

RESEARCH ARTICLE

Small RNA Sequencing for Profiling MicroRNAs in Long-Term Preserved Formalin-Fixed and Paraffin-Embedded Non-Small Cell Lung Cancer Tumor Specimens

Daniel H. Buitrago^{1☯‡}, Santosh K. Patnaik^{4☯‡}, Kyuichi Kadota^{1,2}, Eric Kannisto⁴, David R. Jones¹, Prasad S. Adusumilli^{1,3*}

1 Thoracic Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **2** Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **3** Center of Cell Engineering, Memorial Sloan Kettering Cancer Center, New York, New York, United States of America, **4** Department of Thoracic Surgery, Roswell Park Cancer Institute, Buffalo, New York, United States of America

☯ These authors contributed equally to this work.

‡ These authors are first authors on this work.

* adusumip@mskcc.org



OPEN ACCESS

Citation: Buitrago DH, Patnaik SK, Kadota K, Kannisto E, Jones DR, Adusumilli PS (2015) Small RNA Sequencing for Profiling MicroRNAs in Long-Term Preserved Formalin-Fixed and Paraffin-Embedded Non-Small Cell Lung Cancer Tumor Specimens. *PLoS ONE* 10(3): e0121521. doi:10.1371/journal.pone.0121521

Academic Editor: Soheil S. Dadras, University of Connecticut Health Center, UNITED STATES

Received: August 27, 2014

Accepted: February 3, 2015

Published: March 26, 2015

Copyright: © 2015 Buitrago et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The authors' laboratory work is supported by grants from the National Institutes of Health (R21 CA164568-01A1, R21 CA164585-01A1, R01 CA136705-06, U54 CA137788, P30 CA008748, and P50 CA086438-13), the US Department of Defense (PR101053 and LC110202), and Mr. William H. Goodwin and Mrs. Alice Goodwin, the Commonwealth Foundation for Cancer Research, and the Experimental Therapeutics Center of

Abstract

Background

The preservation of microRNAs in formalin-fixed and paraffin-embedded (FFPE) tissue makes them particularly useful for biomarker studies. The utility of small RNA sequencing for microRNA expression profiling of FFPE samples has yet to be determined.

Methods

Total RNA was extracted from de-paraffinized and proteinase K-treated FFPE specimens (15–20 years old) of 8 human lung adenocarcinoma tumors by affinity chromatography on silica columns. MicroRNAs in the RNA preparations were quantified by the Illumina HiSeq 2000 sequencing platform with sequencing libraries prepared with the TruSeq Small RNA Sample Preparation Kit (version 2.0) to obtain unpaired reads of 50 b for small RNA fragments. MicroRNAs were also quantified using Agilent Human miRNA (release 16.0) microarrays that can detect 1,205 mature microRNAs and by quantitative reverse transcription (RT)-PCR assays.

Results

Between 9.1–16.9 million reads were obtained by small RNA sequencing of extracted RNA samples. Of these, only 0.6–2.3% (mean = 1.5%) represented microRNAs. The sequencing method detected 454–625 microRNAs/sample (mean = 550) compared with 200–349 (mean = 286) microRNAs detected by microarray. In Spearman correlation analyses, the average correlation coefficient for the 126 microRNAs detected in all samples by both methods was 0.37, and >0.5 for 63 microRNAs. In correlation analyses of the sequencing- and

Memorial Sloan Kettering Cancer Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Two of the co-authors of this manuscript (P.A. and S.K.P.) are PLOS ONE Editorial Board members. This does not alter the authors' adherence to PLOS ONE editorial policies and criteria.

RT-PCR-based measurements, the coefficients were 0.19–0.95 (mean = 0.73) and >0.7, respectively, for 7 of 9 examined microRNAs. The average inter-replicate Spearman correlation coefficient for the sequencing method was 0.81.

Conclusions

Small RNA sequencing can be used to obtain microRNA profiles of FFPE tissue specimens with performance characteristics similar to those of microarrays, in spite of the fragmentation of ribosomal and messenger RNAs that reduces the method's informative capacity. The accuracy of the method can conceivably be improved by increasing sequencing depth and/or depleting FFPE tissue RNAs of ribosomal RNA fragments.

Introduction

MicroRNAs are single-stranded, non-coding RNAs of 18–25 nucleotides in length that function as epigenetic regulators by inhibiting protein translation from or inducing degradation of messenger RNA (mRNA) transcripts that they target [1, 2]. MicroRNAs are involved in crucial biological processes such as cell proliferation, differentiation, and apoptosis [3], and the dysregulation of microRNAs has been demonstrated in initiation and progression of a variety of human malignancies [4–8]. Characterization of microRNA expression has been useful in the classification, diagnosis, and prognosis of several malignancies [9–12].

Tissue specimens that have been preserved via fixation in formalin are a valuable resource for retrospective studies. Because of its simplicity and low cost, vast amounts of clinical specimens are routinely archived after formalin fixation. Unlike mRNAs, microRNAs preserve well in formalin-fixed tissue because of their ultra-short length [13, 14]; this allows the use of such tissues for microRNA measurements for clinical and research purposes. MicroRNA profiles of formalin-fixed and paraffin-embedded (FFPE) tissues are comparable to those obtained from matched, fresh-frozen tissue samples (e.g., [15, 16–19]), underscoring the suitability of FFPE specimens as appropriate resources for microRNA expression analyses. A large number of studies on the biological role as well as biomarker utility of microRNAs have consequently been performed using FFPE specimens (e.g., [20, 21]).

Hybridization microarrays are routinely used for global microRNA expression profiling of FFPE samples. Small RNA sequencing has emerged as a new technology for microRNA expression profiling. It offers the advantages of high sensitivity and specificity, detection of both novel and known microRNAs, identification of microRNA sequence polymorphisms and editing, and simpler data normalization strategies for comparing samples [22]. With the cost of adequately informative RNA sequencing assays approaching that of microarrays and with improvements to user-friendliness and robustness of methods for RNA sequencing data analysis, small RNA sequencing is likely to be used more commonly than microarrays in the near future. RNA in FFPE samples is chemically modified through crosslinking because of the formaldehyde and it is significantly fragmented [23] by oxidation from exposure to air, activity of endogenous ribonucleases during fixation, and the use of high temperatures during the paraffin-embedding process. Degradation of RNA also occurs during its isolation from FFPE samples, which involves prolonged exposure to high temperatures (55°C–70°C). Small fragments of ribosomal RNAs (rRNAs) and mRNAs that result from their degradation compete with endogenous small RNAs (e.g., microRNAs) during RNA sequencing. For instance, Sanger DNA sequencing of cDNAs cloned from FFPE tissue RNA shows that more than 80% of 18–25

nucleotide-sized small RNAs in FFPE tissue are fragments of rRNAs and transfer RNAs [24]. Despite these adverse effects of formalin fixation, it is theoretically possible to profile microRNAs in FFPE tissues if RNA sequencing is performed with an adequate depth. Therefore, we sought to examine the feasibility of small RNA sequencing for quantification of microRNAs in FFPE samples. When we initiated this work, there was only one published study that had attempted to address this question [25].

Materials and Methods

Ethics statement

This study was conducted with implied written and informed consent of participants and approved by the Institutional Review Board of Memorial Sloan Kettering Cancer Center (MSK) in New York, NY (IRB number WA0269–08).

Tumor specimens

FFPE specimens of 8 therapy-naïve solitary pathologic stage I lung adenocarcinoma tumors resected at MSK between January 1995 and December 1999 were used. The 8 tumors were from a set of stage I lung adenocarcinoma cases that have been previously described [26, 27]. Rolled sections of 4 μm and 20 μm thickness were obtained from the specimens for RNA extraction. The 4 μm sections were stained with hematoxylin and eosin for histologic evaluation under a BX51 microscope (Olympus, Tokyo, Japan) to ensure that RNA was extracted from tissue with >70% tumor content.

RNA extraction and quantification

Total RNA was isolated from 4 sections, or approximately 3 mm^3 of each tumor, using the affinity spin column-based High Pure FFPE RNA Isolation Kit (Roche, Indianapolis, USA). RNA was quantified via absorbance spectrophotometry on a Nanodrop 2000 instrument (Thermo Scientific, Waltham, USA) and via RiboGreen dye fluorescence on a Qubit fluorometer (Life Technologies, Carlsbad, USA). Integrity of RNA was assessed using electrophoresis on Eukaryote Total RNA Pico chip on Bioanalyzer 2100 (Agilent, Santa Clara, USA). RNA aliquots were stored at -80°C .

Microarray assay of microRNAs

Eight RNA samples were assayed in duplicate and in separate microarray experiments for the two sets of replicates. The Agilent SurePrint G3 Human miRNA 8x60k (release 16.0) microarray platform [28] was used. This platform can detect 1,205 human and 287 human viral mature microRNAs. For each of these microRNAs and control RNAs, the microarray has 1–4 different DNA probes of 40–60 b; each of these was synthesized on the microarray at 10–40 spots of 30 μm diameter. RNA (120 ng) was labeled with 3',5'-cytidine bisphosphate with cyanine dye attached to the 3' phosphate and they were hybridized in 45 μl to a microarray for 20 h at 45°C under rotation at 1/3 Hz, using reagents and methods provided with the Agilent miRNA Complete Labeling and Hybridization Kit. After post-hybridization washing, microarray slides were scanned on an Agilent G2505C microarray scanner and data from the images was extracted using Agilent Feature Extraction software (version 9.5.3). Raw microarray data was deposited in the NCBI Gene Expression Omnibus repository [29] (accession number GSE57835).

Small RNA sequencing

One μg RNA was used to generate a small RNA sequencing library using reagents and methods provided with TruSeq Small RNA Sample Prep Kit version 2 (Illumina, San Diego, USA). Briefly, T4 RNA ligase was used to ligate RA5 and RA3 RNA oligonucleotides to 5' and 3' ends of RNA, respectively. Adapter-ligated RNA was reverse-transcribed using a RTP primer and the resulting cDNA was amplified in an 11-cycle PCR that used RP1 and indexed RP1 primers. PCR products of 140–160 bp were isolated following electrophoresis through a 6% Novex Tris-borate polyacrylamide gel (Life Technologies). Quality of the generated small RNA sequencing library was confirmed using Agilent High Sensitivity DNA Analysis Kit on Bioanalyzer 2100 instrument. Quadruplexed sequencing of libraries to generate single-end reads of 50 b was performed on Illumina HiSeq 2000 instrument using HiSeq clustering and sequencing reagents (version 2.0). Illumina HCS (version 1.4.8), RTA (version 1.12.4.2), and CASAVA (version 1.8.2) software were used for base-calling and the generation of raw, de-multiplexed sequencing data in FASTQ format. Raw sequencing data was deposited in the NCBI Sequence Read Archive [30] (accession number SRP047429).

Reverse transcription (RT)-PCR assays of small RNAs

Briefly, 50 ng of RNA was reverse transcribed in 15 μl using a microRNA-specific oligonucleotide; reagents and methods were provided in TaqMan MicroRNA Reverse Transcription kit (Life Technologies). The cDNA (1.33 μl) was used as template in real time PCR of 40 cycles and 15 μl that was performed in triplicate on a 7900HT machine (Life Technologies) with denaturation for 15 s at 95°C and combined annealing and extension for 60 s at 60°C. SDS software (version 2.4, Life technologies) was used to calculate quantification cycle (C_q) values with automatic baseline and threshold detection. DNA oligonucleotides provided with TaqMan MicroRNA RT-PCR assays [31] with identification numbers 397, 509, 1097, 398, 526, and 2340 (Life Technologies), were used to quantify human mature microRNAs *miR-21-5p*, *-205-5p*, *-146b-5p*, *-22-3p*, *-223-3p*, and *-423-5p*, respectively. Mature microRNAs *miR-16-5p*, *-210-3p*, and *-486-5p* were measured using custom assays that were designed based on the principle of TaqMan microRNA RT-PCR assays and were subsequently validated (references [31, 32] and S1 Table). All RT-PCR assays were performed in a single experiment and the data-set of raw C_q values were normalized using the global mean method [33].

Processing and analyses of microarray data

Raw data from all microarrays were processed together using the AgiMicroRna Bioconductor package [34] (version 2.10.0) in R (version 3.0.2). Briefly, and as per optimum workflow identified for the microarray platform [35], the robust multi-chip average (RMA) method [36] was first used for probe-set summarization without background correction; the data across arrays was then normalized with the quantile method. A single outlier microarray was identified by examination of normalized data using relative log expression, unsupervised hierarchical clustering, inter-array correlation, and principal component plots; the raw microarray data was re-processed after excluding the outlier. Mature microRNA identification numbers (MIMAT IDs) of microRNAs detectable with the microarray platform were obtained from the miRBase microRNA sequence repository [37] (version 20). The value of IsGeneDetected flag, which was determined by the Agilent Feature Extraction software, was used to determine whether microRNAs were detected by the platform.

Processing and analyses of RNA sequencing data

Trimmomatic [38] (version 0.32) was used to trim reads of adapter and poor-quality bases with these criteria, in order: (1) remove read segments that matched sequences of adapters and primers used for sequencing library preparation; (2) remove leading/trailing bases with Phred₃₃ base-quality score <3; (3) using a sliding window of 4 bases, remove the 5' terminal base if the average Phred₃₃ score of the 4 bases was <15; and (4) completely discard trimmed reads with <16 remaining bases. For determining mature microRNA expression, miRExpress [39] (version 2.0) was used to map the processed reads against human pre-microRNA and mature microRNA sequences (miRBase release 20). The default settings for miRExpress was used during the process (100% identity between read and pre-microRNA sequence for alignment, read length \geq 80% of mature microRNA length, and \leq 4 extra leading/trailing bases in a read compared to mature microRNA sequence). To summarize count values of reads that mapped to a single mature microRNA but multiple pre-microRNAs, the rounded mean or median value of pre-microRNA-mapping counts was used when the maximum value was <50% or \geq 50% of minimum, respectively. The summarized raw mature microRNA count data was then normalized across samples using the trimmed mean of M-values (TMM) normalization method, as indicated in the edgeR Bioconductor package [40] (version 3.2.4). To determine the biotypes of RNA (e.g., mRNA, rRNA, microRNA, etc.) that the sequencing reads represented, subread-align function of the Subread aligner software [41] (version 1.4.4) was used to align processed reads—with multi-mapping permitted and an ungapped genome index—to the human genome (Ensembl GRCh37.75 primary assembly). The featureCounts function [42] of Subread was then used, with multi-overlapping permitted, to identify the RNAs' biotype that the sequencing reads represented via examination of annotations of genome locations that the reads were mapped to. The Ensembl's GRCh37.75 gene annotation data with >25 different gene_biotype attribute values was used.

Other

For analyses that involved comparison of replicates, normalized microarray, and normalized and log₂-transformed (with offset = 0.1) sequencing data were additionally processed with the parametric ComBat function of the SVA Bioconductor package [43] (version 3.6.0) to adjust for the batch effect of separate microarray or sequencing experiments used for replicates. For correlation analyses, normalized microarray- and sequencing-based microRNA measurements were log₂-transformed and RT-PCR-based normalized C_q measurements for microRNAs were negative-transformed. Prism software (version 6.0c; GraphPad, La Jolla, USA) was used for plotting data. A *P*-value <0.05 was associated with statistical significance. RT-PCR assays and data analyses were performed at Roswell Park Cancer Institute, Buffalo, USA; all other work was completed at MSK.

Results and Discussion

We extracted total RNA from FFPE specimens of resected stage I lung adenocarcinoma tumors. RNA was isolated using silica matrix-containing spin columns from approximately 3 mm³ of FFPE tissue (n = 8) after overnight digestion with proteinase K. RNA yields ranged from 8.0–59.7 μ g, as quantified by absorbance at 260 nm, but 6.1–34.6 μ g, as quantified with RiboGreen dye. This has more specificity for RNA than it does for DNA (Fig. 1A) and it suggests the presence of DNA in the preparations, as has been noted by others [32]. As expected, the gel electrophoresis assay on Bioanalyzer instrument showed that the RNA isolated from the FFPE specimens was highly degraded (Fig. 1B), with values of RNA integrity ranging from 1.9–2.5 [44].

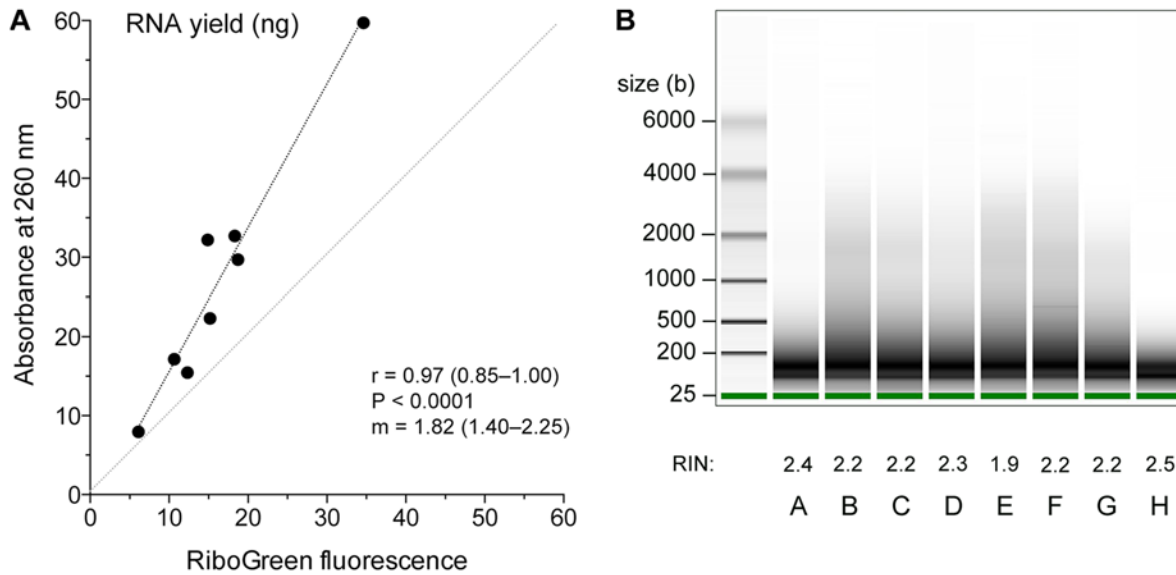


Fig 1. Assessment of total RNA extracted from FFPE tissues. A. Correlation between measurements of RNA yield (μg) obtained using absorbance at 260 nm or fluorescence with RiboGreen dye ($n = 8$). The lines of identity and linear regression (least squares method) and values and 95% confidence intervals of the Pearson coefficient (r) and slope of linear regression (m); P -values are also noted. B. Electrophoretogram of the 8 RNA samples (A–H). Samples were run on an Agilent Eukaryote Total RNA Pico chip on Bioanalyzer 2100. Sizes of molecular weight markers and the RNA integrity numbers are noted.

doi:10.1371/journal.pone.0121521.g001

Sequencing libraries were generated from 1 μg each of the total RNA samples by ligation of adapter RNAs at 5' and 3' ends, followed by reverse transcription and PCR using the Illumina TruSeq Small RNA Sample Preparation Kit. The libraries were size-selected for sequencing of RNA fragments of 15–25 nucleotides. Sequencing was performed on the Illumina HiSeq 2000 platform to obtain single-end reads of 50 bases. Between 9.1 and 16.9 million reads were obtained for each of the 8 samples (mean = 13.2 million; SD = 3.5 million; [Table 1](#)). After the

Table 1. Characteristics of small RNA sequencing data of the eight samples.

	Median	Mean	Range	SD
Total reads (10^3)	14,039	13,220	9,095–16,936	3,474
Reads used for mapping (10^3):	13,046	12,112	7,601–16139	3,674
% with length 16–25 b	25.23	27.15	19.96–37.36	5.83
% mapped by miRExpress to mature microRNAs	1.65	1.54	0.60–2.27	0.50
% mapped by Subread to genome	77.39	76.81	67.55–83.37	4.79
% mapped by Subread to transcribed genome:	68.08	68.63	65.16–73.51	3.37
<i>RNA biotype of mapped region (%)^a</i>				
MicroRNA	0.29	0.29	0.14–0.53	0.13
Small nucleolar RNA	0.83	0.77	0.31–0.97	0.23
Small nuclear RNA	0.42	0.43	0.17–0.76	0.17
Long intergenic non-coding RNA	5.49	5.64	5.42–6.07	0.26
Protein-coding RNA	36.93	37.09	34.82–39.26	1.57
Ribosomal RNA	52.56	52.46	49.36–54.63	1.48
Other	3.39	3.31	1.69–4.10	0.80

^aPercentage among reads mapped by Subread to transcribed region of the genome

doi:10.1371/journal.pone.0121521.t001

reads were processed to remove nucleotide sequences of adapters used for library preparation and bases of poor quality, 67.6–83.4% (mean = 76.8; SD = 4.5) of the reads could be mapped against the hg19 reference human genome assembly (Ensembl GRCh37.75 primary assembly). More than 90% of the mapped reads aligned against transcribed regions of the genome. However, only 0.14%–0.53% (mean = 0.29; SD = 0.13) of reads mapped to loci for microRNAs, whereas 49.4%–54.6% and 34.8%–39.3% mapped to regions that encoded for rRNAs and mRNAs, respectively (Table 1). Thus, only a very small fraction (approximately 0.1%–0.5%) of small RNAs of the 15–25 nucleotides that were isolated from the FFPE tissue specimens were microRNAs; the overwhelming majority of those were rRNA and mRNA fragments.

MicroRNA expression profiles were extracted from the small RNA sequencing data using the miRExpress tool[39]. MiRExpress mapped 0.6%–2.2% of the processed sequencing reads to mature human microRNAs. Per miRBase microRNA sequence database release 20 (June 2013), 2,620 mature human microRNAs were known [37]. A total of 1,021 microRNAs were detected (mapped read count ≥ 1) among the 8 samples, with 454–625 detected per sample (mean = 550; SD = 69), and 283 detected in all samples. To validate the small RNA sequencing-based microRNA measurements, quantitative RT-PCR assays were used to quantify 9 arbitrarily chosen microRNAs that were considered as detected by both the microarray and small RNA sequencing platforms in all of the 8 RNA samples. In Spearman correlation analyses of the microRNA measurements obtained by RT-PCR and small RNA sequencing, the correlation coefficients were 0.19–0.95 (mean = 0.73; SD = 0.24). The coefficient value was >0.7 for 7 of the 9 microRNAs (Table 2), thereby indicating a modest accuracy of the RNA sequencing method. To estimate technical replicability of microRNA expression profiling via the sequencing method, 2 of the 8 RNA samples were subjected to small RNA sequencing in a duplicate experiment. The inter-replicate Spearman correlation coefficients for the duplicate microRNA measurements of the 2 samples were 0.75 and 0.88 (Fig. 2).

Quantification of microRNAs via hybridization of RNA to DNA probes on microarrays is currently the standard method for global microRNA profiling of FFPE tissues. To more thoroughly assess the performance of the small RNA sequencing method, we measured microRNAs in 120 ng of each of the 8 FFPE tissue RNA samples using the Agilent 8x60k Human miRNA microarray platform, which is capable of detecting 1,205 mature human microRNAs. A total of 368 microRNAs were detected (IsGeneDetected flag value = 1) among the 8 samples, with 200–349 detected per sample (mean = 286; SD = 45) and 175 detected in all samples. Of these 175

Table 2. Spearman correlation coefficients of normalized microRNA measurements obtained by reverse-transcription (RT)-PCR and the small RNA sequencing or microarray platforms^a.

<i>MicroRNA</i>	<i>Small RNA sequencing</i>	<i>Microarray</i>
<i>miR-146b-5p</i>	0.83	-0.30
<i>miR-16-5p</i>	0.71	0.84
<i>miR-205-5p</i>	0.98	0.82
<i>miR-21-5p</i>	0.71	0.40
<i>miR-210-3p</i>	0.81	0.39
<i>miR-22-3p</i>	0.57	-0.36
<i>miR-223-3p</i>	0.19	0.43
<i>miR-423-5p</i>	0.79	-0.71
<i>miR-486-5p</i>	0.95	0.47

^aGlobal mean-normalized C_q values obtained by RT-PCR assays were negative-transformed for correlation analysis

doi:10.1371/journal.pone.0121521.t002

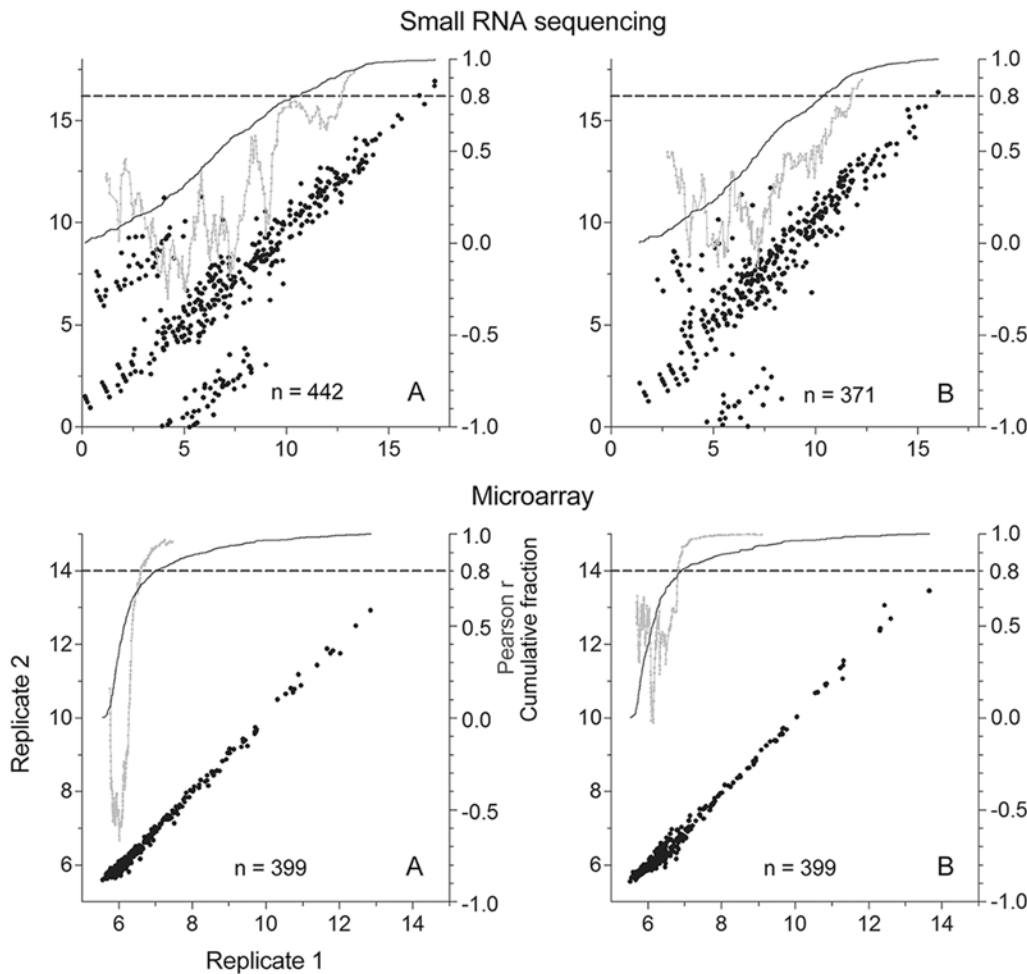


Fig 2. Inter-replicate correlations. MicroRNAs in 2 samples (A and B) were profiled in duplicate by both the small RNA sequencing and microarray platforms. Scatter-plots depict the inter-replicate correlation of normalized microRNA measurements. The cumulative fraction of microRNAs along the X axis (black), and the values of Pearson correlation coefficient of the microRNA measurements in a sliding window along the X axis of size 51 at mid-window (gray) are also shown.

doi:10.1371/journal.pone.0121521.g002

Table 3. Comparison of the small RNA sequencing and microarray platforms.

	Small RNA sequencing	Microarray
Detection of microRNAs		
Platform capability	2,620 ^a	1,205
Detected in any sample	1,021 (39.0%)	368 (30.5%)
Detected in all 8 samples	283 (10.8%)	175 (8.5%)
Detected per sample (mean, range, SD)	550, 454–625, 69	286, 200–349, 45
Inter-replicate Spearman correlation ^b (mean, range, SD)	0.81, 0.75–0.88, 0.09 (n = 2)	0.99, 0.98–0.99, 0.01 (n = 7)
Spearman correlation with RT-PCR measurements ^c (mean, range, SD; n = 9)	0.73, 0.19–0.98, 0.24	0.22, -0.71–0.84, 0.55

^aKnown human mature microRNAs in miRBase release 20

^bCoefficient value of Spearman correlation analyses

^cGlobal mean-normalized C_q values were negative-transformed for correlation analysis

doi:10.1371/journal.pone.0121521.t003

microRNAs, 126 were also detected in all 8 samples by the small RNA sequencing method. In comparisons with the RNA sequencing- and microarray-based microRNA measurements, the average Spearman correlation coefficient for the 126 microRNAs was 0.37 and for the 63 microRNAs was >0.5 . Correlation between FFPE tissue microRNA measurements obtained by small RNA sequencing and microarray platforms has been confirmed by other groups [45–47].

Microarray-based microRNA assays were replicated for 7 of the 8 RNA samples. Measured against an average inter-replicate Spearman correlation coefficient of 0.81, seen with the small RNA sequencing method, the inter-replicate correlation coefficients for the microarray platform were in the 0.98–0.99 range (mean = 0.99; SD = 0.01). However, in the Spearman correlation analyses of the microRNA measurements obtained by RT-PCR and microarray assays, the correlation coefficients were -0.36–0.84 (mean = 0.22; SD = 0.55) and were smaller than those observed in the RNA sequencing-based measurements (Table 2). Thus, whereas the small RNA sequencing platform detected a greater quantity of microRNAs with greater accuracy than the microarray platform, its precision was not as good. Table 3 summarizes the performance characteristics that were observed in the 2 microRNA profiling platforms used in this report.

Overall, our findings demonstrate the feasibility of small RNA sequencing for microRNA profiling of FFPE tissue RNA in spite of its compromised quality and integrity. Even though only 1.54% of 12.1 million small RNA sequencing reads of FFPE tissues were identified as representing 550 known microRNAs in this study, the number of microRNAs that were detected and quantified was more than that by the microarray platform (Tables 1 and 3). The Spearman correlation coefficients for measurements obtained by the sequencing and microarray platforms for half of the 126 microRNAs that were detected in all samples by both platforms was >0.5 . In correlation analyses of the sequencing- and RT-PCR-based measurements of nine microRNAs that were examined, the coefficients were >0.7 for 7 (Table 2). The technical replicability of the small RNA sequencing method was, however, relatively poor compared to the microarray method (average inter-replicate Spearman correlation coefficients of 0.81 and 0.99, respectively; Table 3). It is conceivable that the precision of the sequencing method could be improved by increasing the sequencing depth to obtain a greater number of reads and by utilizing methods to deplete rRNA from the RNA preparations [48–50] or to reduce its representation in the sequencing library using strategies such as 'poison primer' blocking[24].

Only one study [25] had reported that the feasibility of using small RNA sequencing for microRNA profiling of FFPE tissue RNA when we initiated the work described in this paper. Since then within the past year, at least 5 other studies have shown that RNA sequencing can be used to accurately profile microRNAs in FFPE samples (S2 Table) [45–47, 51, 52]. These studies mostly performed from cell lines and human FFPE tissues of 2–9 years old have shown that microRNA expression profiles obtained by small RNA sequencing are highly concordant between FFPE and fresh-frozen tissues [25, 47, 51, 52], and that small RNA sequencing and microarray platforms generate similar microRNA profiles of FFPE tissues [25, 47]. The findings of these and our studies performed on NSCLC specimens of 15 to 20 years old therefore demonstrate that small RNA sequencing can be used to obtain microRNA profiles of FFPE tissue specimens with performance characteristics similar to those of microarrays.

Supporting Information

S1 Table. Custom TaqMan microRNA reverse transcription (RT)-PCR assays for human mature microRNAs *miR-210-3p* and *miR-486-5p*.

(DOCX)

S2 Table. Six studies that have assessed RNA sequencing for profiling microRNAs in formalin-fixed, paraffin-embedded tissues.

(DOCX)

Acknowledgments

We thank Alex Torres of the Thoracic Surgery Service at MSK for his editorial assistance.

Author Contributions

Conceived and designed the experiments: PSA. Performed the experiments: DHB KK EK. Analyzed the data: SKP DHB DRJ PSA. Contributed reagents/materials/analysis tools: KK DRJ PSA. Wrote the paper: KK DRJ PSA.

References

1. Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell*. 2004; 116: 281–297. PMID: [14744438](#)
2. Ambros V. The functions of animal microRNAs. *Nature*. 2004; 431: 350–355. PMID: [15372042](#)
3. Calin GA, Croce CM. MicroRNA signatures in human cancers. *Nat Rev Cancer*. 2006; 6: 857–866. PMID: [17060945](#)
4. Alvarez-Garcia I, Miska EA. MicroRNA functions in animal development and human disease. *Development*. 2005; 132: 4653–4662. PMID: [16224045](#)
5. Sassen S, Miska EA, Caldas C. MicroRNA: implications for cancer. *Virchows Arch*. 2008; 452: 1–10. PMID: [18040713](#)
6. Quesne JL, Jones J, Warren J, Dawson SJ, Ali HR, Bardwell H, et al. Biological and prognostic associations of miR-205 and let-7b in breast cancer revealed by in situ hybridization analysis of micro-RNA expression in arrays of archival tumour tissue. *J Pathol*. 2012; 227: 306–314. doi: [10.1002/path.3983](#) PMID: [22294324](#)
7. Slaby O, Svoboda M, Fabian P, Smerdova T, Knoflickova D, Bednarikova M, et al. Altered expression of miR-21, miR-31, miR-143 and miR-145 is related to clinicopathologic features of colorectal cancer. *Oncology*. 2007; 72: 397–402. doi: [10.1159/000113489](#) PMID: [18196926](#)
8. Costinean S, Zaneni N, Pekarsky Y, Tili E, Volinia S, Heerema N, et al. Pre-B cell proliferation and lymphoblastic leukemia/high-grade lymphoma in E(mu)-miR155 transgenic mice. *Proc Natl Acad Sci U S A*. 2006; 103: 7024–7029. PMID: [16641092](#)
9. Patnaik SK, Kannisto E, Knudsen S, Yendamuri S. Evaluation of microRNA expression profiles that may predict recurrence of localized stage I non-small cell lung cancer after surgical resection. *Cancer Res*. 2010; 70: 36–45. doi: [10.1158/0008-5472.CAN-09-3153](#) PMID: [20028859](#)
10. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med*. 2005; 353: 1793–1801. PMID: [16251535](#)
11. Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D, et al. MicroRNA expression profiles classify human cancers. *Nature*. 2005; 435: 834–838. PMID: [15944708](#)
12. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, et al. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A*. 2011; 108: 3713–3718. doi: [10.1073/pnas.1100048108](#) PMID: [21300873](#)
13. Peiro-Chova L, Pena-Chilet M, Lopez-Guerrero JA, Garcia-Gimenez JL, Alonso-Yuste E, Burgues O, et al. High stability of microRNAs in tissue samples of compromised quality. *Virchows Arch*. 2013; 463: 765–774. doi: [10.1007/s00428-013-1485-2](#) PMID: [24197449](#)
14. Li J, Smyth P, Flavin R, Cahill S, Denning K, Aherne S, et al. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnol*. 2007; 7: 36. PMID: [17603869](#)
15. Liu A, Tetzlaff MT, Vanbelle P, Elder D, Feldman M, Tobias JW, et al. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. *Int J Clin Exp Pathol*. 2009; 2: 519–527. PMID: [19636399](#)
16. Liu A, Xu X. MicroRNA isolation from formalin-fixed, paraffin-embedded tissues. *Methods Mol Biol*. 2011; 724: 259–267. doi: [10.1007/978-1-61779-055-3_16](#) PMID: [21370018](#)

17. Xi Y, Nakajima G, Gavin E, Morris CG, Kudo K, Hayashi K, et al. Systematic analysis of microRNA expression of RNA extracted from fresh frozen and formalin-fixed paraffin-embedded samples. *RNA*. 2007; 13: 1668–1674. PMID: [17698639](#)
18. Zhang X, Chen J, Radcliffe T, Lebrun DP, Tron VA, Feilotter H. An array-based analysis of microRNA expression comparing matched frozen and formalin-fixed paraffin-embedded human tissue samples. *J Mol Diagn*. 2008; 10: 513–519. doi: [10.2353/jmoldx.2008.080077](#) PMID: [18832457](#)
19. Siebolts U, Varnholt H, Drebbler U, Dienes HP, Wickenhauser C, Odenthal M. Tissues from routine pathology archives are suitable for microRNA analyses by quantitative PCR. *J Clin Pathol*. 2009; 62: 84–88. doi: [10.1136/jcp.2008.058339](#) PMID: [18755714](#)
20. Streichert T, Otto B, Lehmann U. MicroRNA expression profiling in archival tissue specimens: methods and data processing. *Mol Biotechnol*. 2012; 50: 159–169. doi: [10.1007/s12033-011-9427-1](#) PMID: [21691932](#)
21. Nuovo GJ. In situ detection of microRNAs in paraffin embedded, formalin fixed tissues and the co-localization of their putative targets. *Methods*. 2010; 52: 307–315. doi: [10.1016/j.ymeth.2010.08.009](#) PMID: [20723602](#)
22. Pritchard CC, Cheng HH, Tewari M. MicroRNA profiling: approaches and considerations. *Nat Rev Genet*. 2012; 13: 358–369. doi: [10.1038/nrg3198](#) PMID: [22510765](#)
23. von Ahlhen S, Missel A, Bendrat K, Schlumpberger M. Determinants of RNA quality from FFPE samples. *PLoS One*. 2007; 2: e1261. PMID: [18060057](#)
24. Ma Z, Lui WO, Fire A, Dadras SS. Profiling and discovery of novel miRNAs from formalin-fixed, paraffin-embedded melanoma and nodal specimens. *J Mol Diagn*. 2009; 11: 420–429. doi: [10.2353/jmoldx.2009.090041](#) PMID: [19710399](#)
25. Weng L, Wu X, Gao H, Mu B, Li X, Wang JH, et al. MicroRNA profiling of clear cell renal cell carcinoma by whole-genome small RNA deep sequencing of paired frozen and formalin-fixed, paraffin-embedded tissue specimens. *J Pathol*. 2010; 222: 41–51. doi: [10.1002/path.2736](#) PMID: [20593407](#)
26. Yoshizawa A, Motoi N, Riely GJ, Sima CS, Gerald WL, Kris MG, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. *Mod Pathol*. 2011; 24: 653–664. doi: [10.1038/modpathol.2010.232](#) PMID: [21252858](#)
27. Sica G, Yoshizawa A, Sima CS, Azzoli CG, Downey RJ, Rusch VW, et al. A grading system of lung adenocarcinomas based on histologic pattern is predictive of disease recurrence in stage I tumors. *Am J Surg Pathol*. 2010; 34: 1155–1162. doi: [10.1097/PAS.0b013e3181e4ee32](#) PMID: [20551825](#)
28. Wang H, Ach RA, Curry B. Direct and sensitive miRNA profiling from low-input total RNA. *RNA*. 2007; 13: 151–159. PMID: [17105992](#)
29. Barrett T, Suzek TO, Troup DB, Wilhite SE, Ngau WC, Ledoux P, et al. NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res*. 2005; 33: D562–566. PMID: [15608262](#)
30. Leinonen R, Sugawara H, Shumway M. The sequence read archive. *Nucleic Acids Res*. 2011; 39: D19–21. doi: [10.1093/nar/gkq1019](#) PMID: [21062823](#)
31. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, et al. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. 2005; 33: e179. PMID: [16314309](#)
32. Patnaik SK, Kannisto E, Yendamuri S. Factors affecting the yield of microRNAs from laser microdissectates of formalin-fixed tissue sections. *BMC Res Notes*. 2012; 5: 40. doi: [10.1186/1756-0500-5-40](#) PMID: [22260539](#)
33. Mestdagh P, Van Vlierberghe P, De Weer A, Muth D, Westermann F, Speleman F, et al. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol*. 2009; 10: R64. doi: [10.1186/gb-2009-10-6-r64](#) PMID: [19531210](#)
34. Lopez-Romero P. Pre-processing and differential expression analysis of Agilent microRNA arrays using the AgiMicroRna Bioconductor library. *BMC Genomics*. 2011; 12: 64. doi: [10.1186/1471-2164-12-64](#) PMID: [21269452](#)
35. Lopez-Romero P, Gonzalez MA, Callejas S, Dopazo A, Irizarry RA. Processing of Agilent microRNA array data. *BMC Res Notes*. 2010; 3: 18. doi: [10.1186/1756-0500-3-18](#) PMID: [20205787](#)
36. Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003; 4: 249–264. PMID: [12925520](#)
37. Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. *Nucleic Acids Res*. 2008; 36: D154–158. PMID: [17991681](#)
38. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014.

39. Wang WC, Lin FM, Chang WC, Lin KY, Huang HD, Lin NS. miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression. *BMC Bioinformatics*. 2009; 10: 328. doi: [10.1186/1471-2105-10-328](https://doi.org/10.1186/1471-2105-10-328) PMID: [19821977](https://pubmed.ncbi.nlm.nih.gov/19821977/)
40. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. 2010; 26: 139–140. doi: [10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616) PMID: [19910308](https://pubmed.ncbi.nlm.nih.gov/19910308/)
41. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*. 2013; 41: e108. doi: [10.1093/nar/gkt214](https://doi.org/10.1093/nar/gkt214) PMID: [23558742](https://pubmed.ncbi.nlm.nih.gov/23558742/)
42. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30: 923–930. doi: [10.1093/bioinformatics/btt656](https://doi.org/10.1093/bioinformatics/btt656) PMID: [24227677](https://pubmed.ncbi.nlm.nih.gov/24227677/)
43. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012; 28: 882–883. doi: [10.1093/bioinformatics/bts034](https://doi.org/10.1093/bioinformatics/bts034) PMID: [22257669](https://pubmed.ncbi.nlm.nih.gov/22257669/)
44. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol*. 2006; 7: 3. PMID: [16448564](https://pubmed.ncbi.nlm.nih.gov/16448564/)
45. Plieskatt JL, Rinaldi G, Feng Y, Levine PH, Easley S, Martinez E, et al. Methods and matrices: approaches to identifying miRNAs for Nasopharyngeal carcinoma. *J Transl Med*. 2014; 12: 3. doi: [10.1186/1479-5876-12-3](https://doi.org/10.1186/1479-5876-12-3) PMID: [24393330](https://pubmed.ncbi.nlm.nih.gov/24393330/)
46. Kelly AD, Hill KE, Correll M, Hu L, Wang YE, Rubio R, et al. Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: preliminary assessment of cross-platform concordance. *Genomics*. 2013; 102: 8–14. doi: [10.1016/j.ygeno.2013.03.008](https://doi.org/10.1016/j.ygeno.2013.03.008) PMID: [23562991](https://pubmed.ncbi.nlm.nih.gov/23562991/)
47. Kolbert CP, Feddersen RM, Rakhshan F, Grill DE, Simon G, Middha S, et al. Multi-platform analysis of microRNA expression measurements in RNA from fresh frozen and FFPE tissues. *PLoS One*. 2013; 8: e52517. doi: [10.1371/journal.pone.0052517](https://doi.org/10.1371/journal.pone.0052517) PMID: [23382819](https://pubmed.ncbi.nlm.nih.gov/23382819/)
48. Li SK, Zhou JW, Yim AK, Leung AK, Tsui SK, Chan TF, et al. Organism-specific rRNA capture system for application in next-generation sequencing. *PLoS One*. 2013; 8: e74286. doi: [10.1371/journal.pone.0074286](https://doi.org/10.1371/journal.pone.0074286) PMID: [24073205](https://pubmed.ncbi.nlm.nih.gov/24073205/)
49. Chen Z, Duan X. Ribosomal RNA depletion for massively parallel bacterial RNA-sequencing applications. *Methods Mol Biol*. 2011; 733: 93–103. doi: [10.1007/978-1-61779-089-8_7](https://doi.org/10.1007/978-1-61779-089-8_7) PMID: [21431765](https://pubmed.ncbi.nlm.nih.gov/21431765/)
50. O'Neil D, Glowatz H, Schlumpberger M. Ribosomal RNA depletion for efficient use of RNA-seq capacity. *Curr Protoc Mol Biol*. 2013; Chapter 4: Unit 4 19.
51. Meng W, McElroy JP, Volinia S, Palatini J, Warner S, Ayers LW, et al. Comparison of microRNA deep sequencing of matched formalin-fixed paraffin-embedded and fresh frozen cancer tissues. *PLoS One*. 2013; 8: e64393. doi: [10.1371/journal.pone.0064393](https://doi.org/10.1371/journal.pone.0064393) PMID: [23696889](https://pubmed.ncbi.nlm.nih.gov/23696889/)
52. Tam S, de Borja R, Tsao MS, McPherson JD. Robust global microRNA expression profiling using next-generation sequencing technologies. *Lab Invest*. 2014; 94: 350–358. doi: [10.1038/labinvest.2013.157](https://doi.org/10.1038/labinvest.2013.157) PMID: [24445778](https://pubmed.ncbi.nlm.nih.gov/24445778/)