Genomics 99 (2012) 275–281

Contents lists available at SciVerse ScienceDirect

Genomics

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

Characterization of the uterine leiomyoma microRNAome by deep sequencing

Boryana Georgieva [a,1], Ivan Milev [a,1], Ivan Minkov [a], Irina Dimitrova [b], Andrew P. Bradford [b], Vesselin Baev [a,*]

[a] Department of Plant Physiology and Molecular Biology, University of Plovdiv, 24 TsarAssen St, 4000 Plovdiv, Bulgaria
[b] Department of Obstetrics and Gynecology, University of Colorado School of Medicine, Aurora, Colorado 80045, USA

ARTICLE INFO

Article history:
Received 6 January 2012
Accepted 9 March 2012
Available online 16 March 2012

Keywords:
miRNAs
Small RNAs
Deep sequencing
Uterine leiomyoma
Myometrium

ABSTRACT

MicroRNAs (miRNAs) are a class of small non-coding RNAs, which are negative regulators of gene expression. Many genes in human uterine leiomyoma (ULM) are aberrantly expressed and in some cases this can be due to dysregulation of miRNAs. Here we present the first study to determine genome-wide miRNA expression patterns in uterine leiomyoma and myometrium using Solexa high-throughput sequencing. We found more than 50 miRNAs, which were differentially expressed, and furthermore we extend the list of putative new miRNA genes. The top five significantly de-regulated miRNAs in ULMs that we found in our libraries were miR-363, miR-490, miR-137, miR-217 and miR-4792. We also observed "isomiRs" with higher copy number than referenced mature miRNA specific for the leiomyoma libraries, which have a potential role in tumorigenesis. The microRNA transcriptomes obtained in this study deliver insights and further expand our understanding the role of small RNAs in uterine leiomyoma development.

1. Introduction

MicroRNAs (miRNAs) are a new class of negative regulators that repress gene expression by binding with their target messenger RNAs (mRNAs). Through this pairing to the 3′UTR of mRNAs, miRNAs trigger degradation of the target transcript or inhibition of protein translation [1,2]. Biogenesis of miRNA involves several steps. Firstly, miRNA genes are transcribed by RNA polymerase II into primary RNA (pri-miRNA) molecule. The pri-miRNA is subsequently cleaved in the nucleus by an RNase III endonuclease, Drosha, forming a 70–100 nucleotide (nt) hairpin loop structure known as precursor miRNA. Next, the pre-miRNA is exported into the cytoplasm and further processed, by Dicer, into a double stranded RNA duplex with 3′-overhangs [3–5]. Based on the thermodynamic stability of this duplex, one of the strands is selected as the biologically active mature miRNA whilst the other, known as the passenger strand or miRNA star, is typically degraded [6].

The mature miRNAs are short (20–24 nt) single stranded RNAs that, together with Argonaute proteins, mediate direct post-transcriptional regulation. miRNAs are essential for normal mammalian development and regulate genes involved in cell division and differentiation, metabolism, stress response and apoptosis [7]. miRNAs have been shown to regulate oncogenes, tumor-suppressors and a number of cancer-related genes, therefore their deregulation can predispose to disease and malignancies [8,9]. Most importantly, different cancer types, and stages exhibit unique miRNA expression patterns, suggesting that miRNA genes can function as novel biomarkers for cancer diagnosis [8,10,11]. Moreover, miRNA signatures are also altered in variety of benign tumors such as human uterine leiomyomas (ULMs) [12–14].

Leiomyomata uteri or fibroids are the most common neoplasm of the female genital tract developing primarily during the reproductive years and becoming symptomatic during perimenopause [15–18]. Tumors occur in approximately 80% of women, and approximately 25% of Caucasians have clinically significant, symptomatic lesions. Most affected women have multiple tumors with the average number of tumors per uterus estimated to be 6.5 [19]. The relative risk of fibroids is two to threefold greater in black women than white women and clinical disease is more severe in the former [20]. The biological sequelae of fibroids manifest as a spectrum of clinical symptoms, primarily pain and excessive menstrual bleeding. Reproductive issues are also a concern, as fibroids are associated with infertility and, if present during pregnancy, may contribute to second trimester pregnancy losses, premature labor, fetal malpresentations and/or distocia. Despite the significant impact on women's health, the etiology of fibroids remains poorly understood and only a few specific genes have been identified to be associated with development of this tumor.

Little is known about the repertoire and function of miRNA in human uterine leiomyoma. Several studies have assessed levels of subsets of miRNAs through microarray expression analysis [12–14], demonstrating that many miRNAs are deregulated in leiomyoma compared to normal tissue. To profile the small RNAome in leiomyomas and myometrium issues and to reveal deeper insights into miRNAs expression changes between diseased and normal states, we used Solexa high-throughput sequencing technology. This method has significant advantages over microarray and PCR-based assays in characterizing the miRNA

* Corresponding author.
E-mail address: vebaev@gmail.com (V. Baev).
[1] Contributed equally to this work.

0888-7543/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2012.03.003

transcriptome [21–23]. Unlike microarray, deep-sequencing is not limited by array content and thus allows identification of novel miRNAs. Moreover, parallel sequencing has the ability to identify low abundance miRNAs and affords an opportunity for detection of small differences in miRNAs expression between samples [21].

To the best of our knowledge, the differential expression of miRNAs has not been studied via deep-sequencing in ULMs. In this manuscript we used next-generation sequencing to characterize small RNA transcriptome, to compare expression signatures of known miRNAs in leiomyoma versus matched normal myometrium and to discover novel miRNAs. We were able to identify more than 50 miRNAs exhibiting altered expression and 25 genomic loci that encode putative novel miRNA genes, thus expanding the list of miRNAs representing the unique signature of uterine leiomyomas.

2. Results

2.1. Overview — sequencing statistics, lengths and reads

By using Solexa technology, a newly developed method for high-throughput sequencing producing highly accurate and quantitative read-outs of small RNA, four libraries of small RNAs from myometrium (B25N and B27N) and leiomyoma (B25L and B27L) were established and sequenced. After applying reads filters and discarding low quality reads, adaptor sequences were trimmed. Additionally, reads attributed to ligation contaminants or self-ligation of the adaptors were also discarded. We obtained 7,284,165 (B25N) and 5,506,769 (B27N) clean reads from normal myometrium and −7,629,120 (B25L) and 7,423,663 (B27L) from leiomyoma libraries. For the small RNA analysis we further clustered all raw reads into unique sequences with associate copy numbers generating 74,547 (B25N), 48,114 (B27N), 45,538 (B25L) and 59,148 (B27L) unique tags from normal myometrium and leiomyoma samples, respectively.

The length distribution of the high quality tags was consistent among the four RNA libraries. In our analysis of small RNAs we selected sequences in the range of 15 to 30 nt. Results showed that this resulted in enrichment of miRNAs since the fraction of reads within the range 20–22 nt (considered as a typical miRNA length) comprised the majority (80–90%) in all four samples (Fig. 1). Interestingly, a significant 23 nt peak representing 1,237,791 of 7,423,663 reads (~17%) was observed in sample B27L but not in the other libraries.

2.2. Mapping and annotation of small RNAs

The 15–30 nt clean reads from four libraries were subsequently aligned and mapped to the human genome allowing at most one mismatch. Total mappable reads were approximately 99% in all sequencing libraries. After clustering these mappable reads to unique sequences they represented 67.51% (B25N), 66.58% (B27N), 59.34% (B25L) and 61.96% (B27L) of the sequences in myometrium and leiomyoma samples, respectively. Based on the UCSC human genome annotation and additional well-characterized RNA datasets (see Materials and Methods) all small RNA sequences were annotated as known miRNAs, degradation fragments from non-coding RNAs (tRNA, rRNA, snRNA/snoRNA), genomic repeats, mRNAs (exons, introns) or unclassified. The fraction of small RNAs derived from UTR regions of mRNA was very small and therefore not included in our results.

As expected, the most abundant RNA category from the four libraries was known miRNAs: 95% (B25N and B27N); 96% (B25L and B27L). The remaining, less abundant categories were non-coding RNAs and repeat elements. A minor fraction of small RNAs were mapped to coding-sequences, which likely represent degradation products from mRNAs. Read counts of the classified categories are summarized in Table 1.

Recent studies showed that some genomes can generate repeat associated small RNAs, which can regulate retrotransposon expression. Yang et al., demonstrated that small interfering RNAs, which are endogenously coded, can suppress the L1 retrotransposon in human cell cultures [24]. Therefore, we additionally explored the repertoire of small RNAs that are aligned to genome repeats, and represent a set of putative RNAs associated with transposable elements. In the mammalian genome, transposon elements are mostly composed of three major classes: LTR, LINE and SINE retrotransposons. Our results revealed that the majority of reads mapping to repeat elements were

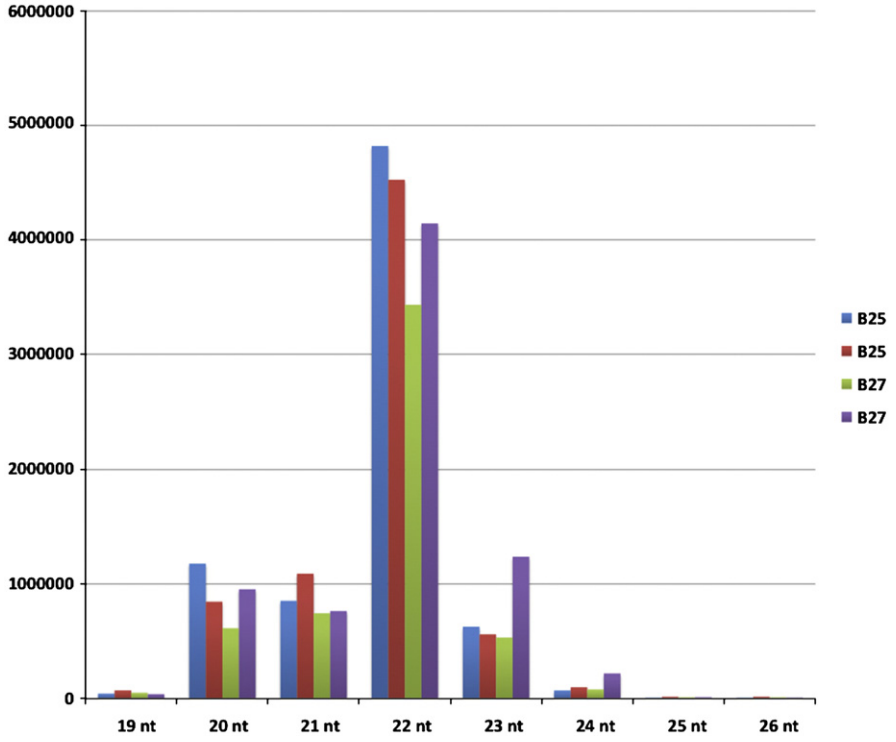

Fig. 1. The length distribution of small RNAs in leiomyoma (B25L and B27L) and myometrial (B25N and B27N) libraries.

Table 1
Classified small RNA species in four sequencing libraries.

Category	B25N		B25L		B27N		B27L	
	Unique	Total	Unique	Total	Unique	Total	Unique	Total
Exon sense	5262	5895	1966	2302	2713	3396	3073	3796
Exon antisense	324	400	129	156	220	264	233	319
Intron sense	6477	14637	3858	5886	3730	6241	5325	9288
Intron antisense	1383	3617	764	1262	723	1427	1143	2782
lincRNA	2275	5553	1066	3658	1123	2097	1796	3845
miRNA	6363	6906338	4963	7304242	4344	5233252	5629	7084478
otherRNA	768	4417	455	3226	372	1235	587	2529
rRNA	6071	19538	1614	3984	3824	21272	3206	9171
snRNA	459	944	189	302	114	232	287	623
snoRNA	2317	15152	1305	6969	1242	7084	3240	15045
tRNA	1876	9162	1364	6985	1369	5565	1690	7684
Repeats sense	5199	10286	2924	5248	2970	5776	4213	9151
Repeats antisense	2656	8480	1337	3287	1325	3049	1967	4717
Unannotated	33117	279746	23604	281613	24045	215879	26759	270235
Total	74547	7284165	45538	7629120	48114	5506769	59148	7423663

annotated as LINE derived, with the second most common category being LTR-derived small RNAs (see Appendix A).

2.3. MicroRNA expression patterns

There are 1424 human miRNAs annotated in release 17 of miRBase (Sept 2010). We detected approximately 25% of these miRNAs in small RNA libraries derived from normal myometrium and leiomyoma. A total number of 321 and 258 miRNAs were detected (with at least 5 read counts and not allowing mismatches in the mature sequence) in normal tissue samples (B25N and B27N) and 283 and 295 across leiomyoma samples (B25L and B27L), with 218 miRNA being common to all libraries.

Expression abundance analysis, based on raw read copy numbers (sequencing frequency) from sequencing, revealed that, in all four samples miR-143 represented the most abundant tag in the libraries. Other miRNAs detected at high copy number, based upon sequencing frequency were miR-145, miR-27b, miR-21, let-7a, let-7f, miR-26a and miR-125b (see Appendix C).

2.4. IsomiRs — variations in mature miRNA processing

miRNAs frequently exhibit differences from their "reference" mature sequences, generating multiple variations that are called as "isomiRs". In many cases both mature miRNA and corresponding isomiRs were present in the sequencing libraries. IsomiRs variability can be explained by the imprecise and alternative cleavage of Dicer and Drosha during pre-miRNA hairpin processing. Such events have been observed in several previous studies using high-throughput sequencing data [25–27]. Our results revealed that many of the mature miRNAs in myometrium and leiomyoma sequencing libraries possess isomiRs sequences. Moreover, our analysis shows that several isomiRs have higher raw read copy numbers, relative to their mature reference sequences indexed in miRBase (Fig. 2). In the myometrium datasets we observed 49 (B25N) and 43 (B27N) isomiRs that are more abundant than their mature reference miRNA and 44 (B25L) and 52 (B27L) in leiomyoma, with 23 isomiRs being common to all libraries. 13 of the identified isomiRs were uniquely over expressed in leiomyoma samples, whereas in control samples their referenced mature miRNAs were present at a higher frequency (Appendix D).

2.5. Differential expression of known miRNAs between leiomyoma and myometrium sequencing libraries

Microarray based methods have been extensively used to profile differential expression of known miRNAs at the genomic scale. To date, the miRNA repertoire of uterine leiomyomas and their aberrant expression, compared to normal myometrium, have been explored mainly with such microarray technology [12–14]. There are few reports describing analysis of differentially expressed miRNAs by deep-sequencing technology and currently this approach has not been used in uterine leiomyoma. Moreover, there are no standard

Fig. 2. Example of mature miRNA (miR-10b) in B25L showing 3′ and 5′ sequence variations (isomiRs). One of the observed isomiRs (in green) is showing considerably higher read counts than the published mature sequence in miRbase (in red).

methods for profiling miRNA expression using next-generation sequencing of small RNAs. In most cases expression level of miRNAs is estimated by the frequency (copy number) of the mature miRNA tag (referenced in miRBase). We have shown that many of the miRNAs from our sequencing libraries have corresponding isomiR sequences with higher copy number than their referenced mature sequence, implying that these mature miRNA variations can be specific for different tissues types or states. Recent studies suggest that expression divergence, detected by analyzing the most abundant isomiRs and all isomiRs, respectively, can be a better marker of the relative expression in deep-sequencing data [26,28]. Therefore, we used the intersection of two methods to identify differentially expressed miRNAs: firstly, absolute sequence read counts of the mature miRNA and star sequences (miRbase v17) from each sample were used as a input data for the statistical software Deseq [29] specializing in normalization and differential expression analysis for NGS data (Table 2B); secondly, we took into account all sequence variants generated from miRNA precursors (pre-miRNA) along with referenced mature sequence, thus assigning total number of read counts from all isomiRs; Subsequently, the total counts for each miRNA precursor were again loaded in Deseq module for analysis of expression divergence (Appendix B). The list of the differentially expressed miRNAs (considering isomiR variations) is shown in Table 2A. We identified 35 deregulated miRNAs by the first method and 52 by the second (with P-value<0.05 and $|\log_2 \text{ratio}| \geq 1$). Twenty-four miRNAs were found to be differentially expressed by both approaches.

2.6. Prediction of novel miRNA genes

One of the important advantages of deep-sequencing of the small RNA transcriptome is that it allows identification of novel miRNAs. A total of 24 novel miRNA candidates were predicted in a stepwise manner using widely adopted software for identification of miRNAs from deep-sequencing datasets MIREAP and miRDeep (see Materials and methods). We analyzed our libraries separately with the two programs to gain independent results. We observed these 24 miRNAs have a size range of 21–24 nt and the lengths of their predicted precursors vary from 73 to 112, which is consistent with known human miRNAs. More than a half of the newly predicted candidates (14 miRNAs) were categorized as intronic, having genomic location within an intron of host genes (Appendix E). Thirteen candidates were captured in more than one sequencing library, and several were identified in all four datasets. Notably many of the novel candidate miRNAs showed homology to other mammalian mature miRNAs derived from canine and rodent tissues (*Cannisfamiliaris, Susscrofa* and *Ratusnorvegicus*).

Piriyapongsa and colleagues demonstrated that members of a recently described family of human miRNA genes, miR-548, are derived from transposable elements. We observed 8 novel miRNAs that exhibit sequence similarity to this family and their genomic loci overlapped with transposon track in human genome (UCSC database) [30].

3. Discussion

In the present study we used next-generation sequencing approaches and bioinformatics methods to determine the differentially expressed miRNAs in human leiomyoma and normal myometrium. So far, this is the first profiling study of miRNA changes associated with leiomyoma uteri via high-throughput sequencing. Moreover we predicted 24 novel miRNAs in the sequenced samples.

Recent studies using deep-sequencing of small RNAs from the female reproductive tract showed that the let-7 family is one of the most abundant in the miRNA pool [31]. Based on sequenced frequency in our samples, the most expressed member of let-7 family was let-7a. Interestingly, we observed several miRNAs with higher expression levels. For example, in all four samples, miR-143 represented the most abundant tag in the libraries, with expression levels several fold higher than let-7a. MiRNAs are often located in gene clusters and it is generally believed that in these clusters miRNAs are co-regulated and co-transcribed [32]. Accordingly, we found very high relative expression levels of miR-145, which forms part of the miR-143/145 cluster.

In addition to expression profiling in sequenced libraries, we observed a diverse population of mature miRNA variants, known as "isomiRs". Although the functional role of isomiRs remains unknown, they have the potential to broaden the range of the miRNA regulatory network. Since variation in isomiR sequence occurs at the 3′ and 5′-ends, they could potentially bind to a different repertoire of targets relative

Table 2
Differential expression analysis of miRNAs detected by deep sequencing in leiomyoma compared to myometrium samples. (A) A collection of significantly deregulated miRNAs in leiomyoma, taking into account all isomiR variations generated from the miRNA precursor. (B) Differentially expressed analysis of miRNAs by mature sequences referenced in miRBase. miRNAs found by both methods are highlighted.

A		B	
miRNA name	log2FoldChange	miRNA name	log2FoldChange
Up-regulated		Up-regulated	
hsa-mir-363	6.53	hsa-miR-363	6.13
hsa-mir-490	6.21	hsa-miR-135a	5.51
hsa-mir-137	6.01	hsa-miR-135b	4.69
hsa-mir-548y	5.14	hsa-miR-490-3p	4.48
hsa-mir-135b	4.75	hsa-miR-3607-3p	4.23
hsa-mir-135a-1	4.62	hsa-miR-885-5p	4.12
hsa-mir-135a-2	4.62	hsa-miR-136	3.76
hsa-mir-885	4.42	hsa-miR-106a	3.68
hsa-mir-187	4.34	hsa-miR-154*	3.67
hsa-mir-483	4.06	hsa-miR-153	3.33
hsa-mir-2682	3.87	hsa-miR-222	3.26
hsa-mir-153-1	3.61	hsa-miR-9	3.16
hsa-mir-488	3.56	hsa-miR-20b	3.15
hsa-mir-624	3.54	hsa-miR-200b	3.02
hsa-mir-200b	3.37	hsa-miR-488	3.00
hsa-mir-549	3.12	hsa-miR-378d	2.98
hsa-mir-429	3.09	hsa-miR-378*	2.98
hsa-mir-9-1	2.97	hsa-miR-582-5p	2.92
hsa-mir-9-2	2.97	hsa-miR-204	2.78
hsa-mir-9-3	2.97	hsa-miR-501-5p	2.22
hsa-mir-153-2	2.96	hsa-miR-542-5p	1.80
hsa-mir-219-2	2.83		
hsa-mir-2964a	2.83		
hsa-mir-4443	2.78		
hsa-mir-200a	2.73		
hsa-mir-204	2.68		
hsa-mir-222	2.55		
hsa-mir-582	2.43		
hsa-mir-378b	2.33		
hsa-mir-378h	2.32		
hsa-mir-494	2.31		
hsa-mir-422a	2.28		
hsa-mir-376c	2.23		
hsa-mir-378e	2.11		
hsa-mir-376b	1.94		
hsa-mir-21	1.26		
Down-regulated		Down-regulated	
hsa-mir-320c-2	-1.93	hsa-miR-103a	-2.20
hsa-mir-320c-1	-1.98	hsa-miR-100*	-2.34
hsa-mir-877	-2.52	hsa-miR-101*	-3.18
hsa-mir-183	-2.85	hsa-miR-320a	-3.25
hsa-mir-935	-2.94	hsa-miR-29b-1*	-3.25
hsa-mir-129-2	-2.99	hsa-miR-22*	-3.47
hsa-mir-129-1	-3.01	hsa-miR-182	-3.61
hsa-mir-320a	-3.05	hsa-miR-144*	-3.66
hsa-mir-320b-2	-3.09	hsa-miR-27a*	-3.67
hsa-mir-320b-1	-3.12	hsa-miR-497	-3.92
hsa-mir-497	-3.35	hsa-miR-320b	-4.04
hsa-mir-182	-3.46	hsa-miR-451	-4.18
hsa-mir-144	-3.50	hsa-miR-590-3p	-4.48
hsa-mir-3926-1	-3.55	hsa-miR-486-5p	-4.85
hsa-mir-486	-3.58		
hsa-mir-451	-3.92		
hsa-mir-451b	-3.92		
hsa-mir-590	-4.39		
hsa-mir-4792	-4.99		
hsa-mir-217	-5.36		

to their mature reference counterparts. Moreover, in cases when isomiRs are expressed at higher levels than their reference mature miRNA, the target network of that particular miRNA can be altered. In this study we were able to detect isomiRs of approximately 1/6 of all identified miRNAs in each library. Notably, 13 miRNAs exhibited overexpression of their cognate isomiRs, uniquely in leiomyoma. Therefore the observed specific signature overabundance of isomiRs in the tumor samples, implies a potential role in the aberrant regulation of protein-coding genes that underlie leiomyoma pathogenesis and progression.

Like Guo and Lu [28], we think that analysis of isomiRs copy number can contribute to the estimation of the miRNA relative expression level. On that account, we evaluate the expression divergence of miRNAs by copy number of mature miRNA sequence and additionally based on sum of all isomiRs sequence counts. Our data showed several miRNAs with a significant up-regulation confirming published studies. We observed increased levels of miR-21 in leiomyoma, consistent with previous reports [12,14,33]. Moreover, miR-21 stands out as the only miRNA overexpressed in the vast majority of tumor types analyzed, indicating that miR-21 is a key regulator of cell proliferation, survival and migration and tumorigenesis [34]. Other miRNAs also reported previously to be differentially up-regulated are miR-483, miR-20b, miR-494, miR-376b and miR-582 [12–14]. MiR-483, which is located in an intron within the IGF2 gene, is overexpressed not only in leiomyoma, but also in other tumors (pancreatic, colon, breast and liver cancers), consistent with this miRNA being an important "oncomiR" [35–38]. Similarly, recent microarray analysis of miRNAs expression in uterine leiomyoma also demonstrated overlap with our observations with respect to down-regulated miRNAs. These include miR-217, miR-486, miR-451, miR-144 and miR-320 [12–14]. Interestingly, we found three members from the miR-320 cluster (miR-320a, miR-320b and miR-320c) to exhibit a similar pattern of down-regulation in leiomyoma libraries.

Our study identified many miRNAs that are reported here for the first time as differentially regulated in uterine leiomyoma comparing to normal myometrium. Some of them are mentioned in the literature as important tumor-related miRNAs, and we postulate that they may also play an essential role in leiomyoma formation. For example, we found miR-200a/b and miR-429 (belonging to the same family) to be up-regulated in leiomyoma. Snowdon and co-workers reported that the entire miR-200 family was overexpressed in endometrial carcinoma [39]. Moreover increased expression of miR-200b was also observed in ovarian, pancreatic, prostate and breast cancer [40–43].

In summary, we have investigated the miRNA expression profiles in human uterine leiomyoma and normal myometrium using a small RNA deep-sequencing approach. We identified a number of miRNAs that were significantly differentially expressed in fibroids, consistent with previous studies. In addition, we detected several novel putative miRNAs and unique patterns of isomiRs expression, which may extend the repertoire of miRNA regulated genes involved in the development and pathogenesis of uterine leiomyoma.

4. Materials and methods

4.1. Tissue samples

Portions of leiomyomata, approximately 2 × 3-cm sections from the periphery distinct from normal tissue, and matched unaffected myometrium were collected from 2 women who were undergoing hysterectomy for indications related to symptomatic leiomyomata. Tissue samples were obtained from each leiomyoma and adjacent matched normal myometrium. None of the patients had received any medical treatment for their fibroids. The tissues were collected at the University of Colorado Hospital with prior approval by the Colorado Multiple Institutional Review Board, under protocol number 03-642. Immediately after collection in the operating room, a portion of the tissue was snap frozen and stored in liquid nitrogen.

4.2. RNA isolation

Aliquots (~5 g) frozen tissue sections were pulverized under liquid nitrogen. Powdered tissue (50–100 mg) was placed in 1 mL TRIzol (Invitrogen Life Technologies, Inc., Carlsbad, CA) and was then homogenized with a Polytron probe (Brinkmann Instruments, Westbury, NY). Total cellular RNA was isolated from the tissues using TRIzol per manufacturer's instructions. The isolated RNA was quantitated with the NanoDrop ND1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE), so that no greater than 100 μg was loaded on the column during the DNase treatment. The RNA was DNase treated (twice) using QiagenRNeasy Mini kit and QiagenRNase-Free DNase s (Qiagen, Inc., Chatsworth, CA). Quality of purified RNA was assessed on a RNA 6000 Nano Chip using an Agilent 2100 bioanalyzer (Agilent Technologies, Germany).

4.3. Small RNA library construction and sequencing

Small RNA libraries and sequencing for each sample (control samples B25N, B27N and leiomyoma samples B25L, B27L) were performed by LC Sciences (Houston, TX). Briefly, the small RNA libraries were generated according to Illumina's sample preparation instruction. Total RNA samples were size-fractionated on a 15% tris-borate-EDTA-Urea polyacrylamide gel. RNA fragments of length 15–50 nt were isolated, quantified following gel elution, and ethanol precipitated. The SRA 5′ adapter (Illumina) was ligated to the aforementioned RNA fragments with T4 RNA ligase (Promega). The ligated RNAs were size-fractionated on a 15% tris-borate-EDTA-Urea polyacrylamide gel and RNA fragments of ~41–76 nt were isolated. The SRA 3′ adapter (Illumina) ligation was then performed followed by a second size-fractionation, as described above. RNA fragments of size ~64–99 nt were isolated through gel elution and ethanol precipitation. The ligated RNA fragments were reverse transcribed to single-stranded cDNAs using M-MLV (Invitrogen) with RT-primers recommended by Illumina. The cDNAs were amplified with pfx DNA polymerase (Invitrogen) in 20 cycles of PCR using Illumina's small RNA primer set. The resulting PCR products were purified on a 12% TBE polyacrylamide gel and a slice of gel between ~80 and 115 bps was excised. This fraction was eluted and the recovered cDNAs were precipitated and quantified on Nanodrop (Thermo Scientific) and on TBS-380 mini-fluorometer (Turner Biosystems) using Picogreen® dsDNA quantization reagent (Invitrogen). The concentrations of the samples were adjusted to ~10 nM and a total of 10 μL used in sequencing reactions. The purified cDNA libraries were used for cluster generation on Illumina's Cluster Station and then sequenced on IlluminaGAIIx following the vendor's instructions. Raw sequencing reads were obtained using Illumina's Pipeline v1.5 software following sequencing image analysis by Pipeline Firecrest Module and base-calling by Pipeline Bustard Module. The extracted sequencing reads were stored in separate files for each library and were then used in the following data analysis.

4.4. Reads filters and sequence analysis

After the raw sequence reads or sequenced sequences were extracted from image data into a raw data file, tags were counted and a unique family of sequences identified. In this step, the "impurity" sequences due to sample preparation, sequencing chemistry and processing, and the optical digital resolution of the sequencer detector were removed. After the low quality filters were applied, adaptor sequences were trimmed into a clean full length reads and formatted into a non-redundant FASTA format (using FASTQ Toolkit; http://hannonlab.cshl.edu/fastx_toolkit/download.html). Since each small RNA was represented by a unique sequence, the number of reads of this tag reflects relative expression level of that RNA. The FASTA headers were comprised of the running number along with the count number of the corresponding sequence. Only small RNA sequences of 15 to 30 nt were retained for further analysis.

4.5. Small RNA annotation

All unique sequence tags that passed the read filters were mapped onto the reference human genome (hg19) using Bowtie [44] software allowing at most one mismatch. Moreover, the unique sequences tags were aligned against miRbase v17 [45,46], computationally predicted human ncRNA and Rfam 9.1 (rfam.sanger.ac uk/), the Repeat Masker pre-build (hg19) annotation and human genes UCSC annotation hg19 (genome.ucsc.edu/) to generate tag classifications for known miRNA, degradation fragments of non-coding RNA, genomic repeats and mRNA respectively. To ensure every unique small RNA tag mapped to only one annotation class, we used the following priority rule: ncRNA (Rfam) > miRNA(miRBase) > Repeats (RepeatMasker) > mRNA (RefSeq) [47]. The fraction of total ribosomal RNA served as a quality control for each sample. Usually it should be less than 40% in animal samples of high quality. Tags that did not map to any of the above reference annotation sets or tags that could not be aligned to the genome were termed as unclassified.

4.6. Statistical analysis of differential miRNA expression

In order to determine differential miRNA expression between leiomyoma (samples B25L and B27L) and myometrium (samples B25N and B27N), expression data for known miRNAs produced by reads count was used as input for the Bioconductor DESeq package [29]. DESeq uses a negative binomial distribution model to test for differential expression in deep sequencing datasets. The list of differentially expressed miRNAs produced by DESeq was further filtered to remove miRNAs with less than 5 reads in all samples and retain those with significant differences $p<0.05$ and $|\log_2 \text{ratio}| \geq 1$.

4.7. Novel miRNA prediction

The novel miRNA prediction strategy is based on initial scanning of read libraries with publically available software for prediction of novel miRNAs from deep sequencing datasets — MIREAP (http://sourceforge.net/projects/mireap/) and miRDeep [48] using their default parameters. We further analyzed the output of the programs by in-house Perl script (unpublished) that annotates all newly found genomic loci and discards those that overlap with known miRNAs or known human ncRNA loci from Rfam 9.1 (rfam.sanger.ac uk/). Furthermore, we discarded predicted miRNAs that had less than 5 total mapped reads.

Supplementary materials related to this article can be found online at doi:10.1016/j.ygeno.2012.03.003.

Acknowledgments

This work was supported by EU FP7 (REGPOT, BIOSUPPORT), the Bulgarian National Science Fund (DOO2-071-2008) and the Department of Obstetrics and Gynecology, University of Colorado School of Medicine.

References

[1] D.P. Bartel, MicroRNAs: genomics, biogenesis, mechanism, and function, Cell 116 (2004) 281–297.
[2] M. Lagos-Quintana, R. Rauhut, W. Lendeckel, T. Tuschl, Identification of novel genes coding for small expressed RNAs, Science 294 (2001) 853–858.
[3] C.M. Hammell, The microRNA-argonaute complex: a platform for mRNA modulation, RNA Biol. 5 (2008) 123–127.
[4] Y. Lee, J. Han, K.H. Yeom, H. Jin, V.N. Kim, Drosha in primary microRNA processing, Cold Spring Harb. Symp. Quant. Biol. 71 (2006) 51–57.
[5] M. Tijsterman, R.H. Plasterk, Dicers at RISC; the mechanism of RNAi, Cell 117 (2004) 1–3.
[6] K. Okamura, A. Ishizuka, H. Siomi, M.C. Siomi, Distinct roles for Argonaute proteins in small RNA-directed RNA cleavage pathways, Genes Dev. 18 (2004) 1655–1666.
[7] I. Alvarez-Garcia, E.A. Miska, MicroRNA functions in animal development and human disease, Development 132 (2005) 4653–4662.
[8] J. Lu, G. Getz, E.A. Miska, E. Alvarez-Saavedra, J. Lamb, D. Peck, A. Sweet-Cordero, B.L. Ebert, R.H. Mak, A.A. Ferrando, J.R. Downing, T. Jacks, H.R. Horvitz, T.R. Golub, MicroRNA expression profiles classify human cancers, Nature 435 (2005) 834–838.
[9] S. Volinia, G.A. Calin, C.G. Liu, S. Ambs, A. Cimmino, F. Petrocca, R. Visone, M. Iorio, C. Roldo, M. Ferracin, R.L. Prueitt, N. Yanaihara, G. Lanza, A. Scarpa, A. Vecchione, M. Negrini, C.C. Harris, C.M. Croce, A microRNA expression signature of human solid tumors defines cancer gene targets, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 2257–2261.
[10] J. Stenvang, M. Lindow, S. Kauppinen, Targeting of microRNAs for therapeutics, Biochem. Soc. Trans. 36 (2008) 1197–1200.
[11] J. Stenvang, S. Kauppinen, MicroRNAs as targets for antisense-based therapeutics, Expert. Opin. Biol. Ther. 8 (2008) 59–81.
[12] E.E. Marsh, Z. Lin, P. Yin, M. Milad, D. Chakravarti, S.E. Bulun, Differential expression of microRNA species in human uterine leiomyoma versus normal myometrium, Fertil. Steril. 89 (2008) 1771–1776.
[13] Q. Pan, X. Luo, N. Chegini, Differential expression of microRNAs in myometrium and leiomyomas and regulation by ovarian steroids, J. Cell. Mol. Med. 12 (2008) 227–240.
[14] T. Wang, X. Zhang, L. Obijuru, J. Laser, V. Aris, P. Lee, K. Mittal, P. Soteropoulos, J.J. Wei, A micro-RNA signature associated with race, tumor size, and target gene activity in human uterine leiomyomas, Genes Chromosom. Cancer 46 (2007) 336–347.
[15] K. Cesen-Cummings, J.A. Copland, J.C. Barrett, C.L. Walker, B.J. Davis, Pregnancy, parturition, and prostaglandins: defining uterine leiomyomas, Environ. Health Perspect. 108 (Suppl. 5) (2000) 817–820.
[16] A.H. Ligon, C.C. Morton, Leiomyomata: heritability and cytogenetic studies, Hum. Reprod. Update 7 (2001) 8–14.
[17] M.K. Whiteman, E. Kuklina, D.J. Jamieson, S.D. Hillis, P.A. Marchbanks, Inpatient hospitalization for gynecologic disorders in the United States, Am. J. Obstet. Gynecol. 202 (541) (2010) e541–546.
[18] S. Okolo, Incidence, aetiology and epidemiology of uterine fibroids, Best Pract. Res. Clin. Obstet. Gynaecol. 22 (2008) 571–588.
[19] S.F. Cramer, A. Patel, The frequency of uterine leiomyomas, Am. J. Clin. Pathol. 94 (1990) 435–438.
[20] V.C. Buttram Jr., Uterine leiomyomata–aetiology, symptomatology and management, Prog. Clin. Biol. Res. 225 (1986) 275–296.
[21] C.J. Creighton, J.G. Reid, P.H. Gunaratne, Expression profiling of microRNAs by deep sequencing, Brief. Bioinform. 10 (2009) 490–497.
[22] A. Git, H. Dvinge, M. Salmon-Divon, M. Osborne, C. Kutter, J. Hadfield, P. Bertone, C. Caldas, Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression, RNA 16 (2010) 991–1006.
[23] P.A. t Hoen, Y. Ariyurek, H.H. Thygesen, E. Vreugdenhil, R.H. Vossen, R.X. de Menezes, J.M. Boer, G.J. van Ommen, J.T. den Dunnen, Deep sequencing-based expression analysis shows major advances in robustness, resolution and inter-lab portability over five microarray platforms, Nucleic Acids Res. 36 (2008) e141.
[24] N. Yang, L. Zhang, H.H. Kazazian Jr., L1 retrotransposon-mediated stable gene silencing, Nucleic Acids Res. 33 (2005) e57.
[25] P. Landgraf, M. Rusu, R. Sheridan, A. Sewer, N. Iovino, A. Aravin, S. Pfeffer, A. Rice, A.O. Kamphorst, M. Landthaler, C. Lin, N.D. Socci, L. Hermida, V. Fulci, S. Chiaretti, R. Foa, J. Schliwka, U. Fuchs, A. Novosel, R.U. Muller, B. Schermer, U. Bissels, J. Inman, Q. Phan, M. Chien, D.B. Weir, R. Choksi, G. De Vita, D. Frezzetti, H.I. Trompeter, V. Hornung, G. Teng, G. Hartmann, M. Palkovits, R. Di Lauro, P. Wernet, G. Macino, C.E. Rogler, J.W. Nagle, J. Ju, F.N. Papavasiliou, T. Benzing, P. Lichter, W. Tam, M.J. Brownstein, A. Bosio, A. Borkhardt, J.J. Russo, C. Sander, M. Zavolan, T. Tuschl, A mammalian microRNA expression atlas based on small RNA library sequencing, Cell 129 (2007) 1401–1414.
[26] R.D. Morin, M.D. O'Connor, M. Griffith, F. Kuchenbauer, A. Delaney, A.L. Prabhu, Y. Zhao, H. McDonald, T. Zeng, M. Hirst, C.J. Eaves, M.A. Marra, Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells, Genome Res. 18 (2008) 610–621.
[27] J.G. Ruby, C. Jan, C. Player, M.J. Axtell, W. Lee, C. Nusbaum, H. Ge, D.P. Bartel, Large-scale sequencing reveals 21U-RNAs and additional microRNAs and endogenous siRNAs in *C. elegans*, Cell 127 (2006) 1193–1207.
[28] L. Guo, Z. Lu, Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data, Comput. Biol. Chem. 34 (2010) 165–171.
[29] S. Anders, W. Huber, Differential expression analysis for sequence count data, Genome Biol. 11 (2010) R106.
[30] J. Piriyapongsa, I.K. Jordan, A family of human microRNA genes from miniature inverted-repeat transposable elements, PLoS One 2 (2007) e203.
[31] C.J. Creighton, A.L. Benham, H. Zhu, M.F. Khan, J.G. Reid, A.K. Nagaraja, M.D. Fountain, O. Dziadek, D. Han, L. Ma, J. Kim, S.M. Hawkins, M.L. Anderson, M.M. Matzuk, P.H. Gunaratne, Discovery of novel microRNAs in female reproductive tract using next generation sequencing, PLoS One 5 (2010) e9637.
[32] Y. Altuvia, P. Landgraf, G. Lithwick, N. Elefant, S. Pfeffer, A. Aravin, M.J. Brownstein, T. Tuschl, H. Margalit, Clustering and conservation patterns of human microRNAs, Nucleic Acids Res. 33 (2005) 2697–2706.
[33] Q. Pan, X. Luo, N. Chegini, microRNA 21: response to hormonal therapies and regulatory function in leiomyoma, transformed leiomyoma and leiomyosarcoma cells, Mol. Hum. Reprod. 16 (2010) 215–227.
[34] K. Schramedei, N. Morbt, G. Pfeifer, J. Lauter, M. Rosolowski, J.M. Tomm, M. von Bergen, F. Horn, K. Brocke-Heidrich, MicroRNA-21 targets tumor suppressor genes ANP32A and SMARCA4, Oncogene 30 (2011) 2975–2985.
[35] A.A. Arslan, L.I. Gold, K. Mittal, T.C. Suen, I. Belitskaya-Levy, M.S. Tang, P. Toniolo, Gene expression studies provide clues to the pathogenesis of uterine leiomyoma: new evidence and a systematic review, Hum. Reprod. 20 (2005) 852–863.
[36] J. Hao, S. Zhang, Y. Zhou, X. Hu, C. Shao, MicroRNA 483-3p suppresses the expression of DPC4/Smad4 in pancreatic cancer, FEBS Lett. 585 (2011) 207–213.

[37] E.J. Lee, G. Kong, S.H. Lee, S.B. Rho, C.S. Park, B.G. Kim, D.S. Bae, J.J. Kavanagh, J.H. Lee, Profiling of differentially expressed genes in human uterine leiomyomas, Int. J. Gynecol. Cancer 15 (2005) 146–154.
[38] A. Veronese, L. Lupini, J. Consiglio, R. Visone, M. Ferracin, F. Fornari, N. Zanesi, H. Alder, G. D'Elia, L. Gramantieri, L. Bolondi, G. Lanza, P. Querzoli, A. Angioni, C.M. Croce, M. Negrini, Oncogenic role of miR-483-3p at the IGF2/483 locus, Cancer Res. 70 (2010) 3140–3149.
[39] J. Snowdon, X. Zhang, T. Childs, V.A. Tron, H. Feilotter, The microRNA-200 family is upregulated in endometrial carcinoma, PLoS One 6 (2011) e22828.
[40] J.C. Brase, M. Johannes, T. Schlomm, M. Falth, A. Haese, T. Steuber, T. Beissbarth, R. Kuner, H. Sultmann, Circulating miRNAs are correlated with tumor progression in prostate cancer, Int. J. Cancer 128 (2011) 608–616.
[41] A. Li, N. Omura, S.M. Hong, A. Vincent, K. Walter, M. Griffith, M. Borges, M. Goggins, Pancreatic cancers epigenetically silence SIP1 and hypomethylate and overexpress miR-200a/200b in association with elevated circulating miR-200a and miR-200b levels, Cancer Res. 70 (2010) 5226–5237.
[42] S.K. Wyman, R.K. Parkin, P.S. Mitchell, B.R. Fritz, K. O'Briant, A.K. Godwin, N. Urban, C.W. Drescher, B.S. Knudsen, M. Tewari, Repertoire of microRNAs in epithelial ovarian cancer as determined by next generation sequencing of small RNA cDNA libraries, PLoS One 4 (2009) e5311.
[43] H. Zhang, S.B. Su, Q.M. Zhou, Y.Y. Lu, Differential expression profiles of microRNAs between breast cancer cells and mammary epithelial cells, Ai Zheng 28 (2009) 493–499.
[44] B. Langmead, C. Trapnell, M. Pop, S.L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.
[45] S. Griffiths-Jones, miRBase: microRNA sequences and annotation Chapter, Curr. Protoc. Bioinform. 12 (12) (2010) 19 11–10 Unit.
[46] S. Griffiths-Jones, miRBase: the microRNA sequence database, Meth. Mol Biol 342 (2006) 129–138.
[47] E. Zhu, F. Zhao, G. Xu, H. Hou, L. Zhou, X. Li, Z. Sun, J. Wu, mirTools: microRNA profiling and discovery based on high-throughput sequencing, Nucleic Acids Res. 38 (2010) W392–W397.
[48] M.R. Friedlander, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, N. Rajewsky, Discovering microRNAs from deep sequencing data using miRDeep, Nat. Biotechnol. 26 (2008) 407–415.