Genomics 102 (2013) 8–14

Contents lists available at SciVerse ScienceDirect

Genomics

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

Next-generation sequencing and microarray-based interrogation of microRNAs from formalin-fixed, paraffin-embedded tissue: Preliminary assessment of cross-platform concordance

Andrew D. Kelly [a], Katherine E. Hill [a], Mick Correll [b], Lan Hu [b], Yaoyu E. Wang [b], Renee Rubio [b], Shenghua Duan [b], John Quackenbush [b], Dimitrios Spentzos [a,*]

[a] Division of Hematology/Oncology, Sarcoma Program, Department of Medicine, Beth Israel Deaconess Medical Center, Harvard Medical School, 330 Brookline Avenue, Boston, MA 02215, USA
[b] Center for Cancer Computational Biology, Department of Biostatistics and Computational Biology, Dana Farber Cancer Institute, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Received 26 July 2012
Accepted 22 March 2013
Available online 3 April 2013

Keywords:
RNA-sequencing
Microarrays
Paraffin-tissue

ABSTRACT

Next-generation sequencing is increasingly employed in biomedical investigations. Strong concordance between microarray and mRNA-seq levels has been reported in high quality specimens but information is lacking on formalin-fixed, paraffin-embedded (FFPE) tissues, and particularly for microRNA (miRNA) analysis. We conducted a preliminary examination of the concordance between miRNA-seq and cDNA-mediated annealing, selection, extension, and ligation (DASL) miRNA assays. Quantitative agreement between platforms is moderate (Spearman correlation 0.514–0.596) and there is discordance of detection calls on a subset of miRNAs. Quantitative PCR (q-RT-PCR) performed for several discordant miRNAs confirmed the presence of most sequences detected by miRNA-seq but not by DASL but also that miRNA-seq did not detect some sequences, which DASL confidently detected. Our results suggest that miRNA-seq is specific, with few false positive calls, but it may not detect certain abundant miRNAs in FFPE tissue. Further work is necessary to fully address these issues that are pertinent for translational research.

1. Introduction

Next-generation sequencing (NGS) technologies have become faster, more accurate, and less expensive in recent years, leading to their widespread application in diverse fields [1,2]. Before the development of next-generation RNA-sequencing (RNA-seq) platforms, probe-based microarrays were widely utilized high-throughput transcriptomic profiling technologies and have yielded numerous significant findings in clinical and basic research [3,4]. Despite the obvious contributions that microarrays have made – and continue to make – to understanding the human genome, NGS methods are becoming increasingly prevalent technologies, which have certain advantages (and disadvantages) relative to microarrays [5–7]. While microarray technologies have been developed that require lower amounts of input RNA and can reliably detect low abundance transcripts, RNA-seq has the capacity to uncover novel transcript variants and is not limited by the potential for cross-hybridization [7].

A major challenge in clinical and translational research has been obtaining reliable genomic data from the degraded RNA present in formalin-fixed, paraffin-embedded (FFPE) tissue specimens. Because these samples are collected and archived in the process of routine medical care, robust methods of interrogating their genomes would greatly accelerate the rate of health-improving discovery. This unmet need was partially tackled – at least for transcriptome profiling – with the development of the cDNA-mediated annealing, selection, extension, and ligation (DASL) assay. This platform has been shown to reproducibly quantitate the expression levels of known genes and microRNAs (miRNAs) in FFPE tissue [8,9]. Furthermore, microarray-based expression profiling methods exhibit a substantial degree of concordance with RNA-sequencing when using high quality RNA [10]. However, to our knowledge, no study has reported findings on the concordance between the DASL assay and NGS when profiling the highly degraded RNA extracted from FFPE tissue. Because of the increasing utilization of NGS methods and the importance of research utilizing archival specimens, the question of cross-platform accuracy and agreement is critically important.

In this preliminary study we examine the consistency across a miRNA-seq platform and the miRNA DASL assay with respect to detecting the presence (or absence) of known miRNAs in five FFPE tissue samples. We demonstrate that both technologies are internally reproducible, and that quantitative expression levels are moderately correlated for

* Corresponding author.
E-mail addresses: andrew.kelly2010@gmail.com (A.D. Kelly), khill5@bidmc.harvard.edu (K.E. Hill), mcorrell@jimmy.harvard.edu (M. Correll), lanhu@jimmy.harvard.edu (L. Hu), yewang@jimmy.harvard.edu (Y.E. Wang), rrubio@jimmy.harvard.edu (R. Rubio), duan@jimmy.harvard.edu (S. Duan), johnq@jimmy.harvard.edu (J. Quackenbush), dspentzo@bidmc.harvard.edu (D. Spentzos).

0888-7543/$ – see front matter © 2013 Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.ygeno.2013.03.008

sequences detected by both technologies. There are, however, a number of sequences we identified that are detected by only one of the two platforms. To explore this discrepancy, we performed quantitative real-time-polymerase chain reaction (q-RT-PCR) amplification of discordantly detected miRNAs. This analysis indicated that the miRNA-seq method reports few false-positive detections; however, it potentially misses the detection of some sequences which are robustly picked up by the DASL assay. There are several plausible explanations for the observed discrepancies, which merit further investigation to improve the reliability of miRNA sequencing methods applied to FFPE tissue.

2. Materials and methods

2.1. FFPE tissue collection

FFPE tissue samples were retrieved from the pathology archives of Beth Israel Deaconess Medical Center and Boston Children's Hospital. A protocol for archival tissue collection was approved by Institutional Review Board at both institutions with a waiver of consent. The five FFPE blocks contained tissue from one liposarcoma of the thigh (5y-old), one leiomyosarcoma of the abdomen (4y-old), and three osteosarcomas (8y-old, 4y-old, and 17y-old, respectively). The first two samples (Leiomyosarcoma and liposarcoma) were run in duplicate. Each of the additional 3 osteosarcoma samples was run as a single assay.

2.2. RNA extraction procedure and miRNA DASL assay protocol

FFPE samples were cut into 1–3 mm cores. Total RNA was isolated using the Qiagen RNeasy FFPE protocol according to manufacturer instructions. RNA purity was assessed by spectrophotometer readings; the A260/A280 ratios were 2.03, 1.87, 1.71, 1.75, and 1.77 for the five samples, which indicates a high level of purity. miRNA DASL (c-DNA-mediated, Annealing, Selection and Ligation) arrays, (Illumina, CA) containing probes for 1146 miRNAs were used for profiling. The DASL assay is a bead-based method for expression profiling of degraded RNA, such as that found in FFPE samples [8,9]. The expression profiling experiments were performed at the Molecular Genetics Core, Boston Children's Hospital and Harvard Medical School. Raw data were processed using the lumi package in R [11,12]. A variance-stabilizing transformation and quantile normalization were applied before data were analyzed. After processing, signal intensities were averaged for miRNAs with p-values smaller than 0.01 in both replicates. The raw data from these experiments have been deposited in the GEO repository (GSE35851 and GSE39040) [13].

2.3. miRNA-sequencing protocol

Total RNA samples were prepared for smRNA sequencing using Illumina's Small RNA v1.5 Sample Preparation Guide. Total RNA input ranged from 1 to 10 μg and first underwent 3′ and 5′ adaptor ligation followed by reverse transcription and 12 cycles of amplification on a Bio-Rad iCycler. cDNA constructs were then purified using a 6% Novex TBE PAGE gel on Invitrogen's XCell SureLock Novex Mini-Cell System. Band sizes ranging from 80 to 100 bp were cut from the gel and purified. cDNA constructs were eluted from the gel and purified by ethanol precipitation according to Illumina's protocol. Libraries were analyzed on Agilent's 2100 Bioanalyzer with a High Sensitivity DNA Chip specific for next generation sequencing. Final libraries were immobilized onto a single read Illumina flowcell at a concentration of 12 pM and underwent cluster amplification on Illumina's Cluster Station using their DGE Small RNA Cluster Generation Kit. The amplified flowcell was then sequenced on Illumina's GAIIx with 36 cycles of sequencing.

2.4. MiRNA read mapping and quantification

The leading 21 bases were trimmed from the 36-bp reads based on the quality score and the length of mature miRNAs. The trimmed reads were mapped to miRNA precursor sequences using the software miRExpress [14,15]. The first two specimens were mapped to miRBase 16.0 and the three subsequent samples were mapped to miRBase 19.0. The two different versions of the database did not materially affect the output as the number of sequences on miRBase 16.0 that were “dead sequences” on miRBase 19.0 was only 19. No differences between the reads and the miRNA precursor sequences were allowed, which indicates exact matches only. The number of reads mapped to a miRNA sequence was taken to represent the expression level of that particular miRNA. Raw data were log2 transformed before performing analysis. miRNAseq count data from these experiments have been deposited in the GEO repository (GSE36147) [13].

2.5. q-RT-PCR protocol

Two small RNA samples were reverse transcribed using the TaqMan microRNA Reverse Transcription Kit (Applied Biosystems) according to the manufacturer's protocol. A total of 10 ng for each RNA sample was used as input in a total reaction volume of 15 μL. Reverse transcription was performed at 16 °C for 30 min, 42 °C for 30 min and 85 °C for 5 min followed by a 4 °C hold on an iCycler thermal cycler (BioRad). Each cDNA sample was then assayed against 9 TaqMan Small RNA Assays (Applied Biosystems) according to the manufacturer's protocol. Two control small RNA assays were also included as well as an 18S endogenous control for each cDNA sample. For each cDNA sample, a total of 1.33 μL was used in a total reaction volume of 20 μL, and each reaction was performed in triplicate for the first two samples and in quadruplicate for the last three specimens. qPCR assays were run using Relative Quantification on a 7900HT Fast instrument (Applied Biosystems) with a 95 °C hold for 10 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. None of the discordant miRNAs that we chose to study with RTPCR was one of the 19 defunct entries (“dead sequences”) between the two different miRbase versions.

2.6. Analytical methods

For the cross-platform concordance analysis we considered a miRNA to be detected by the DASL assay if the detection p-value was smaller than 0.01 in both technical replicates. Detection by sequencing was defined as mapping an average of at least six reads to a known miRNA in the human genome with both replicates reporting positive counts. We chose six reads as the cutoff because six was the highest number of reads in one replicate which gave zero reads in the other replicate from the same sample. We computed two correlation coefficients across the platforms for each sample (Pearson and Spearman). The Spearman correlation coefficient (non-parametric) measures the similarity between rankings of miRNA expression levels across the platforms and therefore does not depend on absolute numerical values. The Pearson coefficient (parametric) does depend on absolute numerical values.

3. Results

3.1. Reproducibility assessment and sequencing yield

We ran the miRNA DASL assay and miRNA-seq on five FFPE tissue specimens (two soft tissue sarcoma specimens, each in duplicate, and three additional osteosarcoma single specimens). Technical reproducibility of DASL has been previously demonstrated by Chen et al. [9]. However, to confirm reproducibility in our dataset, we computed correlation coefficients for the replicates of the two of the aforementioned samples (Pearson correlations: 0.880 and 0.818). These values were

lower than those previously reported in the literature, although for 80% of DASL probes the correlation coefficients were very high (0.971 and 0.970). The internal reproducibility of the sequencing platform was exceptional for both samples (Pearson correlations: 0.983 and 0.990). For each duplicate of the first two samples, there were respectively 26,456,978; 26,804,099; 24,001,082; and 24,022,445 sequencing reads generated. For samples 3, 4, and 5, the sequencing reads generated were 34,611,338, 44,201,733, and 36,109,739, respectively. After removing Illumina adaptor-dimer sequences and other "too short to read" sequences, the remaining totals were respectively 15,305,718; 15,065,728; 10,925,885; and 10,715,805. The remaining totals for samples 3, 4, and 5 were 19,211,152, 39,017,723, and 13,418,289. The percentages of these reads that could be quantified by mapping to miRBase 16.0 were 6.4% and 6.8% for sample 1, 16.1% and 16.9% for sample 2, 13.5% for sample 3, 14.1% for sample 4, and 6.5% for sample 5 (the three later samples mapped to miRBase 19.0). These percentages were expected because we mapped reads only to known miRNAs, and because we sequenced RNA derived from FFPE tissue.

3.2. Cross-platform concordance

Next we assessed the cross-platform concordance by computing correlation coefficients between miRNA levels for sequences detected by both technologies. First, we computed Pearson correlations between DASL signal intensities (averaged for the first two samples that were run in duplicate) and miRNA-seq reads for miRNAs expressed in the bottom third, middle third, and top third of measurements as ranked by each platform. We also computed non-parametric Spearman rank correlation coefficients which do not depend on absolute numerical values (Table 1). Lastly, we computed summary correlations, which aggregate all miRNAs detected on both platforms. The summary Pearson correlations between DASL signal intensities and miRNA-seq reads ranged from 0.524 to 0.594. The Spearman rank correlations ranged from 0.514 to 0.596 (Table 2).

We then assessed the pair-wise correlation between fold change (sample1/sample2, sample3/sample4, etc.) values for miRNAs present in both samples of each pair. There were 139 miRNAs detected in samples 1 and 2 by both platforms, 236 in samples 3 and 4, 219 in samples 3 and 5, and 221 in samples 4 and 5. These overlapping miRNAs yielded fold-change correlations that ranged from 0.268 to 0.651 (Pearson) and 0.397 to 0.646 (Table 2). The two technologies agreed on the direction of the cross-sample fold-change for 68.3% of miRNAs in samples 1 and 2, 75.8% of miRNAs in samples 3 and 4, 68.9% of miRNAs in samples 3 and 5, and 76.9% in samples 4 and 5. Given that we are restricting our analysis to only miRNAs that were robustly detected by both technologies, the correlations obtained likely reflect true non-trivial cross-platform discordance.

Table 1
Cross-platform concordance at different miRNA expression levels. For all samples the cross-platform Pearson (parametric) and Spearman (non-parametric) correlation coefficients are presented for miRNAs expressed in the bottom, middle, and top third as determined by each platform.

	Ranked by DASL intensities			Ranked by RNA-seq reads		
	Bottom 33%	Middle 33%	Top 33%	Bottom 33%	Middle 33%	Top 33%
Sample 1						
Pearson	0.000	−0.014	0.109	−0.001	0.258	0.287
Spearman	0.010	0.011	0.214	−0.043	0.233	0.335
Sample 2						
Pearson	0.196	0.184	0.181	0.059	−0.011	0.485
Spearman	0.186	0.143	0.329	0.118	−0.042	0.536
Sample 3						
Pearson	0.336	0.322	0.274	0.115	0.156	0.330
Spearman	0.329	0.312	0.286	0.150	0.186	0.364
Sample 4						
Pearson	0.332	0.287	0.314	0.136	0.097	0.478
Spearman	0.328	0.265	0.307	0.132	0.084	0.509
Sample 5						
Pearson	0.311	0.331	0.284	0.192	0.218	0.335
Spearman	0.355	0.312	0.281	0.178	0.227	0.370

In addition to the moderate degree of quantitative discordance observed, we examined whether there were substantial differences between the platforms with respect to "present versus absent" calls of miRNAs according to our detection criteria. We compared the miRNAs detected by sequencing – for which there are probes on the DASL array – to the set of miRNAs detected by the DASL assay. For sample 1 there were 449 miRNAs detected by DASL, 334 detected by miRNA-seq, and 169 detected on both platforms. For sample 2, DASL detected 473 miRNAs, miRNA-seq detected 366, and 205 miRNAs were detected by both. For sample 3, DASL detected 501 miRNAs, miRNA-seq detected 454, and 263 were detected by both. Sample 4 had 568 detected by DASL, 417 detected by miRNA-seq, and 277 detected by both. Lastly, for Sample 5 there were 528 miRNAs detected by DASL, 393 by miRNA-seq, and 236 detected by both. These results – summarized in Table 2 – suggested a moderate degree of discordance with respect to simply calling a miRNA species present or absent from a given sample. Upon further analysis, we determined that the large majority of these discordantly detected miRNAs (although not all) were clustered at the low to very low expression range on the platform where they were called "present". In addition, we found that 146 miRNAs included on the DASL array were non-human or could not be attributed to a known miRNA, and that a small number of miRNAs called "present" by miRNA-seq, did not have representative probes on the DASL array. Given these confounders, the moderate degree of detection discordance, and the ambiguity regarding which platform is correct, we further investigated this phenomenon with q-RT-PCR amplification of a small subset of miRNAs.

3.3. q-RT-PCR analysis confirmation and quantitative comparison

From each sample, we examined three possible scenarios. We selected five miRNA sequences for which both DASL and miRNA-seq reported very high abundance levels to serve as positive controls. Then, for both platforms, we chose miRNA sequences which were detected at high or middle expression levels by one of the two technologies and not detected by the other. We performed q-RT-PCR quantitation of these sequences in each sample (results presented in Table 3 and the q-RT-PCR data available online). For the first two samples, three of the sequences detected by DASL and RNA-seq were confirmed as present by q-RT-PCR, as expected. For sample 1, all six of the miRNAs detected by miRNA-seq but not by DASL were all confirmed to be present by q-RT-PCR. Three of the six miRNAs in this category were also detected by miRNA-seq but not by DASL in sample 2. Out of the six miRNAs detected by DASL and missed by miRNA-seq overall in the first two samples, five were confirmed to be present in both samples and one was confirmed to be absent in both samples. For samples 3–5, each miRNA that was detected by either DASL (but absent in miRNA-seq) or miRNA-seq (but absent in DASL) was confirmed to be present in all 3 specimens by q-RT-PCR.

4. Discussion

Next-generation RNA sequencing technologies hold great promise to accelerate translational and basic research due to their high-throughput, quality data generation, ability to generate novel sequencing information, and decreasing cost [5]. Traditionally, microarrays were the predominant tools to investigate genome-wide research questions, and many studies utilizing these data have been published [3,4]. In the coming years sequencing technologies will become increasingly utilized to answer major questions pertaining to genomics, epigenomics, and

Table 2
Summary of cross-platform concordance. For all samples the cross-platform Pearson (parametric) and Spearman (non-parametric) correlation coefficients are presented. Also, the numbers of miRNAs passing detection criteria using each technology are reported along with the number of discordant detections. The number of miRNAs detected by RNA-seq should be considered in the context of the number of miRNAs which DASL is capable of detecting (number of probes spotted on the array, by design).

Quantitative correlations	Fold-change correlations	Detection concordance		
Sample 1				
Pearson	Pearson	Detected by DASL	Detected by RNA-seq	Detected by both
0.535	0.268	449	355	169
Spearman	Spearman	Detected by DASL, not RNA-seq	Detected by RNA-seq, not DASL	
0.557	0.397	280	165	
		Detected by RNA-seq but no probe on DASL	miRNA-seq detected with DASL probes	
		20	334	
Sample 2				
Pearson	Pearson	Detected by DASL	Detected by RNA-seq	Detected by both
0.545	0.268	473	402	205
Spearman	Spearman	Detected by DASL, not RNA-seq	Detected by RNA-seq, not DASL	
0.55	0.397	268	161	
		Detected by RNA-seq but no probe on DASL	miRNA-seq detected with DASL probes	
		36	366	
Sample 3				
Pearson	Pearson	Detected by DASL	Detected by RNA-seq	Detected by both
0.594	0.573 (0.541–0.605)	501	454	263
Spearman	Spearman	Detected by DASL, not RNA-seq	Detected by RNA-seq, not DASL	
0.596	0.631 (0.626–0.635)	238	191	
		Detected by RNA-seq but no probe on DASL	miRNA-seq detected with DASL probes	
		45	354	
Sample 4				
Pearson	Pearson	Detected by DASL	Detected by RNA-seq	Detected by both
0.556	0.628 (0.605–0.651)	568	417	277
Spearman	Spearman	Detected by DASL, not RNA-seq	Detected by RNA-seq, not DASL	
0.548	0.636 (0.626–0.646)	291	140	
		Detected by RNA-seq but no probe on DASL	miRNA-seq detected with DASL probes	
		47	331	
Sample 5				
Pearson	Pearson	Detected by DASL	Detected by RNA-seq	Detected by both
0.524	0.596 (0.541–0.651)	528	393	236
Spearman	Spearman	Detected by DASL, not RNA-seq	Detected by RNA-seq, not DASL	
0.514	0.641 (0.635–0.646)	292	157	
		Detected by RNA-seq but no probe on DASL	miRNA-seq detected with DASL probes	
		57	306	

transcriptomics [6]. At the same time new microarray technologies have been developed, which allow for high throughput interrogation of even low-abundance transcripts using very little input RNA [7], and it has been shown that microarray-based capture in conjunction with RNA-seq can yield deeper insights into species expressed at low levels than RNA-seq alone [16]. Given the evolving nature of NGS methods, an important question to investigators in diverse fields is their degree of concordance with microarray-based expression profiling. Studies demonstrating a high degree of concordance between arrays and sequencing have been published, but they use very high quality RNA with little expected degradation [5,10]. To our knowledge, there have been no investigations comparing miRNA expression data generated by microarrays to those from sequencing that use RNA extracted from FFPE tissues. These challenging samples are particularly important to many clinical and translational investigators due to their widespread acquisition during routine medical care. A platform (DASL assay) has been developed and utilized to date and it has been shown to generate highly reproducible gene and miRNA expression data from FFPE tissue, but comparative data with NGS profiling miRNAs using degraded RNA is lacking.

In this preliminary, pilot study, we attempted to tackle this problem by interrogating RNA from 5 FFPE samples in technical duplicate on both platforms and evaluated how well miRNA quantitation correlated between the technologies, and how often the same miRNAs were detected. Then we attempted to determine which technology was more likely to be accurate in cases of discordant detection using q-RT-PCR.

The most important results arise from the comparison of detection and quantitation levels across the two platforms. We found that quantitative measures of miRNA expression were only moderately correlated. These correlation coefficients are similar to one published cross-platform comparison of moderately expressed gene signals (not measured using DASL) using intact RNA from yeast cells [5]. Interestingly, we found that correlation metrics in the highest expression tertiles were higher than those in the middle and low expression tertiles. We also found only a modest correlation across fold-changes in miRNA levels obtained from both technologies. A possible explanation for these observations is the documented loss of quantitative precision of RNA-seq read counts and microarray intensity measurement for

Table 3
Summary of q-RT-PCR findings. A subset of miRNAs found to be discordantly detected by either RNA-seq or DASL was chosen for q-RT-PCR analysis. For each sequence, the number of sequencing reads, DASL signal intensity, and DASL detection p-value are presented. The last two columns report if q-RT-PCR detected the miRNA and which platform q-RT-PCR data is concordant with.

	Sequencing reads	Normalized DASL Intensity	DASL detection p-value	qRT-PCR detects	Platform q-RT-PCR confirms
Sample 1					
Detected by DASL, not RNA-seq					
hsa-miR-544	0	13.93	0.00	No	RNA-seq
hsa-miR-198	0	11.54	0.00	Yes	DASL
hsa-miR-1260a	0	13.55	0.00	Yes	DASL
hsa-miR-939	0	11.42	0.00	Yes	DASL
hsa-miR-1281	0	11.16	0.00	Yes	DASL
Detected by RNA-seq, not DASL					
hsa-miR-148b	575	6.83	0.79	Yes	RNA-seq
hsa-miR-378	8157	6.92	0.56	Yes	RNA-seq
hsa-miR-151-3p	2881	6.96	0.44	Yes	RNA-seq
hsa-miR-301b	40	6.84	0.76	Yes	RNA-seq
hsa-miR-181d	286	7.07	0.23	Yes	RNA-seq
hsa-miR-345-5p	48	8.23	0.12	Yes	RNA-seq
Detected by both					
hsa-miR-21	332838	13.66	0.00	Yes	Both
hsa-miR-143	250362	11.96	0.00	Yes	Both
hsa-miR-1180	8	10.68	0.00	Yes	Both
Detected by neither					
hsa-miR-410	1	8.78	0.65	Yes	Neither
hsa-miR-494	3	8.19	0.02	Yes	Neither
Sample 2					
Detected by DASL, not RNA-seq					
hsa-miR-544	0	13.78	0.00	No	RNA-seq
hsa-miR-198	1	11.72	0.00	Yes	DASL
hsa-miR-1260a	1	13.37	0.00	Yes	DASL
hsa-miR-494	2	11.22	0.00	Yes	DASL
hsa-miR-939	3	9.74	0.00	Yes	DASL
Detected by RNA-seq, not DASL					
hsa-miR-18a	141	6.84	0.76	Yes	RNA-seq
hsa-miR-301b	189	6.84	0.76	Yes	RNA-seq
has-miR-33a	92	6.96	0.40	No	DASL
hsa-miR-181d	143	6.96	0.38	Yes	RNA-seq
hsa-miR-345-5p	14	6.97	0.33	Yes	RNA-seq
Detected by both					
hsa-miR-21	40156	13.72	0.00	Yes	Both
hsa-miR-143	97639	11.10	0.00	Yes	Both
hsa-miR-1180	10	11.39	0	Yes	Both
Detected by neither					
hsa-miR-1281	0	7.25	0.03	Yes	Neither
hsa-miR-410	5	6.9	0.56	Yes	Neither
Sample 3					
Detected by DASL, not RNA-seq					
hsa-miR-544	0	12.96	0	Yes	DASL
hsa-miR-1260a	3	14.3	0	Yes	DASL
hsa-miR-494	3	13.4	0	Yes	DASL
hsa-miR-939	3	11.86	0	Yes	DASL
hsa-miR-1281	0	10.93	0	Yes	DASL
Detected by RNA-seq, not DASL					
hsa-miR-301b	265	7.76	0.69	Yes	RNA-seq
hsa-miR-181d	517	7.85	0.64	Yes	RNA-seq
hsa-miR-345-5p	204	8.1	0.46	Yes	RNA-seq
hsa-miR-410	50	8	0.55	Yes	RNA-seq
hsa-miR-1180	37	8.18	0.39	Yes	RNA-seq
Sample 4					
Detected by DASL, not RNA-seq					
hsa-miR-544	0	12.48	0	Yes	DASL
hsa-mir-1260a	0	13.87	0	Yes	DASL
hsa-miR-494	1	12.29	0	Yes	DASL
hsa-miR-939	1	11.77	0	Yes	DASL
hsa-miR-1281	0	12.08	0	Yes	DASL
Detected by RNA-seq, not DASL					
hsa-mir-301b	173	7.53	0.79	Yes	RNA-seq
hsa-mir-181d	678	8.07	0.36	Yes	RNA-seq
hsa-mir-345-5p	95	7.93	0.46	Yes	RNA-seq
hsa-mir-410	72	8.56	0.06	Yes	RNA-seq
hsa-mir-1180	65	8.48	0.09	Yes	RNA-seq
Sample 5					
Detected by DASL, not RNA-seq					
hsa-miR-544	0	13.46	0	Yes	DASL
hsa-mir-1260a	2	14.5	0	Yes	DASL
hsa-miR-494	2	14.25	0	Yes	DASL
hsa-miR-939	0	11.86	0	Yes	DASL
hsa-miR-1281	0	12.61	0	Yes	DASL
Detected by RNA-seq, not DASL					
hsa-mir-301b	113	7.51	0.81	Yes	RNA-seq
hsa-mir-181d	267	7.8	0.64	Yes	RNA-seq
hsa-mir-345-5p	98	8.15	0.41	Yes	RNA-seq
hsa-mir-410	36	8.21	0.36	Yes	RNA-seq
hsa-mir-1180	21	8.03	0.48	Yes	RNA-seq

species expressed at low levels; high relative measurement errors may diminish quantitative comparability between DASL and RNA-seq [17].

Another interesting observation is the number of miRNAs which appear to be discordantly detected, suggesting that either technology (or both) harbors some deficiency in sensitivity or specificity. It seems possible that increasing sequencing depth may increase the detection concordance between the technologies for miRNAs expressed at low levels, but would not alter our results for abundant miRNAs. McIntyre et al. propose that the low proportion of RNA molecules sequenced relative to the total pool of RNA present in an initial tissue sample may account for technical variability in RNA-seq experiments, and that increased coverage may reduce this variability [18]. This explanation may be particularly relevant for miRNAs which are naturally present at low levels. In addition, the observation that the majority of discordantly detected miRNAs were clustered in the low to very low expression range seems natural and perhaps expected. Nonetheless, several of the discordant miRNAs were highly expressed by one technology and completely absent by the other suggesting that low expression level cannot totally account for this discrepancy and increased coverage depth may not totally eliminate it.

We performed q-RT-PCR quantitation of a carefully selected set of miRNAs from both samples. Because of the potential problems with cross-hybridization on microarrays, we could reasonably expect that DASL may make false positive detection calls. However, this was not entirely the case; several of the discordant patterns of detection were confirmed by q-RT-PCR for either platform. This suggests that perhaps certain miRNAs are inherently difficult to sequence, but still may be detectable by microarray. The difficulty could be due to a variety of factors – including base composition or secondary structure stability – of which we can only speculate. We considered alternative explanations for the observed detection discrepancies such as increased RNA degradation over time or a greater number of freeze–thaw cycles between the microarray experiment and the RNA-seq run. Both of these possibilities are unlikely since the q-RT-PCR experiments that were run after DASL and miRNA-seq confirmed the presence of the positive control miRNAs at

very high levels. This being said, it is conceivable that hsa-miR-544 may have been degraded through freeze–thaw cycles. It is unlikely that this miRNA would have been falsely detected by a non-specific DASL probe because it has been shown that cross-hybridized miRNAs differing by a single base would have a measured signal intensity reduced by a factor of up to 60 relative to a true positive on the miRNA DASL assay [9]. In light of the apparent non-detection by RNA-seq of present and highly abundant species, we see another plausible explanation that certain miRNAs are lost during sequencing library preparation. Importantly, our q-RT-PCR experiments validated the presence of all but one miRNAs detected by RNA-seq but missed by DASL, thereby supporting the notion that sequencing-based methods produce few false positive calls when studying FFPE tissue.

Our findings were not seriously confounded by a number of specimen-specific or platform-specific technical parameters. We found that older specimen age did not correlate with inferior concordance between the two platforms, although older RNA extractions perhaps could. Older specimen age could lead to lower efficiency (higher fraction of unmappable short reads) and this could have an effect in mutation and sequence analyses, but it did not affect our quantitative and detection data. Similarly, although the magnitude of discordance is partly diminished by DASL design limitations (such as non human or "dead" probes, or a predefined limit of ~1000 array probes), or by limitations of earlier miRNA-seq platforms, our data still suggests a residual amount of true discordance in the performance of the two technologies.

It is important to consider various limitations of this pilot study. We studied only one specific platform for RNA sequencing as an example. Other sequencing platforms have been developed, including some which are thought to be more specifically designed for FFPE tissues [19]. We cannot know with certainty if our findings would be applicable to these other platforms but we believe that they provide the basis for further evaluation in the future. We utilized both miRBase 16.0 and miRBase 19.0 in successive steps of our study. That being said, our analysis suggested that this did not have material impact on the main conclusions and we also took care not to use any of the defunct sequences for the RT-PCR validation. Also, we used a relatively small number of samples (5). A larger study would have been preferable, but was precluded due to the high cost of these experiments. We feel that this initial pilot experiment on a limited number of samples provided clear and novel insights that would be better studied in larger cohorts. Furthermore, we would like to underscore that our results are not meant to prove that one technology is better overall since we could not perform comprehensive RT-PCR comparison on all discordant miRNAs in this pilot analysis; rather, we wish to elucidate some of the differences in performance and perhaps suggest their complementary value. From the sample of 18 discordant miRNAs studied with RT-PCR, we can reasonably conclude that neither of the two technologies is likely to be always right or always wrong in this degraded tissue material, although an emerging theme appears to be that miRNA seq may be more specific. Lastly, it should be noted that our study did not attempt to directly address the question of whether FFPE specimens are adequately preserved for high-throughput profiling compared to frozen tissue. Future larger studies that would include paired frozen/FFPE specimens, ideally encompassing enough samples from different histologic types, coupled with large-scale RTPCR experiments would be required in order to definitively assess the comparative merits and shortcomings of these technologies.

5. Conclusions

In summary, we have directly compared the detection and quantitation of miRNAs present in FFPE tissue samples using the DASL assay and miRNA-seq. Our data indicate that for miRNAs robustly detected by both technologies, there is a moderate degree of correlation between quantitative expression levels, and that highly abundant species are more highly correlated. We also conclude that miRNA-seq is highly specific in that it produces few false positive miRNA detections; however we caution investigators that certain present miRNAs – including highly abundant ones – may be missed by sequencing. We believe that these findings elucidate an important technological advance and merit further research into the potential causes for the non-detection of certain miRNAs to improve the applicability of NGS methods to archival tissue-based studies. Further studies building on these results will also help evaluate how NGS and microarray based technologies can be of complementary value for clinical investigation.

Author contributions

ADK, MC, JQ, and DS conceived the study. ADK, KEH, MC, LH, YEW, RR, SD, and DS conducted experiments and performed data analysis. ADK, KEH, JQ, MC, YW, and DS wrote the manuscript.

Acknowledgments

The authors would like to thank the following institutions for the funding provided for this study:

The National Institutes of Health [K22 CA138716-01A2 to D.S., U19 CA148065 to J.Q.]; the Dana-Farber Cancer Institute Strategic Plan Fund [to M.C., J.Q.]; and a donation by Dr Richard and Virginia Clemmer to the sarcoma program at Beth Israel Deaconess Medical Center [to D.S.]. The authors also wish to thank Antonio R. Perez-Atayde and Kyle Kurek (Boston Children's Hospital, Department of Pathology) and Jeffrey Goldsmith (BIDMC Department of Pathology) for providing FFPE sample for this study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.03.008.

References

[1] M. Meyerson, S. Gabriel, G. Getz, Advances in understanding cancer genomes through second-generation sequencing, Nat. Rev. Genet. 11 (2010) 685–696.
[2] R.D. Hawkins, G.C. Hon, B. Ren, Next-generation genomics: an integrative approach, Nat. Rev. Genet. 11 (2010) 476–486.
[3] H.S. Chon, J.M. Lancaster, Microarray-based gene expression studies in ovarian cancer, Cancer Control 18 (2011) 8–15.
[4] B. Lallemant, A. Evrard, G. Chambon, O. Sabra, S. Kacha, J.G. Lallemant, S. Lumbroso, J.P. Brouillet, Gene expression profiling in head and neck squamous cell carcinoma: clinical perspectives, Head Neck 32 (2010) 1712–1719.
[5] Z. Wang, M. Gerstein, M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, Nat. Rev. Genet. 10 (2009) 57–63.
[6] P.J. Hurd, C.J. Nelson, Advantages of next-generation sequencing versus the microarray in epigenetic research, Brief Funct. Genomic Proteomic 8 (2009) 174–183.
[7] W. Xu, J. Seok, M.N. Mindrinos, A.C. Schweitzer, H. Jiang, J. Wilhelmy, T.A. Clark, K. Kapur, Y. Xing, M. Faham, J.D. Storey, L.L. Moldawer, R.V. Maier, R.G. Tompkins, W.H. Wong, R.W. Davis, W. Xiao, Human transcriptome array for high-throughput clinical studies, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 3707–3712.
[8] J.B. Fan, J.M. Yeakley, M. Bibikova, E. Chudin, E. Wickham, J. Chen, D. Doucet, P. Rigault, B. Zhang, R. Shen, C. McBride, H.R. Li, X.D. Fu, A. Oliphant, D.L. Barker, M.S. Chee, A versatile assay for high-throughput gene expression profiling on universal array matrices, Genome Res. 14 (2004) 878–885.
[9] J. Chen, J. Lozach, E.W. Garcia, B. Barnes, S. Luo, I. Mikoulitch, L. Zhou, G. Schroth, J.B. Fan, Highly sensitive and specific microRNA expression profiling using BeadArray technology, Nucleic Acids Res. 36 (2008) e87.
[10] J.C. Marioni, C.E. Mason, S.M. Mane, M. Stephens, Y. Gilad, RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays, Genome Res. 18 (2008) 1509–1517.
[11] P. Du, W.A. Kibbe, S.M. Lin, lumi: a pipeline for processing Illumina microarray, Bioinformatics 24 (2008) 1547–1548.
[12] S.M. Lin, P. Du, W. Huber, W.A. Kibbe, Model-based variance-stabilizing transformation for Illumina microarray data, Nucleic Acids Res. 36 (2008) e11.
[13] R. Edgar, M. Domrachev, A.E. Lash, Gene Expression Omnibus: NCBI gene expression and hybridization array data repository, Nucleic Acids Res. 30 (2002) 207–210.
[14] A. Kozomara, S. Griffiths-Jones, miRBase: integrating microRNA annotation and deep-sequencing data, Nucleic Acids Res. 39 (2010) D152–D157.
[15] W.C. Wang, F.M. Lin, W.C. Chang, K.Y. Lin, H.D. Huang, N.S. Lin, miRExpress: analyzing high-throughput sequencing data for profiling microRNA expression, BMC Bioinforma. 10 (2009) 328.

[16] A. Roberts, L. Pachter, RNA-Seq and find: entering the RNA deep field, Genome Med. 3 (2011) 74.
[17] P.P. Labaj, G.G. Leparc, B.E. Linggi, L.M. Markillie, H.S. Wiley, D.P. Kreil, Characterization and improvement of RNA-Seq precision in quantitative transcript expression profiling, Bioinformatics 27 (2011) i383–i391.
[18] L.M. McIntyre, K.K. Lopiano, A.M. Morse, V. Amin, A.L. Oberg, L.J. Young, S.V. Nuzhdin, RNA-seq: technical variability and sampling, BMC Genomics 12 (2011) 293.
[19] A.H. Beck, Z. Weng, D.M. Witten, S. Zhu, J.W. Foley, P. Lacroute, C.L. Smith, R. Tibshirani, M. van de Rijn, A. Sidow, R.B. West, 3′-end sequencing for expression quantification (3SEQ) from archival tumor samples, PLoS One 5 (2010) e8768.