

ORIGINAL ARTICLE

MicroRNA expression signature of human sarcomas

S Subramanian¹, WO Lui¹, CH Lee^{1,2}, I Espinosa¹, TO Nielsen², MC Heinrich³, CL Corless⁴, AZ Fire^{1,5} and M van de Rijn¹¹Department of Pathology, Stanford University, Stanford, CA, USA; ²Genetic Pathology Evaluation Centre, University of British Columbia, Vancouver, Canada; ³Division of Hematology/Oncology, Oregon Health and Science University, Portland, OR, USA; ⁴Department of Pathology, Oregon Health and Science University, Portland, OR, USA and ⁵Department of Genetics, Stanford University, Stanford, CA, USA

MicroRNAs (miRNAs) are ~22 nucleotide-long noncoding RNAs involved in several biological processes including development, differentiation and proliferation. Recent studies suggest that knowledge of miRNA expression patterns in cancer may have substantial value for diagnostic and prognostic determinations as well as for eventual therapeutic intervention. We performed comprehensive analysis of miRNA expression profiles of 27 sarcomas, 5 normal smooth muscle and 2 normal skeletal muscle tissues using microarray technology and/or small RNA cloning approaches. The miRNA expression profiles are distinct among the tumor types as demonstrated by an unsupervised hierarchical clustering, and unique miRNA expression signatures were identified in each tumor class. Remarkably, the miRNA expression patterns suggested that two of the sarcomas had been misdiagnosed and this was confirmed by reevaluation of the tumors using histopathologic and molecular analyses. Using the cloning approach, we also identified 31 novel miRNAs or other small RNA effectors in the sarcomas and normal skeletal muscle tissues examined. Our data show that different histological types of sarcoma have distinct miRNA expression patterns, reflecting the apparent lineage and differentiation status of the tumors. The identification of unique miRNA signatures in each tumor type may indicate their role in tumorigenesis and may aid in diagnosis of soft tissue sarcomas.

Oncogene (2008) 27, 2015–2026; doi:10.1038/sj.onc.1210836; published online 8 October 2007

Keywords: miRNA; microarrays; sarcoma

Introduction

Sarcomas are a heterogeneous group of malignant mesenchymal tumors that can occur in a wide range of age groups. A large number of different diagnoses have been described within this tumor group and pathologic diagnosis can be challenging. Currently few markers

exist to help distinguish sarcoma subtypes, yet the recent advent of targeted drug therapies—as in the case of gastrointestinal stromal tumor (GIST) and dermatofibrosarcoma protuberans—makes accurate diagnosis imperative (Weiss and Goldblum, 2001).

MicroRNAs (miRNAs) are short, processed, RNA molecules ~22 nucleotides in length that can control gene function through mRNA degradation, translation inhibition or chromatin-based silencing mechanisms (Doench and Sharp, 2004). In humans, about 500 miRNAs have been discovered so far (miRBase, Release 9.1; <http://microRNA.sanger.ac.uk/sequences>) (Griffiths-Jones *et al.*, 2006). In 2004, Calin *et al.* (2004) reported that more than half of the then known miRNAs were located in fragile sites on chromosomes. DNA copy number abnormalities in these fragile sites can have a direct impact on the miRNA expression levels (Calin *et al.*, 2004). miRNAs can influence a wide variety of biological processes, including development, proliferation and differentiation (He and Hannon, 2004). Accumulating evidence suggests that miRNAs may act as either tumor suppressors or oncogenes that control growth and apoptosis (Esquela-Kerscher and Slack, 2006). In addition, a number of recent studies have highlighted the potential of miRNA profiles for diagnosis and prognosis of some epithelial tumors (Lu *et al.*, 2005) and hematological malignancies (Calin *et al.*, 2005). In this study we demonstrate that miRNA expression patterns correlate with known major sarcoma subtypes and can serve as a new tool in defining their biologic differences.

Results

Global miRNA expression profilings were performed on a series of 27 sarcomas, 5 normal smooth muscle and 2 normal skeletal muscle tissues using microarrays and individual molecule sequencing. The 27 sarcomas analysed represent seven different histological types (Table 1).

Distinct miRNA expression profiles in sarcomas

As a first step in the analysis, we asked whether the miRNA expression signatures of the sarcomas were molecularly distinct. The 87 miRNAs that met the filtering criteria were subjected to hierarchical clustering

Correspondence: Professor M van de Rijn, Department of Pathology, Stanford University Medical Center, 300 Pasteur Drive, L-235, Stanford, CA 94305, USA.

E-mail: mrjijn@stanford.edu

Received 10 May 2007; revised 10 August 2007; accepted 1 September 2007; published online 8 October 2007

Table 1 Summary of clinical, histopathologic and molecular data for the 28 sarcomas, 2 normal skeletal muscle and 5 normal smooth muscle tissues used in this study

Case no.	Diagnosis	Age (year)	Sex	Tumor status	Tumor site	Immunohistochemistry	Molecular confirmation
638	SS	NA	Male	NA	NA	NA	<i>SYT-SSX</i> ^d
TB-5321	SS	43	Female	LR	Head	CD99+, CAM5.2+	<i>SYT-SSX</i> ^d
TB-5363	SS	41	Female	P	Right arm	CK-, CD99-, DES-, CD34-, SMA+, S100-	<i>SYT-SSX</i> ^d
TB-2500	SS	32	Male	M	Chest (axilla primary)	CK+, EMA-, DES+	<i>SYT-SSX</i> ^d
3217	SS	29	Female	P	Chest wall/lung	NA	<i>SYT-SSX</i> ^d
3705	SS	32	Female	P	Left arm	NA	<i>SYT-SSX</i> ^d
1739	SS	NA	NA	NA	NA	NA	<i>SYT-SSX</i> ^d
4728	ERMS	4	Male	P	Pelvis	DES+, MYO+	NA
1433	ARMS ^a	50	Male	P	Left sinus	DES+, SMA+, CD99	<i>PAX3-FKHR</i> ^d
932	ARMS	9 (months)	Male	P	Right groin	DES+, CD99+	<i>PAX3-FKHR</i> ^d
110	ARMS	8	Female	P	Right groin	MYO+	<i>PAX3-FKHR</i> ^d
4814	PRMS	71	Male	P	Left thigh	DES+, SMA-, MG+, S100-, KER-	NA
4813	PRMS	84	Male	P	Right thigh	VIM+, DES+, MG-, SMA-, CAM5.2-, S100-	NA
TB-5918	DDLPS ^b	92	Male	P	Left colon	CD34+, CD117-, DOG1-	ND
TB-2798	LMS	49	Male	M	Lung (neck primary)	DES+, SMA+, EGFR+, CD117-, S100-, KI67 95%	NA
TB-5792	LMS	81	Male	LR	Thigh	DES+, SMA+	NA
516	LMS	68	Female	P	Retroperitoneum	SMA+, DES+	NA
4608	LMS	47	Female	P	Retroperitoneum	SMA+, DES+	NA
4612	LMS	66	Female	M	Lung (uterine primary)	SMA+, DES+	NA
TB-2384	LMS ^c	67	Female	P	Retroperitoneum	SMA+, DES+, CD34-, KER-	NA
TB-1081	GIST	42	Male	P	Stomach	CD117+, CD34-, DOG1 weak	NA
TB-1296	GIST	61	Female	P	Stomach	CD117+, CD34+, S100-, CK-	ND
335	GIST	41	Male	P	Rectum	CD117+, CD34+, S100-	<i>KIT</i> exon 11 mutant ^e
623	GIST	73	Male	P	Stomach	CD34+, S100-	<i>KIT</i> exon 11 mutant ^e
TB-265	GIST	65	Male	P	Small bowel	CD117+	ND
111	GIST	69	Female	P	Stomach	NA	<i>KIT</i> exon 11 mutant ^e
2000	GIST	70	Male	P	Stomach	NA	<i>KIT</i> exon 11 mutant ^e
TB-40	GIST	34	Male	LR	Pelvis	CD117+, CD34-, DOG1+	ND
4865	Normal SKM	51	Male	NA	NA	NA	NA
31	Normal SKM	41	Male	NA	Abdomen	NA	NA
5159	Normal SM	NA	Female	NA	Myometrium	NA	NA
5160	Normal SM	NA	Female	NA	Myometrium	NA	NA
5161	Normal SM	42	Female	NA	Small bowel	NA	NA
5162	Normal SM	NA	NA	NA	Stomach	NA	NA
5163	Normal SM	NA	Female	NA	Myometrium	NA	NA

Abbreviations: ARMS, alveolar rhabdomyosarcoma; DDLPS, dedifferentiated liposarcoma; ERMS, embryonal rhabdomyosarcoma; GIST, gastrointestinal stromal tumor; LMS, leiomyosarcoma; LR, local recurrence; M, metastasis; P, primary; PRMS, pleiomorphic rhabdomyosarcoma; SKM, skeletal muscle; SM, smooth muscle; SS, synovial sarcoma. ^aARMS case that has been previously diagnosed as ERMS. ^bDDLPS case that has been previously diagnosed as GIST. ^cLMS case analysed by northern blot only. Both PRMS had support for striated muscle differentiation by electron microscopy. ^dBy RT-PCR and sequencing. ^eBy sequencing alone. 'ND' and 'NA' denote not done and not available, respectively. Note that mutation results of case nos. 335, 623, 111 and 2000 have been published in Subramanian *et al.* (2004); cases 638 and 516 were included in Nielsen *et al.* (2002).

among the 27 sarcomas, 5 normal smooth muscles and 2 normal skeletal muscles in an unsupervised manner. The clustering algorithm grouped both miRNAs and samples into clusters based on overall similarity in miRNA expression pattern without prior knowledge of sample identity. Clustering based on the 87 miRNAs revealed substantive distinctions in overrepresented and underrepresented miRNAs among the tumors (Figure 1). As is evident from the dendrogram at the top of the cluster pattern, the sarcomas and normal muscles samples clustered into five main groups, whereby almost all synovial sarcoma (SS), rhabdomyosarcomas (RMS), leiomyosarcomas (LMS), normal smooth muscles (NSM) and GIST were grouped according to their diagnosis. Two exceptions occurred, with case 4728 (embryonal RMS; ERMS) being loosely related to SS and case 2798 (LMS) loosely related to the RMS branch. Notably, the two normal skeletal muscle tissues fell into the muscle-derived tumors cluster, that is, RMS. Likewise, the normal smooth muscle samples closely clustered with the LMS samples (branch C and D; Figure 1).

Of the 28 sarcomas used in this study, 5 have previously been analysed for mRNA expression profiling by gene microarray analysis, and another 7 are part of two ongoing studies. As part of those studies the diagnosis was confirmed by a variety of techniques including immunohistochemistry (IHC) and molecular diagnostic tests (Table 1). Remarkably, in the remaining 16 tumors that were not previously analysed for mRNA expression, the miRNA expression patterns suggested that two cases had been misdiagnosed. Case 1433 was initially diagnosed as ERMS based on its histology and reactivity for desmin, smooth muscle actin and CD99. Only a small sample was available for evaluation and showed significant crush artifact. As a result the typical alveolar growth pattern of alveolar RMS (ARMS) was not appreciated in the initial analysis. Upon profiling of the miRNA expression, this case clustered tightly with the other ARMS included in the study. Subsequent demonstration of the fusion transcript *PAX3-FKHR* by reverse transcription (RT)-PCR and sequencing in this sample confirms that this case is indeed an ARMS (Supplementary Figure I). The second misdiagnosed case 5918 was initially diagnosed as a GIST based on its origin in the wall of the colon and reactivity for CD34. However this case clustered away from the eight other GISTs and instead clustered loosely with pleomorphic RMS (PRMS) (Figure 1). To reevaluate the diagnosis, we analysed *KIT* and *DOG1* expression by immunostaining, and performed mutation analyses for exon 11 of the *KIT* gene. Both *KIT* and *DOG1* are highly expressed in the vast majority of GIST (West *et al.*, 2004; Espinosa *et al.*, 2007). Case 5918 did not have immunoreactivity for *KIT* or *DOG1* and no mutations were found in *KIT* exon 11, the most common mutation in GIST (Fletcher and Rubin, 2007). Histologic review of the tumor suggested a new diagnosis of dedifferentiated liposarcoma. This diagnosis was supported at the molecular level by FISH analysis showing amplification for *MDM2* and *CDK4*, which are abnormalities that are seen frequently

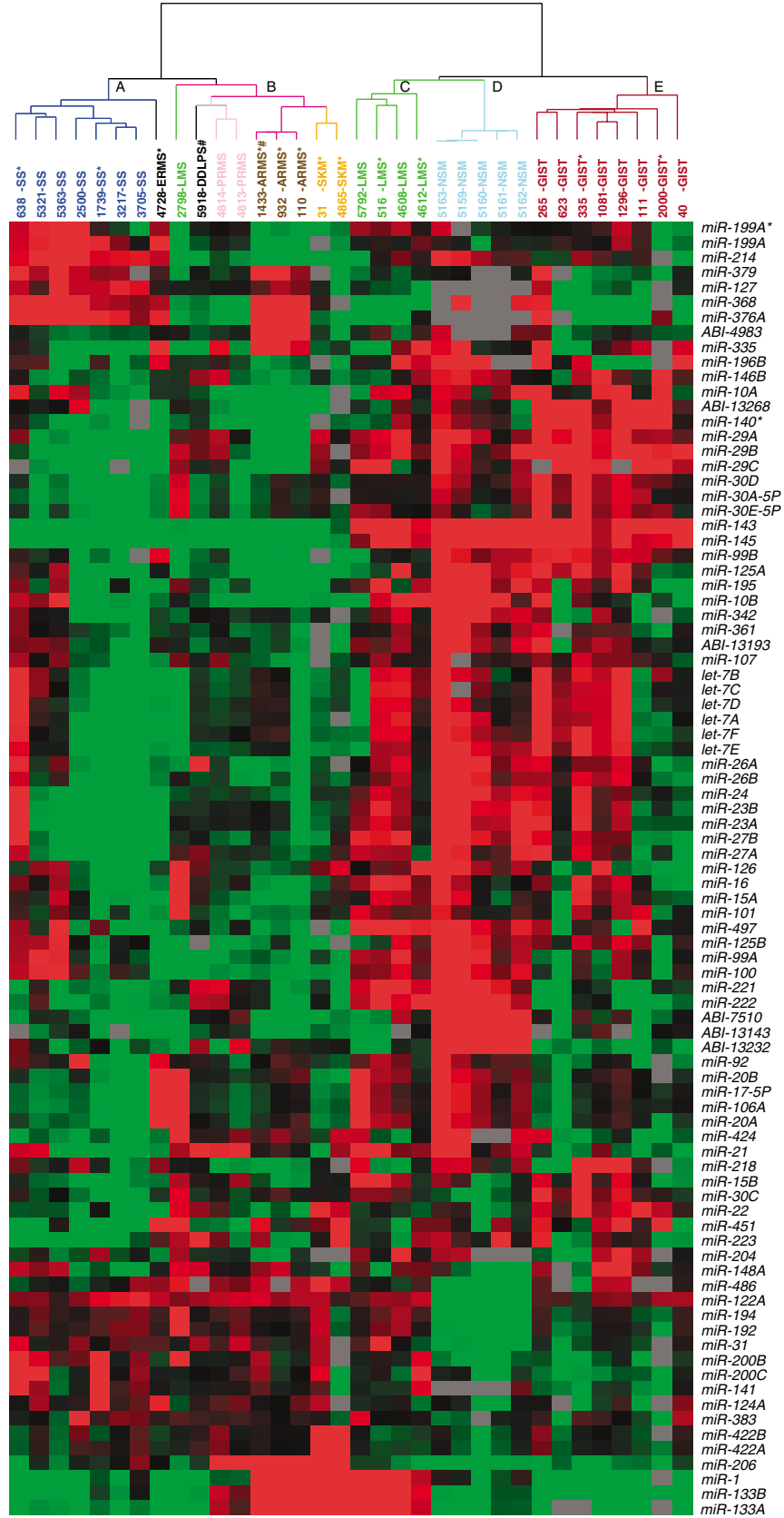
in well differentiated and dedifferentiated liposarcoma but rarely in GIST (Binh *et al.*, 2005) (Supplementary Figure I).

Tumor type-specific miRNA expression profiles

Using significance analysis of microarrays (SAM) analysis and permutation tests, we identified the miRNAs that correlated with each tumor class. The results for the top miRNAs of each class with <2% FDR (false discovery rate) are detailed in Table 2a–f. A total of 16 and 12 miRNAs showed significant relative overrepresentation in GIST and LMS, respectively compared to rest of the samples. In GISTs, 10 miRNAs showed relative down regulation. However, no miRNAs were found in LMS that showed relative underrepresentation. LMS has smooth muscle origin; uterine LMS originates from myometrium and nonuterine LMS probably has an origin from smooth muscle of big vessels. SAM analysis identified 10 overrepresented miRNAs and 11 underrepresented miRNA in LMS compared to the normal smooth muscle samples. For SS, 9 miRNAs showed relative overrepresentation and 28 miRNAs showed relative underrepresentation compared to the rest of the samples. Given that ARMS and PRMS are believed to be derived from skeletal muscle progenitor cells, their miRNA expression profiles were compared to that of the normal skeletal muscle in an attempt to identify potential etiologically relevant miRNAs. As only a single ERMS sample was studied in this study, it was excluded from the SAM analysis. Furthermore, only miRNAs with adequate measurements in both skeletal muscle samples were included in this series of SAM analysis. Several miRNAs were relatively underrepresented in both PRMS and ARMS compared to normal skeletal muscle: 3 in PRMS and 11 in ARMS. However, three miRNAs had significant higher expression in ARMS than in the normal skeletal muscle samples.

Identification of known and novel miRNAs by cloning

To validate the miRNA profiles determined by microarray and to search for novel candidate miRNAs or other small RNAs in human sarcomas and normal skeletal muscle tissues, we cloned and sequenced small RNA libraries from 10 sarcomas and 2 normal skeletal muscle tissues. A total of 8134 small RNA clones were sequenced, of which 6350 (78%) could be annotated as known miRNAs (Supplementary Table I). A total of 37 clones (0.5%) could be aligned to the human genome sequences but did not correspond to any annotated RNA species. Eighteen of these small RNA clones were designated as novel candidate miRNAs, based on a predicted hairpin precursor with the following properties (Ambros *et al.*, 2003) (Figure 2): (1) complete containment of the cDNA sequence within one arm of a hairpin, (2) at least 16 nucleotides of the cDNA sequence involved in base-pairing and (3) identification as the lowest free energy structure by mfold. Another five clones fulfilled criteria 1 and 3 but had somewhat fewer duplex base pairs in the hairpin region



(Supplementary Table II). The remaining aligned but nonannotated clones either showed poor secondary fold-back hairpin structure predicted by mfold or could not fulfill the criteria listed above (Supplementary Table III).

From 18 relevant clones, 14 novel candidate miRNAs were identified (Figure 2). Of these, *candidate-3* appears to be close homolog to the known mouse and rat *miR-543*. In addition, the sequence of two miRNAs is similar to two previously identified human miRNAs: *candidate-1* is similar to *miR-28*, and *candidate-12* is very similar to *miR-374*. Notably, *candidate-12* is located ~150 bp upstream of *miR-421*, and *candidate-14* is located in the miRNA cluster within the human imprinted 14q32 locus. All candidate miRNAs are conserved in mammals (including chimpanzee, mouse, rat and dog) but not in invertebrates. Interestingly, *candidate-13* is mapped to a bidirectional transcript pair, that is, sense strand of intron 22 of coagulation factor VIII (*F8*; NM_000132) and antisense of coding strand of coagulation factor VIII-associated protein (*F8A1*; NM_012151).

A comprehensive list of miRNA clones is detailed in Supplementary Table IV. Of 177 known miRNAs expressed in this sample group, some (for example, the *let-7* family) are highly represented in all samples analysed. Other miRNAs were found to have substantial expression variation among the tumor types. One striking example, *miR-143* was found to be highly abundant in GISTs (172/634 (27%) in case 335 and 156/973 (16%) in case 2000) and LMSs (122/606 (14%) in case 516 and 154/1129 (14%) in case 4612), while only 0–2 clones were found in the other sarcoma types and normal muscle samples that were analysed. Two miRNAs seem to be restricted to specific tumor type: *miR-200c* was only found in SS and *miR-140* and its antisense strand *miR-140** were only found in GISTs. In addition, known muscle-specific miRNAs (*miR-1*, *miR-133* and *miR-206*) were only identified in muscle-derived tumors and normal skeletal muscles. Notably, both *miR-1* and *miR-133a* were more abundant in normal skeletal muscles than the tumors.

Comparison of miRNA expression between microarray and cloning approaches

To determine the consistency of miRNA expression profiles obtained by microarray and cloning methods, we selected all samples that were analysed by both approaches and generated heat maps of the miRNA expression profiles based on the mean ratio or mean cloning frequency of each tumor types. As illustrated in Figure 3, the miRNA expression profiles are strikingly consistent between the two methods.

Differential expression of miR-143 in sarcomas

As *miR-143* expression varied significantly among the tumor types by both microarray and cloning, we further analysed its expression by northern analysis. Consistent with the microarray and cloning results, *miR-143* was highly expressed in the majority of LMS and all GIST, with little or no *miR-143* signal detected in RMS and SS, as well as the case of dedifferentiated liposarcoma (DDLPS) (Figure 4). Notably, expression of *miR-143* was barely detected in two LMS. One of these (case 2798) was loosely associated with the skeletal muscle/RMS cluster by clustering (Figure 1) and the other case (2384) was only studied by northern blot and not included in the microarray or cloning analyses.

Discussion

Using a microarray approach, we characterized miRNA expression profiles in a series of 27 sarcomas from 7 different histological types. We identified four major groups based on the miRNA expression patterns by clustering. Three groups consisted predominantly of the same tumor type, that is, SS, GIST and LMS. It is known that the oncogenesis and cellular origin of these sarcoma types are different. SS is characterized by a t(X;18) translocation that leads to the formation of a fusion oncoprotein SYT–SSX, which is believed to underlie its transformed phenotype (Pretto *et al.*, 2006). Most GISTs are driven by an activating mutation in *KIT* or *PDGFRA* (Fletcher and Rubin, 2007), while no unifying molecular event has been uncovered for LMS. Phenotypically, there is evidence to indicate that GIST is derived from interstitial cells of Cajal in the myenteric plexus, while LMS are thought to arise from smooth muscle cells in the uterus or from extra-uterine sites. In contrast, SS belongs to a class of sarcomas for which the cell of origin is unknown; biphasic mesenchymal and epithelial differentiation is common in these tumors (Weiss and Goldblum, 2001). It is therefore plausible that the distinct miRNA expression patterns observed among these sarcoma types are likely reflective of the differences in cell lineage, differentiation and/or oncogenic pathways.

The fourth expression cluster consisted predominantly of normal skeletal muscle tissues and muscle-related sarcomas (ARMS and PRMS) showing histologic, ultrastructural and/or immunophenotypic evidence of skeletal muscle differentiation. Within this cluster, ARMS, PRMS and normal skeletal muscles are separated into their individual subclusters, indicating that each of them possesses a miRNA expression

Figure 1 Unsupervised hierarchical clustering analysis of 27 sarcomas, 5 normal smooth muscles and 2 normal skeletal muscle tissues. Each row represents the relative levels of expression for a single miRNA and each column shows the expression levels for a single sample. The red or green color indicates relatively high or low expression, respectively, while gray indicates absent data points. The five main groups of the dendrogram labeled as A (synovial sarcoma; SS), B (rhabdomyosarcomas; RMS), C (leiomyosarcomas; LMS), D (normal smooth muscles; NSM) and E (gastrointestinal stromal tumors; GIST) predominantly separated the tumors and normal muscle samples. Asterisk (*) refers to those samples included in small RNA cloning. Note that case 1433 and 5918 were initially misdiagnosed as embryonal RMS (ERMS) and GIST respectively, as indicated by '#'. For miRNA species, * represents the antisense strand.

Table 2 SAM results for the top miRNAs of each class with <2% FDR

Gene ID	Score (d)	Fold change	q-value (%)
<i>(a) GIST vs rest of the samples</i>			
<i>miR-140*</i>	3.610	9.417	0.00
<i>ABI-13268</i>	3.248	7.327	0.00
<i>miR-29C</i>	2.661	4.297	0.00
<i>miR-29B</i>	2.462	3.592	0.00
<i>miR-22</i>	2.073	2.686	0.00
<i>miR-30A-5P</i>	2.021	3.123	0.00
<i>miR-30D</i>	1.940	2.856	0.00
<i>miR-99B</i>	1.924	2.300	0.00
<i>miR-30E-5P</i>	1.883	2.737	0.00
<i>miR-143</i>	1.794	0.570	0.00
<i>miR-29A</i>	1.741	2.039	0.00
<i>miR-30C</i>	1.647	2.516	0.00
<i>miR-145</i>	1.598	0.415	0.00
<i>miR-125A</i>	1.553	2.732	0.00
<i>let-7B</i>	1.392	2.131	0.00
<i>miR-10A</i>	1.294	2.099	1.92
<i>Underrepresented miRNAs</i>			
<i>miR-368</i>	-2.115	0.206	0.00
<i>miR-133B</i>	-2.016	0.017	0.00
<i>miR-1</i>	-1.783	0.033	0.00
<i>miR-376A</i>	-1.778	0.279	0.00
<i>miR-133A</i>	-1.664	0.020	0.00
<i>miR-200B</i>	-1.573	0.206	1.92
<i>miR-221</i>	-1.544	0.157	1.92
<i>ABI-13232</i>	-1.524	0.062	1.92
<i>miR-222</i>	-1.497	0.134	1.92
<i>miR-92</i>	-1.484	0.385	1.92
<i>(b) LMS vs NSM</i>			
<i>miR-122A</i>	9.874	93.054	0
<i>miR-194</i>	4.193	7.267	0
<i>miR-133B</i>	4.035	14.398	0
<i>miR-133A</i>	3.987	15.071	0
<i>miR-192</i>	3.286	5.693	0
<i>miR-1</i>	3.175	9.662	0
<i>miR-486</i>	2.602	4.469	0
<i>miR-31</i>	2.148	10.229	0
<i>miR-200B</i>	1.994	3.809	0
<i>miR-30C</i>	1.777	2.750	0
<i>Underrepresented miRNAs</i>			
<i>ABI-13232</i>	-6.366	0.018	0
<i>ABI-13143</i>	-5.852	0.018	0
<i>ABI-7510</i>	-5.669	0.031	0
<i>miR-145</i>	-5.552	0.057	0
<i>miR-368</i>	-4.486	0.084	0
<i>miR-143</i>	-3.313	0.103	0
<i>miR-99B</i>	-2.446	0.240	0
<i>miR-376A</i>	-2.389	0.231	0
<i>miR-127</i>	-2.107	0.303	0
<i>miR-342</i>	-1.908	0.275	0
<i>miR-24</i>	-1.673	0.329	2
<i>(c) LMS vs rest of samples</i>			
<i>miR-20A</i>	1.737	3.756	0
<i>miR-16</i>	1.733	3.166	0
<i>miR-15B</i>	1.614	3.043	0
<i>miR-17-5P</i>	1.505	2.852	0
<i>miR-194</i>	1.492	5.617	0
<i>miR-106A</i>	1.474	2.581	0
<i>miR-101</i>	1.440	2.392	0
<i>miR-192</i>	1.420	4.982	0
<i>miR-221</i>	1.403	1.753	0
<i>miR-15A</i>	1.385	2.544	0
<i>miR-20B</i>	1.358	2.598	0
<i>miR-204</i>	1.333	4.298	0
<i>(d) SS vs rest of samples</i>			
<i>miR-376A</i>	2.776	5.175	0
<i>miR-214</i>	2.537	3.915	0

Table 2 (continued)

Gene ID	Score (d)	Fold change	q-value (%)
<i>miR-127</i>	2.493	3.903	0
<i>miR-368</i>	2.307	3.993	0
<i>miR-199A</i>	2.189	3.314	0
<i>miR-200B</i>	2.124	5.152	0
<i>miR-200C</i>	2.013	6.756	0
<i>miR-379</i>	1.873	2.660	0.91
<i>miR-141</i>	1.659	3.339	1.52
<i>Underrepresented miRNAs</i>			
<i>miR-29B</i>	-3.029	0.070	0
<i>miR-143</i>	-2.783	0.007	0
<i>miR-145</i>	-2.676	0.007	0
<i>miR-29A</i>	-2.565	0.163	0
<i>miR-30D</i>	-2.500	0.216	0
<i>miR-30A-5P</i>	-2.328	0.216	0
<i>miR-29C</i>	-2.273	0.144	0
<i>miR-451</i>	-2.261	0.103	0
<i>miR-23B</i>	-2.235	0.379	0
<i>miR-23A</i>	-2.234	0.356	0
<i>miR-30E-5P</i>	-2.021	0.290	0
<i>miR-17-5P</i>	-1.962	0.235	0
<i>miR-106A</i>	-1.960	0.237	0
<i>miR-27B</i>	-1.958	0.320	0
<i>miR-27A</i>	-1.878	0.365	0
<i>miR-24</i>	-1.821	0.414	0
<i>miR-20A</i>	-1.771	0.223	0
<i>miR-21</i>	-1.758	0.393	0
<i>Let-7F</i>	-1.737	0.496	0
<i>miR-223</i>	-1.726	0.169	0
<i>miR-20B</i>	-1.664	0.266	0
<i>miR-335</i>	-1.582	0.236	0
<i>miR-424</i>	-1.559	0.212	0
<i>miR-146B</i>	-1.474	0.349	0
<i>miR-15B</i>	-1.452	0.260	0
<i>miR-16</i>	-1.413	0.464	0
<i>miR-126</i>	-1.385	0.430	0
<i>miR-222</i>	-1.345	0.154	0
<i>(e) PRMS vs SKM</i>			
<i>Underrepresented miRNAs</i>			
<i>miR-1</i>	-17.408	0.007	0
<i>miR-133B</i>	-6.718	0.014	0
<i>miR-133A</i>	-6.416	0.012	0
<i>(f) ARMS vs SKM</i>			
<i>miR-376A</i>	5.009	24.270	0
<i>ABI-4983</i>	4.185	19.636	0
<i>miR-335</i>	4.000	6.682	0
<i>Underrepresented miRNAs</i>			
<i>miR-422B</i>	-4.941	0.157	0
<i>miR-1</i>	-4.930	0.090	0
<i>miR-29A</i>	-4.489	0.067	0
<i>miR-22</i>	-4.382	0.144	0
<i>miR-133B</i>	-4.217	0.083	0
<i>miR-133A</i>	-3.930	0.081	0
<i>ABI-13143</i>	-3.458	0.361	0
<i>miR-126</i>	-3.403	0.258	0
<i>miR-422A</i>	-3.396	0.131	0
<i>miR-15A</i>	-3.298	0.315	0
<i>miR-29C</i>	-3.039	0.158	0

Abbreviations: ARMS, alveolar rhabdomyosarcoma; GIST, gastrointestinal stromal tumor; LMS, leiomyosarcoma; NSM, normal smooth muscle; PRMS, pleiomorphic rhabdomyosarcoma; SKM, skeletal muscle; SS, synovial sarcoma.

candidate miRNA	sequence (5'->3') ^a	sample (no. of clones)	size (nt)	stem-loop structures of putative miRNA precursors ^b	dG	location	Gene/intergenic ^c exon/intron/UTR (S/AS)
candidate-1 miR - 708	AAGGAGCUUACAAUCUAGCUGG	4728 (2) 1739 (1)	17-22	<pre> aa A A C Ggg aa acu ggu cugccucc AGGAGCUUACA UCUAG UG gu aug \ cca gacggggag ucuucgagugu agauc ac ca uac u gg a c a a-- ag acg </pre>	-47.7	11q14.1	intergenic
candidate-2 miR - 1178	UUGUCACAGUUCUUCUCCUAG	4728 (1)	21	<pre> g gagga uccaggggu c u c gcguug cuggca agggaaaggg cag ugagca gcc u uguagu gacugu UCCUUCUUU GUC ACUCGU Ugg c g aaGA- ----- - - a </pre>	-37.9	12q24.23	<i>CTT</i> , intron (S)
candidate-3 miR - 543	AUUCGCGGUCACUUCUU	1739 (1)	18	<pre> gc--- u u u - u--- u uuu gguacu aa gagaagu gc ccgug uuuu uegc \ cuaua uu UUCUACA CG GCGCG AAAG agug a uuauc c u - U UUA C c uuu </pre>	-30.3	14q32.31	intergenic
candidate-4 miR - 1179	AAGCAUUCUUCUUCUUGG	4728 (1)	21	<pre> ----- a -- C Gu ugua ggcugg aaagga gAA GCAUUCUUU AUUGGUUG g u ucgauc uuuccu uuu cguaggaga uaaccaac u u agucu a ac a ug uccg </pre>	-34.5	15q26.1	intergenic
candidate-5 miR - 1180	UUUCGCGUCGCGUGGGUGUGU	4728 (1) 110 (1)	20-22	<pre> --- c - gu uc gcugcugg acccac cg gccgggaaua gc \ cggcggc<u>U</u> UGGGUG GC CGGCCUU<u>U</u>gu ug c GUG C U -- gu </pre>	-57.8	17p11.2	<i>EPPB9</i> , intron (S)
candidate-6 miR - 1181	CCGUCGCCACCCGAGCCG	4728 (1)	21	<pre> - ---- -- --- CA G Gg uccac ugcug ccgCCGUC GCC GC CCC AGCC a gggug gcgac ggugguag cgg cg ggg ucgg g u auca aa aac cc - gc </pre>	-42.8	19p13.2	<i>CDC37</i> , 5'UTR (S)
candidate-7 miR - 935	CAGUACCGUCCGCUACCG	932 (1)	21	<pre> gcg a ccc c caucc ggcgggggc ggcggc guggcgggagcgg cu gcc u ucgcccucg ccgccG CAUCGCCUUCGCC GA ccg c --- C AUU C ucugc </pre>	-61.8	19q13.42	<i>CACNG8</i> , exon (S)
candidate-8 miR - 1182	GAGGUCUUGGGAGGAGUGAC	4728 (1)	23	<pre> a g cu cu ucc u cagc gga gu ggg cuu uca gc guc ucccuc ccag gacu uuucgga \ ccc gaa ggu ug CAG AGGGAG GGUU CUGG GAGauca c - g -- u- UGU - --- --- ac </pre>	-38.2	1q42.2	<i>FAM89A</i> , 3'UTR (S)
candidate-9 miR - 885-5p	UCCAUAACACUACCCUGCCUC	31 (1)	21	<pre> ca c A - cu ggcccg cuu UCCAUAACACU CC CUGCCU<u>u</u> c uuuggc gaga agguagugug gg gacggagag c ac u - c ua </pre>	-50.2	3p25.3	<i>ATP2B2</i> , intron (S)
candidate-10 miR - 874	CUGCCUGGCCGAGGACCGAC	1739 (1)	23	<pre> u g cug c a ca a a- ac gu ggu uuagcc cgg ccc cg ccagggua gag g \ cg ccg ggucgg GCC GGG GC GGUCCGU cuu c u u - uCA A A CC C cg uc </pre>	-41.6	5q31.2	<i>KLHL3</i> , intron (S)
candidate-11 miR - 1183	CACUGUAGGUGAUGGAGAGUGGGCA	1739 (1)	27	<pre> - a auuuuuucaa gaga aga- a g aa a gg auucc gcu augcucg cac ac uua ag gac g cc ugagg cga uACGGGU GUG UG GAU UC uug g g a gg----- GAGA GUAG - G AC a </pre>	-23.3	7p15.3	<i>SP4</i> , intron (S)
candidate-12 miR - 374b	AUAUAAUACAACCCUGCUAAGUG	1739 (1)	22	<pre> ug AU c cgga g AUAUACAACCCUGCUAAGUG c gccu u uuuuuuuuaggacgauuacag u gu ac a </pre>	-44.7	Xq13.2	intergenic
candidate-13 miR - 1184	CCUGCAGCGACUUGAUGGCUCC	110 (1)	23	<pre> ug a c ugaa g c--- ca ucaaca cca cu c gaa gagg ggaggu guu ugcu gcag gugg c gg g cuu uuucc CCUUGG UAG GCGA CGUC Cacc a gu - - ugug G UUCA -- ----- ucu </pre>	-30.2	Xq28	<i>F8</i> , intron (S) <i>F8A1</i> , exon (AS)
candidate-14 miR - 1185-1	AGAGGAUACCCUUGUAGUU	638 (2)	21	<pre> u - GAUA a uga uuugguac uga gAGAG CCCUUUGUAGUU c cu u aaacuaug guuu uuuc gggggacauuaag gg u c a aga- ----- c uaa </pre>	-35.3	14q32.31	intergenic
miR - 1185-2				<pre> u - GAUA a uga uuugguac uaaa gAGAG CCCUUUGUAGUU c cu u aaacuaug guuu cucuc gggggacauuaag gg u c a aga- ----- c uaa </pre>	-39.1	14q32.31	intergenic

^aSequences listed represent the observed full length sequence of each miRNA cloned. The extra bases at the termini of some miRNAs are denoted in bold.
^bRNA secondary structure prediction was performed using mfold version 3.2. The miRNA sequence is in uppercase and red. The actual size of the stem-loop has not been experimentally determined.
^cUTR, untranslated region; S, sense; AS, antisense.

Figure 2 Novel candidate miRNAs identified from human sarcomas and normal skeletal muscle tissues.

signature that is measurably different from the others. Correspondingly, ARMS and PRMS differ in their oncogenic origin. ARMS is characterized by a *PAX3*-

FKHR or a *PAX7-FKHR* fusion, while no recurrent molecular event has been found for PRMS or ERMS (Weiss and Goldblum, 2001).

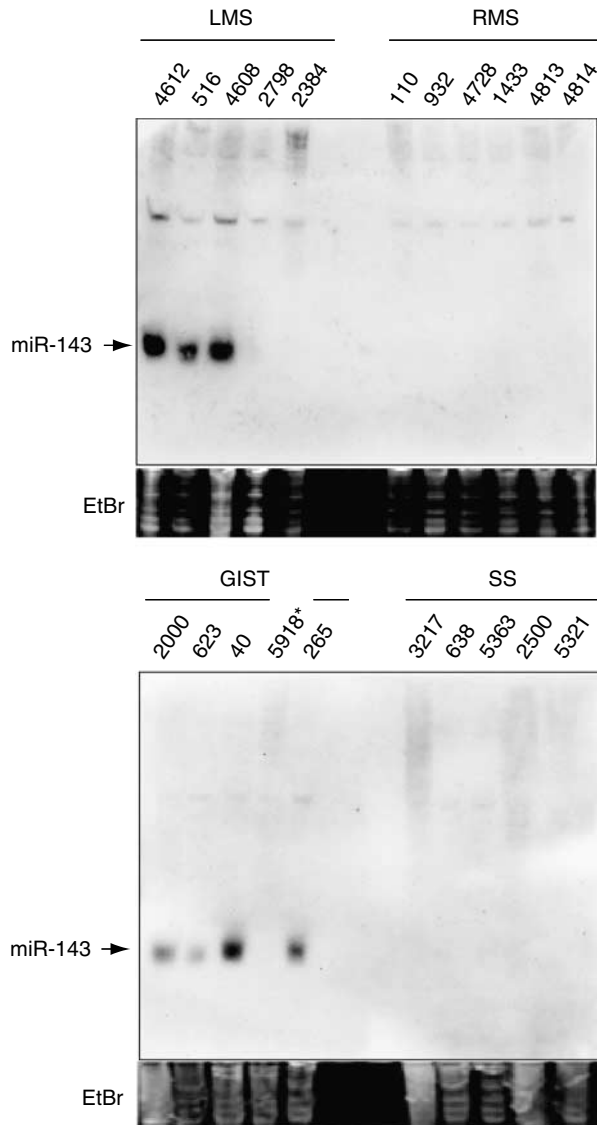


Figure 4 Northern analysis of *miR-143* expression. Mature *miR-143* is noticed in all gastrointestinal stromal tumors (GISTs) and most leiomyosarcomas (LMS) except for cases 2798 and 2384, but absent/barely detected in all synovial sarcomas (SS) and rhabdomyosarcomas (RMS) cases. Ethidium bromide-stained rRNA bands are shown as loading controls. Asterisk (*) refers to the case, which was initially diagnosed as GIST and reclassified as dedifferentiated liposarcoma (DDLPS).

Leiomyosarcoma

SAM comparison of LMS and normal smooth muscle samples identified significant overrepresentation of *miR-1*, *miR-133A* and *miR-133B* in LMS. These miRNAs play a major role in myogenesis and myoblast proliferation (Chen *et al.*, 2006). Both LMS and NSM showed underrepresentation of *miR-206*, an miRNA that is highly expressed in skeletal muscles and implicated in myogenic differentiation (Kim *et al.*, 2006; Politz *et al.*, 2006). Subclassification of LMS has been a challenge. Interestingly, *miR-143* was expressed at low levels in 2 of 6 LMS (as seen in case 2798 by miRNA profiling and northern blot analysis and in case

2384 by northern blot analysis alone). Case 2798 also separated from those LMS with high *miR-143* expression by clustering, suggesting that LMS is molecularly heterogeneous.

Synovial sarcoma

In SS, *miR-143* was expressed at very low levels relative to GIST and LMS, as demonstrated by microarray, cloning and northern analyses. Similarly, *miR-143* is also reduced in expression in several cancer types, including colorectal neoplasia (Michael *et al.*, 2003) and cervical cancer (Lui *et al.*, 2007). *ERK5* (also known as *MAPK7*) is the only experimentally verified target for *miR-143*, which is known to promote cell growth and proliferation in response to tyrosine kinase signaling (Esau *et al.*, 2004), however, its role in sarcomagenesis remains unclear. *SSX1*, a common 3'-fusion partner gene resulting from a t(X;18) in SS (Weiss and Goldbum, 2001), is predicted to be a target for *miR-143* by miRBase (<http://microrna.sanger.ac.uk/targets/v4/>) and Target Scan 3.1 (<http://www.targetscan.org/>). Since miRNAs target the 3'-UTR region of mRNA transcripts, it is tempting to speculate that underrepresentation of *miR-143* in SS tumor cells enables the production of the SYT-SSX1 oncoprotein.

Rhabdomyosarcoma

SAM analysis comparing PRMS to normal skeletal muscle revealed that two of the known muscle-specific miRNAs (*miR-1* and *miR-133*) are relatively underexpressed in PRMS. Several myogenic genes are regulated by these two miRNAs (Chen *et al.*, 2006). For ARMS, *miR-335* seems to be of particular interest as it is overrepresented miRNA compared to normal skeletal muscle and notably, *miR-335* resides in the intron 2 of *MEST* (also known as *PEG1*). *MEST* has been indicated to play a role in muscle differentiation (Yan *et al.*, 2003), and its mRNA expression is high in ARMS (Baer *et al.*, 2004). *MEST* is a downstream target of *PAX3*, the gene involved in the *PAX3-FKHR* fusion that is typical for alveolar rhabdomyosarcomas (Mayanil *et al.*, 2001). It thus appears that the *PAX3-FKHR* fusion may influence the transcription of *miR-335* that has several predicted targets. These predicted targets for *miR-335* (by Target Scan 3.1; <http://www.targetscan.org>) include *CHFR*, which is lost in many tumors, and *HAND1* and *SP1*, which function in mesoderm or muscle differentiation (Supplementary Table V).

Novel candidate miRNAs

Using a cloning approach, we discovered a significant number of novel miRNAs and other small RNAs, indicating that some of these small RNAs are unique to a specific tumor type or are expressed at levels, that escaped detection in previous cloning experiments. Although one might expect some novel miRNAs to be expressed specifically in one tumor type, it is conceivable that others would be expressed in a range of tumor types. For example, *candidate-1*, *candidate-7*, *candidate-10* and *candidate-12* were also found in cervical cancer

samples (Lui *et al.*, 2007); and *candidate-5* was also observed in colon cancer cell line (WOL, unpublished data). Analysis for conservation of fold-back structure revealed that all the 14 novel candidate miRNAs were conserved in mammals including chimpanzee, dog, mouse and rat. Together with their structural characteristics they fulfill the criteria for novel miRNA species as described by Kim (Kim, 2005). The final distribution profile of these miRNAs awaits the testing of many more samples. Novel small RNAs that do not fulfill current miRNA criteria were identified and these could represent: (1) miRNAs that fail to meet the arbitrary miRNA criteria, (2) miRNA-like molecules formed by slightly divergent synthetic mechanisms and (3) other small RNAs such as natural siRNAs that might be formed by completely different mechanisms. Alternatively, these could also represent spurious single-stranded RNA transcripts or common degradation products of longer cellular RNAs.

Our dual approach to miRNA profiling has provided a comprehensive analysis of miRNA expression patterns in different histological types of soft tissue sarcomas. The miRNA expression signatures are clearly distinct among the tumor types studied, implicating their role in tumorigenesis in these tumors and their potential as diagnostic markers or even therapeutic targets.

Materials and methods

Clinical specimens

Tissue samples were as follows: eight GIST, seven SS, six LMS, one DDLPS, six RMS, five NSM and two normal skeletal muscle samples. The RMS included three different histological subtypes, that is, three ARMS, two PRMS and one ERMS. The fresh frozen tissues were collected from surgical specimens at Stanford University Medical Center and Vancouver General Hospital. All cases were centrally reviewed and the diagnoses were further confirmed by ancillary IHC and molecular studies (Table 1). The study was approved by the local ethical committees of Stanford University and the British Columbia Cancer Agency.

miRNA array printing

miRNA microarrays used in the study were printed at the Stanford functional genomics facility (www.microarray.org). The arrays contained a total of 668 probes spotted in duplicate. The 668 probes represent 328 known human miRNAs, 113 mouse miRNAs, 45 rat miRNAs, 154 predicted human miRNAs and 28 control probes (Ambion, Austin, TX, USA). The complete list of probes is given in Supplementary Table VI.

RNA isolation and miRNA array experiments

Total RNAs were extracted from frozen tumor samples using mirVana miRNA isolation kit (Ambion). Reference RNA (XpressRef Universal Total RNA) was obtained from Super-Array (Frederick, MD, USA). miRNA was further enriched from 25 µg of total RNA using a microcon YM-100 column (Millipore, Billerica, MA, USA), and indirectly labeled with Cy3 or Cy5 amine reactive dyes (Amersham Biosciences, Buckinghamshire, UK) using mirVana miRNA labeling kit (Ambion). Hybridization was at 42 °C for 12–16 h. Arrays

were washed and immediately scanned using a GenePix 4000B array scanner (Axon Instruments, Foster City, CA, USA).

Microarray data analysis

Fluorescence ratios (sample/reference) were calculated using GenePix software. miRNA arrays were normalized and data was uploaded to Stanford Microarray Database (<http://genome-www5.stanford.edu/>). To limit the measurement errors, only miRNA spots with a ratio of signal over background of at least 2.5 in either Cy3 or Cy5 channels were included. Further, miRNA spots were filtered based on those where expression levels differed by at least fourfold in at least three arrays. Finally miRNA spots with >80% good data were selected. A total of 87 miRNAs passed the filtering criteria and were used for further analysis. Unsupervised hierarchical clustering analysis and SAM were then performed as described previously (Nielsen *et al.*, 2002).

Small RNA isolation and cloning

Small RNA extracted using mirVana miRNA isolation kit (Ambion) was used as starting material for cloning procedure of Lau *et al.* (2001) with slight modifications. Purified small RNAs were ligated with pre-adenylated 3'-adaptor oligonucleotide, gel purified and subjected to a second ligase reaction with a 5'-adaptor oligonucleotide. The gel-purified, doubly ligated RNA was reverse transcribed using Superscript II (Invitrogen, Carlsbad, CA, USA) and RT primer, followed by PCR amplification using the RT primer and a forward primer. A second PCR was performed using the RT primer and a second forward primer (Supplementary Table VII). The PCR product was purified by phenol/chloroform extractions and then digested with Ban I (NEB, Beverly, MA, USA) for concatemerization using T4 DNA ligase (NEB). Concatamers ranging from 600 to 1000 bp were isolated from a low-melting-point agarose gel, processed with Taq polymerase, and cloned into the pCR4-TOPO vector using the TOPO TA cloning kit (Invitrogen). Colony PCR was performed using the M13 forward and reverse primers, and the PCR products were prepared for sequencing using shrimp alkaline phosphatase and exonuclease I (USB Corporation, Cleveland, OH, USA).

Sequence analysis

Small RNAs obtained by cloning were compared with functionally annotated sequences using BLAST ([blastn](http://blastn.ncbi.nlm.nih.gov/), <http://www.ncbi.nlm.nih.gov/blast/>), BLAT (<http://genome.ucsc.edu>), miRBase Release 9.1 (<http://microrna.sanger.ac.uk/sequences/search.shtml>) and simple text searches. For each cloned small RNA, the best alignments to a functionally annotated sequence (not more than one error) were used to assign a functional category to the small RNA. Putative novel small RNAs were analysed using mfold version 3.2 (Zuker, 2003) to identify potential precursor structures.

Northern analysis

Total RNA (20 µg) from snap-frozen tissues was fractionated on a denaturing 15% polyacrylamide gel. The gels were then transferred to Hybond-N+ membranes (Amersham Biosciences), fixed by ultraviolet cross-linking at 1200 µJ and baked (80 °C) for 1 h. Membranes were then hybridized overnight at 55 °C in PerfectHyb Plus hybridization buffer (Sigma, St Louis, MO, USA), together with a locked nucleic acid (LNA)-modified oligonucleotide probe complementary to the mature *miR-143* (Supplementary Table VII) that was labeled with terminal transferase (NEB) and biotin-16-dUTP (Roche Diagnostics, Indianapolis, IL, USA). Subsequently, the blots were washed at 55 °C for 15 min each in 2 × standard saline

citrate/0.1% SDS and $0.2 \times$ SSC/0.1% SDS. After washes, the blots were incubated in blocking solution ($1 \times$ Phosphate buffered saline, pH 7.4/0.05% Tween 20/0.1% SDS/0.5% blocking reagent, Roche Diagnostics) for 1 h and then in streptavidin–alkaline phosphatase conjugate (USB Corporation) for 1 h, followed by three washes each in buffer A ($1 \times$ PBS, pH 7.4/0.05% Tween 20/0.1% SDS) and buffer B (0.1 M Tris–HCl, pH 9.5, 0.1 M NaCl) at room temperature. The blots were then incubated with chemiluminescent substrate CDP-Star (GE Healthcare, Piscataway, NJ, USA) and exposed to Kodak BioMax XAR film (New Haven, CT, USA).

Immunohistochemistry

Sections were cut at 4 μ m, deparaffinized in xylene and hydrated in a graded ethanol series. The primary antibodies used were DOG1 (mouse monoclonal, 1/50; clone DOG1.1 unpublished data), CD117 (rabbit polyclonal, 1/200; Dako, Carpinteria, CA, USA), and CD34 (mouse monoclonal, 1/80; clone 581/CD34, BD Biosciences, San Jose, CA, USA). The antigen retrieval and IHC were performed as previously described (West *et al.*, 2004).

RT–PCR detection of fusion transcript

Total RNA (1 μ g) was reverse transcribed with thermoscript II reverse transcriptase and random hexamers. The resulting cDNA

was subjected to PCR amplification using specific primers for *SYT–SSX* and *PAX–FKHR*, as detailed in Supplementary Table VII.

Fluorescence in situ hybridization

Fluorescence *in situ* hybridization for copy number detection of *MDM2/CDK4* was performed on a paraffin section of case 5918 using standard techniques (Shimada *et al.*, 2006) to confirm the diagnosis of dedifferentiated liposarcoma.

Mutation analyses of *KIT* and *PDGFRA* genes

GIST tumors were analysed for mutations in exons 9, 11, 13 and 17 of the *KIT* gene and exons 12 and 18 of *PDGFRA* using a combination of denaturing high-performance liquid chromatography and direct sequencing, as previously described (Heinrich *et al.*, 2003).

Acknowledgements

This work was supported by NIH grant CA112270, a grant from LifeRaft and the Department of Pathology, Stanford University. SS was supported by a postdoctoral fellowship from the US National Institutes of Health.

References

- Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X *et al.* (2003). A uniform system for microRNA annotation. *RNA* **9**: 277–279.
- Baer C, Nees M, Breit S, Selle B, Kulozik AE, Schaefer KL *et al.* (2004). Profiling and functional annotation of mRNA gene expression in pediatric rhabdomyosarcoma and Ewing's sarcoma. *Int J Cancer* **110**: 687–694.
- Binh MB, Sastre-Garau X, Guillou L, de Pinieux G, Terrier P, Lagace R *et al.* (2005). MDM2 and CDK4 immunostainings are useful adjuncts in diagnosing well-differentiated and dedifferentiated liposarcoma subtypes: a comparative analysis of 559 soft tissue neoplasms with genetic data. *Am J Surg Pathol* **29**: 1340–1347.
- Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE *et al.* (2005). A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. *N Engl J Med* **353**: 1793–1801.
- Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S *et al.* (2004). Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* **101**: 2999–3004.
- Chen JF, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM *et al.* (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* **38**: 228–233.
- Doench JG, Sharp PA. (2004). Specificity of microRNA target selection in translational repression. *Genes Dev* **18**: 504–511.
- Esau C, Kang X, Peralta E, Hanson E, Marcusson EG, Ravichandran LV *et al.* (2004). MicroRNA-143 regulates adipocyte differentiation. *J Biol Chem* **279**: 52361–52365.
- Espinosa I, Lee CH, Kim M, Rouse B, Subramanian S, Montgomery K *et al.* (2007). A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. *Am J Surg Pathol* (in press).
- Esquela-Kerscher A, Slack FJ. (2006). Oncomirs—microRNAs with a role in cancer. *Nat Rev Cancer* **6**: 259–269.
- Felli N, Fontana L, Pelosi E, Botta R, Bonci D, Facchiano F *et al.* (2005). MicroRNAs 221 and 222 inhibit normal erythropoiesis and erythroleukemic cell growth via kit receptor down-modulation. *Proc Natl Acad Sci USA* **102**: 18081–18086.
- Fletcher JA, Rubin BP. (2007). KIT Mutations in GIST. *Curr Opin Genet Dev* **17**: 3–7.
- Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ. (2006). miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**: D140–D144.
- He L, Hannon GJ. (2004). MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**: 522–531.
- Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen CJ, Joseph N *et al.* (2003). PDGFRA activating mutations in gastrointestinal stromal tumors. *Science* **299**: 708–710.
- Kim HK, Lee YS, Sivaprasad U, Malhotra A, Dutta A. (2006). Muscle-specific microRNA miR-206 promotes muscle differentiation. *J Cell Biol* **174**: 677–687.
- Kim VN. (2005). MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* **6**: 376–385.
- Lau NC, Lim LP, Weinstein EG, Bartel DP. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* **294**: 858–862.
- Lu J, Getz G, Miska EA, Alvarez-Saavedra E, Lamb J, Peck D *et al.* (2005). MicroRNA expression profiles classify human cancers. *Nature* **435**: 834–838.
- Lui WO, Pourmand N, Patterson BK, Fire A. (2007). Patterns of known and novel small RNAs in human cervical cancer. *Cancer Res* **67**: 6031–6043.
- Mayanil CS, George D, Freilich L, Miljan EJ, Mania-Farnell B, McLone DG *et al.* (2001). Microarray analysis detects novel Pax3 downstream target genes. *J Biol Chem* **276**: 49299–49309.
- Michael MZ, O'Connor SM, van Holst Pellekaan NG, Young GP, James RJ. (2003). Reduced accumulation of specific microRNAs in colorectal neoplasia. *Mol Cancer Res* **1**: 882–891.
- Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX *et al.* (2002). Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* **359**: 1301–1307.
- Politz JC, Zhang F, Pederson T. (2006). MicroRNA-206 colocalizes with ribosome-rich regions in both the nucleolus and cytoplasm of rat myogenic cells. *Proc Natl Acad Sci USA* **103**: 18957–18962.
- Pretto D, Barco R, Rivera J, Neel N, Gustavson MD, Eid JE. (2006). The synovial sarcoma translocation protein SYT–SSX2 recruits beta-catenin to the nucleus and associates with it in an active complex. *Oncogene* **25**: 3661–3669.

- Shimada S, Ishizawa T, Ishizawa K, Matsumura T, Hasegawa T, Hirose T. (2006). The value of MDM2 and CDK4 amplification levels using real-time polymerase chain reaction for the differential diagnosis of liposarcomas and their histologic mimickers. *Hum Pathol* **37**: 1123–1129.
- Subramanian S, West RB, Corless CL, Ou W, Rubin BP, Chu KM *et al.* (2004). Gastrointestinal stromal tumors (GISTs) with KIT and PDGFRA mutations have distinct gene expression profiles. *Oncogene* **23**: 7780–7790.
- Weiss SW, Goldburn JR. (2001). *Soft tissue tumors*. Mosby: St Louis.
- West RB, Corless CL, Chen X, Rubin BP, Subramanian S, Montgomery K *et al.* (2004). The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. *Am J Pathol* **165**: 107–113.
- Yan Z, Choi S, Liu X, Zhang M, Schageman JJ, Lee SY *et al.* (2003). Highly coordinated gene regulation in mouse skeletal muscle regeneration. *J Biol Chem* **278**: 8826–8836.
- Zuker M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**: 3406–3415.

Supplementary Information accompanies the paper on the Oncogene website (<http://www.nature.com/onc>).