Issues and Prospects of microRNA-Based Biomarkers in Blood and Other Body Fluids

Cell-free circulating microRNAs (miRNAs) in the blood are good diagnostic biomarker candidates for various physiopathological conditions, including cancer, neurodegeneration, diabetes and other diseases. Since their discovery in 2008 as blood biomarkers, the field has expanded rapidly with a number of important findings. Despite the initial optimistic views of their potential for clinical application, there are currently no circulating miRNA-based diagnostics in use. In this article, we review the status of circulating miRNAs, examine different analytical approaches, and address some of the challenges and opportunities.
ABSTRACT

Cell-free circulating microRNAs (miRNAs) in the blood are good diagnostic biomarker candidates for various physiopathological conditions, including cancer, neurodegeneration, diabetes and other diseases. Since their discovery in 2008 as blood biomarkers, the field has expanded rapidly with a number of important findings. Despite the initial optimistic views of their potential for clinical application, there are currently no circulating miRNA-based diagnostics in use. In this article, we review the status of circulating miRNAs, examine different analytical approaches, and address some of the challenges and opportunities.
Issues and Prospects of microRNA-Based Biomarkers in Blood and Other Body Fluids
Review

Issues and Prospects of microRNA-Based Biomarkers in Blood and Other Body Fluids

John R. Chevillet 1, Inyoul Lee 1, Hilary A. Briggs 2, Yuqing He 2,* and Kai Wang 1,*

1 Institute for Systems Biology, Seattle, WA 98109, USA
2 Institute of Medical Systems Biology, Guangdong Medical College, Dongguan, Guangdong 523808, China
3 Laboratory of the Biology of Addictive Diseases, Rockefeller University, New York, NY 10065, USA

* Authors to whom correspondence should be addressed; E-Mails: dr.hyq@hotmail.com (Y.H.); kwang@systemsbiology.org (K.W.).

Received: 17 March 2014; in revised form: 7 May 2014 / Accepted: 12 May 2014 / Published: 14 May 2014

Abstract: Cell-free circulating microRNAs (miRNAs) in the blood are good diagnostic biomarker candidates for various physiopathological conditions, including cancer, neurodegeneration, diabetes and other diseases. Since their discovery in 2008 as blood biomarkers, the field has expanded rapidly with a number of important findings. Despite the initial optimistic views of their potential for clinical application, there are currently no circulating miRNA-based diagnostics in use. In this article, we review the status of circulating miRNAs, examine different analytical approaches, and address some of the challenges and opportunities.

Keywords: diagnostic; exosomes; microvesicles; quantitative methods

1. Introduction

MicroRNAs (miRNAs) were originally identified as small, non-coding RNA mediators of temporal pattern formation in C. elegans [1–5], which regulated messenger RNAs (mRNAs) by base pairing with their 3'-untranslated regions (3'-UTR). Later research showed that miRNAs (and this type of RNA mediated regulatory processes) are conserved in metazoan species [6–9] and have also been found in viruses [10]. MiRNAs are currently differentiated from other small RNAs by: (1) a hairpin
fold-back precursor structure in their primary transcript, as predicted by base composition (2) and an approximately 18–22 nucleotide long mature sequence [11]. At present, 1,872 precursor sequences and 2,578 mature human sequences have been catalogued by the online repository miRBase (v20) [12].

Primary miRNA transcripts (pri-miRNAs) are transcribed in the nucleus and processed by an RNase III enzyme, Drosha [13]. The resulting smaller hairpin pre-miRNA is transported to the cytoplasm via a RanGTP-dependent double stranded RNA binding protein, exportin 5. The pre-miRNA is then further processed by another RNase III, Dicer [14], into an approximately 22 nucleotide long duplex RNA with 3’ overhangs. The duplex is subsequently unwound by RNA helicase and usually one strand (the guide strand) is preferentially associated with the effector-protein Argonaute [15,16], as part of the RNA Induced Silencing Complex (RISC). The other strand (the passenger strand) is degraded. The process for strand selection is yet to be fully understood. The miRNA-RISC complex can then interact with mRNA targets to inhibit translation or induce degradation of the mRNA, typically through binding at the 3’-untranslated region (3’-UTR) with partial sequence complementarity. The RNase III cleavages are often imprecise, resulting in end region sequence variations (especially the 3’ end) of mature miRNAs. In addition to this, miRNAs can be post-transcriptionally modified (e.g., via non-templated additions), resulting in sequence variants (iso miRs) [17–20], which may have functionally distinct properties [21–23], expanding the miRNA repertoire.

Through interacting with mRNA targets, miRNAs predominantly act to modulate the transcriptome of cells [24] and have been implicated in various physiopathological processes, including development, homeostasis and cellular pathology [25,26]. Based on various prediction algorithms, the majority of mRNA transcripts are potential targets for miRNAs [27]. MicroRNA-target binding is usually mediated through the initial interaction of the seed region (nucleotides 2-8 of the 5’ end of the miRNA) [13], but a significant fraction (approximately 60% based on one report [28]) of seed interactions are non-canonical (i.e., contain bulging or mismatches) and are generally associated with additional, non-seed base pairing. These flexible constraints on specificity permit a single miRNA to target multiple (possibly hundreds) of mRNAs and elicit effects on multiple biological processes. A single mRNA can also contain binding sites for multiple different miRNA sequences. This makes miRNA-mediated transcriptome/proteome regulation a complicated process and difficult to decipher.

The expression of certain miRNAs is restricted to specific cell types [29] or developmental stages; therefore, they are frequently dysregulated in pathologies including cancer [30–33], cardiovascular diseases [34–41] and neurological conditions [42–46]. In the case of cancer, analyzing the miRNA spectrum of tumors yields qualitative diagnostic information and allows possible identification of the tissue of origin and in some cases, provides molecular stratification of tumor subtypes [47–49].

2. Early Studies of Circulating miRNA

The presence of cell-free nucleic acids in blood circulation has been established for almost 60 years [50–54]. Tumor derived DNA and RNA are commonly observed in cancer patient plasma samples [55–58]. RNA was historically thought to be unsuitable as a practical blood-based biomarker, due to the high level of nuclease activity in human plasma [59], but this skeptical position was reconsidered after the observation that miRNAs are stable in fixed tissues [60] and that the
concentration of miR-21 in serum could distinguish patients with B-cell lymphoma from control individuals [61]. Subsequent work detected specific circulating miRNA signatures in healthy individuals and patients with prostate, breast and colorectal cancers [62–67], in addition to other disorders. Detailed examination [62] revealed that the detected species were bona fide cell-free miRNAs and explicitly demonstrated their stability outside of the cells (tolerating prolonged incubation at room temperature in addition to multiple cycles of freezing and thawing of the plasma samples). The release of miRNA from tumors was further demonstrated by detecting human miRNA sequences in plasma drawn from mice harboring human xenografts. The field expanded rapidly, describing numerous circulating miRNA biomarkers for cancer (reviewed elsewhere in this issue [68], and other conditions including neurodegeneration, cardiovascular disease and metabolic disease (Table 1).

Although it has been demonstrated that miRNAs can be released into circulation from cells specific to the disease (e.g., miR-122 from liver toxicity studies [69,70]) the cellular origin of many miRNA in biofluids is unclear. For example, blood levels of miRNAs highly expressed in heart tissues, such as miR-208a and miR-499, were included as a part of signatures to differentiate patients with cardiovascular disease in several studies [71–74]. This suggests that increased levels of heart-specific blood miRNAs may derive from the heart and directly reflect the heart condition of patients. However whether all reported miRNAs are actually directly associated with the disease or are the product of nonspecific, systemic or secondary responses remains to be further evaluated.

**Table 1. Circulating miRNA biomarkers for neurodegeneration, cardiovascular and metabolic diseases.** Red text: increased abundance relative to controls, green text: reduced abundance relative to controls. (AD) Alzheimer Disease, (CAD) Coronary Artery Disease, (GDM) Gestational Diabetes Mellitus, (HD) Huntington Disease, (IFG) impaired fasting glucose, (IGT) impaired glucose tolerance, (MCI) Mild Cognitive Impairment, (NEC) Normal Elderly Controls, (NGS) Next Generation Sequencing, (PBMC) Peripheral Blood Mononuclear Cells, (PD) Parkinson Disease, (T1D) Type 1 Diabetes, (T2D) Type 2 Diabetes, (RT-qPCR) Reverse Transcription Quantitative PCR.

<table>
<thead>
<tr>
<th>Disease</th>
<th>miRNA</th>
<th>Detection Method</th>
<th>Specimen</th>
<th>Population</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neurodegenerative Disease</td>
<td>miR-34a, -181b, -200a, let-7f</td>
<td>Microarray, RT-qPCR</td>
<td>PBMC</td>
<td>16 AD patients and 16 NEC matched for ethnicity, age, gender and education</td>
<td>[75]</td>
</tr>
<tr>
<td>Alzheimer Disease</td>
<td>60 miRNAs (37 higher abundance, 23 lower abundance)</td>
<td>RT-qPCR</td>
<td>CSF</td>
<td>10 AD, 10 controls</td>
<td>[76]</td>
</tr>
<tr>
<td>miRNA-9, -125b, -146a, -155</td>
<td>Microarray, Northern dot blot</td>
<td>CSF</td>
<td>6 AD, 6 controls</td>
<td>[77]</td>
<td></td>
</tr>
<tr>
<td>miR-137, -181c, -9, -29a, -29b</td>
<td>RT-qPCR</td>
<td>Serum</td>
<td>7 AD, 7 controls</td>
<td>[78]</td>
<td></td>
</tr>
<tr>
<td>miR-29b</td>
<td>RT-qPCR</td>
<td>PBMC</td>
<td>393 AD, 412 controls</td>
<td>[79]</td>
<td></td>
</tr>
</tbody>
</table>
Table 1. Cont.

<table>
<thead>
<tr>
<th>Disease</th>
<th>miRNA</th>
<th>Detection Method</th>
<th>Specimen</th>
<th>Population</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alzheimer Disease</td>
<td>miR-15a</td>
<td>RT-qPCR</td>
<td>Plasma, CSF</td>
<td>11 AD, 9 MCI, 20 NC; 20 AD, 17 NC</td>
<td>[80]</td>
</tr>
<tr>
<td></td>
<td>let-7d-5p, let-7g-5p, miR-15b-5p, -142-3p, -191-5p, -301a-3p, -545-3p</td>
<td>nCounter, RT-qPCR</td>
<td>Plasma</td>
<td>Screening: 11 AD, 20 NC; validation 20 AD, 17 NC</td>
<td>[81]</td>
</tr>
<tr>
<td></td>
<td>miR-34c</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>110 AD, 123 NEC</td>
<td>[82]</td>
</tr>
<tr>
<td></td>
<td>miR-146a</td>
<td>RT-qPCR</td>
<td>CSF</td>
<td>10 AD, 10 early AD, 11 controls</td>
<td>[83]</td>
</tr>
<tr>
<td>Parkinson Disease</td>
<td>miR-1, -22*, -29a, -16-2*, -26a2*, -30a</td>
<td>RT-qPCR</td>
<td>Whole blood</td>
<td>15 PD (8 untreated PD), 8 controls</td>
<td>[84]</td>
</tr>
<tr>
<td></td>
<td>18 miRNAs</td>
<td>Microarray</td>
<td>PBMC</td>
<td>19 PD, 13 controls</td>
<td>[85]</td>
</tr>
<tr>
<td></td>
<td>miR-222, -626, -505</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>32 PD, 34 controls; 42 PD, 30 controls; 30 PD, 8 controls</td>
<td>[86]</td>
</tr>
<tr>
<td></td>
<td>miR-331-5p</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>31 PD, 25 controls</td>
<td>[87]</td>
</tr>
<tr>
<td></td>
<td>16 miRNAs (10 higher abundance, 6 lower abundance)</td>
<td>NGS</td>
<td>Blood leukocyte</td>
<td>7 PD, 6 controls</td>
<td>[88]</td>
</tr>
<tr>
<td>Huntington Disease</td>
<td>miR-34b</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>27 HD, 12 controls</td>
<td>[89]</td>
</tr>
<tr>
<td>Coronary Artery Disease</td>
<td>miR-126, -17, -92a, -155, -145; -133a, -208a</td>
<td>RT-qPCR</td>
<td>Plasma, serum</td>
<td>8 CAD, 8 controls; 36 CAD, 17 controls</td>
<td>[71]</td>
</tr>
<tr>
<td></td>
<td>miR-135a and miR-147</td>
<td>RT-qPCR</td>
<td>PBMC</td>
<td>50 CAD, 20 controls</td>
<td>[90]</td>
</tr>
<tr>
<td></td>
<td>miR-146a and miR-146b</td>
<td>RT-qPCR</td>
<td>PBMC</td>
<td>41 CAD, 15 controls</td>
<td>[91]</td>
</tr>
<tr>
<td></td>
<td>miR-19a, -584, -155, -222, -145, -29a, -378, -342, -181d, -30e-5p, -150)</td>
<td>Microarray, RT-qPCR</td>
<td>Whole blood</td>
<td>5 CAD, 5 controls for initial; 10 CAD, 15 controls for validation</td>
<td>[92]</td>
</tr>
<tr>
<td></td>
<td>miR-214</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>40 CAD, 15 controls</td>
<td>[93]</td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>miR-1, -133a, -499, -208a</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>33 AMI, 30 controls</td>
<td>[72]</td>
</tr>
<tr>
<td></td>
<td>miR-1</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>93 AMI, 66 controls</td>
<td>[94]</td>
</tr>
<tr>
<td></td>
<td>miR-1</td>
<td>RT-qPCR</td>
<td>Serum</td>
<td>31 AMI, 20 controls</td>
<td>[95]</td>
</tr>
<tr>
<td></td>
<td>miR-499</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>14 AMI, 15 heart failure patients, 10 controls</td>
<td>[73]</td>
</tr>
<tr>
<td></td>
<td>miR-126, -197, -223</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>820 Bruneck cohort</td>
<td>[96]</td>
</tr>
</tbody>
</table>
## Table 1. Cont.

<table>
<thead>
<tr>
<th>Disease</th>
<th>miRNA</th>
<th>Detection Method</th>
<th>Specimen</th>
<th>Population</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiovascular Disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acute Myocardial Infarction</td>
<td>miR-1915, -181</td>
<td>RT-qPCR</td>
<td>Whole blood</td>
<td>60 AMI, 21 controls, 5 time points (0–24 h)</td>
<td>[98]</td>
</tr>
<tr>
<td></td>
<td>miR-133a</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>13 AMI patients, 176 angina pectoris patients, 127 controls</td>
<td>[99]</td>
</tr>
<tr>
<td></td>
<td>miR-1, -134, -186, -208, -223 and -499</td>
<td>NGS, RT-qPCR</td>
<td>Serum</td>
<td>117 AMI patients, 182 AP patients, 100 controls</td>
<td>[74]</td>
</tr>
<tr>
<td>Congestive Heart Failure</td>
<td>miR-210</td>
<td>RT-qPCR</td>
<td>PBMC</td>
<td>13 patients, 6 controls</td>
<td>[100]</td>
</tr>
<tr>
<td>Aortic Aneurysm</td>
<td>miR-126</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>33 patients, 17 controls</td>
<td>[101]</td>
</tr>
<tr>
<td></td>
<td>miR-29b, -124, -155, -223</td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>23 patients, 12 healthy controls, 17 coronary artery disease patients</td>
<td>[102]</td>
</tr>
<tr>
<td>Stroke</td>
<td>miR-125b-2*, -27a, -422a, -488, -627</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>169 stroke patients, 94 metabolic syndrome patients, 24 healthy controls</td>
<td>[103]</td>
</tr>
<tr>
<td></td>
<td>miR-145</td>
<td>RT-qPCR</td>
<td>Whole Blood</td>
<td>32 ischemic stroke patients, 14 healthy controls</td>
<td>[104]</td>
</tr>
<tr>
<td>Atherosclerosis</td>
<td>miR-130a, -27b, -210</td>
<td>RT-qPCR</td>
<td>Serum</td>
<td>104 patients, 105 controls</td>
<td>[105]</td>
</tr>
<tr>
<td><strong>Metabolic Disease</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Type 1 Diabetes</td>
<td>miR-152, -30a-5p, -181a, -24, -148a, -210, -29a, -26a, -27b, -25, -200a</td>
<td>NGS, RT-qPCR</td>
<td>Serum</td>
<td>pooled from 2 T1D groups (275, 129) and one control group (n = 151)</td>
<td>[106]</td>
</tr>
<tr>
<td></td>
<td>miR-20b, -21, -24, -15a, -126, -191, -197, -223, -320, -486, -150, -29b, -28-3p</td>
<td>Microarray, RT-qPCR</td>
<td>Plasma</td>
<td>80 patients, 80 controls</td>
<td>[107]</td>
</tr>
<tr>
<td></td>
<td>miR-9, -29a, -30d, -34a, -124a, -146a and -375</td>
<td>RT-qPCR</td>
<td>Serum</td>
<td>18 T2D, 19 pre-diabetes (IGT and/or IFG), 19 controls</td>
<td>[108]</td>
</tr>
<tr>
<td>Disease</td>
<td>miRNA</td>
<td>Detection Method</td>
<td>Specimen</td>
<td>Population</td>
<td>Ref.</td>
</tr>
<tr>
<td>-----------------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>----------</td>
<td>-------------------------------------------------</td>
<td>------------</td>
</tr>
<tr>
<td><strong>Type 2 Diabetes</strong></td>
<td><strong>miR-146a</strong></td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>90 patients, 90 controls</td>
<td>[109]</td>
</tr>
<tr>
<td><strong>miR-29a</strong></td>
<td></td>
<td>RT-qPCR</td>
<td>Urine</td>
<td>83 patients (42 with albuminuria, 41 with normoalbuminuria)</td>
<td>[110]</td>
</tr>
<tr>
<td><strong>miR-24 and miR-29b, (miR-144 in Swedish T2D)</strong></td>
<td>RT-qPCR</td>
<td>Plasma</td>
<td>33 patients (14 Swedes, 19 Iraqis), 119 controls</td>
<td>[111]</td>
<td></td>
</tr>
<tr>
<td><strong>Gestational Diabetes Mellitus</strong></td>
<td><strong>miR-132, -29a and miR-222</strong></td>
<td>RT-qPCR</td>
<td>Serum</td>
<td>24 GDM, 24 controls</td>
<td>[112]</td>
</tr>
</tbody>
</table>

3. Physical State and Biological Function

The stability of cell-free circulating miRNA in the extracellular environment (e.g., plasma) had been established, but the mechanistic basis of this resistance to degradation remained unclear. Synthetic miRNA oligonucleotides and miRNA purified from plasma samples were rapidly degraded when combined with non-denatured plasma [62], indicating that miRNAs were not stabilized by intrinsic factors such as secondary structure or chemical modification. This observation led to the hypothesis in the field that cell-free circulating miRNAs were packaged in a way to protect them from RNase degradation, possibly in lipid vesicles or as part of a protein complex.

3.1. Exosomes and Other Vesicles

Among extracellular lipid vesicles, exosomes have received considerable attention [113–116]. Exosomes are 50–100 nm diameter membrane vesicles, secreted by diverse cell types in vivo and in vitro. These vesicles are present in many bodily fluids (e.g., blood, saliva, breast milk, seminal fluid, ascites, urine, etc.) [115,117], and have been shown to contain characteristic protein, mRNA and miRNA molecules derived from their cell of origin. In addition to exosomes, cells also produce larger lipid vesicles, microvesicles (1–10 μm in diameter). Like exosomes, microvesicles also contain protein, miRNA and mRNA. Recently, prostate cancer cells have been observed to produce microvesicles, which have been termed large oncosomes (LOs) [118]. LOs contain miRNAs associated with cell migration and invasion. Apoptotic bodies derived from normal cell turnover or from specific pathologies (e.g., cancer) may also contribute to RNA in circulation.

Circulating miRNAs may also be derived from platelets, which are abundant in typical plasma preparations, contain a large quantity of miRNA [119,120] and have been reported to contain tumor-derived mRNA when isolated from the plasma of cancer patients [121]. The authors of this study
suggest that platelets acquire tumor-derived mRNA through exosome-mediated transfer, although it is possible that platelet preparations may be contaminated with circulating tumor cell debris [122].

Several studies have observed the uptake of exosomes and other miRNA-containing vesicles into cells in vitro [116,118,123–125], leading to the hypothesis of miRNA-based intercellular communication mediated by exosomes (like a “message in a bottle” [126]) but it remains unclear if this is a general phenomenon.

3.2. Protein-miRNA Complexes

Multiple studies have also reported cell-free circulating miRNAs in non-lipid vesicle forms [127,128]. Using differential centrifugation, size-exclusion chromatography, filtration and immunoprecipitation approaches, investigators observed that a large fraction of miRNAs in platelet-poor plasma/serum from healthy human donors and conditioned cell medium were not associated with lipid vesicles and were sensitive to protease digestion. This observation suggested that miRNA was protected from plasma RNase activity by association with protein. A substantial amount of this soluble miRNA was associated with the endogenous effector protein Argonaute-2 (Ago2). Independent studies have also observed miRNA associated with other proteins such as high-density lipoprotein [129] and nucleophosmin [130].

3.3. Exogenous miRNAs in Circulation

Recent reports describing exogenous miRNA and other RNA species in the blood of healthy individuals [131,132] have raised some interesting questions about the genesis, function, and complexity of cell-free RNA in circulation. To accurately determine the origin of these RNAs derived from exogenous species is a challenge, due to the lack of genomic sequence information for many different species. Nevertheless, the preliminary results indicate these RNA molecules are probably derived from numerous bacteria, fungi and food, which suggest that gut is the primary source of these sequences. Functional delivery of small RNAs through ingestion has been very well established in nematodes [133]. However, tissue barrier systems in mammals are much more elaborate and it is unclear how RNA might penetrate them. Limited experimental data also showed that some of these exogenous RNAs are packaged in lipid vesicles and can affect the cellular transcriptome upon entry into the cells. How these RNAs permeated the gut lining, how they are packaged, how they are selected and how did cells recognize and take up these exogenous RNA species are just some of the interesting issues are waiting to be answered. The finding of exogenous miRNAs in normal human blood also raises some interesting concerns. Since some miRNA sequences are extremely conserved throughout evolution, the levels of those miRNAs in circulation may not be completely contributed from human cells. In addition, pathogenic or adverse effects induced by exogenous RNAs are a possibility.

4. Status of Circulating miRNA Biomarkers

Since the establishment of circulating miRNAs as biomarkers for cancer in 2008 [61,62], there have been over 500 reports describing their biology and diagnostic utility for various conditions. Relative to
other types of diagnostic biomolecules in circulation (e.g., proteins), miRNA has many advantages including tissue or cell type/stage specific expression, mature nucleic acid detection technologies, lower cost and shorter time required to develop assays (compared to the development of proteomic tools such as antibodies) and amplifiable signal. Despite these advantages, the rapid expansion of the literature, and significant commercial interest, circulating miRNA biomarkers face an unclear path toward clinical application. For example, an initial study of a circulating miRNA panel on 120 patients provided 91 percent sensitivity and 72 percent specificity to diagnose colon cancer. However, later studies yielded inconsistent results [134]. The inconsistencies were attributed to variation in specimen collection and handling, which are well known to alter miRNA spectrum [120]. Additional factors might also have been involved.

5. Considerations for Circulating MiRNA Studies

5.1. Prerequisites

Ideal prerequisites for biomarker discovery/development processes should include: (1) the need for such biomarker, (2) the intended use (e.g., type of tumor vs. tumor progression or response to treatment), (3) well characterized patient and matched control specimens in acceptable quantity, (4) samples have been banked or can be prospectively collected, processed and stored in a consistent manner and (5) a specific, reliable and easy to implement detection method. This process has been done for other clinical tests such as pathogen screening and protein-based biomarker measurement, but there is no commonly accepted standard for miRNA. In part this is due to the lack of understanding of both intrinsic and extrinsic factors that may affect the miRNA spectrum. Therefore, fundamental and comprehensive studies are urgently needed to address these issues. This need is well recognized in the community. For example, the National Institutes of Health recently established the Extracellular RNA Communication (ERC) program [135] to explore various properties and applications for extracellular miRNA in circulation.

5.2. Sample type, Collection and Processing

Donor-related factors such as the feeding state (e.g., fasting vs. non-fasting, food lipid content, etc.), time of blood draw, gender, and female hormone cycle may all affect the spectrum of circulating miRNA. However, the impact of these factors is largely unknown at this point. The type of blood product analyzed (serum vs. plasma) has also been shown to affect the spectrum of circulating miRNA [136]. Several known issues associated with the practice of blood collection and serum/plasma preparation are listed below:

- The use of small diameter needles (23 gauge or above) should be avoided due to shearing induced hemolysis of red blood cells (RBCs) [137], which contain abundant miRNAs and can alter profiling results [119,138,139].
- Lot-to-lot variation in the production of blood collection tubes may influence results [140]; if possible, purchase draw tubes from the same lot. Discard expired tubes, as these may have lost vacuum and cause variation on the final concentration of anticoagulant in the blood.
• Skin contains abundant epithelial miRNAs; thus a precursor blood draw or a blood tube that is drawn and discarded [141] using the same needle and tubing can prevent skin cells from contaminating a blood sample (especially for studies of epithelial cancers).

• Anticoagulant choice is also critical; this should be based on the requirements for analysis and uniform across the study (using only one type of anticoagulant tube with the part number and vendor specified in the protocol). Heparin is a well-established inhibitor of PCR and should be avoided [142]; EDTA tubes have been widely used instead. In the event heparin-containing samples are required, heparinase treatment prior to analysis has been shown to increase miRNA detection [143] but is likely to introduce additional variability.

• Anticoagulants and blood samples should be gently mixed in the tube, as shaking can cause hemolysis [144]. Serum coagulation conditions (time, temperature) and use of serum-separator polymers should be standardized between cases and controls.

• Blood should be expediently processed and the time allowed between draw and processing should be stated in the protocol. Specimens exceeding this time limitation should be flagged and noted in the data analysis.

• Insulated containers with uniform temperature should be used in packing specimens for transit. If environmental conditions include possible extreme heat or cold exposure, devices that can indicate if the specimen has exceeded a threshold temperature should be used [140].

• Standardized centrifugation conditions used to prepare cell-free blood fractions (time, temperature, g-force, rotor type, acceleration/deceleration conditions) are also important as residual platelets, cell debris, etc. can alter miRNA abundance [120]. Blood fractions require expedient separation from cell pellets to prevent contamination with cellular debris and contents [138, 145].

• As blood cells contain higher concentrations of miRNAs than plasma, care should be taken in aspirating plasma and serum to prevent cellular carryover [146].

• Centrifugation to remove debris or precipitates from body fluid samples prior to RNA extraction needs to be standardized, as it will alter miRNA profiles.

• To avoid unnecessary freeze/thaw cycles, if possible samples should be stored as single use aliquots, with volumes corresponding to those intended for analysis.

• Freezing and storage conditions should be standardized (snap or slow freezing). In addition, case/control specimens should be matched as closely as possible for storage time in the freezer.

5.3. RNA Isolation

The method used to purify RNA impacts yield and the spectrum of RNA isolated [142, 147]. Methods based on phenol-chloroform extraction and silica-membrane based columns have been observed to enhance detection of circulating miRNAs relative to methods that omit either of these steps. Omission of chloroform extraction may be especially problematic for blood samples containing high lipid content (e.g., blood drawn after high-fat meals, etc.). miRNA-specific methods are important, as buffers for miRNA extraction often utilize higher ethanol concentrations than their mRNA counterparts. Whether or not to enrich small RNA fraction is yet to be determined, since there are reports indicating the alteration of small RNA profile after enrichment. Addition of carrier RNA (e.g., MS2 bacteriophage RNA) [148] has been observed to improve miRNA quantification, but more
data is needed to demonstrate its reliability. In addition, carrier RNA may also affect downstream profiling methods (e.g., next-generation sequencing). The use of low-adhesion plasticware is critical, as standard plastics can adsorb miRNAs and skew quantification.

MiRNA derived from exosomes and other vesicles has been a significant focus of study, however; the purification method used to isolate vesicles will impact the type of vesicles that are collected and therefore their miRNA content. Isolation of lipid vesicles is generally performed via differential centrifugation [149–151]. The centrifugation force, type of rotor (e.g., fixed angle vs. swinging bucket) and viscosity of the sample are important variables. Exoquick (EXOQ5A-1, System Biosciences, Mountain View, CA), a polymer solvent, facilitates precipitation of exosomes [152,153] but may also precipitate non-exosomal vesicles and protein-miRNA complexes [154].

5.4. Data Correction, Normalization, Standards and Avoiding Contamination

Controls for variation in sample quality, processing and analysis can be introduced throughout the workflow. Hemolysis can be estimated using colorimetry (less sensitive) of the blood product specimens or by analysis of miRNAs abundant in RBCs (e.g., miR-451) [119,155]. Normalization to correct for variability during sample purification is typically achieved through the use of synthetic oligoribonucleotides (with low homology to those expressed in the species of interest) spiked-in to the samples after thorough denaturation [146]. Additional spike-in standards (e.g., UniSp6 and UniSp3, Exiqon) can be introduced prior to cDNA synthesis and PCR to estimate efficiency during these steps and normalize results for comparison.

Standard curves are typically prepared by serial dilution of synthetic oligoribonucleotide standards. To ensure reproducible quantification, these standards should be prepared as single use aliquots in low-adhesion plastics. The choice of diluent in which to prepare the standard curve is important, as the solution matrix alters the performance of the assay significantly [156] through effects on analyte recovery from liquid transfers, enzyme and primer efficiencies, etc. The most appropriate choice is a diluent that approximates the complexity of the samples (e.g., total RNA prepared from plasma) or the use of an abundant carrier RNA to dominate matrix effects (e.g., MS/2 bacteriophage RNA) [148]. It is possible that assays may cross-react with sequences present in the carrier RNA; thus, the tolerance of the assay for the carrier should be empirically determined.

As the concentration of standards and analytes span several orders of magnitude, care must be taken to avoid cross contamination, especially with concentrated stocks of synthetic standards, target-rich biologic samples and post-PCR products. Contamination events have proven a serious distraction for blood-based molecular diagnostics [157–159]. Analytic layouts should include multiple no-template controls distributed over the analysis plate as sentinel wells, to detect possible aerosol contamination of wells during plate manipulation. Dedicated pre-PCR workspaces, equipment and protocols should be used [144,160,161].

5.5. Technologies for miRNA Profiling

A variety of methods have been developed for miRNA profiling [162,163], but due to their short sequence length, end region sequence variation (e.g., isomiRs) and high sequence conservation among family members, accurate measurement for miRNA is nontrivial. In addition, the wide range of
individual miRNA concentrations in body fluids and number of clinical specimens needed to generate statistically meaningful results, require that methods selected need to have sufficient-sensitivity and dynamic range with reasonable throughputs. Commonly used miRNA measurement methods include: hybridization-based approaches (e.g., Agilent microarrays, Nanostring nCounter) [164] reverse transcription quantitative PCR arrays (RT-qPCR, e.g., TaqMan TLDA microfluidic cards, Exiqon microRNA Ready-to-Use PCR panels, etc.) and next generation sequencing (NGS, e.g., HiSeq 2000, SOLiD, Ion Torrent, MiSeq) [165]. Each technique presents relative strengths and weaknesses, and the choice of approach is dependent upon the research need.

Hybridization-based methods are well established, have considerable throughput. However, array-based approaches have limited sequence specificity, sensitivity (requiring ng-μg of total input RNA), relatively limited linear and dynamic ranges (four orders of magnitude) and are difficult to use for absolute quantification and cannot identify novel miRNAs

RT-qPCR arrays are also well established and are more sensitive than hybridization-based approaches (requiring < ng-μg of total input RNA), have wider dynamic range (six orders of magnitude), and can be followed by absolute quantification on the same instrument, but with lower throughput. However, like hybridization-based approaches, prior knowledge regarding the miRNA to be measured is required, the specific assay design is non-trivial and this approach cannot be used for novel miRNA discovery.

Next generation sequencing-based methods can identify novel miRNA sequences, distinguish isomiRs, and show substantial dynamic range (five or more orders of magnitude) although, this approach requires significant starting material (ng-μg) and computational support. The analysis of multiple samples during a single run is possibly through “barcoding” in which each RNA sample is tagged with a 6-nucleotide long identifying sequence [166]. Relative quantification between datasets can be performed using a digital gene expression profiling approach [167]. A major limitation of NGS is that it is subject to critical sequence-specific bias, amplifying the detection of some miRNAs and reducing others. Systematic investigation has revealed that this bias is primarily derived from cDNA library preparation and not the NGS platform used [168–172]. Using a 473-member synthetic miRNA test set, researchers observed variability in miRNA abundance of up to four orders of magnitude by digital gene expression, depending on the method used to construct the library [168]. Variability in library preparation has been attributed to sequence preferences by ligation enzymes that are used to attach adapter oligonucleotides [168,170,172], differences in RNA secondary structure [173] and amplification by PCR [174–176]. Remarkable sequence-specific differences in product yields have been observed using both T4Rnl1 and T4Rnl2tr-catalyzed ligation. In the case of T4Rnl2tr-based adapter ligations, sequence bias was substantially alleviated by the use of a pooled mixture of 5’- and 3’- adapter sequences that varied at the two terminal nucleotides proximal to the ligation [172]. RNA secondary structure may also impact analysis by sequencing, as RNAs with 3’-stem structures are underrepresented in sequencing reads [173]. Adding further complexity to the problem, RNA may co-fold with adapter sequences, sequestering ends and making them unavailable for sequencing. PCR induced bias in NGS library construction has been observed [174] and is a function of G/C content in the sequence. Additional denaturation [177] and optimized PCR buffers [174] have been reported to reduce this bias. The use of non-amplification dependent, third generation sequencing technologies
Molecules 2014, 19 6091

(tSMS, Helicos, Cambridge, MA) [178] may circumvent NGS limitations attributable to PCR, but these technologies are not widely available.

5.6. Quantification and Validation

Lead miRNA biomarkers candidates identified through initial profiling are often directly quantified individually using RT-qPCR, an analog technology based on a fluorescent signal output relative to external standard curves [179]. This approach introduces multiple significant sources of intra- and inter-assay variation (due to differences in preparation, unequal amplification efficiency between standard and target, Monte-Carlo effect, variations in thresholding, etc. [179–181]). Digital PCR [182] is an absolute method of nucleic-acid quantification, which partitions target molecules across many replicate reactions, followed by end-point PCR. In the presence of amplification-dependent fluorescent probes, each reaction generates a highly resolved, digital output signal corresponding to the presence or absence of starting template. The starting concentration of template is then determined by Poisson statistical analysis of the number of positive and total reactions. Although this concept was introduced early in the history of PCR [183] and presents theoretical advantages over RT-qPCR based methods, practical limitations on dynamic range and throughput subsequently limited the use of dPCR. However, recent advancements in instrumentation and chemistry [184,185] circumvent these obstacles by using rapid microfluidic analysis of nano- to picoliter-sized droplet partitions (BioRad QX-200 and RainDance RainDrop, respectively), enabling the practical analysis of sizable sample sets. Droplet digital PCR (ddPCR) is highly precise, because as an absolute method of quantification it is not reliant on reference standard curves. ddPCR has recently been shown [156] to reduce analytic variability between sample preparations 37%–86% and to improve day-to-day reproducibility in the analysis of miRNAs from prostate cancer patient serum by a factor of 7 compared to RT-qPCR. As an endpoint method, ddPCR is also resistant to residual PCR inhibitors that may be present in samples [142,186], providing further robustness.

A basic set of assay performance parameters should be included in each assay such as: limit of detection (LOD) [187], linear dynamic range [188], PCR-efficiency (where applicable) and precision. The determination of these parameters would aid in study reproducibility by eliminating spurious results based on data generated outside the operating range of the assay and provide quality assurance that the analysis is based on valid, reproducible results. The Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE Guidelines) provides a detailed framework for the analysis and reporting of miRNA quantification results (derived from RT-qPCR or other methods) [189]. In addition, a concise set of criteria to be reported in cell-free miRNA studies [190] has been proposed for use with MIQE:

- Sample collection: collection site, gauge and type of needle, elapsed time between collection and processing (including clotting time, if applicable), processing conditions, storage conditions
- Sample quality control: Age of sample, hemolysis measurement and cut-off criteria
- RNA isolation: sample volume, isolation kit/reagent, carrier used, spike-in used, elution/resuspension volume
6. Future Prospects and Recommendations

The discovery and analysis of miRNA biomarkers from biofluids entails substantial experimental detail and complexity. Although substantial knowledge has been acquired regarding circulating miRNAs and their applications for various pathologies, their clinical diagnostic potential has yet to be realized. In part this is due to the lack of fundamental understanding of various intrinsic and extrinsic factors that may affect the spectrum of circulating RNA, but assay variability and the lack of standard protocols for miRNA measurement also play a role. To further advance the field, we will need to gain knowledge on the state of miRNA in circulation, develop commonly accepted standard operating procedures, introduce controls to assess the consistency of the results at different stages and develop proper normalization strategies. In order to improve the accuracy and reproducibility in the field, and with the aim of realizing their clinical potential, we make the following general suggestions for future studies of miRNA biomarkers from biofluids:

- To the extent that is practical, match patient with control samples as closely as possible with regard to all relevant information (age, sex, smoking status, etc.) except the disease status in question (e.g., cancer or benign). Minimize possible confounding variables by explicitly specifying detailed inclusion and exclusion criteria for studies and collecting specimens from donors in the most consistent manner possible (time of day, feeding status, etc.).
- Estimate the practical variability associated with collection (time before processing, etc.), storage (freezing conditions, time in freezer, etc.), purification (e.g., precision of extraction methods) and analysis by performing basic pilot experiments with similar specimens and report the results as supplemental materials.
- miRNAs that are highly expressed in abundant blood cells (e.g., miR-451 in RBCs, [119,138]) are likely to be derived from these cells and not from the diseased tissue, in addition to being sensitive to hemolysis, etc. They are therefore unlikely to be suitable as robust biomarkers and should be filtered from the data or at a minimum interpreted with caution.
- Validate measurement results using different platforms (e.g., NGS followed by RT-qPCR).
- Report the essential variables outlined in the MIQE Guidelines [189] and those recommended for cell-free miRNA studies [190].

It is also important to note that the combination of miRNA and protein-based biomarkers may provide more robust and sensitive measurement as they are susceptible to different types of errors. Despite the difficulties of getting consistent miRNA measurement results with our current technologies, we believe implementing and following a better sample preparation guideline, and new measurement technologies such as isothermal amplification will facilitate the development of miRNA based biomarkers in clinical setting.
Acknowledgments

We would like to thank I. Ruf (Fred Hutchinson Cancer Research Center) for critical evaluation of this manuscript. The authors gratefully acknowledge support from DOD awards W911NF-10-2-0111 and HDTRA1-13-C-0055, grants from National Science Foundation of China 81201223 (Y.H.), Guangdong Province Medical Research Foundation A2012533 (Y.H.), Guangdong Medical College Research Startup Fund B2013001 (Y.H.) and Guangdong Medical College Integration of Industry, Education and Research C2013004 (Y.H.).

Conflicts of Interest

The authors declare no conflict of interest.

References


131. Semenov, D.V.; Baryakin, D.N.; Brenner, E.V.; Kurilshikov, A.M.; Vasiliev, G.V.; Bryzgalov, L.A.; Chikova, E.D.; Filippova, J.A.; Kuligina, E.V.; Richter, V.A. Unbiased approach to profile the


© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (http://creativecommons.org/licenses/by/3.0/).