqon)	miRCURY RNA Isolation Kit – Biofluids (Exiqon)
Size-exclusion chromatography	Exo-spin Midi Columns (Cell Guidance Systems)	miRCURY RNA Isolation Kit – Biofluids (Exiqon)
Size-exclusion chromatography	qEV Columns (Izon Science)	miRCURY RNA Isolation Kit – Biofluids (Exiqon)
Membrane affinity	exoRNeasy Serum-Plasma Midi Kit (Qiagen)	exoRNeasy Serum/Plasma Midi Kit (Qiagen)
Sedimentation	Differential ultracentrifugation (Beckman Coulter Optima LE-80K)	exoRNeasy Serum/Plasma Midi Kit (Qiagen)

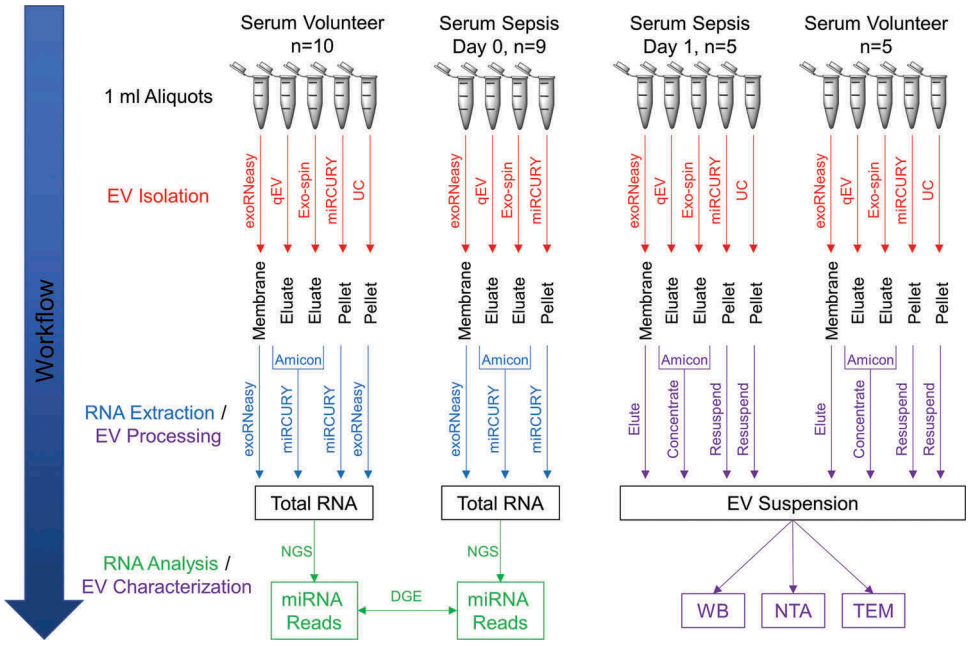

Figure 1. Schematic summary of EV isolation, RNA extraction and downstream analyses. EVs were isolated from human serum using five (healthy donors) or four (sepsis patients) different methods. After extracting total RNA from EV isolates, small RNA species were profiled by NGS. Differential expression of miRNAs between volunteers and patients was assessed to identify potential biomarker candidates. Sera from a subset of volunteers and patients were used to additionally characterize isolates from each method by Western blot (WB), nanoparticle tracking analysis (NTA) and transmission electron microscopy (TEM).

lysis solution. Pellets for biological characterization were resuspended in phosphate-buffered saline (PBS).

Size-exclusion chromatography

For Exo-spin (Cell Guidance Systems), EVs from 1 ml serum were purified on the provided columns and eluted in 3 ml particle-free PBS according to the manufacturer's quick start protocol. For qEV (Izon Science), columns were equilibrated, overlaid with 1 ml serum and flushed with particle-free PBS, collecting sequential fractions of 0.5 ml. Fractions 7–9 were pooled to maximize EV yield.

Membrane affinity

Pre-cleared serum was applied to exoEasy columns (Qiagen) as per the manufacturer's protocol. EVs were captured and washed using reagents provided in the kit. For RNA extraction, EVs bound to the membrane were lysed by adding QIAzol (Qiagen). Intact EVs for biological characterization were eluted from the column by addition of the provided buffer XE (analogous to the procedures in Qiagen's exoEasy kit).

Sedimentation

Serum was diluted 1:4 in PBS and subjected to low-speed centrifugation (12,000 *g*, 1 h, k-factor: 1401.3). EVs from the pre-cleared supernatant were then pelleted at 120,000 *g* for 14 h (k-factor: 139.7). All centrifugation steps were carried out at 4°C using an Optima LE-80K ultracentrifuge (Beckman Coulter) and a SW60 rotor. Pellets were lysed in QIAzol for RNA extraction or resuspended in PBS for EV characterization.

RNA extraction and characterization

Total RNA was extracted from day 0 serum EVs using commercial column-based kits listed in Table 1. For all extraction methods, RNA eluates were reapplied to the membrane for a second elution.

Precipitation

RNA was extracted from EV lysates using the corresponding miRCURY RNA Isolation Kit for biofluids. Procedures were carried out according to the manufacturer's protocol, and RNA was eluted in 30 µl nuclease-free water.

Size-exclusion chromatography

Eluted EVs (Exo-spin: 3 ml; qEV: 1.5 ml) were concentrated to 200 µl on Amicon Ultra-4 30 kDa NMWL spin filters. RNA was subsequently extracted from the concentrate using the miRCURY biofluids kit as described above.

Membrane affinity

RNA was extracted from EVs lysed in QIAzol using reagents provided in the exoRNeasy kit. Procedures were carried out according to the manufacturer's protocol, and RNA was eluted in 14 µl nuclease-free water.

Differential UC

Following lysis of pellets in QIAzol, RNA was extracted using the exoRNeasy kit as described above.

In order to compensate for the varying elution volumes, all RNAs were gently dried in a centrifugal evaporator and resuspended in 10 µl nuclease-free water. Yield and size distribution of EV-RNA were assessed by capillary electrophoresis on the 2100 Bioanalyzer (Agilent Technologies). We used the RNA 6000 Pico Assay (Agilent Technologies) to assess the total RNA profile including potential contaminations with cellular RNA.

Next-Generation Sequencing

EV-RNA from sepsis patients and healthy volunteers was profiled by small RNA-Seq. For all isolation methods, we used 60% (6 µl) of eluted total RNA as starting material. Library preparation was performed as described in Reithmair et al. [34], using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs Inc.). To compensate for the low RNA input, all adaptors and primers were diluted 1:2 in nuclease-free water. Size selection of PCR products was performed by high-resolution 4% agarose gel electrophoresis, selecting bands of 130–150 base pairs. Fragment sizes of purified libraries were assessed using capillary electrophoresis prior to 50 cycles of single-end sequencing on the HiSeq2500 (Illumina Inc.).

Data analysis

Sequencing data were processed as described elsewhere [35]. Briefly, FastQC (version 0.10.1) [36] was used to assess sequence length distribution and quality. Adaptor sequences were trimmed using Btrim [37], and all reads without adaptors were discarded. Additionally, reads shorter than 16 nt, probably degradation products from longer coding and non-coding RNA species, were excluded from the data set before proceeding to alignment [4]. To avoid false-positive hits during miRNA analysis, reads that mapped to sequences from human rRNA, tRNA, snRNA and snoRNA (obtained from RNAcentral) were initially removed from the data set [38]. Remaining reads were then aligned to human miRNA sequences in the most recent version (21) of miRBase [39]. Mapping was performed using Bowtie [40] and the "best" alignment algorithm, allowing one mismatch for alignment to both RNAcentral and miRBase. For all RNA classes, final read count tables were generated directly from Bowtie output by summing up all hits per sequence. Differential gene expression (DGE) analysis was subsequently performed via the Bioconductor Package DESeq2 (version 1.8.1) [41] using the included normalization strategy based on median ratios of mean miRNA expression and the Benjamini–Hochberg method to correct for false discovery. A $\log_2$ fold change $\geq|1|$ and an adjusted p-value of ≤ 0.05 were set as thresholds to identify significantly regulated miRNAs. Only transcripts with a baseMean ≥ 50 were included in the analysis. Hierarchical clustering (Euclidean distances, Ward's method), principal component analysis (regularized log-transformed, sizefactor-corrected counts obtained from DESeq2) and visualization of significantly regulated miRNAs in Venn diagrams were carried out in R (version 3.4.0) using the packages gplots, ggplots2, RColorBrewer, dendextend, ggfortify and VennDiagram [42–48]. Trimmed sequence reads were deposited in the European Nucleotide Archive under accession number PRJEB24913 (http://www.ebi.ac.uk/ena/data/view/PRJEB24913).

Nanoparticle tracking analysis

EV suspensions were diluted in particle-free PBS (prepared by a 120,000 g spin at 4°C for 14 h, k-factor: 231.6) and analysed using a NanoSight LM10 (Malvern Instruments GmbH) equipped with a 405-nm laser and a high-sensitivity sCMOS camera. Samples were introduced manually, and six videos of 45 s each were captured at a frame rate of 25 frames/second. With sample temperatures monitored manually, individual particles were tracked using NTA 3.0 software (Malvern Instruments GmbH) at camera level 10 and the Finite Track Length Adjustment (FTLA) algorithm. For analysis, we used a conservative detection threshold with blur and minimum track length set to auto and only considered captures with at least 2000 completed tracks. Starting from concentrations measured by NTA, initial particle concentrations in serum were calculated using the respective dilution factors for each sample as described elsewhere [49].

Transmission electron microscopy

EVs were fixed in 2% paraformaldehyde and adsorbed onto formvar/carbon-coated 200-mesh nickel grids (Electron Microscopy Sciences) for 15 min. Grids

were then washed with PBS, fixed in 2.5% glutaraldehyde for 5 min and washed with milliQ water. After performing negative staining with 2% uranyl acetate for 1 min, grids were washed again and air-dried overnight. Images were acquired on a Zeiss EM900 (Carl Zeiss Microscopy GmbH) with a wide-angle dual-speed 2K-CCD camera at 80 kV.

Western blot

EV samples were lysed in ice-cold Radioimmunoprecipitation Assay (RIPA) buffer on ice for 15 min intermitted by three bouts of sonication in a water bath. After centrifugation at 13,000 *g* for 10 min, protein concentration in the supernatant was analysed using Bicinchoninic Acid (BCA) assay (Sigma Aldrich). Input for exoRNeasy, Exo-spin, miRCURY and UC was normalized to 25 µg total protein. Due to very low protein concentrations, maximum volumes were loaded on the gel for qEV. For electrophoresis, samples were reduced in Laemmli buffer and heated at 70 °C for 10 min. Protein lysates for analysis of CD63 were incubated with non-reducing sample buffer at RT for 20 min. Proteins were separated using NuPAGE 4–12% Bis-Tris Gels (Invitrogen) prior to transfer to a 0.45 µm nitrocellulose membrane (GE Healthcare Life Sciences). Membranes were blocked with 1% non-fat milk powder in Phosphate Buffered Saline with Tween (PBST) for 1 h at RT and incubated with primary antibodies at 4°C overnight. Secondary antibodies were added for 1 h at RT. After washing with blocking buffer, blots were developed using the Clarity Western ECL Blotting Substrate Kit (Bio-Rad). Primary antibodies were from Abcam (mouse anti-TSG101 clone 4A10, ab83, 1:800, rabbit anti-Syntenin clone EPR8102, 1:5000, ab57113, 1:250, mouse anti-CD63, clone TS63, ab59479, 1:500, mouse anti-Human Serum Albumin clone 1A9, ab37989, 1:250), OriGene (rabbit anti-CD81, TA343598, 1:500) and Biomol (goat anti-Calnexin, WA-AF1179a, 1:2500). All marker proteins except CD63 were analysed using reducing conditions. HRP-conjugated secondary antibodies were purchased from Abcam (goat anti-Mouse, ab97040, 1:10,000, goat anti-Rabbit, ab97080, 1:10,000, rabbit anti-Goat, ab97105, 1:10,000).

Results

Analysis of isolation-specific EV-RNA composition by small RNA-Seq

Total EV-RNA was characterized by capillary electrophoresis, revealing major differences in quantity and size distribution across EV isolation strategies (Supplemental Figure 1). Similarly, sequencing of small RNA resulted in vastly differing total library sizes, ranging from 3.68E6 ± 1.72E6 reads (qEV sepsis) to 1.17E7 ± 3.76E6 reads (UC volunteer). Two EV samples precipitated from sepsis patients did not properly amplify during sequencing and were excluded from subsequent analyses. Method-dependent capture of miRNAs was assessed by aligning reads to miRBase and expressing mapped miRNAs as percentages of library size (Figure 2). miRNA enrichment was highest for precipitation-based EV isolation, followed by UC, membrane affinity and SEC. Even though library sizes were similar for UC, exoRNeasy and miRCURY, the latter displayed a 3.5–5-fold higher percentage of mapped miRNAs, respectively. For all isolation methods, relative frequencies of mapped miRNAs for sepsis patients were slightly lower than for volunteers (Figure 2). The top 10 most highly expressed miRNAs for each method are provided in Supplemental Table 2.

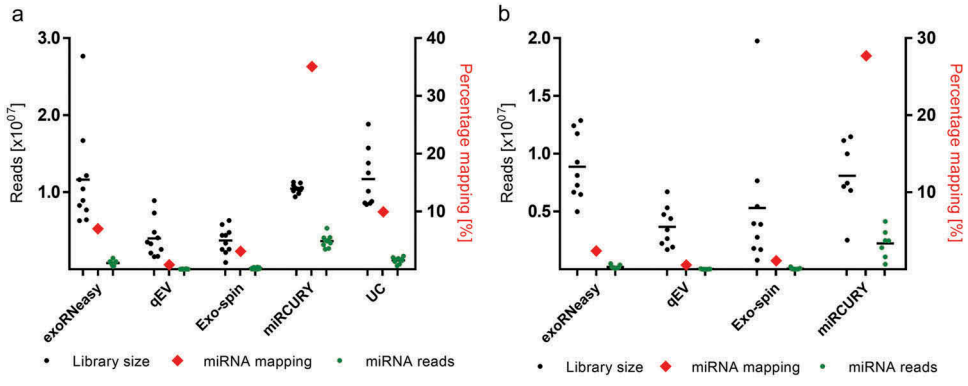

Figure 2. Mean library size and mapped miRNAs for EVs isolated from healthy volunteers (a) and sepsis patients (b). miRNA mapping frequencies (red diamonds) are expressed as percentages of total library size and plotted against the right x-axes. Enrichment of miRNA reads was highest for miRCURY (35.08% and 27.56% for volunteers and patients, respectively) and lowest for qEV (0.79% for volunteers and 0.57% for patients). All data are mean ± SD for 10 volunteers and 9 sepsis patients.

Similar differences were found when mapping reads to further classes of small non-coding RNA (Figure 3). Expressed as the ratio of non-target reads to miRNA reads, both SEC-based methods tended to isolate more rRNA fragments than other methods (Supplemental Figure 2). Increased frequencies of rRNA reads were also observed in sepsis EVs isolated by membrane affinity. Additionally, membrane affinity captured significantly more tRNA fragments than other methods from both septic and healthy EVs. SEC-based methods, particularly qEV, also isolated large numbers of fragments shorter than 15 nt. Mean library sizes, mapped miRNAs and results from DGE are provided in Supplemental Table 3.

EV-miRNAs from precipitation and membrane affinity separate volunteers and patients

DESeq2 was used to assess differential regulation of miRNA levels between sepsis patients and volunteers for commercial isolation kits. After applying stringent filtering criteria (baseMean ≥50, $\log_2$ fold change ≥|1|, adjusted p-value ≤0.05), we found 6 (qEV), 14 (Exospin), 60 (exoRNeasy) and 90 (miRCURY) miRNAs to be significantly regulated. While there was minimal overlap between all EV isolation strategies, most regulated miRNAs were unique for a specific isolation method (Figure 4). A common set of two significantly regulated miRNAs was detected for all EV isolation methods. Data for unfiltered differential expression analysis are provided in Supplemental Figure 3.

Similarities between miRNA patterns from each patient and isolation method were assessed by hierarchical clustering analysis (HCA) (Figure 5). Based on all miRNA reads, HCA separated isolation by precipitation, UC and membrane affinity from both SEC-based methods. Within these principal clusters, precipitation and membrane affinity flawlessly separated sepsis patients from healthy volunteers. Even though samples from precipitation and UC showed a high degree of similarity, UC volunteers were more closely related to miRCURY sepsis patients. Clustering of miRNAs from SEC isolation revealed substantial heterogeneity within and overlap between qEV and Exospin. Subsequently, these methods did not accurately distinguish volunteers from patients. This was also demonstrated by principal component analysis (Supplemental Figure 4), where separation of patient groups was achieved exclusively by miRCURY and exoRNeasy.

The number of differentially regulated miRNAs detected in DESeq2 analysis varied significantly

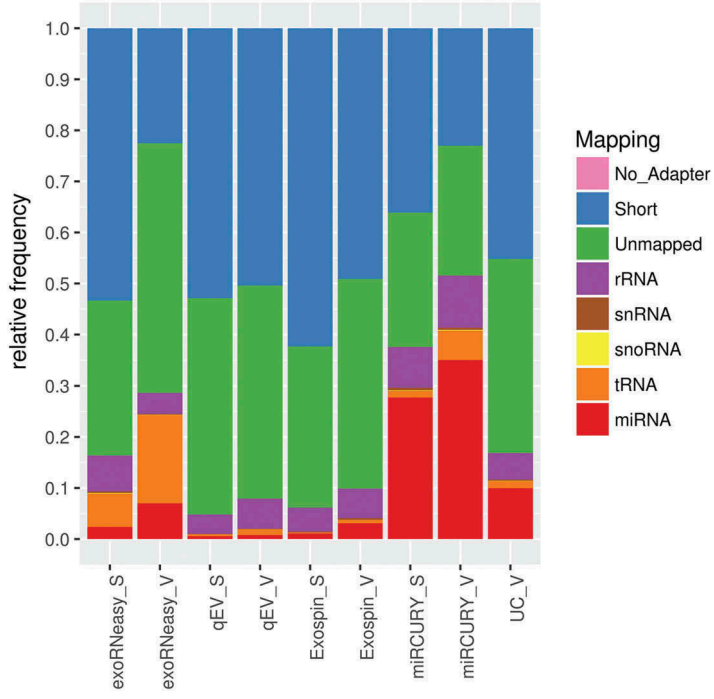

Figure 3. Mapping statistics for various classes of small non-coding RNA. Highest frequencies of miRNA mapping were observed in isolates from precipitation, sedimentation and membrane affinity. Both SEC-based methods were prone to capture short sequences, while libraries from membrane affinity-derived samples contained an increased share of tRNA fragments. Short: sequence is shorter than 15 nt; unmapped: sequence did not align to human rRNA, snRNA, snoRNA, tRNA or miRNA. Data are expressed as mean mapping percentages for 10 volunteers (V) and 9 sepsis patients (S).

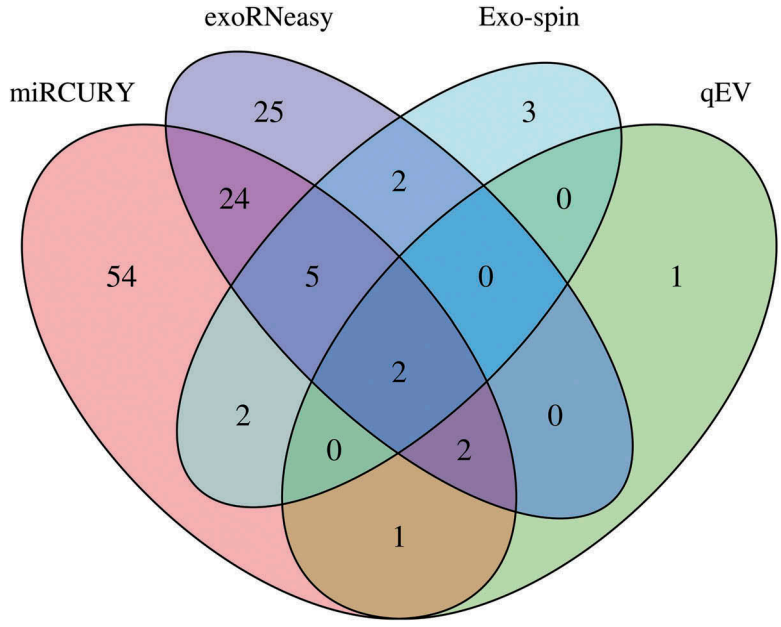

Figure 4. Differential expression of miRNAs in EVs isolated by commercial methods. Precipitation and membrane affinity yielded high numbers of differentially regulated miRNAs (miRCURY: 90; exoRNeasy: 60). Far fewer regulated miRNAs were detected in SEC-derived samples (Exo-spin: 14; qEV: 6). Two differentially regulated miRNAs were detected in EVs isolated by all methods. Data are filtered for baseMean ≥50, $\log_2$ fold change ≥|1| and adjusted *p*-value ≤0.05.

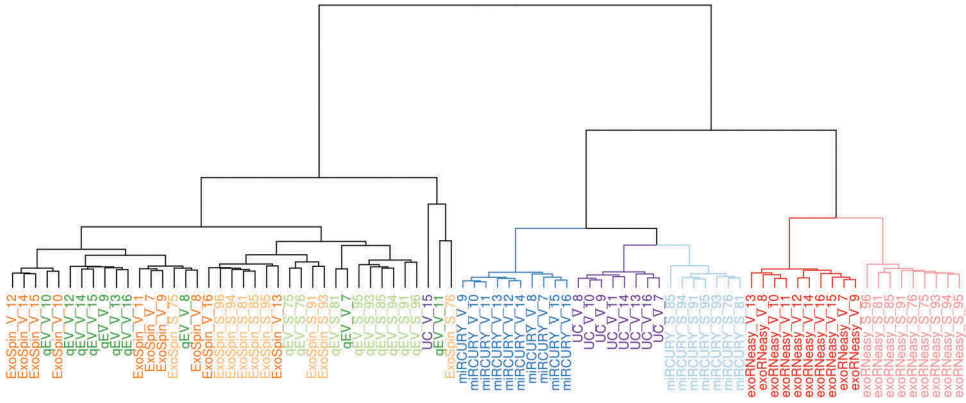

Figure 5. Hierarchical clustering analysis of miRNAs in EVs isolated by commercial methods. Samples split up into two clusters, separating precipitation and membrane affinity from both SEC-based methods. miRCURY (blue) and exoRNeasy (red) accurately distinguished between healthy volunteers (darker shades, V) and sepsis patients (lighter shades, S). miRNAs isolated from SEC-EVs (Exo-spin, qEV) showed noticeable heterogeneity and were less capable of separating volunteers and patients.

between isolation methods (Figure 4). Differentially expressed miRNAs as well as corresponding $\log_2$ fold changes and adjusted *p*-values for each method are provided in Supplemental Table 4. As predicted by sequencing output, methods yielding larger libraries also tended to result in more dysregulated miRNAs and greater fold changes. A common set of two miRNAs was found to be differentially expressed in EVs isolated by all methods. In EVs from sepsis patients, miR-122-5p was upregulated with $\log_2$ fold changes of 1.86 (Exo-spin) to 4.53 (exoRNeasy). miR-151a-3p, on the other hand, was downregulated in septic EVs, displaying $\log_2$ fold changes of −1.18 (miRCURY) to −1.65 (exoRNeasy) (Table 2).

EV populations isolated by divergent methods differ in size, concentration and purity

EVs captured by all isolation methods were analysed by NTA. Mean and mode particle diameters ranged from 104.46 ± 11.96 nm and 80.02 ± 10.12 nm (miRCURY volunteer) to 202.86 ± 10.70 nm and 174.48 ± 18.20 nm (exoRNeasy volunteer), respectively (Figure 6(a)). Size distributions for sepsis patients were slightly broader for all

Table 2. Common set of miRNAs differentially regulated between sepsis and healthy controls for all EV isolation methods.

	miR-122-5p	
Isolation method	log_2FC	p-adj
exoRNeasy	4.53	6.72E-17
qEV	2.11	2.73E-04
Exo-spin	1.86	2.72E-04
miRCURY	2.88	1.42E-07
	miR-151a-3p	
Isolation method	log_2FC	p-adj
exoRNeasy	−1.65	4.61E-10
qEV	−1.55	4.97E-06
Exo-spin	−1.19	3.92E-02
miRCURY	−1.18	5.72E-03

Log_2FC: log_2 fold change; p-adj: DESeq2-adjusted *p*-value.

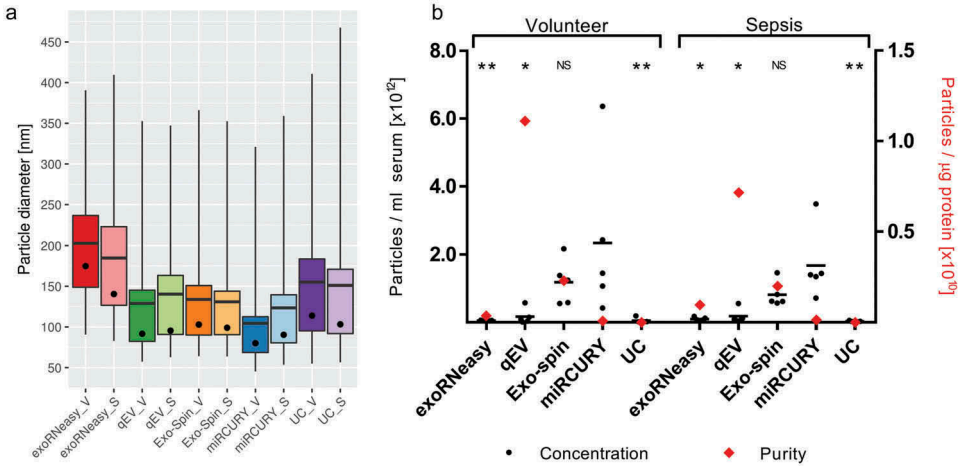

Figure 6. Analysis of EVs by NTA demonstrates differences in size distribution (a). Whiskers indicate 1st and 99th percentiles; line: mean diameter; dot: modal diameter; V: volunteer; S: sepsis patient. Precipitation- and membrane affinity-based methods isolated the smallest and largest EVs, respectively. Concentration and purity of isolated EVs differed depending on isolation strategies (b). Black bars indicate the absolute number of vesicles isolated from 1 ml serum; red diamonds plotted against the right x-axis represent vesicle purity defined as the particle to protein ratio. While precipitation most efficiently isolated EVs from serum, SEC-based isolation yielded fewer but highly pure vesicles. Asterisks indicate significant differences in particle numbers compared to miRCURY. *$p < 0.05$; **$p < 0.01$; NS: not significant. All data are mean ± SD for five volunteers and five sepsis patients.

isolation methods except qEV, but no significant differences in particle diameter were detected between volunteers and patients. The total number of particles isolated from 1 ml serum was highest for miRCURY, followed by Exo-spin, qEV, UC and exoRNeasy (Figure 6(b)). Additional plots for particle diameter and concentration are provided in Supplemental Figure 5.

Estimates for sample purity were calculated as ratios between NTA particle counts and protein concentrations [50]. While calculating these ratios does not necessarily help characterize a sample's EV fraction and their homogeneity, it provides a useful metric for assessing to which degree a sample is contaminated with non-EV protein. SEC-based isolation yielded isolates with significantly higher particle to protein ratios than all other methods, indicating less co-isolation of soluble protein (Figure 6(b)). Isolates derived from precipitation and UC, on the other hand, displayed the lowest ratios due to increased protein contamination. Additional data on particle size, concentration and purity are included in Supplemental Table 5.

Next, we assessed particle morphology by TEM. Confirming our findings from NTA, we detected particles with EV morphology and size for all isolation methods (Figure 7). While the majority of vesicles were less than 200 nm in diameter for all methods, precipitation-derived EVs seem to be additionally enriched for particles smaller than 100 nm.

Enrichment of contaminating soluble protein in EVs isolated by precipitation and UC

Prior to immunoblotting, total protein in EV lysates from each method was quantified by BCA assay.

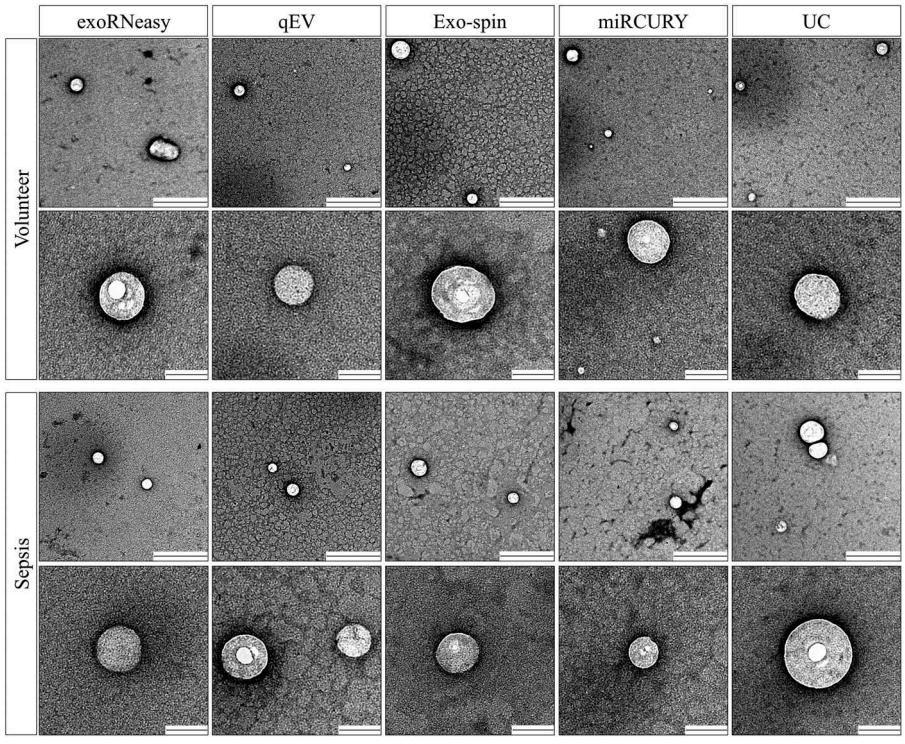

Figure 7. Morphology of serum EVs by transmission electron microscopy. Images are representative for three separate biological replicates for both volunteers (top panel) and sepsis patients (bottom panel). Scale bars are 500 nm (top row) and 100 nm (bottom row).

Similar to initial RNA concentrations and sequencing library sizes, striking differences in protein yield were observed (Supplemental Table 6). The amount of total protein in EV isolates ranged from 11.73 ± 5.18 µg (qEV, volunteer) to 26,202.95 ± 3904.31 µg (precipitation, volunteer). On average, precipitation and UC captured 50–80 times the amount of protein derived from SEC and membrane affinity isolations. Protein recovery from volunteer samples was significantly higher ($p < 0.05$) for exoRNeasy, miRCURY and UC, but failed to reach significance for Exo-spin ($p = 0.87$). Isolation by qEV captured significantly more ($p = 0.01$) protein from sepsis patients.

EV-specific proteins as well as negative markers were assessed by Western blot (Figure 8). CD63, a commonly used vesicle marker, was detected as a broad smear between 30 and 60 kDa, indicating differentially glycosylated forms of the protein. EVs isolated by membrane affinity showed high signal intensities for CD63, while both SEC-based methods resulted in weaker bands. No CD63 was detected for isolation by precipitation and UC. A similar pattern was observed for syntenin, showing clear signals for exoRNeasy, qEV and Exo-spin, but not for miRCURY and UC. EV markers CD81 and TSG101 were not detected for any isolation strategy.

Nonspecific staining of total EV protein by Ponceau S revealed a very prominent band at 60–70 kDa for Exo-spin, miRCURY and UC (Supplemental Figure 6), potentially indicating co-isolation of non-vesicular material. Human serum albumin (HSA), the most abundant blood protein, was selected as a likely candidate for protein contamination in EV preparations. Indeed, Western blot analysis revealed extraordinarily high HSA levels for miRCURY and UC, but also exoRNeasy and Exo-spin (Figure 8). Only minor amounts of HSA were detected for qEV isolations.

Protein lysates were also analysed for contamination with cellular fragments as indicated by the endoplasmic reticulum protein calnexin. In contrast to HSA, no calnexin signal was detected for any of the isolation methods. These findings hint at a contamination with soluble proteins, but not with non-vesicular membrane fragments. None of the detected protein markers showed significant enrichment for either volunteers or sepsis patients.

Increased contamination with soluble proteins such as HSA leads to an underrepresentation of marker proteins in EV lysates. As no EV markers were detected for miRCURY and UC, we increased the input for immunoblotting to 50 µg total protein. Additionally, EVs isolated by these techniques were further purified

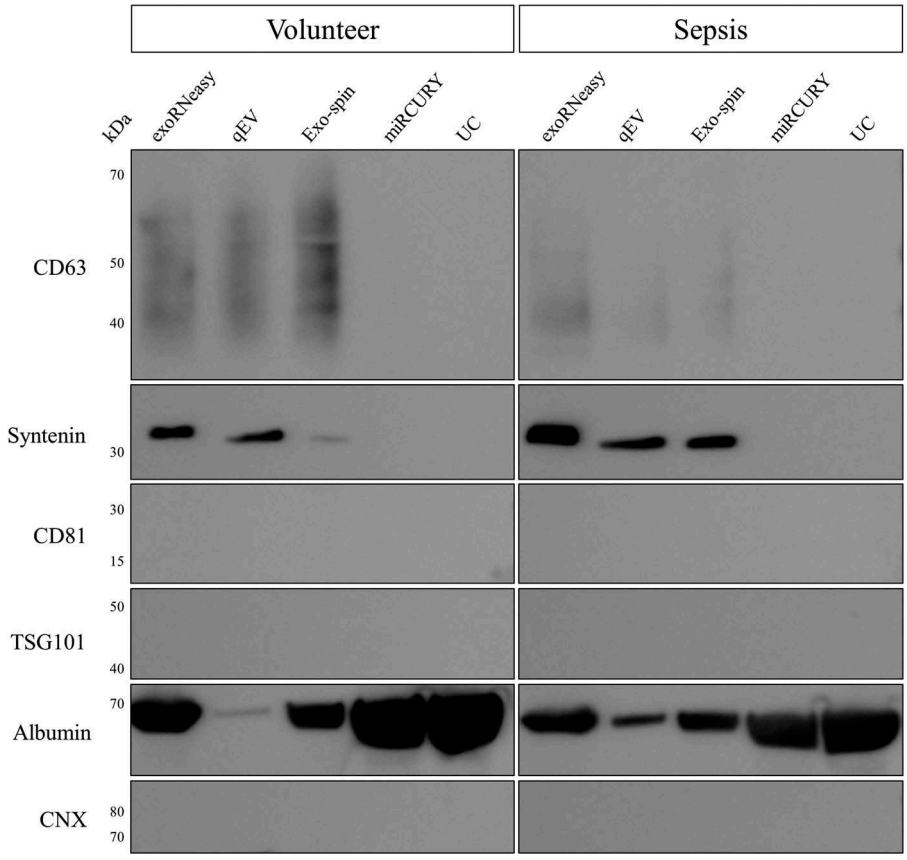

Figure 8. Analysis of marker proteins in EVs from volunteers (left) and sepsis patients (right). EV markers CD63 and syntenin were detected in vesicles isolated by membrane affinity (exoRNeasy) and SEC (qEV, Exo-spin), but not precipitation (miRCURY) and UC. All EV isolates were negative for TSG101, CD81 and calnexin. Significant albumin contamination of EVs was found for non-SEC isolation methods. Results are representative for three separate biological replicates for both volunteers and sepsis patients.

by iodixanol density gradient centrifugation to remove soluble proteins. Even though increasing total protein did not lead to the detection of protein markers (Supplemental Figure 7), floatation into a density gradient effectively separated contaminating HSA from EVs captured by sedimentation and precipitation. While the majority of HSA was retained in fractions of 1.02–1.07 g/ml, EV markers syntenin and CD63 were identified in a fraction of 1.18 g/ml, corresponding to previously reported floatation densities of blood-derived EVs [51] (Figure 9).

Discussion

EVs fascinate researchers in basic science and translational applications alike, but our understanding of EV biogenesis, secretion, tissue retention and potential therapeutic use depends on the ability to isolate and characterize specific, well-defined populations of vesicles. The question as to which EV isolation method to utilize for a given downstream application is a frequent subject of controversial debate that has yet to be settled. In this study, we qualitatively and quantitatively compared EV isolation strategies based on different physiochemical mechanisms ranging from sedimentation and precipitation to membrane affinity and SEC. Importantly, we used serum as a biofluid relevant to clinical applications and included diseased patients as well as healthy volunteers. As isolation methods need to be validated using clinical samples, we opted for sepsis patients, who represent a prime example for both interindividual variability and complex aberrations in blood parameters.

High-throughput sequencing has evolved into a mainstream method of analysing nucleic acids. It allows precise quantification of miRNAs and sheds light on RNA composition, co-isolation of non-target molecules and novel classes of non-coding RNA. Using Illumina small RNA-Seq, we found that vesicular RNA profiles greatly depend on the respective EV isolation strategy. While the methods less specific for EVs (precipitation, sedimentation and membrane affinity) resulted in higher absolute and relative numbers of mapped miRNAs, a more stringent size selection on EVs (SEC) led to lower mapping rates and an abundance of short RNA fragments in preparations

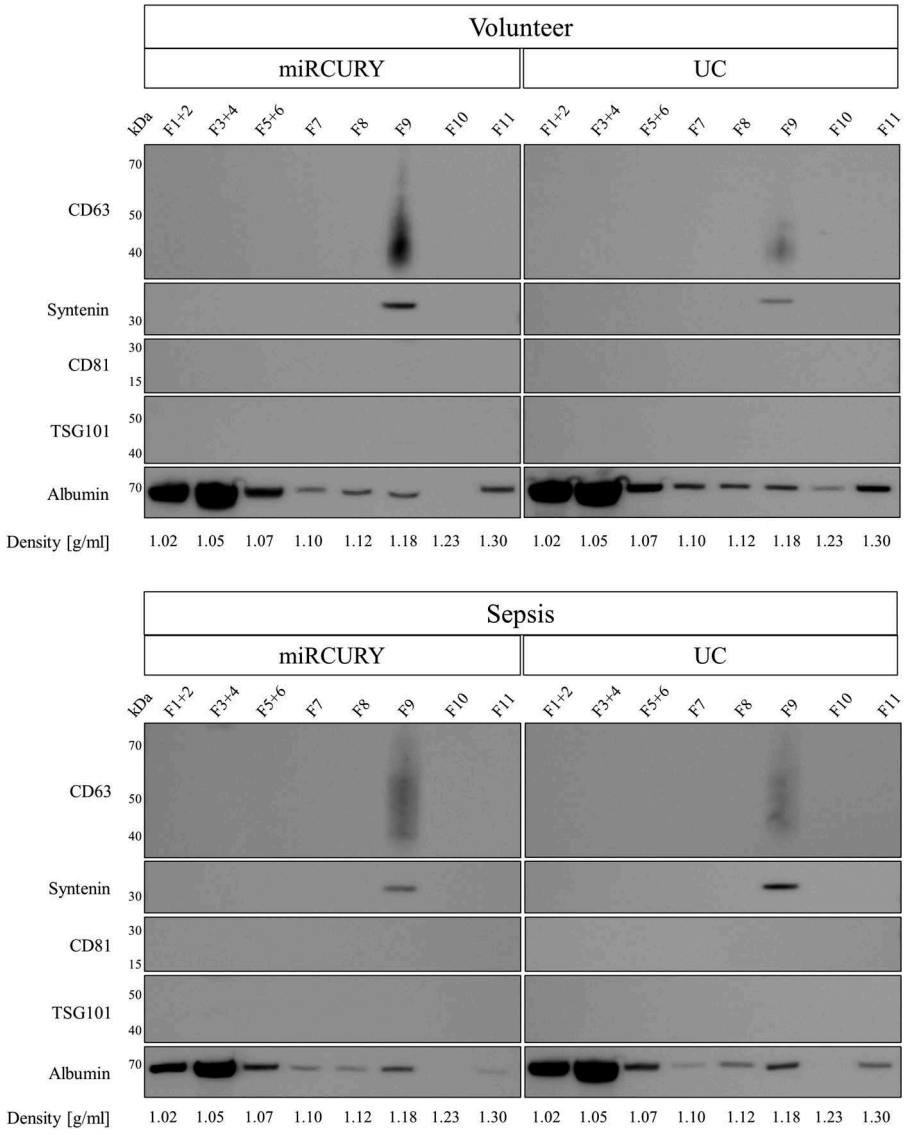

Figure 9. Analysis of EV markers and human serum albumin in EVs isolated by precipitation and sedimentation and further purified by iodixanol density gradient centrifugation. CD63 and syntenin were detected in a density fraction of 1.18 g/ml, while the majority of albumin floated in fractions of 1.02–1.05 g/ml. Results are representative for two separate biological replicates for both volunteers (top panel) and patients (bottom panel).

(Figure 3). Furthermore, isolates from different methods were reproducibly enriched in fragments of additional short non-coding RNA classes such as tRNA. We have used two different RNA extraction kits in this study, and varying combinations of EV isolation and RNA extraction methods might yield slightly different results [5,52]. However, based on the vastly different characteristics of the material captured by each method, we believe that EV isolation itself has a far greater impact on downstream RNA analysis than the respective extraction method (Figure 6).

The prime objective of this study was to assess EV isolation methods regarding their suitability for miRNA-based biomarker studies. In our data, the ability to separate healthy individuals from diseased patients strongly correlated with sequencing output for a given method: EV isolation based on precipitation and membrane affinity resulted in higher absolute numbers of mapped miRNA reads, more candidates in DGE analysis and enhanced separation of groups in hierarchical clustering. In contrast, low-output methods (SEC) were also able to identify a core set of two miRNAs differentially regulated regardless of isolation strategy but did not reliably assign individual samples to the correct study population (Figure 5). Interestingly, only SEC-based methods generated similar or greater numbers of reads from diseased samples, while precipitation and membrane affinity seemed to

work more effectively for healthy individuals (Supplemental Table 3). This correlated only partially with particle data from NTA, which indicate that both qEV and exoRNeasy recover more particles from sepsis sera. Even though assessing sequencing library size as a standalone metric is of limited use, results from differential expression analysis correlated with higher sequencing output and more diverse libraries in our data. Library composition might, however, differ depending on sample preparation, as demonstrated by Huang et al. [28]. As different library preparation kits tend to preferentially capture specific RNA sequences, NGS data in different experiments might be biased for certain transcripts. Highly abundant miRNAs are less affected by library preparation-induced biases, which might be more problematic for low-abundance transcripts or biomarker studies in diseases with less extreme alterations in miRNA expression. In conclusion, isolation methods less specific for EVs yielded more RNA, better libraries and, therefore, increased separation of patient groups. More specific methods, which purify EVs rather than enrich cell-free material in general, resulted in less complex libraries, fewer miRNA reads and poor performance in clustering. It is worth noting that increasing sequencing depth for RNA associated with pure vesicle preparations might in turn improve results from DGE analysis.

Observed variations in RNA composition could be attributed to a number of factors including a method's efficiency of isolating EVs, isolation of non-overlapping subpopulations of vesicles and co-isolation of non-vesicular RNA. As blood samples from sepsis patients and healthy donors were drawn using slightly different methods, potential sampling-related batch effects might confound genuine disease effects. Yet, as shown in hierarchical clustering (Figure 5), poor separation of sample groups for SEC-based isolation indicates the absence of a systematic batch effect caused by blood sampling. Additionally, in a separate study on paired samples collected by arterial and venous catheters from the same donors, we couldn't detect any significant sampling-dependent differences in EV morphology and associated miRNA profiles (manuscript in preparation). Differential expression of miRNAs in this study is therefore likely to be caused by sepsis itself, rather than by collection methods.

Several of the miRNAs we found to be differentially expressed (Supplemental Table 4) have previously been associated with sepsis and inflammation. Reithmair et al. and Wang et al. reported increased levels of circulating miR-193b-5p in EVs from sepsis patients and a strong association with disease mortality [34,53]. miR-30a-5p, upregulated in samples from exoRNeasy, miRCURY and Exo-spin, is induced by inflammatory stimuli and discriminates sepsis from non-infectious systemic inflammatory response syndrome [54,55]. Several groups reported that treating macrophages with lipopolysaccharides (LPS) increased expression of miR-155-5p, which, in turn, dampened the immune response, protected septic mice from cardiac dysfunction and improved survival [56–58]. Additionally, circulating levels of miR-155 were shown to correlate with disease severity and poor prognosis in a cohort of 60 sepsis patients [59]. In mice and rats, miR-150-5p, upregulated in samples captured by membrane affinity, was increased by polymicrobial sepsis and LPS treatment, respectively [60,61]. Vasilescu et al. reported plasma levels of miR-150-5p to correlate with sepsis aggressiveness in a cohort of 17 sepsis patients [62]. In concordance with previous findings in critically ill patients, we detected miR-122-5p as upregulated in samples derived from all isolation methods. This miRNA is commonly considered to be liver specific, and increased serum concentrations have been reported in cases of liver injury and hepatotoxicity [63]. Increased serum levels of miR-122-5p were also reported in sepsis patients, correlating with liver damage, coagulation disorders and mortality [53,64,65]. A more recent publication by Roderburg et al., however, demonstrated that miR-122-5p expression in critically ill patients was dysregulated by hepatic injury alone, independent of an infectious state [66]. Exclusively focusing on the septic shock patients in our cohort might have tightened expression patterns of disease-related miRNAs, but as our goal was to assess EV isolation methods capable of also detecting the less severe stage, we did not perform separate analyses on this subgroup. In conclusion, our findings match previous reports about altered profiles of circulating miRNAs in critically ill patients and animal models of sepsis, and crude preparations of cell-free RNA allow for more robust detection of disease-associated differential expression.

In an attempt to shed light on the nature of vesicles isolated by each method, we characterized intact particles using NTA. Unsurprisingly, all methods used in this study isolated particles in the size range of small EVs. Mean particle diameters, however, differed significantly: membrane affinity captured EVs with diameters close to 200 nm, while precipitation isolated vesicles with an exosome-like diameter of 100–120 nm. These differences might be due to capturing different EV populations or manipulation of originally identical EVs during isolation by aggregation [67] or coating with serum proteins [68]. In line with our findings, Stranska et al. recently reported larger particle diameters for EV samples isolated from human plasma by membrane affinity compared to SEC [69]. For most kits, variability of particle

diameters was greater in sepsis samples, indicating disease-specific changes in circulating vesicles, or interferences caused by matrix effects in serum from critically ill patients. It is also conceivable that an increased proportion of immune cell-derived EVs or bacterial outer membrane vesicles, typically ranging from 20 to 300 nm in diameter, might contribute to the broader range of particles recovered from septic sera [70,71].

Quantitative analysis of particles revealed another layer of complexity, as precipitation captured both the smallest and the highest number of EVs, whereas particles isolated by membrane affinity were larger and much less abundant. For a given isolation method, seemingly high standard deviations of particle sizes can most likely be attributed to endogenous variability within patient groups. Recent work by Eitan et al. revealed individual-specific set points for EV concentration and composition, indicating the need for larger cohorts in descriptive and clinical EV studies [49]. The slightly decreased concentration of EVs recovered from sepsis patients by most methods could be due to less efficient capture from patient sera, or genuinely lower EV concentrations in serum caused by decreased secretion, increased clearance from the bloodstream, dilution of EVs by therapeutic blood products administered in the ICU or a combination thereof. It should be noted, however, that lipoproteins such as low-density lipoprotein (LDL) outnumber EVs in cell-free blood by at least one order of magnitude and are known to co-purify with EVs [72]. As LDL and other common contaminants such as protein aggregates mimic characteristics of genuine EVs, particle quantification using NTA might overestimate EV concentrations in low-purity preparations and skew isolation-dependent size profiles [73–75]. Interestingly, the ability of extracellular RNA to separate healthy and diseased individuals did not seem to be tied to the diameter of the corresponding EVs since methods isolating both very small (precipitation) and very large (membrane affinity) EVs performed best in differential expression analysis. As these methods are prone to contamination with soluble material (Figure 8), we cannot rule out the possibility that separation of patient groups is based on RNA not associated with EVs, but co-isolated from the non-vesicular serum compartment. SEC-based methods, on the other hand, isolated EVs contaminated with large quantities of short RNA fragments, rendering them less suitable for robust classification of patient samples. This discovery raises questions about the origin of these fragments and whether they are encapsulated in EVs or co-isolated from non-vesicular blood compartments. While part of the population of short fragments might be derived from RNA degradation and library preparation artefacts such as adaptor dimers, non-human RNA sequences could also contribute to this category. Certain bacteria secrete cell-free RNA, some of it vesicle-associated, which might have been captured from septic sera by the EV isolation methods used here [76,77]. Additionally, reads categorized as unmapped (Figure 3) might be derived from bacterial RNA, even though the frequency of unmapped reads was not significantly increased in sepsis samples. As differential contamination of libraries with non-human sequences would impair normalization to library size or reads per million, we strictly normalized expression values for confirmed human miRNAs. Analysing particle morphology by TEM demonstrated that all methods isolated vesicles in the 100–200 nm size range. Isolates from precipitation did not display significantly more vesicles than other methods, indicating that non-vesicular particles such as protein aggregates might have contributed to increased particle counts in NTA [78]. The number of particles per field shown in Figure 7 does not necessarily correlate to particle concentrations from NTA since different dilution factors were used for samples from each isolation method in TEM imaging. In accordance with recent reports [79], we also observed double vesicles and vesicles containing two or more smaller vesicles.

Further profiling of EV isolates demonstrated an enrichment in CD63 and syntenin for SEC-based and membrane affinity-based isolations, but not for precipitation and UC. Potentially owing to insufficient starting material or technical factors, and in contrast to preexisting publications [23,34,80], we did not detect TSG101 and CD81 for any isolation method. Similar findings were recently presented in a publication by Stranska et al., which demonstrated the absence of CD81 and TSG101 in plasma EVs isolated by membrane affinity [69]. Additionally, recent advances in the field have demonstrated that so-called exosome markers can also be present on other classes of EVs and that EV isolates are a heterogeneous mixture of various subpopulations with specific protein profiles [81,82]. It is therefore conceivable that isolation methods are biased towards only partially overlapping EV populations, resulting in different protein profiles. Additionally, modifications of EVs during isolation, including coating with precipitation polymers or serum proteins, might mask antigens and impede detection of marker proteins as observed elsewhere [83,84]. Even though increasing the input for protein analysis helped other investigators to detect markers in crude EV samples [23], it was not sufficient for samples from precipitation and sedimentation in our study. Additional purification by density gradient centrifugation, however, established the presence of EV markers syntenin and CD63 in fractions with a density of 1.18 g/ml (Figure 9). While serum albumin and other soluble proteins overpowered EV markers in crude isolates, floatation into a density gradient could be used to specifically purify vesicles from pre-enriched samples. The endoplasmic reticulum marker calnexin could

not be detected in lysates from any of the isolation principles, indicating the absence of contaminating cellular fragments and vesicles not originating from endosomes [85]. Contamination of EV preparations with highly abundant blood proteins is a well-known problem, particularly for proteomic analyses [86]. In line with earlier publications, we demonstrate that precipitation- and sedimentation-based isolations co-fractionate significant amounts of serum albumin that mask genuine EV-enriched proteins. Additionally, we herein confirm previous findings [69,83,87] that SEC-based methods represent an efficient way of removing high-abundance serum proteins (Supplemental Figure 6), trading decreased vesicle yield for higher purity [88].

Highly pure and well-defined populations of EVs, however, might not be prerequisite for all research questions. While mechanistic and descriptive scrutinies are crucial for basic research and developing EV therapeutics, biomarker applications rely heavily on pronounced and reproducible changes in the molecules of interest. A recent publication by Quek et al. suggests that impurities in vesicle preparations have little effect on downstream nucleic acid quantification and states the utility of time-efficient, but rather crude EV isolation methods for biomarker discovery [89]. In line with this, we report precipitation-based isolation to yield samples with lower purity and significant protein contamination, but excellent potential for transcriptomics-driven biomarker discovery. We agree with previous publications stating that enriching serum EVs by precipitation might be a viable strategy for biomarker discovery studies [21,23]. Alvarez et al. presented similar findings for profiling RNA biomarkers in urinary EVs [9]. Decreased purity does not have to be a limitation if the objective is enriching cell-free miRNAs rather than purifying specific EV populations. If researchers strive to identify extracellular miRNA signatures that separate patient populations, these could be analysed regardless of their carriers [90]. As long as samples isolated by a given method reproducibly provide strong divisional capabilities for patient populations of interest, comprehensively characterizing isolated vesicles might not be a mandatory requirement for clinical biomarker applications. Given that precipitation is time efficient, inexpensive and demands no specialized equipment, it also seems to conveniently lend itself to integration into clinical usage. However, in a research field as vibrant and international, standardizing reagents and protocols utilized for EV precipitation and characterization are crucial for generating valid and reproducible data across laboratories [22,85,91].

In conclusion, we herein report that enriching cell-free miRNAs by precipitation allows for reliable separation of sepsis patients and healthy volunteers in sequencing-based analyses. As extracellular RNA can be encapsulated in vesicles or stabilized by binding to circulating proteins such as argonaute 2 (Ago2), further investigations using additional purification steps such as density gradient centrifugation or SEC are needed to conclusively verify if miRNAs dysregulated in this study are genuinely encapsulated in EVs [92]. Based on our experiments, we cannot rule out the possibility that miRNAs separating patients and volunteers are associated with non-vesicular carriers rather than EVs [93]. Should this be the case, additional validation of sepsis-related miRNA signatures might be carried out on total cell-free RNA without prior enrichment of EVs, reducing time and cost of analysis. Even though exosomes have been shown to provide an enriched source of miRNA with higher predictive value than total cell-free blood, miRNAs of diagnostic potential might be associated with different carriers in a disease-specific manner, calling for the careful validation of previous findings in each biomarker discovery process [2,94]. In diseases with less drastic clinical manifestation than sepsis, extracellular signalling could be more clearly detectable in pure EVs as opposed to crude preparations of cell-free RNA. Our findings might therefore not be generalizable to all clinical applications, as a different approach may be more appropriate for diseases other than sepsis.

Even though the focus of this work was on transcriptomic profiling of EVs, our findings could be transferred to different routes of analysis as well. It has become increasingly clear that the optimal method of EV isolation differs depending on the respective research setting and downstream analyses. Both failing to choose appropriate isolation methods for a particular experiment and trying to integrate results from multiple studies conducted with inappropriate or incompatible methodology squander resources decrease experimental validity and hamper translation of research findings into practical applications. This work therefore provides valuable guidance for navigating the wide array of EV isolation methods available today.

Acknowledgements

The authors wish to thank Franz Jansen for excellent technical assistance during vesicle isolation and Renate Scheler and PD Dr Ricarda Schumann from the University Eye Hospital LMU Munich for excellent assistance with TEM imaging. The authors are also grateful to Prof. Dr Jörg Kleiber for kindly providing access to the NanoSight LM10. The authors thank Qiagen for providing the exoRNeasy kit used in this study. Qiagen had no role in study design, analysis of data or decisions regarding publication.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the German Federal Ministry of Economy and Energy under Grant ZF4247001MD6.

ORCID

Dominik Buschmann http://orcid.org/0000-0003-0460-6459
Benedikt Kirchner http://orcid.org/0000-0003-3878-0148
Stefanie Hermann http://orcid.org/0000-0002-6274-5919
Melanie Märte http://orcid.org/0000-0002-0741-616X
Christine Wurmser http://orcid.org/0000-0003-0649-6206
Florian Brandes http://orcid.org/0000-0003-3741-287X
Ortrud K. Steinlein http://orcid.org/0000-0003-4311-6276
Michael W. Pfaffl http://orcid.org/0000-0002-3192-1019
Gustav Schelling http://orcid.org/0000-0002-6538-0652
Marlene Reithmair http://orcid.org/0000-0002-9113-9643

References

[1] Gardiner C, Di Vizio D, Sahoo S, et al. Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. J Extracell Vesicles. 2016;5:32945.
[2] Cheng L, Sharples RA, Scicluna BJ, et al. Exosomes provide a protective and enriched source of miRNA for biomarker profiling compared to intracellular and cell-free blood. J Extracell Vesicles. 2014;3:23743
[3] Mustapic M, Eitan E, Werner JK Jr., et al. Plasma extracellular vesicles enriched for neuronal origin: a potential window into brain pathologic processes. Front Neurosci. 2017;11:278.
[4] Buschmann D, Haberberger A, Kirchner B, et al. Toward reliable biomarker signatures in the age of liquid biopsies – how to standardize the small RNA-Seq workflow. Nucleic Acids Res. 2016;44: 5995–6018. Epub 2016/ 06/19.
[5] El-Khoury V, Pierson S, Kaoma T, et al. Assessing cellular and circulating miRNA recovery: the impact of the RNA isolation method and the quantity of input material. Sci Rep. 2016;6:19529.
[6] Guo Y, Vickers K, Xiong Y, et al. Comprehensive evaluation of extracellular small RNA isolation methods from serum in high throughput sequencing. BMC Genomics. 2017;18:50.
[7] Danielson KM, Rubio R, Abderazzaq F, et al. High throughput sequencing of extracellular RNA from human plasma. PLoS One. 2017;12:e0164644.
[8] Van Deun J, Mestdagh P, Sormunen R, et al. The impact of disparate isolation methods for extracellular vesicles on downstream RNA profiling. J Extracell Vesicles. 2014;3:24858.
[9] Alvarez ML, Khosroheidari M, Kanchi Ravi R, et al. Comparison of protein, microRNA, and mRNA yields using different methods of urinary exosome isolation for the discovery of kidney disease biomarkers. Kidney Int. 2012;82:1024–1032.
[10] Royo F, Diwan I, Tackett MR, et al. Comparative miRNA analysis of urine extracellular vesicles isolated through five different methods. Cancers (Basel). 2016;8:112
[11] Blans K, Hansen MS, Sorensen LV, et al. Pellet-free isolation of human and bovine milk extracellular vesicles by size-exclusion chromatography. J Extracell Vesicles. 2017;6:1294340.
[12] Rekker K, Saare M, Roost AM, et al. Comparison of serum exosome isolation methods for microRNA profiling. Clin Biochem. 2014;47: 135–138. Epub 2013/ 11/05.
[13] Muller L, Hong CS, Stolz DB, et al. Isolation of biologically-active exosomes from human plasma. J Immunol Methods. 2014;411:55–65.
[14] Nielsen MH, Beck-Nielsen H, Andersen MN, et al. A flow cytometric method for characterization of circulating cell-derived microparticles in plasma. J Extracell Vesicles. 2017;3:20795.
[15] Lasser C, Shelke GV, Yeri A, et al. Two distinct extracellular RNA signatures released by a single cell type identified by microarray and next-generation sequencing. RNA Biol. 2017;14:58–72.
[16] Caponnetto F, Manini I, Skrap M, et al. Size-dependent cellular uptake of exosomes. Nanomedicine. 2017;13: 1011–1020.
[17] Zlotogorski-Hurvitz A, Dayan D, Chaushu G, et al. Human saliva-derived exosomes: comparing methods of isolation. J Histochem Cytochem. 2015;63:181–189.
[18] Abramowicz A, Widlak P, Pietrowska M. Proteomic analysis of exosomal cargo: the challenge of high purity vesicle isolation. Mol Biosyst. 2016;12: 1407–1419. Epub 2016/ 04/01.
[19] Royo F, Zuniga-Garcia P, Sanchez-Mosquera P, et al. Different EV enrichment methods suitable for clinical settings yield different subpopulations of urinary extracellular vesicles from human samples. J Extracell Vesicles. 2016;5: 29497. Epub 2016/ 02/20.
[20] Mol EA, Goumans MJ, Doevendans PA, et al. Higher functionality of extracellular vesicles isolated using size-exclusion chromatography compared to ultracentrifugation. Nanomedicine. 2017;13:2061–2065.
[21] Andreu Z, Rivas E, Sanguino-Pascual A, et al. Comparative analysis of EV isolation procedures for miRNAs detection in serum samples. J Extracell Vesicles. 2016;5:31655.
[22] Crossland RE, Norden J, Bibby LA, et al. Evaluation of optimal extracellular vesicle small RNA isolation and qRT-PCR normalisation for serum and urine. J Immunol Methods. 2016;429:39–49.
[23] Helwa I, Cai J, Drewry MD, et al. A comparative study of serum exosome isolation using differential ultracentrifugation and three commercial reagents. PLoS One. 2017;12:e0170628.
[24] Bellingham SA, Shambrook M, Hill AF. Quantitative analysis of exosomal miRNA via qPCR and digital PCR. Methods Mol Biol. 2017;1545: 55–70. Epub 2016/ 12/13.
[25] Zeringer E, Li M, Barta T, et al. Methods for the extraction and RNA profiling of exosomes. World J Methodol. 2013;3: 11–18. Epub 2013/ 03/26.
[26] Takahashi K, Yan IK, Kim C, et al. Analysis of extracellular RNA by digital PCR. Front Oncol. 2014;4: 129. Epub 2014/ 06/14.
[27] Li M, Zeringer E, Barta T, et al. Analysis of the RNA content of the exosomes derived from blood serum and urine and its potential as biomarkers. Philos

Trans R Soc Lond B Biol Sci. 2014;369 (Epub 2014/ 08/20).
[28] Huang X, Yuan T, Tschannen M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics. 2013;14: 319. Epub 2013/ 05/15.
[29] Tsang EK, Abell NS, Li X, et al. Small RNA sequencing in cells and exosomes identifies eQTLs and 14q32 as a region of active export. G3 (Bethesda). 2017;7: 31–39. Epub 2016/ 11/04.
[30] Tang YT, Huang YY, Zheng L, et al. Comparison of isolation methods of exosomes and exosomal RNA from cell culture medium and serum. Int J Mol Med. 2017;40: 834–844. Epub 2017/ 07/25.
[31] Nimah M, Brilli RJ. Coagulation dysfunction in sepsis and multiple organ system failure. Crit Care Clin. 2003;19:441–458.
[32] SepNet Critical Care Trials G. Incidence of severe sepsis and septic shock in German intensive care units: the prospective, multicentre INSEP study. Intensive Care Med. 2016;42:1980–1989.
[33] Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373–383.
[34] Reithmair M, Buschmann D, Märte M, et al. Cellular and extracellular miRNAs are blood-compartment-specific diagnostic targets in sepsis. J Cell Mol Med. 2017;21:2403–2411.
[35] Spornraft M, Kirchner B, Haase B, et al. Optimization of extraction of circulating RNAs from plasma – enabling small RNA sequencing. PLoS One. 2014;9:e107259.
[36] Andrews S FastQC: a quality control tool for high throughput sequence data; 2010.
[37] Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics. 2011;98:152–153.
[38] Consortium RN. RNAcentral: an international database of ncRNA sequences. Nucleic Acids Res. 2015;43: D123–D129.
[39] Kozomara A, Griffiths-Jones S. miRBase: annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Research. 2014;42:D68–D73.
[40] Langmead B, Trapnell C, Pop M, et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 2009; 10:R25.
[41] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15:550.
[42] R Core Team. R: a language and environment for statistical computing. Vienna, Austria; 2017.
[43] Gregory R, Warnes BB, Bonebakker L, et al. Various R Programming Tools for Plotting Data; 2016.
[44] Wickham H ggplot2: Elegant Graphics for data analysis; 2009.
[45] Neuwirth E RColorBrewer: ColorBrewer Palettes. 2014.
[46] Galili T. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31:3718–3720.
[47] Yuan Tang MH, Wenxuan L. ggfortify: unified interface to visualize statistical result of popular R Packages; 2016.
[48] Chen H VennDiagram: generate high-resolution Venn and Euler plots; 2016.
[49] Eitan E, Green J, Bodogai M, et al. Age-related changes in plasma extracellular vesicle characteristics and internalization by leukocytes. Sci Rep. 2017;7:1342.
[50] Webber J, Clayton A. How pure are your vesicles? J Extracell Vesicles. 2013;2:19861.
[51] Kalra H, Adda CG, Liem M, et al. Comparative proteomics evaluation of plasma exosome isolation techniques and assessment of the stability of exosomes in normal human blood plasma. Proteomics. 2013;13: 3354–3364. Epub 2013/ 10/12.
[52] Eldh M, Lotvall J, Malmhall C, et al. Importance of RNA isolation methods for analysis of exosomal RNA: evaluation of different methods. Mol Immunol. 2012;50:278–286.
[53] Wang H, Zhang P, Chen W, et al. Serum microRNA signatures identified by Solexa sequencing predict sepsis patients' mortality: a prospective observational study. PLoS One. 2012;7: e38885. Epub 2012/ 06/22.
[54] Caserta S, Kern F, Cohen J, et al. Circulating plasma microRNAs can differentiate human sepsis and systemic inflammatory response syndrome (SIRS). Sci Rep. 2016;6: 28006. Epub 2016/ 06/21.
[55] Jiang X, Xu C, Lei F, et al. MiR-30a targets IL-1alpha and regulates islet functions as an inflammation buffer and response factor. Sci Rep. 2017;7: 5270. Epub 2017/ 07/15.
[56] Cheng Y, Kuang W, Hao Y, et al. Downregulation of miR-27a* and miR-532-5p and upregulation of miR-146a and miR-155 in LPS-induced RAW264.7. Macrophage Cells Inflammation. 2012;35:1308–1313.
[57] Tili E, Michaille JJ, Cimino A, et al. Modulation of miR-155 and miR-125b levels following lipopolysaccharide/ TNF-alpha stimulation and their possible roles in regulating the response to endotoxin shock. J Immunol. 2007;179: 5082–5089. Epub 2007/ 10/04.
[58] Zhou Y, Song Y, Shaikh Z, et al. MicroRNA-155 attenuates late sepsis-induced cardiac dysfunction through JNK and beta-arrestin 2. Oncotarget. 2017;8: 47317–47329. Epub 2017/ 05/20.
[59] Liu J, Shi K, Chen M, et al. Elevated miR-155 expression induces immunosuppression via CD39(+) regulatory T-cells in sepsis patient. Int J Infect Dis. 2015;40: 135–141. Epub 2015/ 10/04.
[60] Tacke F, Roderburg C, Benz F, et al. Levels of circulating miR-133a are elevated in sepsis and predict mortality in critically ill patients. Crit Care Med. 2014;42: 1096–1104. Epub 2014/ 01/15.
[61] Sari AN, Korkmaz B, Serin MS, et al. Effects of 5,14-HEDGE, a 20-HETE mimetic, on lipopolysaccharide-induced changes in MyD88/TAK1/IKKbeta/IkappaB-alpha/NF-kappaB pathway and circulating miR-150, miR-223, and miR-297 levels in a rat model of septic shock. Inflamm Res. 2014;63: 741–756. Epub 2014/ 06/12.
[62] Vasilescu C, Rossi S, Shimizu M, et al. MicroRNA fingerprints identify miR-150 as a plasma prognostic marker in patients with sepsis. PLoS One. 2009;4: e7405. Epub 2009/ 10/14.
[63] Lin H, Ewing LE, Koturbash I, et al. MicroRNAs as biomarkers for liver injury: current knowledge, challenges and future prospects. Food Chem Toxicol. 2017;110: 229–239. Epub 2017/ 10/19.

[64] Wang H, Yu B, Deng J, et al. Serum miR-122 correlates with short-term mortality in sepsis patients. Crit Care. 2014;18: 704. Epub 2015/ 02/13.
[65] Wang HJ, Deng J, Wang JY, et al. Serum miR-122 levels are related to coagulation disorders in sepsis patients. Clin Chem Lab Med. 2014;52: 927–933. Epub 2014/ 01/15.
[66] Roderburg C, Benz F, Vargas Cardenas D, et al. Elevated miR-122 serum levels are an independent marker of liver injury in inflammatory diseases. Liver Int. 2015;35: 1172–1184. Epub 2014/ 07/22.
[67] Linares R, Tan S, Gounou C, et al. High-speed centrifugation induces aggregation of extracellular vesicles. J Extracell Vesicles. 2015;4:29509.
[68] Dominguez-Medina S, McDonough S, Swanglap P, et al. In situ measurement of bovine serum albumin interaction with gold nanospheres. Langmuir. 2012;28:9131–9139.
[69] Stranska R, Gysbrechts L, Wouters J, et al. Comparison of membrane affinity-based method with size-exclusion chromatography for isolation of exosome-like vesicles from human plasma. J Transl Med. 2018;16: 1. Epub 2018/ 01/11.
[70] Kaparakis-Liaskos M, Ferrero RL. Immune modulation by bacterial outer membrane vesicles. Nat Rev Immunol. 2015;15: 375–387. Epub 2015/ 05/16.
[71] Alexander M, Hu R, Runtsch MC, et al. Exosome-delivered microRNAs modulate the inflammatory response to endotoxin. Nat Commun. 2015;6: 7321. Epub 2015/ 06/19.
[72] Sodar BW, Kittel A, Paloczi K, et al. Low-density lipoprotein mimics blood plasma-derived exosomes and microvesicles during isolation and detection. Sci Rep. 2016;6: 24316. Epub 2016/ 04/19.
[73] Gyorgy B, Szabo TG, Turiak L, et al. Improved flow cytometric assessment reveals distinct microvesicle (cell-derived microparticle) signatures in joint diseases. PLoS One. 2012;7: e49726. Epub 2012/ 11/28.
[74] Mork M, Handberg A, Pedersen S, et al. Prospects and limitations of antibody-mediated clearing of lipoproteins from blood plasma prior to nanoparticle tracking analysis of extracellular vesicles. J Extracell Vesicles. 2017;6: 1308779. Epub 2017/ 05/06.
[75] Gardiner C, Ferreira YJ, Dragovic RA, et al. Extracellular vesicle sizing and enumeration by nanoparticle tracking analysis. J Extracell Vesicles. 2013;2: 19671. Epub 2013/ 09/07.
[76] Ghosal A. Importance of secreted bacterial RNA in bacterial-host interactions in the gut. Microb Pathog. 2017;104: 161–163. Epub 2017/ 01/24.
[77] Ghosal A, Upadhyaya BB, Fritz JV, et al. The extracellular RNA complement of *Escherichia coli*. Microbiologyopen. 2015;4: 252–266. Epub 2015/ 01/23.
[78] Soo CY, Song Y, Zheng Y, et al. Nanoparticle tracking analysis monitors microvesicle and exosome secretion from immune cells. Immunology. 2012;136: 192–197. Epub 2012/ 02/22.
[79] Zabeo D, Cvjetkovic A, Lasser C, et al. Exosomes purified from a single cell type have diverse morphology. J Extracell Vesicles. 2017;6: 1329476. Epub 2017/ 07/19.
[80] Enderle D, Spiel A, Coticchia CM, et al. Characterization of RNA from exosomes and other extracellular vesicles isolated by a novel spin column-based method. PLoS One. 2015;10:e0136133.
[81] Koliha N, Wiencek Y, Heider U, et al. A novel multiplex bead-based platform highlights the diversity of extracellular vesicles. J Extracell Vesicles. 2016;5:29975.
[82] Kowal J, Arras G, Colombo M, et al. Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci U S A. 2016;113:E968–77.
[83] Gamez-Valero A, Monguio-Tortajada M, Carreras-Planella L, et al. Size-exclusion chromatography-based isolation minimally alters extracellular vesicles' characteristics compared to precipitating agents. Sci Rep. 2016;6:33641.
[84] Zarovni N, Corrado A, Guazzi P, et al. Integrated isolation and quantitative analysis of exosome shuttled proteins and nucleic acids using immunocapture approaches. Methods. 2015;87:46–58.
[85] Lotvall J, Hill AF, Hochberg F, et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. J Extracell Vesicles. 2014;3: 26913. Epub 2014/ 12/30.
[86] Lobb RJ, Becker M, Wen SW, et al. Optimized exosome isolation protocol for cell culture supernatant and human plasma. J Extracell Vesicles. 2015;4: 27031. Epub 2015/ 07/22.
[87] Welton JL, Webber JP, Botos LA, et al. Ready-made chromatography columns for extracellular vesicle isolation from plasma. J Extracell Vesicles. 2015;4:27269.
[88] Baranyai T, Herczeg K, Onodi Z, et al. Isolation of exosomes from blood plasma: qualitative and quantitative comparison of ultracentrifugation and size exclusion chromatography methods. PLoS One. 2015;10: e0145686.
[89] Quek C, Bellingham SA, Jung CH, et al. Defining the purity of exosomes required for diagnostic profiling of small RNA suitable for biomarker discovery. RNA Biol. 2017;14:245–258.
[90] Deregibus MC, Figliolini F, D'Antico S, et al. Charge-based precipitation of extracellular vesicles. Int J Mol Med. 2016;38:1359–1366.
[91] Witwer KW, Soekmadji C, Hill AF, et al. Updating the MISEV minimal requirements for extracellular vesicle studies: building bridges to reproducibility. J Extracell Vesicles. 2017;6: 1396823. Epub 2017/ 12/01.
[92] Arroyo JD, Chevillet JR, Kroh EM, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci U S A. 2011;108: 5003–5008. Epub 2011/ 03/09.
[93] Mateescu B, Kowal EJ, van Balkom BW, et al. Obstacles and opportunities in the functional analysis of extracellular vesicle RNA – an ISEV position paper. J Extracell Vesicles. 2017;6: 1286095. Epub 2017/ 03/23.
[94] van Eijndhoven MA, Zijlstra JM, Groenewegen NJ, et al. Plasma vesicle miRNAs for therapy response monitoring in Hodgkin lymphoma patients. JCI Insight. 2016;1: e89631. Epub 2016/ 11/25.

Appendix III

Substantial contributions by Dominik Buschmann:

- Execution of experiments
- Curation, analysis and interpretation of data
- Validation and formal analysis
- Drafting of figures
- Writing of the manuscript

Dominik Buschmann

Michael W. Pfaffl

Glucocorticoid receptor overexpression slightly shifts microRNA expression patterns in triple-negative breast cancer

DOMINIK BUSCHMANN[1,2*], RICARDO GONZÁLEZ[1,3*], BENEDIKT KIRCHNER[2], CLAUDIA MAZZONE[4], MICHAEL W. PFAFFL[2], GUSTAV SCHELLING[5], ORTRUD STEINLEIN[1] and MARLENE REITHMAIR[1]

[1]Institute of Human Genetics, University Hospital, LMU Munich, 80336 Munich; [2]Division of Animal Physiology and Immunology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, 85354 Freising, Germany; [3]Faculty of Sciences, University of Porto, 4169-007 Porto, Portugal; [4]Department of Pharmacy and Medicine, Sapienza University of Rome, 00185 Rome, Italy; [5]Department of Anesthesiology, University Hospital, LMU Munich, 81377 Munich, Germany

Received December 15, 2017; Accepted February 15, 2018

DOI: 10.3892/ijo.2018.4336

Abstract. Triple-negative breast cancer (TNBC) is a particularly aggressive subtype of breast cancer with limited options for clinical intervention. As with many solid tumors, TNBC is known to promote invasiveness and metastasis by secreting extracellular vesicles (EVs) capable of modulating the behaviour of recipient cells. Recent investigations have demonstrated that high expression levels of glucocorticoid receptor (GR) in TNBC are linked to therapy resistance, higher recurrence rates and increased mortality. In addition to activating protein-coding genes, GR is also involved in the expression of short non-coding RNAs including microRNAs (miRNAs or miRs). The molecular mechanisms responsible for the oncogenic effects of GR on TNBC have yet to be fully elucidated; however, emerging evidence suggests that miRNAs may play a pivotal role in tumorigenesis and metastasis. Thus, the aim of this study was to identify GR-regulated cellular and vesicular miRNAs that might contribute to the particularly oncogenic phenotype of TNBC with a high GR expression. We analyzed miRNA profiles of three TNBC cell lines using an *in vitro* model of GR overexpression. Next-generation sequencing revealed minor, cell line-specific changes in cellular miRNA expression, whereas vesicular miRNAs were not significantly regulated by GR. Additionally, the analysis of predicted miRNA targets failed to establish a causal link between GR-induced miRNA expression and oncogenic signaling. On the whole, given that GR influences miRNA profiles to only a small degree, other mechanisms are more likely to be responsible for the increased mortality of patients with TNBC with a high GR expression.

Correspondence to: Mr. Dominik Buschmann, Division of Animal Physiology and Immunology, TUM School of Life Sciences Weihenstephan, Technical University of Munich, Weihenstephaner Berg 3, 85354 Freising, Germany
E-mail: dominik.buschmann@wzw.tum.de

*Contributed equally

Key words: triple-negative breast cancer, glucocorticoid receptor, microRNA, miR-203a-3p, MDA-MB-436, MYLK, exosome

Introduction

Breast cancer (BC) is the most prevalent type of cancer affecting women, and the second most common type of cancer as a whole. Globally, the incidence of BC is one in nine women (1). Despite substantial advancements in diagnosis and treatment, the mortality rates for BC are still at 15% (2).

One of the most aggressive variants of BC is triple-negative BC (TNBC). This subgroup of BC cells does not express the receptors for estrogen (ER), progesterone receptor (PR) and human epidermal growth factor 2 (HER2). The absence of these proteins renders commonly used clinical interventions, such as inhibiting aromatase and blocking hormone receptors ineffectual for TNBC therapy (3,4). Treatment is therefore limited to surgical resection and adjuvant chemotherapy, making TNBC a significant unmet clinical need (5).

Synthetic glucocorticoids, such as dexamethasone are commonly used in tumor therapy (6). However, recent data have indicated that mortality is increased in patients with TNBC variants overexpressing glucocorticoid receptor (GR). GR signaling activates oncogenes, inhibits apoptosis and represses tumor suppressor genes in TNBC, leading to unfavorable clinical outcomes (7). A high GR expression is also associated with therapy resistance and increased recurrence (8).

In addition to regulating protein-coding genes, GR is known to activate non-coding RNAs, including microRNAs (miRNAs or miRs) (9). These molecules of18-22 nt in length modulate cellular gene expression by specifically binding complementary mRNA sequences and repressing their translation. Highly malignant tumor cells commonly display dysregulated miRNA profiles, leading to oncogenic and anti-apoptotic signaling (10). For TNBC, previous studies have discovered signatures of altered miRNA expression that distinguish cancer cells from surrounding tissues, and predict the receptor status for ER, PR and HER2 (11,12). Additionally,

the analysis of miRNAs highlights the heterogeneity of TNBC phenotypes and their respective signaling pathways (13). Unraveling miRNA regulation for individual TNBC variants is therefore crucial to understand pathogenesis and draw consequences about potential therapeutic interventions.

Recent studies have indicated that TNBC modulates distant cells by secreting signaling factors, including extracellular vesicles (EVs). EVs shed from tumor cells are enriched in specific miRNAs that might contribute to tumor progression and metastasis (14). Indeed, BC EVs have been found to induce the proliferation, migration and invasion of recipient cells, thus enhancing disease progression (15-17). Exploiting circulating miRNAs for the diagnosis and prognosis of TNBC patients might yield powerful biomarkers that are easily accessible via liquid biopsy (18).

To the best of our knowledge, however, to date, there is no study available investigating the effects of GR signaling on miRNA regulation in TNBC. Thus, in this study, to address this knowledge gap, we analyzed cellular and vesicular miRNA profiles in an *in vitro* model of GR-overexpressing TNBC by high-throughput next-generation sequencing (NGS). Surprisingly, no statistically significant alterations in EV miRNAs were detected in two TNBC cell lines upon GR overexpression. We did, however, detect a small set of cellular miRNAs regulated by GR in a cell line-specific manner. As the validation of miRNA target genes yielded ambiguous results, we concluded that the unfavorable influence of a high GR expression on TNBC phenotypes is not mediated by miRNAs to a significant extent.

Materials and methods

TNBC cell culture and transfection. The human TNBC cell lines, MDA-MB-231, MDA-MB-436 and MDA-MB-468, were purchased from the Leibnitz Institute DSMZ-German Collection of Microorganisms and Cell Culture (Braunschweig, Germany) and Cell line services (Eppelheim, Germany).

The cells were cultured in T75 flasks in a monolayer in DMEM (Sigma-Aldrich, Hamburg, Germany) containing 4.5 g/l glucose, 1% L-glutamine, 10% exosome-depleted fetal bovine serum (FBS) (BioCat GmbH, Heidelberg, Germany), 160 ng/l cortisol and 1% penicillin/streptomycin (Invitrogen, Karlsruhe, Germany). Cultures were maintained in a humidified atmosphere at 37˚C and 5% CO_2. For the experiments studying cellular miRNAs, the MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells were seeded 4.5 h prior to transfection at a concentration of $1x10^5$ cells per well in 24-well plates (Greiner Bio-One, Frickenhausen, Germany) using 0.5 ml culture medium.

GR overexpression was induced by transfecting the TNBC cells with nuclear receptor subfamily 3 group C member 1 (*NR3C1*)-encoding DNA plasmids. The coding sequence (CCDS4278.1) of the predominant glucocorticoid receptor gene *NR3C1* transcript variant 1 (NM_000176.2) was synthesized with the restriction sites *Kpn*I and *Xho*I on the 5' and 3' end, respectively (MWG Eurofins, Ebersberg, Germany), and cloned in frame into the pcDNA6/V5-His A vector (Life Technologies, Darmstadt, Germany). The cells were transfected with the NR3C1 plasmid (0.4 μg) for 24 h using Lipofectamine 2000 transfection reagent (Invitrogen). Following 30 h of cultivation, the cells had reached 90% confluency, and were rinsed once with HBSS before proceeding to total RNA extraction.

For experiments studying EV miRNAs, the MDA-MB-231 and MDA-MB-468 cells were seeded at a concentration of $3x10^6$ cells per well in 6-well plates (Greiner Bio-One). Both parental and transfected cells were seeded in 3 wells with 2.5 ml culturing medium each. Transfection was performed with 2.5 μg of plasmid harboring the coding sequence of *NR3C1*.

For all experiments, untransfected cells with endogenous GR expression were used as control samples. Three independent technical replicates per cell line were analyzed.

Validation of NR3C1 overexpression. To evaluate the effectiveness of transfection, the *NR3C1* mRNA levels were quantified by reverse transcription-quantitative (real-time) PCR (RT-qPCR). Total RNA from each cell line was initially reverse transcribed using the QuantiTect Reverse Transcription kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Quantitative PCR was performed using the Prime PCR Assay *NR3C1* and *GAPDH* human SsoAdvanced universal supermix (Bio-Rad, Munich, Germany) and 10 ng of template cDNA. PCR reactions were run on a MiniOpticon real-time PCR system (Bio-Rad). Additionally, the transcription levels of NR3C1 downstream targets dual specificity phosphatase 1 (*DUSP1*; NM_004417), serum/glucocorticoid regulated kinase 1 (*SGK1*; NM_005627.3) and glucocorticoid-induced leucine zipper protein (*GILZ*; NM_198057.2) were assessed by RT-qPCR. For each cell line, total RNA from three biological replicates was reverse transcribed using the QuantiTect Reverse Transcription kit (Qiagen). Subsequently, 8 ng cDNA were analyzed in a 10 μl reaction volume of SsoFast EvaGreen Supermix (Bio-Rad) and 300 nM primers (Sigma-Aldrich). Real-time PCR was carried out on triplicate samples using a Rotor-Gene Q thermal cycler (Qiagen) and a thermal profile for polymerase activation (95˚C for 1 min) and 45 cycles of amplification (95˚C for 10 sec 60˚C for 15 sec, 65˚C for 45 sec). The primer sequences are provided in Table I. The expression of *NR3C1* and its downstream targets was normalized to *GAPDH*, a stable reference gene for breast cancer cells (19). Relative quantification was carried out using the ΔΔCq method (20). Statistical significance was determined using the Student's t-test. Values of $P<0.05$ were considered to indicate statistically significant differences.

Extracellular vesicle isolation and characterization. For EV isolation, 7.5 ml of cell culture supernatant were collected from the parental and transfected cells after 30 h of cultivation, and centrifuged (3,200 x g, 5 min) to remove the cellular debris. EVs were isolated from pre-cleared supernatant using the miRCURY Exosome Isolation kit - Cells, urine and CSF according to the manufacturer's instructions (Exiqon, Vedbaek, Denmark). EV pellets were resuspended in either lysis buffer for RNA extraction, or PBS for vesicle characterization.

For nanoparticle tracking analysis (NTA), the EVs were diluted in particle-free PBS and analyzed on a NanoSight LM10 (Malvern Instruments GmbH, Herrenberg, Germany) using a 408 nm laser and NTA 3.0 software. Four videos of 30 sec each were captured, and analyzed using default settings

Table I. Primer pairs used for validation of *NR3C1* overexpression.

Name	Forward primer (5'-3')	Reverse primer (5'-3')
DUSP1	GCCATTGACTTCATAGACTCCATC	ATGATGCTTCGCCTCTGCTT
SGK1	GACGGTGAAAACTGAGGCTG	AGAAGGACTTGGTGGAGGAGA
GILZ	TCTTCTTCCACAGTGCCTCC	TCTTCAGGGCTCAGACAGGA

DUSP1, dual specificity phosphatase 1; *SGK1*, serum/glucocorticoid-regulated kinase 1; *GILZ*, glucocorticoid-induced leucine zipper.

for blur and minimum track length, and a detection threshold of two.

RNA extraction and NGS library preparation. Total RNA was isolated from the cells and EVs using the miRCURY RNA Isolation kit – Cell and Plant (Exiqon,) according to the manufacturer's instructoins. Cellular RNA was quantified using a nanophotometer (Implen GmbH, Munich, Germany), and RNA integrity was assessed by capillary electrophoresis on the Bioanalyzer 2100 using the RNA 6000 Nano kit (Agilent Technologies, Waldbronn, Germany). EV RNA was analyzed using the Agilent Small RNA kit (Agilent Technologies).

Sequencing libraries were constructed from 190 ng of cellular RNA, or the entire EV RNA isolated from 7.5 ml conditioned culture medium, respectively. Library preparation was performed as previously described by Spornraft *et al* (21). Briefly, the RNA was adaptor-ligated, reverse-transcribed, amplified by PCR and barcoded using the NEBNext Multiplex Small RNA Library Prep Set for Illumina (New England BioLabs Inc., Frankfurt, Germany). Adaptors and primers were diluted 1:2 in nuclease-free water to accommodate the low RNA input. Size selection of pooled PCR products was performed by agarose gel electrophoresis (4%), cutting out bands with 130 to 150 bp fragments. The purity and concentration of the libraries extracted from the gel were verified by capillary electrophoresis using the High Sensitivity DNA kit on the Bioanalyzer 2100 (Agilent Technologies). Finally, the libraries were subjected to Illumina single-end sequencing-by-synthesis using 50 cycles on the HiSeq 2500 (Illumina Inc., San Diego, CA, USA).

Data processing and differential gene expression analysis. FastQC (version 0.11.5) was used to assess the sequence length distribution and quality of the NGS data, as previously described (22). Adaptor sequences were trimmed using BTRIM, and all reads without adaptors were discarded (23). Additionally, reads shorter than 15 nt were excluded from the data set (24). Prior to miRNA analysis, reads pertaining to ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA) and small nucleolar RNA (snoRNA) were removed by mapping to sequences obtained from RNAcentral (25). The remaining reads were then aligned to miRBase (version 21) (26). Mapping was carried out using Bowtie and the 'best' alignment algorithm, allowing one mismatch for both RNAcentral and miRBase (27). Final read count tables were generated by sorting and indexing aligned reads using SAMtools, and calling the sum of hits per miRNA sequence (28). Differential gene expression analysis was subsequently performed via the bioconductor package DESeq2 (version 1.8.1), using the Benjamini-Hochberg method to correct for false discovery (29). A log2 fold change ≥|1| and an adjusted p-value (Padj) of ≤0.05 were set as thresholds to identify significantly regulated miRNAs. Only miRNAs with a mean expression of at least 50 counts were included in the analysis. Principal component analysis (regularized log-transformed, sizefactor-corrected counts obtained from DESeq2), and data visualization were performed in R (version 3.4.0) using the packages gplots, ggfortify, genefilter and RColorBrewer.

Validation of regulated miRNAs. Based on the NGS data, differentially regulated cellular miRNAs were validated by RT-qPCR. First, 111 ng of RNA were reverse transcribed in triplicate using the miScript II RT kit (Qiagen) according to the manufacturer's instructions. A total of 1 μl cDNA was subjected to real-time PCR in a 10 μl reaction volume using the miScript SYBR-Green PCR kit (Qiagen). Reactions were run on a CFX384 real-time PCR detection system (Bio-Rad) using the recommended protocol of polymerase activation (95˚C for 15 min) and 45 cycles of amplification (94˚C for 15 sec, 55˚C for 30 sec, 70˚C for 30 sec). Quantification cycle (Cq) values were determined automatically using default threshold settings, and Cq values above 37 were manually set to 40. The NGS data was utilized to assess potential reference miRNAs using the geNorm and NormFinder algorithms (30,31). Cq values of regulated miRNAs were subsequently normalized to the geometric mean of the following reference miRNAs: miR-24-3p, miR-25-3p and miR-148b-3p for MDA-MB-231 and MDA-MB-436, and miR-24-3p, miR-25-3p and let-7a-5p for MDA-MB-468. Relative quantification was carried out using the ΔΔCq method (20). Statistical significance was assessed using Student's t-test. Values of $P<0.05$ were considered to indicate statistically significant differences.

Prediction and quantification of miRNA target genes. miRWalk 2.0 was used to predict mRNAs targeted by miR-203a-3p (32). Four target genes known to be associated with metastasis in solid tumors [Actin, gamma 2, smooth muscle, enteric (*ACTG2*), calponin 1 (*CNN1*), major histocompatibility complex, class II, DP beta 1 (*HLA-DPB1*) and myosin light chain kinase (*MYLK*)] were selected for analysis by RT-qPCR (33). The expression of these genes was quantified in the same cellular MDA-MB-436 samples previously used for NGS. Initially, RNA was reverse transcribed in triplicate using the QuantiTect Reverse Transcription kit (Qiagen) according to the manufacturer's instructions. Quantitative PCR was then performed using Prime PCR Assays, SsoAdvanced Universal Supermix (Bio-Rad) and 10 ng of template cDNA. PCR reactions were run on a MiniOpticon real-time PCR

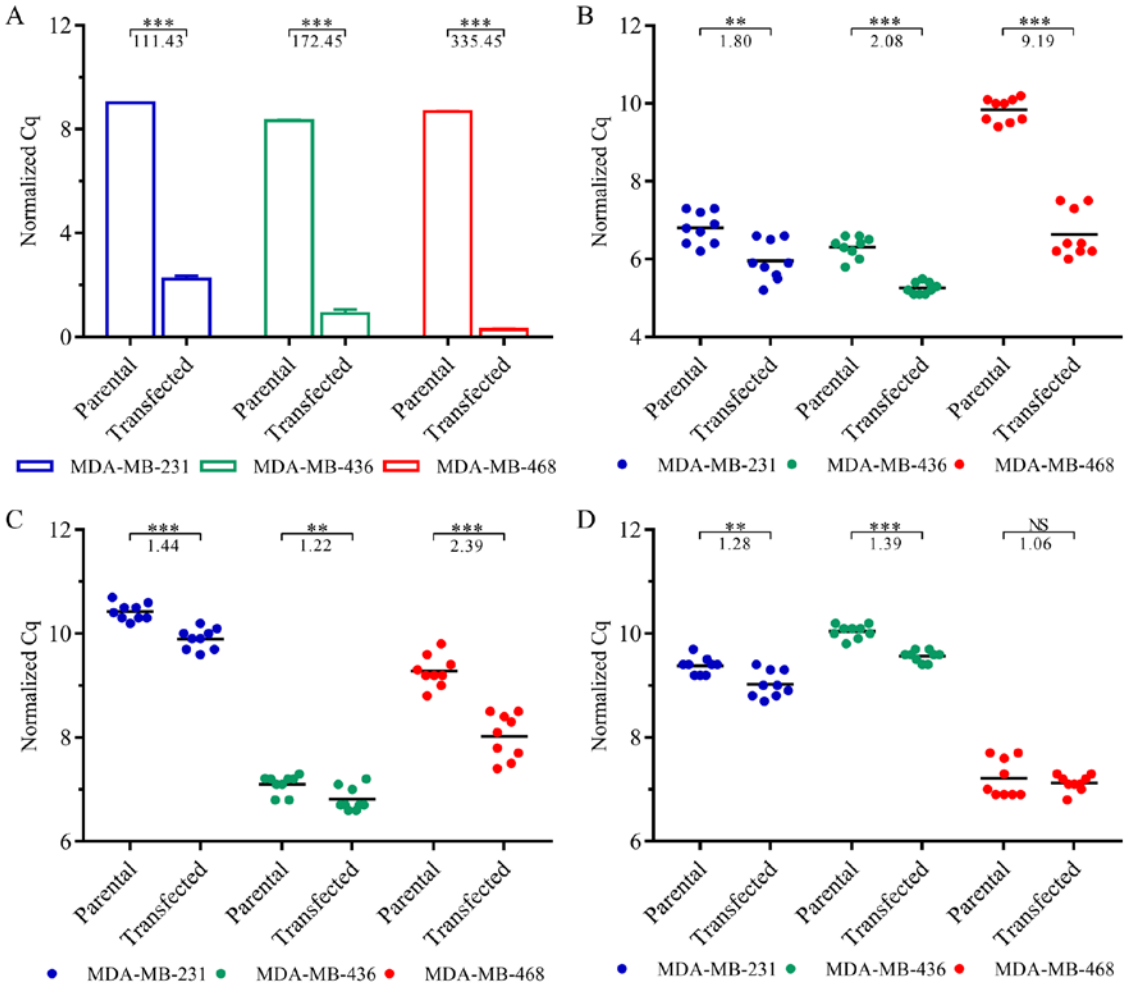

Figure 1. Validation of *NR3C1* overexpression. (A) Transfection with *NR3C1*-coding plasmids significantly increased glucocorticoid receptor (GR) expression in the MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells. Analysis of GR downstream effectors (B) *DUSP1*, (C) *SGK1* and (D) *GILZ*, *DUSP* and *SGK1* were significantly upregulated in all GR-overexpressing TNBC cell lines, while the expression levels of GILZ were not altered significantly in the transfected MDA-MB-468 cells. Numbers below significance indicator display corresponding fold change. **P<0.01; ***P<0.001; NS, not significant; FC, fold change.

system (Bio-Rad). Target gene Cq values were normalized to *GAPDH* (19).

Results

Plasmid transfection induces the expression of NR3C1 and downstream effectors. The transfection of the TNBC cells with *NR3C1*-coding plasmids was validated by RT-qPCR (Fig. 1A). GR expression was significantly increased in the transfected MDA-MB-231 (111-fold, P=1.75E-4), MDA-MB-436 (172-fold, P=2.51E-4) and MDA-MB-468 (335-fold, P=6.39E-6) cells.

To assess overexpression-induced changes in GR signaling, we additionally quantified the expression levels of GR target genes *DUSP1* (Fig. 1B), *SGK1* (Fig. 1C) and *GILZ* (Fig. 1D). While *DUSP1* and *SGK1* were significantly upregulated in all cell lines, GILZ expression was not altered significantly in the MDA-MB-468 cells (P=0.51). In the MDA-MB-231 and MDA-MB-436 cells, however, *GILZ* expression was significantly increased (P=2.79E-3 and P=3.79E-7, respectively).

TNBC secretes extracellular vesicles carrying RNA. EVs isolated from the conditioned media of the parental MDA-MB-231 and MDA-MB-468 cells were characterized by NTA (Fig. 2). Single-particle analysis revealed a narrow size distribution with mean particle diameters of 119.0±74.4 nm (mode, 87.5 nm) and 140.0±116.7 nm (mode, 117.5 nm) for the MDA-MB-231 and MDA-MB-468 cells, respectively. Despite

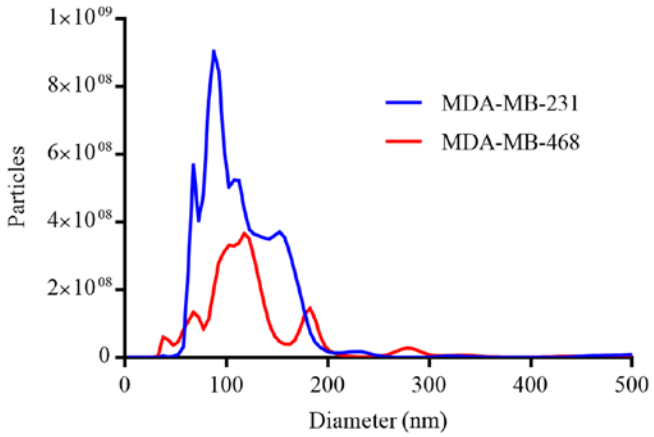

Figure 2. Particle size distribution in extracellular vesicle (EV) preparations from MDA-MB-231 and MDA-MB-468 cells. The area under the curve represents the absolute number of particles isolated from 7.5 ml conditioned media.

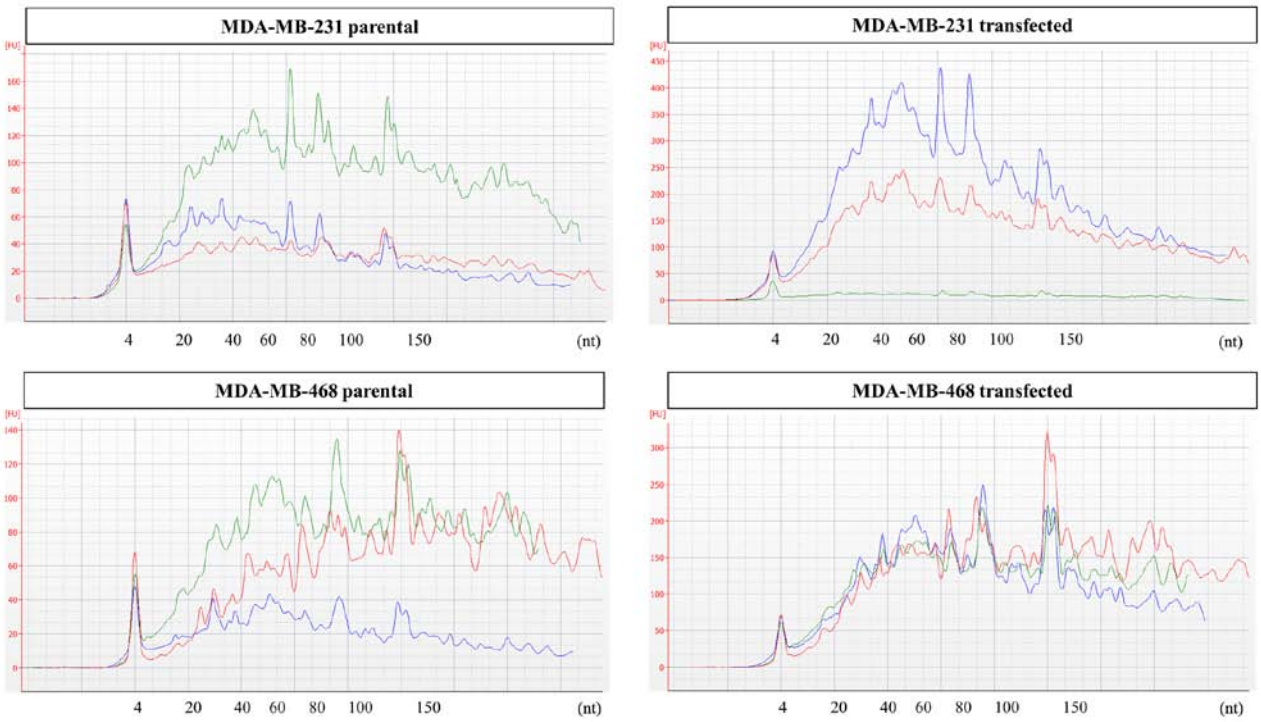

Figure 3. Bioanalyzer electropherograms for small RNA analysis in extracellular vesicles (EVs) from parental (left panels) and transfected (right panels) of MDA-MB-231 and MDA-MB-468 cells. FU, fluorescence unit; nt, nucleotide.

Table II. Library sizes and number of detected miRNA species in MDA-MB-231 and MDA-MB-468 EVs.

	MDA-MB-231 cells	
	Parental	Transfected
Library size ± SD	1.11E7±1.66E6	1.27E7±1.64E6
Distinct miRNAs	796	788
	MDA-MB-468 cells	
	Parental	Transfected
Library size ± SD	1.09E7±1.25E6	1.25E7±1.50E6
Distinct miRNAs	762	736

EV, extracellular vesicle; SD, standard deviation.

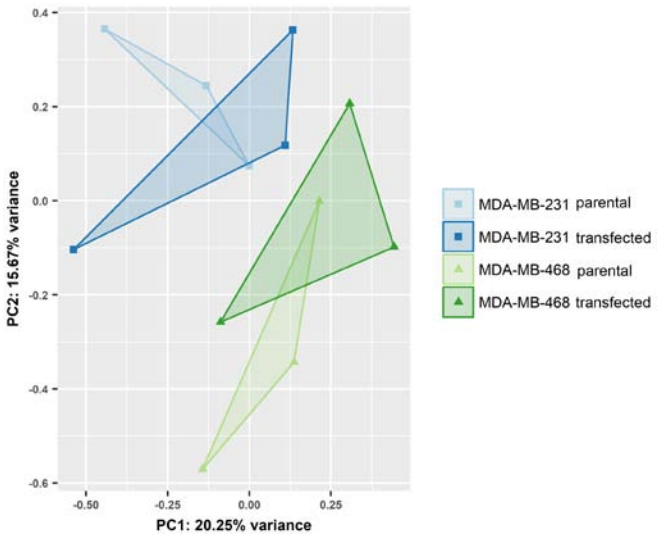

Figure 4. Principal component analysis of miRNA expression in MDA-MB-231 and MDA-MB-468 extracellular vesicles (EVs). Even though the cell lines were clearly separated, expression patterns in EVs from parental and transfected cells overlapped.

the similarities in diameter, significantly more particles were isolated from the MDA-MB-231 cells (P<0.001).

Prior to sequencing, RNA extracted from EV preparations was analyzed by capillary electrophoresis. Samples from both cell lines were found to be enriched in small RNA species <150 nt without obvious differences in size profiles between the parental and transfected cells. Full electropherograms for small RNA analysis are provided in Fig. 3.

GR overexpression does not alter vesicular miRNA profiles. EV RNA from the MDA-MB-231 and MDA-MB-468 cells was profiled by small RNA-Seq. The mean per-replicate library sizes are provided in Table II. After mapping to miRBase, between 736 (MDA-MB-468, transfected) and 796 (MDA-MB-231, parental) distinct miRNA transcripts were detected in at least one sample. For EVs from both cell lines, there was significant overlap in the 10 most highly expressed miRNAs between the treatment groups (Table III).

Differential expression of miRNAs was assessed in EVs from the parental and transfected MDA-MB-231 and MDA-MB-468 cells. While individual cell lines were clearly distinguished by principal component analysis (Fig. 4), the overexpression of GR did not lead to noticeable changes in

Table III. Top 10 most highly expressed miRNAs in EVs from parental and transfected MDA-MB-231 and MDA-MB-468 EVs.

MDA-MB-231 parental		MDA-MB-231 transfected	
miRNA	Count ± SD	miRNA	Count ± SD
miR-100-5p	32,896.06±15,567.28	miR-100-5p	28,093.64±1,746.75
miR-21-5p	20,517.88±6,132.81	miR-21-5p	20,563.67±13,352.19
let-7f-5p	14,534.16±7,389.27	let-7i-5p	16,700.98±7,018.44
let-7i-5p	13,851.26±3,041.99	let-7f-5p	13,224.79±6,215.57
miR-486-5p	12,773.16±9,474.51	let-7a-5p	13,170.83±7,480.14
let-7a-5p	12,044.75±4,144.34	miR-486-5p	12,251.12±9,557.67
miR-92a-3p	9,763.94±5,940.53	miR-451a	11,594.65±13,346.77
let-7g-5p	9,167.48±3,312.17	let-7g-5p	9,583.62±3,853.93
miR-451a	8,899.87±5,035.93	miR-92a-3p	8,909.74±4,096.12
miR-27b-3p	7,298.91±2,952.47	miR-27b-3p	7,004.21±4,888.94
MDA-MB-468 parental		**MDA-MB-468 transfected**	
miRNA	Count ± SD	miRNA	Count ± SD
let-7f-2-3p	20,174.15±3,134.83	miR-505-3p	17,483.16±12,287.71
miR-103b	14,956.91±4,896.68	miR-4742-3p	15,703.77±5,227.01
miR-4742-3p	14,411.10±5,455.42	miR-103b	14,095.48±11,598.41
let-7a-3p	13,768.25±2,884.00	let-7f-2-3p	13,819.32±4,499.01
miR-505-3p	10,494.01±2,170.28	let-7a-3p	9,814.04±2,947.26
let-7f-5p	10,065.30±2,199.66	let-7i-3p	8,987.19±3,015.38
let-7i-3p	9,369.82±1,379.22	let-7f-5p	8,452.97±2,966.55
miR-22-5p	7,455.85±526.85	miR-22-5p	7,692.73±2,962.41
let-7b-3p	5,176.99±1,522.35	miR-196b-5p	4,445.60±5,386.47
miR-196b-5p	3,375.73±1,780.00	miR-27b-3p	3,582.84±1,131.12

Data are mean normalized readcounts for 3 replicates each. EV, extracellular vesicle; SD, standard deviation.

Table IV. Library sizes and number of detected miRNA species in MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells.

	MDA-MB-231 cells	
	Parental	Transfected
Library size ± SD	9.42E6 1.56E6	7.98E6±8.10E5
Distinct miRNAs	1,025	949
	MDA-MB-436 cells	
	Parental	Transfected
Library size ± SD	9.25E6±8.75E5	8.28E6±5.92E5
Distinct miRNAs	1,187	1,216
	MDA-MB-468 cells	
	Parental	Transfected
Library size ± SD	8.32E6±1.23E6	7.27E6±9.70E5
Distinct miRNAs	1,096	1,016

SD, standard deviation.

miRNA expression. None of the miRNAs detected in small RNA-Seq displayed statistically significant regulation between endogenous and artificially induced GR expression.

Changes in cellular RNA profiles upon GR overexpression. Cellular RNA was initially analyzed by capillary electrophoresis to assess its suitability for NGS analysis. For all cell lines, samples from both parental and transfected cells featured excellent RNA integrity, as indicated by the RNA integrity number (RIN) values >9. Bioanalyzer electropherograms for cellular RNA are shown in Fig. 5.

In the NGS data, both the mean size of sequencing libraries and the number of detected miRNAs were higher than in the EV samples (Table IV compared with Table II). The most highly expressed miRNAs in all of the three parental cell lines displayed a high degree of similarity, sharing 8 of the top 10 miRNAs (Table V). Similarly, 7 of the top 10 most highly expressed miRNAs were common to all transfected cell lines.

Differential gene expression analysis revealed slight, yet statistically significant changes in specific miRNAs during GR overexpression (Table VI). Of note, a different set of GR-responsive miRNAs was detected in each of the TNBC cell lines studied herein, highlighting the heterogeneity of molecular signaling. As shown in Fig. 6, miRNA expression

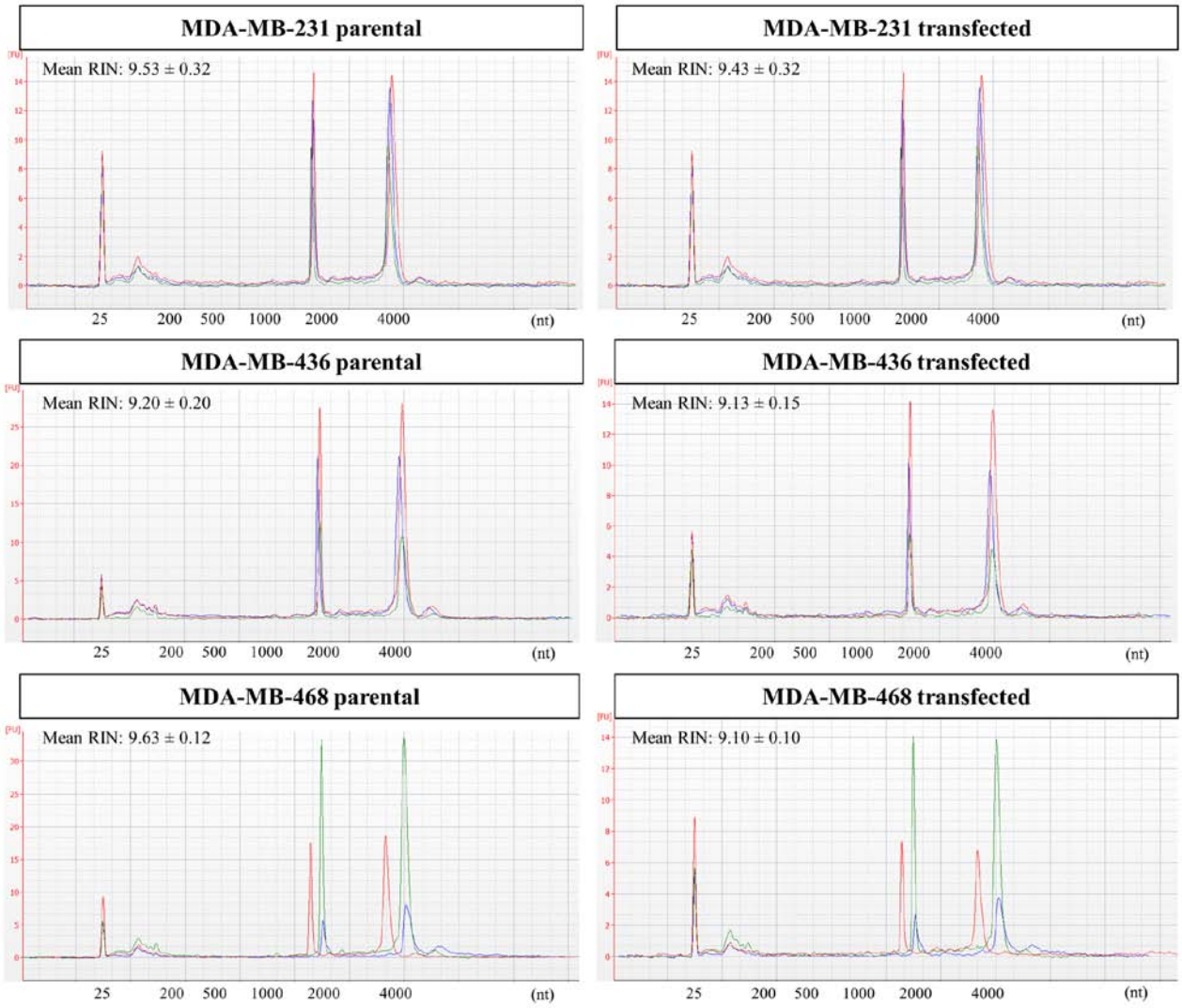

Figure 5. Bioanalyzer electropherograms for total RNA analysis in parental (left panels) and transfected (right panels) MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells. RIN, RNA integrity number; FU, fluorescence unit; nt, nucleotide.

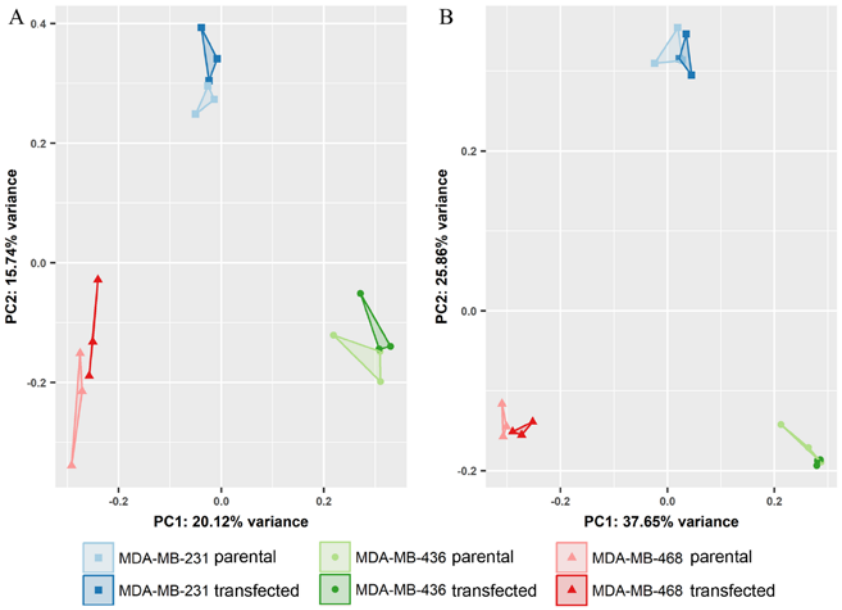

Figure 6. Principal component analysis of intracellular miRNA expression. (A) Based on all miRNAs, individual cell lines were clearly separated on principal components 1 and 2 with subtle differences between parental and transfected cells. (B) Analysis of the top 500 highest variance miRNAs reduced the separation of groups.

Table V. Top 10 most highly expressed miRNAs in parental and transfected MDA-MB-231, MDA-MB-436 and MDA-MB-468 cells.

MDA-MB-231 parental		MDA-MB-231 transfected	
miRNA	Count ± SD	miRNA	Count ± SD
miR-100-5p	963,722.05±111,405.87	miR-100-5p	1,348,718.94±345,916.54
let-7i-5p	663,458.05±68,993.69	let-7i-5p	774,622.60±97,337.13
let-7f-5p	267,434.97±31,684.39	let-7f-5p	324,183.97±17,197.04
let-7a-5p	195,527.39±12,006.56	let-7a-5p	204,938.44±20,414.46
miR-151a-3p	106,021.75±1å1,082.05	miR-151a-3p	163,277.67±56,462.79
let-7g-5p	90,178.84±13,127.37	miR-21-5p	80,090.23±21,445.21
miR-21-5p	67,628.11±16,283.91	let-7g-5p	78,292.35±25,760.59
miR-92a-3p	57,221.86±7,227.13	miR-92a-3p	61,589.95±11,515.77
miR-99b-5p	51,137.53±9,440.27	miR-10a-5p	60,968.56±16,758.93
miR-26a-5p	46,946.37±4,194.59	miR-99b-5p	57,331.95±10,509.88
MDA-MB-436 parental		MDA-MB-436 transfected	
miRNA	Count ± SD	miRNA	Count ± SD
let-7f-5p	374,536.91±26,561.27	let-7f-5p	340,698.27±40,869.66
miR-148a-3p	184,192.97±10,352.18	miR-148a-3p	226,502.28±169,068.63
let-7a-5p	153,782.25±10,424.01	let-7a-5p	163,998.51±24,567.84
let-7i-5p	146,573.76±5,731.89	let-7i-5p	157,923.40±59,299.50
miR-92a-3p	138,702.18±11,122.50	miR-151a-3p	148,161.03±108,764.69
miR-151a-3p	123,436.41±9,688.08	miR-92a-3p	116,819.87±14,412.48
miR-100-5p	71,942.27±10,410.65	miR-100-5p	74,514.77±19,201.37
let-7g-5p	71,876.07±1,749.76	miR-21-5p	66,313.38±31,125.38
miR-21-5p	55,083.27±6,735.91	let-7g-5p	62,873.25± 0,847.78
miR-99b-5p	38,293.24±5,041.04	miR-27a-3p	46,550.47±14,720.23
MDA-MB-468 parental		MDA-MB-468 transfected	
miRNA	Count ± SD	miRNA	Count ± SD
let-7f-5p	265,322.95±28,671.87	let-7f-5p	284,627.19±31,730.98
let-7i-5p	235,059.66±20,017.95	let-7i-5p	187,056.07±30,924.33
let-7a-5p	92,998.25±9,551.52	let-7a-5p	92,984.74±14,147.88
miR-99b-5p	78,640.71±22,892.91	miR-92a-3p	85,018.48±14,067.98
miR-92a-3p	75,727.64±7,097.75	miR-99b-5p	64,627.71±7,761.79
miR-151a-3p	73,088.61±32,556.30	let-7g-5p	62,517.17±6,768.50
let-7g-5p	50,292.17±10,045.56	miR-21-5p	49,269.32±7,542.30
miR-21-5p	43,287.45±14,869.30	miR-151a-3p	43,104.23±13,278.15
miR-25-3p	38,646.33±3,045.93	miR-25-3p	40,277.41±1,076.45
miR-30a-3p	35,921.06±9,828.02	miR-26a-5p	33,429.91±3,036.12

Data are mean normalized readcounts for three replicates each. SD, standard deviation.

clearly separated individual cell lines, but exhibited only minor differences between the transfected and parental cells. Based on the expression profiles of all miRNAs, principal component 3 (PC3), distinguished parental and GR-overexpressing cells (Fig. 7A). Limiting the input for analysis to the 500 miRNAs with highest variance, however, a reduced separation of groups was observed (Fig. 7B).

We then assessed differential miRNA regulation between endogenous and induced GR expression using RT-qPCR and the Student's t-test. Of the 7 miRNAs found to be significantly regulated in the NGS data, only miR-203a-3p was validated with statistical significance. In the transfected MDA-MB-436 cells, it was upregulated with a log2 fold change of 0.63 (Fig. 8).

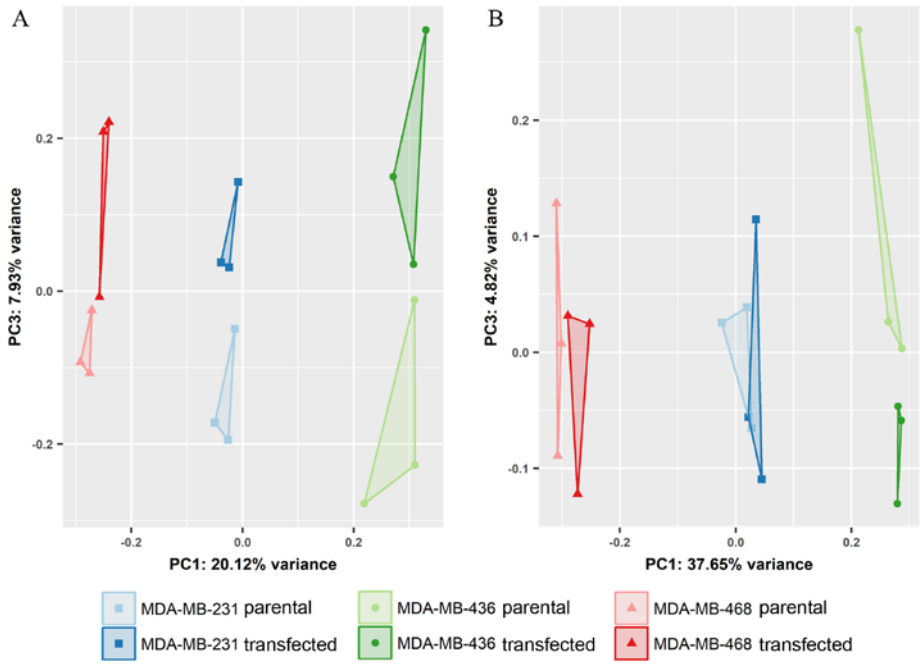

Figure 7. Principal component analysis of intracellular miRNA expression. (A) Analysis of all miRNAs in the dataset separated cell lines (PC1) and treatment groups (PC3). (B) Distances between parental and transfected cells were reduced when limiting the analysis to the top 500 highest variance miRNAs.

Table VI. Cellular miRNAs significantly regulated by GR.

	miRNA	log2FC	baseMean	P-adj
MDA-MB-231	miR-221-5p	1.13	223.75	0.0010
	miR-576-3p	1.11	53.47	0.0071
	let-7b-3p	-1.10	88.78	0.0118
MDA-MB-436	miR-203a-3p	1.35	134.76	0.0301
	miR-4746-5p	-1.07	74.25	0.0444
MDA-MB-468	miR-1260a	-1.54	291.24	0.0003
	miR-1260b	-1.54	335.12	0.0001

Positive fold changes indicate upregulation during GR overexpression. GR, glucocorticoid receptor; log2FC, log2 fold change; P-adj, DESeq2-adjusted P-value.

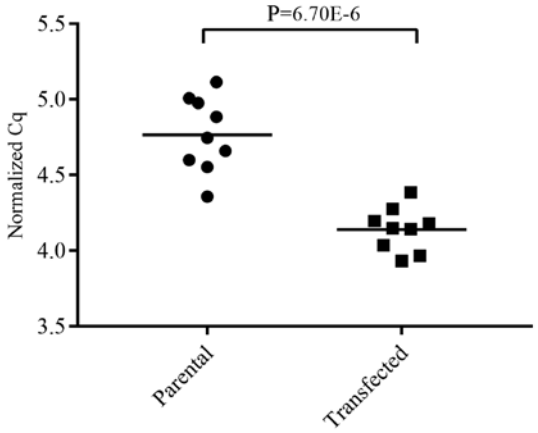

Figure 8. Results from RT-qPCR validation of miR-203a-3p. In transfected MDA-MB-436 cells, miR-203a-3p was upregulated with a log2 fold change of 0.63.

To assess the potential biological functions of miR-203a-3p in the MDA-MB-436 cells, in which it was found to be significantly increased in, we quantified the mRNA levels of 4 of its predicted target genes. Using RT-qPCR, we detected a 3-fold increase in *MYLK* expression during GR overexpression (P=0.03). The expression of *ACTG2, CNN1* and *HLA-DPB1* was not significantly altered between the parental and transfected cells (data not shown).

Discussion

TNBC is a particularly aggressive form of BC, leading to a poor prognosis for patients. Both the absence of hormone receptors and its molecular heterogeneity render TNBC a difficult target for therapeutic intervention. Additionally, a high GR expression was recently linked to therapy failure and worse outcomes in patients with TNBC, as well as other solid tumors. As Chen *et al* have previously reported, dexamethasone-mediated GR activation induced the expression of genes involved in carcinogenesis and tumor progression (7). These pro-oncogenic processes may also be fostered by GR-responsive non-coding RNAs, including miRNAs. This study therefore aimed at deciphering potential alterations in extracellular and intracellular miRNAs during GR overexpression.

GR biology is fascinatingly complex, involving ligand-dependent receptor activation and isoform-specific transcriptional activity (7,34). In this study, we focused on *NR3C1* transcript variant 1, as this is not only the most common isoform in epithelial cells, but also more transcriptionally active than others. Additionally, previous studies have demonstrated notable increases in *NR3C1* variant 1 in

TNBC (35,36). As expected, the transfection of TNBC cell lines with *NR3C1*-coding plasmids induced both a strong overexpression of GR mRNA and the increased the expression of several downstream targets (37-39). In concordance with previous reports, we found that TNBC cell lines secrete EVs in an exosome-like size range (15-17). Small RNA from cells with endogenous and artificially increased GR expression was sequenced, and miRNAs compared in differential expression analyses. Although a total of 419 and 442 miRNAs were detected in the MDA-MB-231 and MDA-MB-468 EVs, respectively, there were no statistically significant changes in miRNA profiles for either cell line. Increased metastasis and the progression of GR-overexpressing TNBC may therefore not be mediated by the secretion of soluble mediators to a significant extent. On the other hand, Harris *et al* recently identified vesicular proteins involved in BC metastasis, indicating a signaling function of vesicle components other than miRNAs (15). Circulating miRNAs are, however, not restricted to EVs, but can also be transported by lipoproteins and protein complexes, such as Argonaute2 (Ago2) (40,41). Given that this study focused exclusively on vesicular miRNAs, no conclusions can be drawn about the impact of GR expression on secreted miRNAs associated with other carrier vehicles.

When assessing the impact of GR overexpression on intracellular expression profiles, we detected a slight, cell line-specific modulation of 7 miRNAs. Even though the impact of GR signaling on TNBC miRNAs has not yet been elucidated, previous studies have reported GR-responsive miRNAs in primary lymphocytes, as well as in liver and spleen cells (42-44). In our data, the non-overlapping profiles of regulated miRNAs in the studied cell lines may be reflective of the inherent transcriptional heterogeneity of TNBC phenotypes (13,45,46). Rainer *et al* reported similar findings for several lymphoma cell lines that, although featuring GR-reactive miRNAs, only displayed moderate and non-uniform changes in miRNA profiles upon GR activation by dexamethasone (47). In our data, the parental and transfected cells were separated by changes in the global expression profile of miRNAs (Fig. 7). Distances between groups decreased when reducing the number of analyzed miRNAs, indicating that GR slightly shifts the expression patterns of many miRNAs, instead of inducing large changes in the abundance of a few specific transcripts.

Of note, miR-203a-3p, upregulated by GR expression in MDA-MB-436 cells, is controversially discussed in BC literature. Several studies have reported its overexpression in BC, as well as an association with a poor prognosis (48-50). Different data, on the other hand, have suggested that miR-203a-3p serves as a tumor suppressor miRNA, and have stated a decreased expression in BC (51-53). In this study, we found miR-203a-3p to be significantly upregulated in the MDA-MB-436 cells upon the overexpression of GR. As GR is known to be associated with tumor progression, this finding may corroborate the postulation of miR-203a-3p as an oncogenic factor in BC. However, considering the magnitude of expression changes, its biological impact may be of minor relevance. In line with our findings, a previous study reported an upregulation of miR-203 in dexamethasone-treated bone cells, potentially indicating a common miRNA response to GR activation across cell types (54).

The myosin light chain kinase (*MYLK, MLCK*) has been shown to interact with PI3K-AKT and p38 signaling, increasing cell motility and inhibiting apoptosis in BC cells (55,56). Furthermore, Sundararajan *et al* pointed out the ability of *MYLK* to promote invasiveness in several BC cell lines (57). Additionally, using LC-MS/MS-based proteomic profiling, Lawrence *et al* reported TNBC to feature particularly high levels of *MYLK* compared to less aggressive BC variants (58). As *MYLK* is a predicted miR-203a-3p target, we quantified its expression in MDA-MB-436 levels using RT-qPCR. Surprisingly, an increased miR-203a-3p expression in the transfected cells was accompanied by a 3-fold increase in *MYLK* mRNA levels. This finding was not in concordance with our expectations, as canonical miRNA regulation involves binding of mRNAs and repressing their translation. Consequently, the observed upregulation indicates that *MYLK* is not directly bound and downregulated by miR-203a-3p.

Taken together, our data suggest that *MYLK* may be regulated by GR, and can be regarded as a candidate gene involved in the poor survival rates of TNBC patients overexpressing GR. GR's mode of action on *MYLK*, however, seems not to be mediated by major alterations in cellular miRNAs.

In conclusion, we did not observe any prominent alterations in cellular or vesicular miRNA profiles upon overexpression of GR. The patterns of miRNA expression seem to be influenced by GR to only a small degree, and other mechanisms may therefore be the primary driver for the higher mortality rates of patients suffering from TNBC with GR overexpression.

Acknowledgements

The authors wish to thank Franz Jansen for excellent technical assistance. We are grateful to Renate Scheler and Dr Ricarda Schumann from the University Eye Hospital LMU Munich for excellent assistance with TEM. We also wish to thank Professor Jörg Kleiber for kindly providing access to the NanoSight LM10.

Funding

This study was supported by the K.L. Weigand'sche Stiftung, Curt-Bohnewands-Fonds, Georg and Traud Gravenhorst Stiftung, as well as by the Friedrich-Baur-Stiftung. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

The analyzed datasets generated during the study are available from the corresponding author on reasonable request.

Authors' contributions

MR, GS and OS conceived and designed the experiments; DB, RG, CM and MR performed the experiments; DB, RG, BK and MR and performed the validation and formal analysis; DB, BK, MWP and MR curated and analyzed the data; DB, BK and MR wrote the manuscript; MWP, GS and OS reviewed and revised the manuscript; MR, OS and GS acquired funding. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Ataollahi MR, Sharifi J, Paknahad MR and Paknahad A: Breast cancer and associated factors: A review. J Med Life 8: 6-11, 2015.
2. Malvezzi M, Bertuccio P, Levi F, La Vecchia C and Negri E: European cancer mortality predictions for the year 2014. Ann Oncol 25: 1650-1656, 2014.
3. Wahba HA and El-Hadaad HA: Current approaches in treatment of triple-negative breast cancer. Cancer Biol Med 12: 106-116, 2015.
4. Yagata H, Kajiura Y and Yamauchi H: Current strategy for triple-negative breast cancer: Appropriate combination of surgery, radiation, and chemotherapy. Breast Cancer 18: 165-173, 2011.
5. Hurvitz S and Mead M: Triple-negative breast cancer: Advancements in characterization and treatment approach. Curr Opin Obstet Gynecol 28: 59-69, 2016.
6. Sundahl N, Clarisse D, Bracke M, Offner F, Berghe WV and Beck IM: Selective glucocorticoid receptor-activating adjuvant therapy in cancer treatments. Oncoscience 3: 188-202, 2016.
7. Chen Z, Lan X, Wu D, Sunkel B, Ye Z, Huang J, Liu Z, Clinton SK, Jin VX and Wang Q: Ligand-dependent genomic function of glucocorticoid receptor in triple-negative breast cancer. Nat Commun 6: 8323, 2015.
8. Skor MN, Wonder EL, Kocherginsky M, Goyal A, Hall BA, Cai Y and Conzen SD: Glucocorticoid receptor antagonism as a novel therapy for triple-negative breast cancer. Clin Cancer Res 19: 6163-6172, 2013.
9. Tessel MA, Krett NL and Rosen ST: Steroid receptor and microRNA regulation in cancer. Curr Opin Oncol 22: 592-597, 2010.
10. Zhang B, Pan X, Cobb GP and Anderson TA: microRNAs as oncogenes and tumor suppressors. Dev Biol 302: 1-12, 2007.
11. Lowery AJ, Miller N, Devaney A, McNeill RE, Davoren PA, Lemetre C, Benes V, Schmidt S, Blake J, Ball G, *et al*: MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast Cancer Res 11: R27, 2009.
12. Chang YY, Kuo WH, Hung JH, Lee CY, Lee YH, Chang YC, Lin WC, Shen CY, Huang CS, Hsieh FJ, *et al*: Deregulated microRNAs in triple-negative breast cancer revealed by deep sequencing. Mol Cancer 14: 36, 2015.
13. Cascione L, Gasparini P, Lovat F, Carasi S, Pulvirenti A, Ferro A, Alder H, He G, Vecchione A, Croce CM, *et al*: Integrated microRNA and mRNA signatures associated with survival in triple negative breast cancer. PLoS One 8: e55910, 2013.
14. Falcone G, Felsani A and D'Agnano I: Signaling by exosomal microRNAs in cancer. J Exp Clin Cancer Res 34: 32, 2015.
15. Harris DA, Patel SH, Gucek M, Hendrix A, Westbroek W and Taraska JW: Exosomes released from breast cancer carcinomas stimulate cell movement. PLoS One 10: e0117495, 2015.
16. O'Brien K, Rani S, Corcoran C, Wallace R, Hughes L, Friel AM, McDonnell S, Crown J, Radomski MW and O'Driscoll L: Exosomes from triple-negative breast cancer cells can transfer phenotypic traits representing their cells of origin to secondary cells. Eur J Cancer 49: 1845-1859, 2013.
17. Singh R, Pochampally R, Watabe K, Lu Z and Mo YY: Exosome-mediated transfer of miR-10b promotes cell invasion in breast cancer. Mol Cancer 13: 256, 2014.
18. Liu Y, Zhang Y, Li Q, Li J, Ma X, Xing J, Rong S, Wu Z, Tian Y, Li J, *et al*: MiRNAs predict the prognosis of patients with triple negative breast cancer: A meta-analysis. PLoS One 12: e0170088, 2017.
19. Liu LL, Zhao H, Ma TF, Ge F, Chen CS and Zhang YP: Identification of valid reference genes for the normalization of RT-qPCR expression studies in human breast cancer cell lines treated with and without transient transfection. PLoS One 10: e0117058, 2015.
20. Livak KJ and Schmittgen TD: Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25: 402-408, 2001.
21. Spornraft M, Kirchner B, Haase B, Benes V, Pfaffl MW and Riedmaier I: Optimization of extraction of circulating RNAs from plasma - enabling small RNA sequencing. PLoS One 9: e107259, 2014.
22. Andrews S: FastQC: a quality control tool for high throughput sequence data. http://www.citeulike.org/user/nailest/article/11583827. Accessed Nov 21, 2017.
23. Kong Y: Btrim: A fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. Genomics 98: 152-153, 2011.
24. Buschmann D, Haberberger A, Kirchner B, Spornraft M, Riedmaier I, Schelling G and Pfaffl MW: Toward reliable biomarker signatures in the age of liquid biopsies - how to standardize the small RNA-Seq workflow. Nucleic Acids Res 44: 5995-6018, 2016.
25. Consortium RN; RNAcentral Consortium: RNAcentral: An international database of ncRNA sequences. Nucleic Acids Res 43D: D123-D129, 2015.
26. Kozomara A and Griffiths-Jones S: miRBase: Annotating high confidence microRNAs using deep sequencing data. Nucleic Acids Res 42 (D1): D68-D73, 2014.
27. Langmead B, Trapnell C, Pop M and Salzberg SL: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25, 2009.
28. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G and Durbin R; 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079, 2009.
29. Love MI, Huber W and Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15: 550, 2014.
30. Vandesompele J, Kubista M and Pfaffl MW: Reference gene validation software for improved normalization. In: Real-Time PCR: Current Technology and Applications. Logan J, Edwards K and Saunders N (eds). Caister Academic Press, London, pp47-64, 2009.
31. Andersen CL, Jensen JL and Ørntoft TF: 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: 5245-5250, 2004.
32. Dweep H, Sticht C, Pandey P and Gretz N: miRWalk - database: Prediction of possible miRNA binding sites by 'walking' the genes of three genomes. J Biomed Inform 44: 839-847, 2011.
33. Ramaswamy S, Ross KN, Lander ES and Golub TR: A molecular signature of metastasis in primary solid tumors. Nat Genet 33: 49-54, 2003.
34. Lu NZ and Cidlowski JA: Glucocorticoid receptor isoforms generate transcription specificity. Trends Cell Biol 16: 301-307, 2006.
35. Reeder A, Attar M, Nazario L, Bathula C, Zhang A, Hochbaum D, Roy E, Cooper KL, Oesterreich S, Davidson NE, *et al*: Stress hormones reduce the efficacy of paclitaxel in triple negative breast cancer through induction of DNA damage. Br J Cancer 112: 1461-1470, 2015.
36. McNamara KM, Kannai A and Sasano H: Possible roles for glucocorticoid signalling in breast cancer. Mol Cell Endocrinol S0303-7207(17)30358-1, 2017.
37. Wu W, Pew T, Zou M, Pang D and Conzen SD: Glucocorticoid receptor-induced MAPK phosphatase-1 (MPK-1) expression inhibits paclitaxel-associated MAPK activation and contributes to breast cancer cell survival. J Biol Chem 280: 4117-4124, 2005.
38. Maiyar AC, Phu PT, Huang AJ and Firestone GL: Repression of glucocorticoid receptor transactivation and DNA binding of a glucocorticoid response element within the serum/glucocorticoid-inducible protein kinase (sgk) gene promoter by the p53 tumor suppressor protein. Mol Endocrinol 11: 312-329, 1997.
39. Ayroldi E and Riccardi C: Glucocorticoid-induced leucine zipper (GILZ): A new important mediator of glucocorticoid action. FASEB J 23: 3649-3658, 2009.
40. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, Stirewalt DL, *et al*: Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci USA 108: 5003-5008, 2011.

41. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD and Remaley AT: MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol 13: 423-433, 2011.
42. Puimège L, Van Hauwermeiren F, Steeland S, Van Ryckeghem S, Vandewalle J, Lodens S, Dejager L, Vandevyver S, Staelens J, Timmermans S, *et al*: Glucocorticoid-induced microRNA-511 protects against TNF by down-regulating TNFR1. EMBO Mol Med 7: 1004-1017, 2015.
43. Smith LK, Tandon A, Shah RR, Mav D, Scoltock AB and Cidlowski JA: Deep sequencing identification of novel glucocorticoid-responsive miRNAs in apoptotic primary lymphocytes. PLoS One 8: e78316, 2013.
44. Wang ZH, Liang YB, Tang H, Chen ZB, Li ZY, Hu XC and Ma ZF: Dexamethasone down-regulates the expression of microRNA-155 in the livers of septic mice. PLoS One 8: e80547, 2013.
45. Mathe A, Scott RJ and Avery-Kiejda KA: MiRNAs and other epigenetic changes as biomarkers in triple negative breast cancer. Int J Mol Sci 16: 28347-28376, 2015.
46. Radojicic J, Zaravinos A, Vrekoussis T, Kafousi M, Spandidos DA and Stathopoulos EN: MicroRNA expression analysis in triple-negative (ER, PR and Her2/neu) breast cancer. Cell Cycle 10: 507-517, 2011.
47. Rainer J, Ploner C, Jesacher S, Ploner A, Eduardoff M, Mansha M, Wasim M, Panzer-Grümayer R, Trajanoski Z, Niederegger H, *et al*: Glucocorticoid-regulated microRNAs and mirtrons in acute lymphoblastic leukemia. Leukemia 23: 746-752, 2009.
48. Shao Y, Gu W, Ning Z, Song X, Pei H and Jiang J: Evaluating the prognostic value of microRNA-203 in solid tumors based on a meta-analysis and the Cancer Genome Atlas (TCGA) Datasets. Cell Physiol Biochem 41: 1468-1480, 2017.
49. He S, Zhang G, Dong H, Ma M and Sun Q: miR-203 facilitates tumor growth and metastasis by targeting fibroblast growth factor 2 in breast cancer. Onco Targets Ther 9: 6203-6210, 2016.
50. Liang Y, Yang W, Zhu Y and Yuan Y: Prognostic role of microRNA-203 in various carcinomas: Evidence from a meta-analysis involving 13 studies. Springerplus 5: 1538, 2016.
51. Ding X, Park SI, McCauley LK and Wang CY: Signaling between transforming growth factor β (TGF-β) and transcription factor SNAI2 represses expression of microRNA miR-203 to promote epithelial-mesenchymal transition and tumor metastasis. J Biol Chem 288: 10241-10253, 2013.
52. Wang C, Zheng X, Shen C and Shi Y: MicroRNA-203 suppresses cell proliferation and migration by targeting BIRC5 and LASP1 in human triple-negative breast cancer cells. J Exp Clin Cancer Res 31: 58, 2012.
53. Zhang Z, Zhang B, Li W, Fu L, Fu L, Zhu Z and Dong JT: Epigenetic silencing of miR-203 upregulates SNAI2 and contributes to the invasiveness of malignant breast cancer cells. Genes Cancer 2: 782-791, 2011.
54. Laxman N, Rubin CJ, Mallmin H, Nilsson O, Tellgren-Roth C and Kindmark A: Second generation sequencing of microRNA in human bone cells treated with parathyroid hormone or dexamethasone. Bone 84: 181-188, 2016.
55. Choi C, Kwon J, Lim S and Helfman DM: Integrin β1, myosin light chain kinase and myosin IIA are required for activation of PI3K-AKT signaling following MEK inhibition in metastatic triple negative breast cancer. Oncotarget 7: 63466-63487, 2016.
56. Cui WJ, Liu Y, Zhou XL, Wang FZ, Zhang XD and Ye LH: Myosin light chain kinase is responsible for high proliferative ability of breast cancer cells via anti-apoptosis involving p38 pathway. Acta Pharmacol Sin 31: 725-732, 2010.
57. Sundararajan V, Gengenbacher N, Stemmler MP, Kleemann JA, Brabletz T and Brabletz S: The ZEB1/miR-200c feedback loop regulates invasion via actin interacting proteins MYLK and TKS5. Oncotarget 6: 27083-27096, 2015.
58. Lawrence RT, Perez EM, Hernández D, Miller CP, Haas KM, Irie HY, Lee SI, Blau CA and Villén J: The proteomic landscape of triple-negative breast cancer. Cell Reports 11: 630-644, 2015.

Appendix IV

Substantial contributions by Dominik Buschmann:

- Execution of experiments
- Curation, analysis and interpretation of data
- Validation and formal analysis
- Drafting of figures
- Writing of the manuscript

Dominik Buschmann D. Buschmann

Michael W. Pfaffl Michael Pfaffl

J. Cell. Mol. Med. Vol XX, No X, 2017 pp. 1-8

Cellular and extracellular miRNAs are blood-compartment-specific diagnostic targets in sepsis

Marlene Reithmair [a, *], Dominik Buschmann [a, b], Melanie Märte [c], Benedikt Kirchner [b], Daniel Hagl [c, d], Ines Kaufmann [d], Martina Pfob [a], Alexander Chouker [c], Ortrud K. Steinlein [a], Michael W. Pfaffl [b], Gustav Schelling [c]

[a] *Institute of Human Genetics, University Hospital, Ludwig-Maximilians-University, Munich, Germany*
[b] *Division of Animal Physiology and Immunology, TUM School of Life Sciences Weihenstephan, Technical University Munich, Munich, Germany*
[c] *Department of Anesthesiology, University Hospital, Ludwig-Maximilians-University, Munich, Germany*
[d] *Department of Anaesthesiology, Neuperlach Hospital, City Hospitals of Munich, Munich, Germany*

Received: December 30, 2016; Accepted: February 16, 2017

Abstract

Septic shock is a common medical condition with a mortality approaching 50% where early diagnosis and treatment are of particular importance for patient survival. Novel biomarkers that serve as prompt indicators of sepsis are urgently needed. High-throughput technologies assessing circulating microRNAs represent an important tool for biomarker identification, but the blood-compartment specificity of these miRNAs has not yet been investigated. We characterized miRNA profiles from serum exosomes, total serum and blood cells (leukocytes, erythrocytes, platelets) of sepsis patients by next-generation sequencing and RT-qPCR ($n = 3 \times 22$) and established differences in miRNA expression between blood compartments. *In silico* analysis was used to identify compartment-specific signalling functions of differentially regulated miRNAs in sepsis-relevant pathways. In septic shock, a total of 77 and 103 miRNAs were down- and up-regulated, respectively. A majority of these regulated miRNAs (14 in serum, 32 in exosomes and 73 in blood cells) had not been previously associated with sepsis. We found a distinctly compartment-specific regulation of miRNAs between sepsis patients and healthy volunteers. Blood cellular miR-199b-5p was identified as a potential early indicator for sepsis and septic shock. miR-125b-5p and miR-26b-5p were uniquely regulated in exosomes and serum, respectively, while one miRNA (miR-27b-3p) was present in all three compartments. The expression of sepsis-associated miRNAs is compartment-specific. Exosome-derived miRNAs contribute significant information regarding sepsis diagnosis and survival prediction and could serve as newly identified targets for the development of novel sepsis biomarkers.

Keywords: sepsis • exosome • blood compartment • liquid biopsy • miRNA • biomarker

Introduction

Sepsis has been described as one of the oldest and most pressing problems in medicine [1]. Even with treatment in modern intensive care units (ICUs), mortality rates of patients with septic shock are in the range between 30% and 44.3% [2] with an even higher in-hospital mortality of up to 50% [3].

Pathobiological models of sepsis emphasize the important role of host immunity in disease development. Current theories postulate that organ injury and death from sepsis are the result of an unabated activation of innate-immunity-driven inflammation accompanied by a down-regulation of genes that modulate the adaptive immune response. On the other hand, there is clear evidence for a prolonged phase of immunosuppression.

Despite these controversies, there is a consensus that rapid cell-to-cell communication between the immune system and afflicted organs plays an important role during the cascade of events leading from initial contact with an invasive pathogen to the fulminant clinical syndrome of septic shock.

A growing body of literature suggests that nano-sized extracellular vesicles, including exosomes, play a critical role in cell-to-cell communication, especially in inflammatory processes and malignancy [4]. Interestingly, small non-coding microRNAs (miRNAs) are highly enriched in exosomes and certain miRNAs are over-represented compared to their fraction in the donor cell [5, 6].

*Correspondence to: Marlene REITHMAIR MD
E-mail: Marlene.Reithmair@med.uni-muenchen.de

doi: 10.1111/jcmm.13162

These miRNAs play a crucial role in exosome-mediated phenotypic modulation of recipient cells. It is well known that miRNAs are capable of influencing gene expression by specifically binding to the 3′-untranslated region (UTR) of complementary mRNA sequences and thus repressing their translation. Previous publications were able to demonstrate that exosomes from platelets under septic insult induce myocardial dysfunction and apoptosis in endothelial cells, and may thus contribute to the vascular abnormalities commonly observed in patients with sepsis [7, 8]. Consequently, cell-to-cell delivery of miRNAs through circulating exosomes could represent an important mechanism of long-range signalling during sepsis and offers promising opportunities for assessing biomarkers in minimally invasive liquid biopsies.

In this study, we examined serum-derived exosomes from patients with septic shock and compared their miRNA expression levels to those from total serum and blood cells (leukocytes, erythrocytes, platelets) using comprehensive high-throughput next-generation sequencing (NGS).

Materials and methods

Patient recruitment and sample collection

In total, 22 critically ill patients with sepsis were included in the study and matched to 23 healthy volunteers (Table S1). Patients were recruited from two anaesthesiological ICUs caring for a mixed medical/surgical patient population of an academic medical centre of the University of Munich and a city hospital of Munich.

Healthy volunteers were recruited from hospital personnel and by advertisement.

Seven patients with septic shock were initially selected for NGS and matched to seven age and gender comparable healthy volunteers. RT-qPCR confirmation of expression levels of selected miRNAs was performed in another 15 patients with varying disease severity and in a matched control group of equal size.

Blood samples were drawn from healthy subjects *via* venipuncture and from patients through intravascular catheters on day 0 and day 4 (if available). Samples to obtain serum were collected into 9 ml serum tubes (S-Monovette; Sarstedt AG&Co, Nümbrecht, Germany), immediately centrifuged at 3400 × *g* for 10 min. and frozen. Whole blood samples designated for extraction of cellular miRNAs were collected in RNA tubes (PAXgene; Qiagen, Hilden, Germany) according to the manufacturer's protocol. Serum aliquots and RNA tubes were stored at −80°C.

Exosome isolation and quality control

3 ml serum was digested with 34 μl thrombin for 5 min. After centrifugation (10,000 × *g*, 5 min.), exosomes were isolated from the supernatant using miRCURY™ Exosome Isolation Kit according to the manufacturer's protocol (Exiqon A/S). The presence and purity of exosomes were confirmed by transmission electron microscopy (TEM) and nanoparticle tracking analysis (NTA). Exosomal marker proteins (CD81, TSG101, syntenin-1) were substantiated by immunoblotting (Fig. 1).

For TEM, exosomes were diluted in PBS and deposited on formvar-coated copper grids (Plano GmbH, Wetzlar, Germany) for 60 sec. Grids were negative stained in 2% uranyl acetate (Sigma Aldrich Co., Taufkirchen, Germany) before imaging on a Zeiss EM900 (Carl Zeiss Microscopy GmbH, Munich, Germany) with a wide-angle dual-speed 2K-CCD camera at 80 kV.

For NTA, exosomes were analysed using a NanoSight LM10 instrument (Malvern Instruments GmbH, Malvern, UK) equipped with a monochromatic 405 nm laser and a high-sensitivity sCMOS camera. Exosome preparations were diluted in PBS and recorded by the NTA software (Malvern Instruments GmbH, Malvern, UK). Five videos of 30 sec. with a frame rate of 25 frames/sec were recorded. Each particle was subsequently tracked on a frame-by-frame basis by the Finite Track Length Adjustment algorithm, and the recorded Brownian motion was used to infer particle sizes by applying the Stokes–Einstein equation.

For Western blot analysis (WB), miRCURY™ precipitates were washed in PBS and pelleted by ultracentrifugation at 100,000 × *g* for 2 hrs (Beckman Coulter Optima LE-80K using a SW60 Ti rotor, k-factor: 167.9, 4°C). The pellets were lysed in ice-cold RIPA buffer with a protease inhibitor cocktail (Roche Deutschland Holding GmbH, Grenzach-Wyhlen, Germany). Lysates were sonicated, and then centrifuged at 13,000 × *g* for 5 min., and protein concentration in the supernatant was measured using a BCA assay (Sigma Aldrich) before separation by SDS-PAGE. Lysates were boiled in reducing sample buffer for 10 min., and fractionated using NuPAGE 4–12% Bis-Tris Gels (Invitrogen, Carlsbad, California, USA). Proteins were transferred to a nitrocellulose membrane (GE Healthcare Life Sciences, Freiburg, Germany). Primary antibodies were from antibodies-online.com (rabbit anti-syntenin-1, ABIN1881779, 1:1000) or OriGene Technologies, Inc., Rockville, Maryland, USA (rabbit anti-CD81, TA343281, 1:1000 and rabbit anti-TSG101, TA343598, 1:500). Secondary antibodies were from Santa Cruz Biotechnology, Dallas, Texas, USA (goat anti-rabbit IgG-HRP, sc-2030, 1:5000).

Extraction of extracellular and cellular RNA

After exosome isolation from 3 ml serum, exosomal RNA was extracted with the miRCURY™ RNA Isolation Kit—Biofluids (Exiqon, Vedbaek, Denmark) and eluted in 30 μl nuclease-free water.

Serum of 600 μl were extracted with the miRCURY™ RNA Isolation Kit—Biofluids (Exiqon) according to the manufacturer's protocol. RNA was eluted in 30 μl nuclease-free water.

For extraction of blood cell RNA, PAXgene blood tubes were processed with the PAXgene blood miRNA kit (Qiagen) according to the manufacturer's protocol. Integrity of total blood cell-derived RNA was assessed with the RNA 6000 Nano assay on the Bioanalyzer 2100 (Agilent Technologies, Inc., Santa Clara, California, USA). Small RNA fractions in cellular and extracellular samples were analysed using the Small RNA assay and RNA yield for cell-derived RNA was quantified on the Qubit 2.0 Fluorometer with the RNA HS Assay Kit (Life Technologies, Carlsbad, California, USA) according to the manufacturer's protocol.

Calculation of haemolysis scores

Blondal *et al.* [9] found that in RT-qPCR a delta Cq of miR-23a–miR-451a >7–8 indicates a high risk for haemolysis bias in the samples. Haemolysis was assessed for each patient sample by calculating the Cq ratio of these miRNAs accordingly. All samples

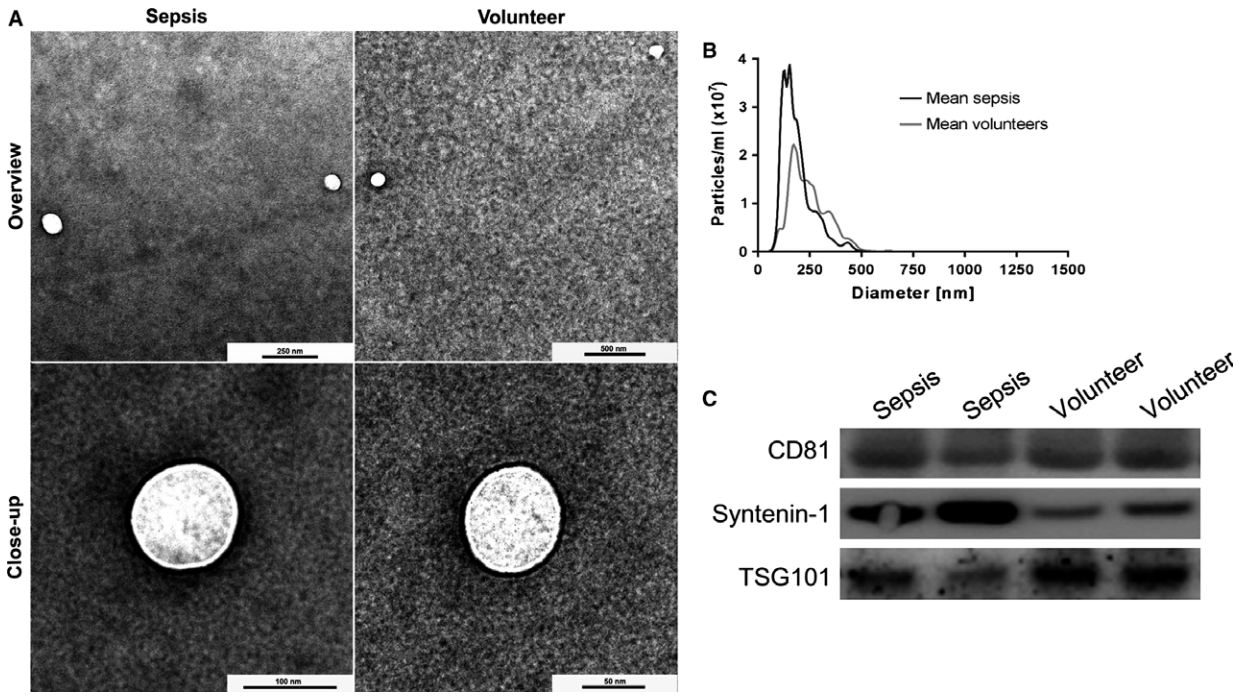

Fig. 1 Exosome characterization by transmission electron microscopy (TEM), nanoparticle tracking analysis (NTA) and WB. TEM revealed vesicles with an average size range of about 100–150 nm (**A**). NTA showed a sharp size distribution with a mean particle diameter of 154.4 ± 40.2 nm for sepsis patients (black) and 225.2 ± 24.3 nm for volunteers (grey) (**B**). Exosomes were detected by WB using the exosomal markers CD81, TSG101 and syntenin-1 (**C**).

included in the study showed a delta Cq < 7 and thus a low risk for haemolysis.

Next-generation sequencing

Blood cell, exosomal and serum RNA from seven patients diagnosed with septic shock as well as seven healthy volunteers were sequenced for small RNA.

Blood cell-derived RNA of 100 ng was used as starting material, while the entire RNA yields from 3 ml, or 600 μl serum were used for exosomal RNA and serum RNA, respectively. Library preparation was performed as described in Spornraft *et al.* [10]. Libraries were subsequently subjected to Illumina single-end sequencing by synthesis, using 50 cycles on the HiSeq 2500 (Illumina Inc., San Diego, California, USA).

Sequencing data were processed as described previously [10]. A $\log_2$ fold change ≥|1| and an adjusted p value (Padj) of ≤0.05 were set as thresholds to identify significantly regulated miRNAs. Only transcripts with a baseMean ≥50 were included in the analysis. For technical NGS data, see Figure S1.

Bioinformatic analysis

Ingenuity Pathway Analysis (IPA®) (Qiagen) was used for the identification of causal networks in our high-throughput miRNA expression data. In our analyses, we only considered pathways with a negative log *P* value >20. Only miRNAs meeting the pre-defined cut-off values mentioned above were entered into IPA®, and only experimentally confirmed or highly predicted relationships were considered for the analysis of miRNA effects.

Significantly regulated miRNAs exclusively present in one of the compartments (exosomes, serum and blood cells) but not in any of the others were entered into *IPA® microRNA Target Filter* to identify target genes of relevance to sepsis. For this purpose, disease filtering was set to *Antimicrobial Response*, *Inflammatory Response* and *Infectious Disease*. miRNAs from all three compartments were then included in a *Core Analyses* step, which identified pathways and causal networks regulated within each compartment. In a final step, *IPA® Comparison Analysis* was used to generate heat maps to visualize and compare the canonical pathways and disease states of relevance to miRNAs regulated within the three compartments.

RT-qPCR validation

In all, 20 miRNAs covering baseMeans from 69 to 279,000 were investigated by RT-qPCR in a new cohort of 15 septic patients on day 0 (= day of hospitalization) and day 4, as well as 16 healthy volunteers. geNorm [11] and NormFinder [12] were used to predict the most stable miRNAs for each of the three compartments based on the NGS data set, and six potential reference miRNA candidates were selected. For RT-qPCR, 10 ng of blood cell-derived RNA was reverse transcribed with the universal cDNA synthesis kit II according to the manufacturer's protocol (Exiqon). Exosomal and serum RNA were normalized to the volume of starting material (3 ml serum and 600 μl serum, respectively),

Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

and 2 μl of a 30 μl eluate were subjected to reverse transcription. RT-qPCR was performed in a 10 μl total reaction volume of ExiLENT SYBR® Green master mix and 1 μl of each miRNA LNA™ PCR primer (Exiqon). PCR reactions were performed on a Rotor-Gene Q (Qiagen) or CFX384 (Bio-Rad) real-time cycler. Relative quantification was carried out using the ΔΔCq method [13].

Statistical analysis

Demographic characteristics between volunteers and sepsis patients were compared using analysis of variance (ANOVA) with *post-hoc* testing. For comparison of demographic and clinical parameters between the NGS group and the RT-qPCR confirmation sample, Student's *t*-test was applied in case of normally distributed data and the Wilcoxon rank-sum test for non-parametric continuous variables. The Chi-square or Fisher's exact test was used for comparison of categorical variables. Data in the text and in tables are given as mean ± SD. To increase clarity in some of the figures, NGS and RT-qPCR data are expressed as mean ± SEM. Statistical calculations were performed using SPSS (version 24.0, IBM GmbH, Ehningen, Germany).

Ethics approval and consent to participate

The study was approved by the Ethics Committee of the Medical Faculty of the University of Munich (protocol #551-14). The study was carried out according to the World Medical Association Declaration of Helsinki and all study samples were anonymized during analysis.

Written informed consent for publication of blinded individual personal data was obtained from each participant or the patient's legal representative.

Results

Exosome isolation and characterization

WB analysis revealed the presence of exosome markers CD81, TSG101 and syntenin-1. Single particle analysis by NTA revealed a narrow size distribution with mean particle diameters of 154.4 ± 40.2 nm (mode: 110.7 ± 23.5 nm) and 225.2 ± 24.3 nm (mode: 169.7 ± 21.8 nm) for septic patients and healthy volunteers, respectively. No significant differences in particle number were observed between the two groups. TEM revealed a homogeneous population, no morphological differences were observed in vesicles from healthy volunteers and sepsis patients (Fig. 1).

Differential gene expression analysis

For each of the three analysed compartments, miRNA reads were compared between septic shock patients and healthy volunteers. Compared to healthy volunteers, 77 distinct miRNAs were down-regulated, whereas 103 distinct miRNAs were up-regulated in septic shock patients. The majority of regulated miRNAs was found exclusively in blood cell-derived samples and amounted to 62.14% for up-regulated and 64.94% for down-regulated miRNAs, respectively. The extracellular compartment accounted for 23.38% of all down-regulated (exosome: 14.29%; serum: 3.90%; both: 5.19 %) and 28.16% of all up-regulated (exosome: 15.53%; serum: 3.88%; both: 8.74%) miRNAs. A small fraction of miRNAs was found to be simultaneously up- or down-regulated in all three compartments (2.91% and 3.90%, respectively) (Table S2).

Newly identified sepsis-associated miRNAs

The literature was searched for previously sepsis-associated miRNAs and compared to significantly regulated miRNAs from this study. For this purpose, each miRNA was searched for in PubMed in combination with the terms 'sepsis'; 'lipopolysaccharide' and 'lps'. miRNAs with ambiguous literature results were considered as PubMed hits and thus categorized as previously sepsis-associated. Applying the mentioned criteria, our NGS dataset provided 32 newly identified sepsis-associated miRNAs in exosomes, 14 in serum and 73 in blood cells (Table S2). Seventeen of these miRNAs were found exclusively in exosomes, respectively, four and 67 miRNAs in serum or blood cells. One miRNA, miR-423-5p, was present in all three compartments, showing down-regulation in the sepsis shock cohort.

RT-qPCR validation

Subsequent to normalization with the geometric mean of the reference miRNAs miR-185-5p and miR-25-3p, differential miRNA expression between healthy volunteers and patients with sepsis or septic shock were assessed. For exosome and serum samples, seven out of 12 miRNAs (58.3%) selected from NGS results could be validated in RT-qPCR, while nine out of 14 miRNAs (64.3%) were validated for cellular samples. Log_2 fold changes of validated miRNAs for septic shock patients in NGS ($n = 7$) and RT-qPCR ($n = 6$) were correlated for each compartment (Fig. S2). One of the validated miRNAs was present in all compartments while one, one and eight were unique to exosomes, serum and blood cells, respectively. Five miRNAs were present in both extracellular compartments, but not in blood cells (Fig. 2).

Disease detection by miRNAs

RT-qPCR data were screened for miRNAs distinguishing patients with both sepsis and septic shock from healthy volunteers on the day of admission to the ICU. One miRNA, cellular miR-199b-5p, also indicated the less severe stage with high confidence ($P < 0.001$) (Fig. 3).

miRNA expression levels across sepsis stages

In order to assess potential regulations of miRNAs across disease stages, expression levels from healthy volunteers ($n = 16$) and patients at study inclusion (day 0) suffering from sepsis ($n = 9$) and septic

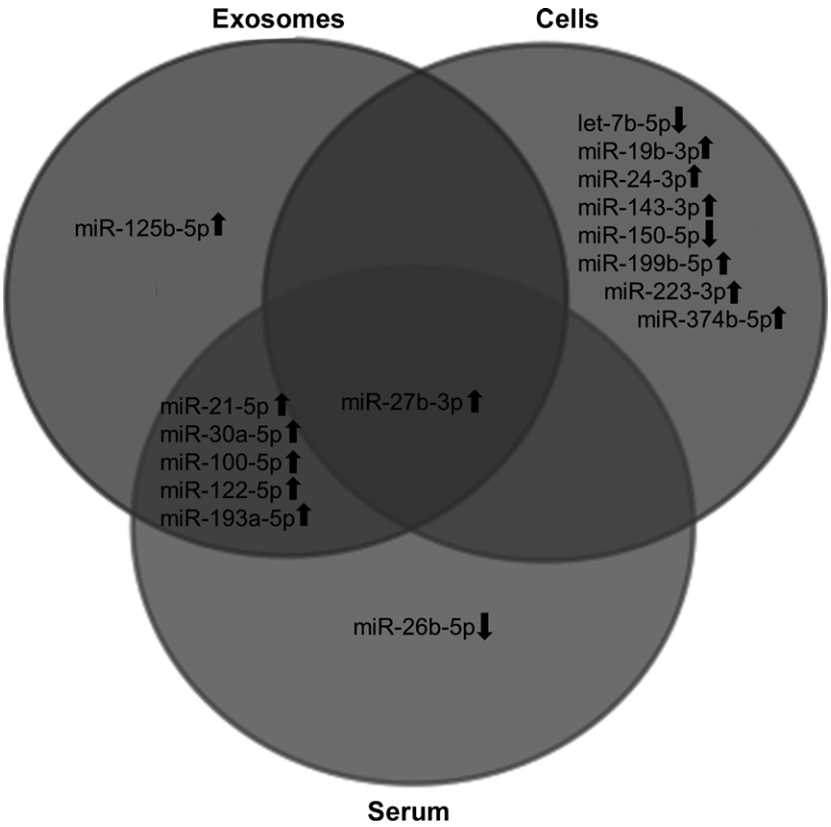

Fig. 2 Venn diagram of differential miRNA expression profiles from septic shock patients validated by RT-qPCR.

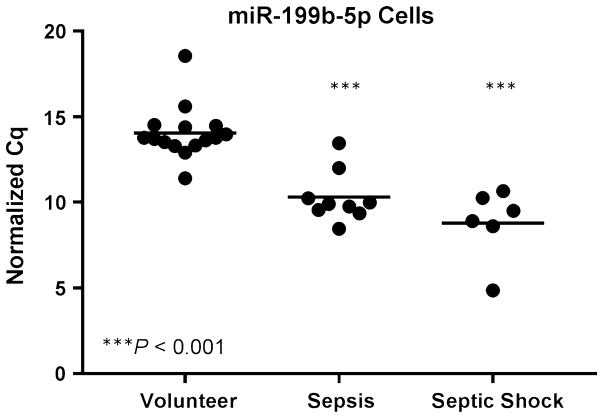

Fig. 3 Sepsis detection by miRNAs: Only cellular miR-199b-5p detected both sepsis and septic shock with high confidence ($n = 6$).

shock ($n = 6$) were compared. Almost linear correlations of expression level and disease severity were found for 11 different miRNAs in at least one compartment (three exosomal, one serum, nine cellular) (Fig. 4). With the exception of cellular miR-150-5p and let-7b-5p, all miRNAs correlating to disease severity were up-regulated. Two miRNAs showed correlations in two compartments: miR-27b-3p was simultaneously up-regulated in exosomes and blood cells from septic patients, while miR-193a-5p was exclusively up-regulated in extracellular samples. All 11 miRNAs showed significant differences ($P < 0.05$) in expression levels between the healthy state and septic shock. Additionally, multiple miRNAs in exosomes (miR-21-5p, miR-193a-5p), serum (miR-26b-5p) and cells (let-7b-5p, miR-27b-3p, miR-143-3p, miR-150-5p, miR-199b-5p, miR-223-3p) were also able to differentiate ($P < 0.05$) between healthy volunteers and septic patients (Fig. S3).

Survival prediction by extracellular miRNAs

RT-qPCR patients were grouped according to outcome (death, $n = 4$; survival, $n = 11$) irrespective of sepsis stage. On the day of admission to the ICU, three extracellular miRNAs (exosomal miR-30a-5p and miR-125b-5p, miR-193a-5p in serum) were able to predict survival with high confidence ($P < 0.001$) (Fig. 5). Septic shock was markedly associated with death (75.0% of non-survivors and 27.3% of survivors were characterized to be in septic shock) with significantly different expression levels between disease stages for miR-193a-5p and miR-125b-5p.

Longitudinal analysis of miRNAs correlating with disease stage

Expression levels of selected miRNAs on day 0 and day 4 were assessed by RT-qPCR in a separate experiment on patients with

Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

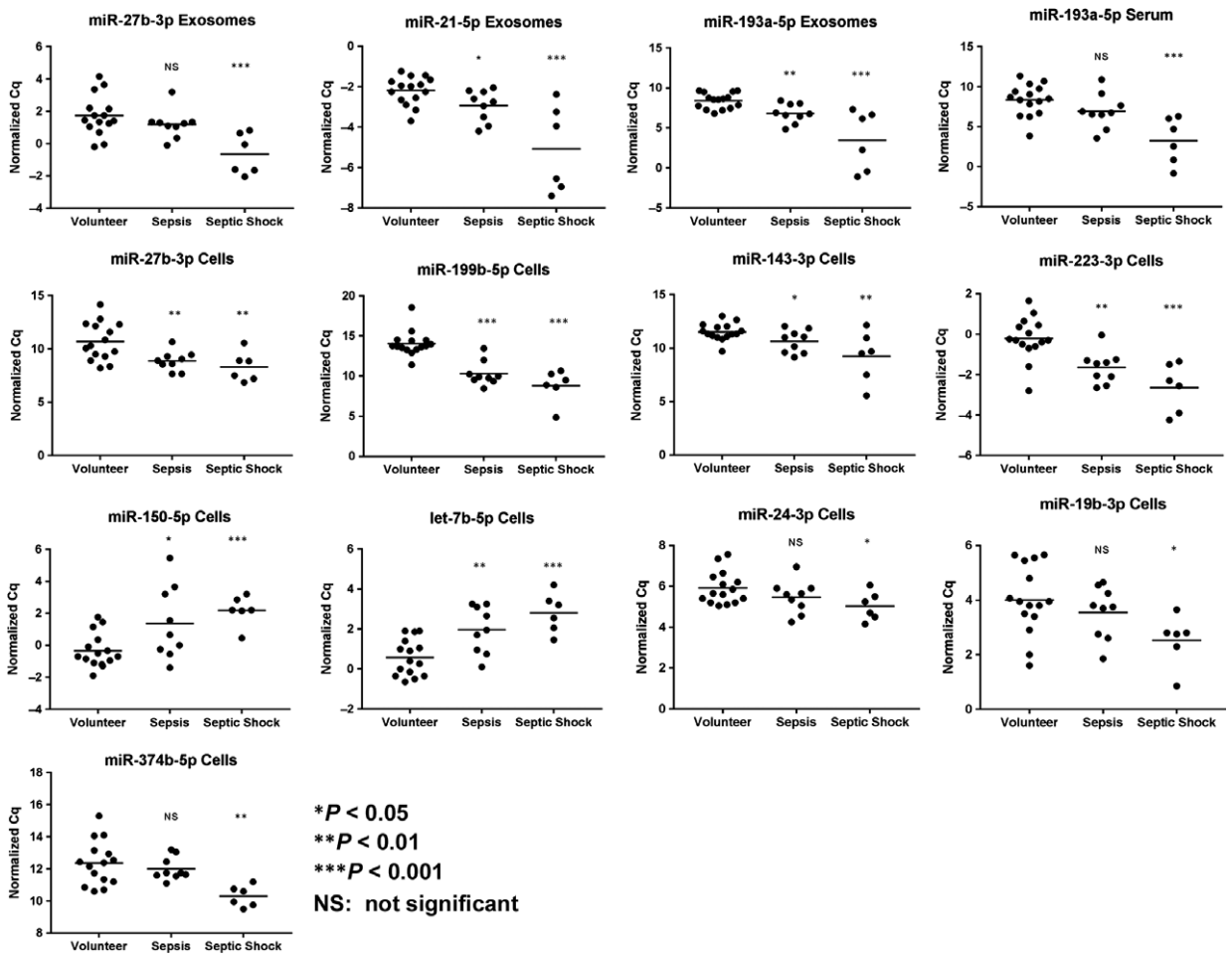

Fig. 4 Relationship between miRNA expression level and disease severity. Eleven miRNAs in at least one compartment display stringent correlation of normalized Cq values and disease severity in septic patients upon admission to the ICU.

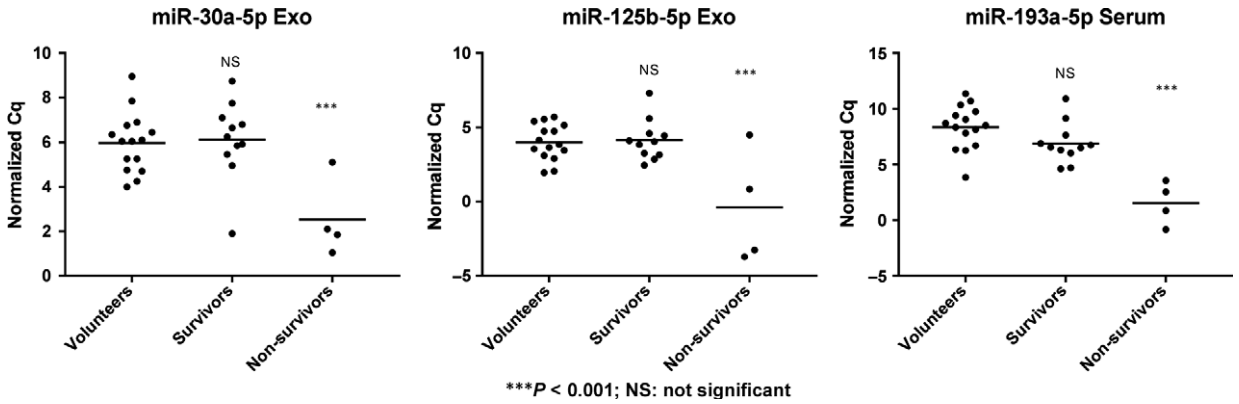

Fig. 5 Survival prediction by miRNAs. Exosomal miR-30a-5p and miR-125b-5p and serum miR-193a-5p predict survival of septic patients with high confidence.

sepsis (n = 3) and septic shock (n = 5). Some miRNAs showed a normalization of expression levels on day 4 (miR-100-5p exosomes, miR-26b-5p serum, miR-143-3p cells), while exosomal miR-199b-5p was found to become more up-regulated compared to healthy volunteers (n = 16) (Fig. 6).

In silico analyses of differentially expressed miRNAs

Top canonical pathways identified by the target disease-driven approach mentioned above were *Pathogen Influenced Signalling, Cellular and Humeral Immune Response* including *Cytokine Signalling, Cellular Stress Response and Injury* and *Cardiovascular Signalling* (Fig. S4).

Discussion

There are a number of earlier studies that have investigated miRNAs in the context of sepsis. However, only two studies took a genome-wide approach for assessing the complete miRNA expression profile in human patients with sepsis in comparison with non-septic controls by means of NGS [14, 15]. Another study by Vasilescu *et al.* [16] profiled miRNAs derived from peripheral blood leukocytes by microarray analysis and validated the most dysregulated miRNA in their sepsis cohort, miR-150, in plasma samples from septic patients. Although their clinical study used a less sensitive approach (microarray instead of NGS), it is the only one investigating at least one miRNA in both cellular and extracellular blood compartments in human sepsis patients. In light of the scarcity of miRNA profiling studies in patients with this disorder, and taking into consideration the limited possibility of inter-study comparison due to different compartments, it is essential to comprehensively profile the blood miRNA signature of critically ill patients based on all compartments.

Comprehensive small RNA-Seq of blood compartments in septic individuals, performed in this study, revealed compartment-specific differences in miRNA profiles. While exosomes and serum were found to share a significant number of miRNAs, there was very limited overlap in the miRNA profiles of cellular and extracellular samples. Individual sampling compartments in blood thus differ not only in RNA quantity, but also display qualitatively distinct profiles of bioactive nucleic acids. Particularly, exosomes, released by various cell populations in the body and thus present in the extracellular blood fraction, play a major role as miRNA carriers and mediators of intercellular communication [17–19]. Considered an enriched sampling fraction, the exosomal miRNA profile may serve as an excellent source of diagnostic biomarkers for identifying sepsis patients early on, discriminating different sepsis stages or even predicting the course of the disease. In this context, it is worth mentioning that, to our knowledge, no clinical study characterizing miRNAs derived from exosomes as a miRNA-relevant compartment in blood has been carried out in sepsis patients before.

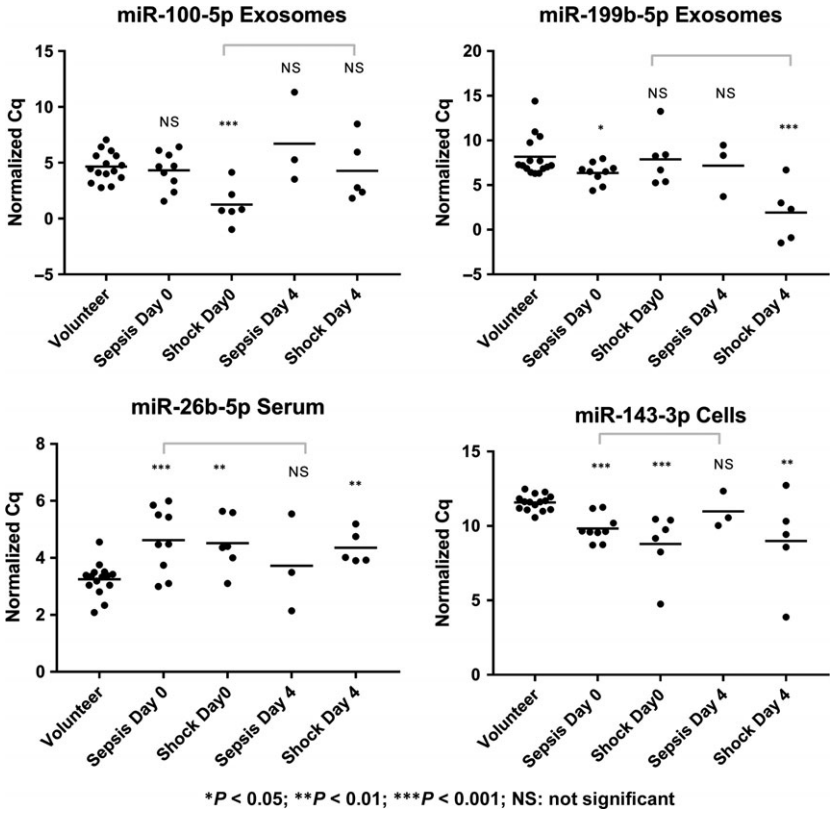

Fig. 6 Significantly different miRNA expression levels between days 0 and 4. Samples on day 4 were available from eight of the patients with sepsis or septic shock previously used in RT-qPCR expression analysis at day 0. Grey bars indicate notable changes in expression within a group of patients. Tested miRNAs were exosomal miR21-5p, miR27b-3p, miR100-5p, miR193a-5p, miR199b-5p, serum miR26b-5p and miR193a-5p, as well as blood cellular miR27b-3p, miR143-3p, miR150-5p, miR199b-5p, miR223-3p and let-7b-5p.

Journal of Cellular and Molecular Medicine published by John Wiley & Sons Ltd and Foundation for Cellular and Molecular Medicine.

When profiling miRNAs differentially regulated in sepsis patients on the day of admission to the ICU by RT-qPCR, only one blood cell-derived miRNA (miR-199b-5p) distinguished both sepsis and septic shock from healthy volunteers. Conversely, the miRNAs significantly associated with survival (miR-30a-5p, miR-125b-5p, miR-193a-5p) were found to be extracellular. Due to this non-uniform distribution of miRNA information across sample types, it seems prudent to choose the appropriate blood compartment for the respective question at hand.

In our data, tracking miRNAs in patients upon admission to the ICU and 96 hrs afterwards revealed varying courses of regulation which were, again, compartment specific. While expression of some up-/down-regulated miRNAs normalized over time, others were found to deviate even further from healthy volunteers. This might be due to heterogeneous patient responses to clinical care, differences in type of treatment and medication received or a combination thereof. Serial measurements of compartment-specific miRNAs might therefore also hold value for sepsis management.

As a conclusion, this study presents the first comprehensive profiling of cellular and extracellular miRNAs in sepsis patients. We found a distinctly compartment-specific regulation of miRNAs between sepsis patients and healthy volunteers. NGS-based profiling allowed the identification of sepsis patients as opposed to healthy volunteers, prognosis of patient survival and elucidation of miRNA disease functions. Additionally, we detected multiple compartment-specific, highly regulated miRNAs that had not previously been associated with sepsis. Most importantly, our data demonstrate the benefit of including exosomal miRNAs in disease assessment protocols. The practicality and applicability of sampling miRNAs in liquid biopsies will enhance biomarker research and eventually the clinical management of sepsis.

Acknowledgements

We thank Franz Jansen for excellent technical assistance. We thank Renate Scheler and PD Dr. Ricarda Schumann from the University Eye Hospital LMU Munich for excellent assistance with TEM. This work was supported by a grant from the German Federal Ministry of Economy and Energy (KF3361701MD4 and WB1317). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. MR, IK, MaP and GS conceived and designed the experiments; MM, IK and DH recruited patients and provided samples; MR and DB performed the experiments; DB, BK, MR, MM and GS performed validation and formal analysis; MR, DB, MM, BK, IK, AC, MiP and GS curated and analysed the data; MR, DB and GS wrote the paper; BK, MiP, AC and OS reviewed and revised the paper. All authors read and approved the final manuscript.

Conflict of interest

The authors confirm that there are no conflicts of interest.

Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

Figure S1 Technical NGS quality and quantity data for each matrix.

Figure S2 Correlation of NGS and RT-qPCR log2 fold changes for validated miRNAs.

Figure S3 miRNAs differentiating between healthy volunteers and sepsis patients.

Figure S4 Heat maps for the 5 Top Canonical Pathway networks.

Table S1 Summary of demographic and clinical characteristics of study groups.

Table S2 Significantly regulated miRNAs in septic shock patients detected by NGS.

References

1. **Angus DC, van der Poll T.** Severe sepsis and septic shock. *N Engl J Med*. 2013; 369: 2063.
2. **Kumar G, Kumar N, Taneja A, *et al.*** Nationwide trends of severe sepsis in the 21st century (2000–2007). *Chest*. 2011; 140: 1223–31.
3. **SepNet Critical Care Trials Group**. Incidence of severe sepsis and septic shock in German intensive care units: the prospective, multicentre INSEP study. *Intensive Care Med*. 2016;42: 1980–1989.
4. **Kalra H, Drummen GP, Mathivanan S.** Focus on extracellular vesicles: introducing the next small big thing. *Int J Mol Sci*. 2016; 17: 170.
5. **Mittelbrunn M, Gutierrez-Vazquez C, Villarroya-Beltri C, *et al.*** Unidirectional transfer of microRNA-loaded exosomes from T cells to antigen-presenting cells. *Nat Commun*. 2011; 2: 282.
6. **O'Driscoll L.** Expanding on exosomes and ectosomes in cancer. *N Engl J Med*. 2015; 372: 2359–62.
7. **Azevedo LC, Janiszewski M, Pontieri V, *et al.*** Platelet-derived exosomes from septic shock patients induce myocardial dysfunction. *Crit Care*. 2007; 11: R120.
8. **Gambim MH, do Carmo Ade O, Marti L, *et al.*** Platelet-derived exosomes induce endothelial cell apoptosis through peroxynitrite generation: experimental evidence for a novel mechanism of septic vascular dysfunction. *Crit Care*. 2007; 11: R107.
9. **Blondal T, Jensby Nielsen S, Baker A, *et al.*** Assessing sample and miRNA profile quality in serum and plasma or other biofluids. *Methods*. 2013; 59: S1–6.
10. **Spornraft M, Kirchner B, Haase B, *et al.*** Optimization of extraction of circulating RNAs from plasma–enabling small RNA sequencing. *PLoS ONE*. 2014; 9: e107259.
11. **Vandesompele J, De Preter K, Pattyn F, *et al.*** Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002; 3: RESEARCH0034.
12. **Andersen CL, Jensen JL, Orntoft TF.** 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*. 2004; 64: 5245–50.

13. **Livak KJ, Schmittgen TD.** Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C (T)) Method. *Methods*. 2001; 25: 402–8.
14. **Caserta S, Kern F, Cohen J, *et al.*** Circulating plasma microRNAs can differentiate human sepsis and systemic inflammatory response syndrome (SIRS). *Sci Rep*. 2016; 6: 28006.
15. **Ma Y, Vilanova D, Atalar K, *et al.*** Genome-wide sequencing of cellular microRNAs identifies a combinatorial expression signature diagnostic of sepsis. *PLoS ONE*. 2013; 8: e75918.
16. **Vasilescu C, Rossi S, Shimizu M, *et al.*** MicroRNA fingerprints identify miR-150 as a plasma prognostic marker in patients with sepsis. *PLoS ONE*. 2009; 4: e7405.
17. **Dreyer F, Baur A.** Biogenesis and functions of exosomes and extracellular vesicles. *Methods Mol Biol*. 2016; 1448: 201–16.
18. **Gyorgy B, Szabo TG, Pasztoi M, *et al.*** Membrane vesicles, current state-of-the-art: emerging role of extracellular vesicles. *Cell Mol Life Sci*. 2011; 68: 2667–88.
19. **Lee Y, El Andaloussi S, Wood MJ.** Exosomes and microvesicles: extracellular vesicles for genetic information transfer and gene therapy. *Hum Mol Genet*. 2012; 21: R125–34.