npg

Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma

Robert Nakayama[1,2,3], Takeshi Nemoto[2,4], Hiro Takahashi[2], Tsutomu Ohta[2], Akira Kawai[5], Kunihiko Seki[6], Teruhiko Yoshida[2], Yoshiaki Toyama[3], Hitoshi Ichikawa[1] and Tadashi Hasegawa[6,7]

[1]Cancer Transcriptome Project, National Cancer Center Research Institute, Tokyo, Japan; [2]Genetics Division, National Cancer Center Research Institute, Tokyo, Japan; [3]Department of Orthopaedic Surgery, Keio University School of Medicine, Tokyo, Japan; [4]Department of Dermatology, Tokyo Medical and Dental University, Tokyo, Japan; [5]Orthopedics Division, National Cancer Center Hospital, Tokyo, Japan; [6]Pathology Division, National Cancer Center Hospital, Tokyo, Japan and [7]Department of Surgical Pathology, Sapporo Medical University School of Medicine, Sapporo, Japan

In soft tissue sarcomas, the diagnosis of malignant fibrous histiocytoma (MFH) has been a very controversial issue, and MFH is now considered to be reclassified into pleomorphic subtypes of other sarcomas. To characterize MFH genetically, we used an oligonucleotide microarray to analyze gene expression in 105 samples from 10 types of soft tissue tumors. Spindle cell and pleomorphic sarcomas, such as dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, malignant peripheral nerve sheath tumor (MPNST), fibrosarcoma and MFH, showed similar gene expression patterns compared to other tumors. Samples from those five sarcoma types could be classified into respective clusters based on gene expression by excluding MFH samples. We calculated distances between MFH samples and other five sarcoma types (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma) based on differentially expressed genes and evaluated similarities. Three of the 21 MFH samples showed marked similarities to one of the five sarcoma types, which were supported by histological findings. Although most of the remaining 18 MFH samples showed little or no histological resemblance to one of the five sarcoma types, 12 of them showed moderate similarities in terms of gene expression. These results explain the heterogeneity of MFH and show that the majority of MFHs could be reclassified into pleomorphic subtypes of other sarcomas. Taken together, gene expression profiling could be a useful tool to unveil the difference in the underlying molecular backgrounds, which leads to a rational taxonomy and diagnosis of a diverse group of soft tissue sarcomas.
Modern Pathology (2007) **20,** 749–759; doi:10.1038/modpathol.3800794; published online 27 April 2007

Keywords: gene expression; malignant fibrous histiocytoma; myxofibrosarcoma; soft tissue sarcoma; reclassification; undifferentiated pleomorphic sarcoma

Malignant soft tissue tumors are a diverse group of tumors of mesenchymal origin, which have generally been classified according to their histological resemblance to normal tissue. Understanding of molecular pathology gained in recent decades shows that some soft tissue tumors exhibit single recurrent genetic aberrations, such as chromosomal translocations resulting in gene fusions (*SYT-SSX* in synovial sarcoma, *TLS-CHOP* in myxoid/round cell liposarcoma) or somatic mutations (*KIT* in gastrointestinal stromal tumors), and they are now classified by these molecular markers specific to each tumor.[1] In contrast, other malignant soft tissue tumors, such as malignant fibrous histiocytoma (MFH), fibrosarcoma and leiomyosarcoma, are characterized by numerous, nonrecurrent complex chromosomal aberrations, and frequently show overlapping histological appearance and immunohistochemical phenotypes that are often difficult to

interpret by pathologists.[2] Among them, diagnosis of MFH has been the most controversial issue.[3–5] MFH has been considered the most common soft tissue sarcoma of adults; it is manifested by a broad range of histological appearances and consists of four subtypes: storiform and pleomorphic type, myxoid type, giant cell type and inflammatory type. Recent clinicopathological, ultrastructural and immunohistochemical studies revealed that MFH shows no evidence of true histiocytic differentiation and that it is not a single entity but rather a heterogeneous collection of pleomorphic subtypes of other sarcomas. Since each type of sarcoma other than MFH shows distinct biological behavior, particularly in local recurrence or metastasis rate, MFH showing a variety of clinicopathological characteristics should be further reclassified to correctly evaluate the malignant potential of each case. In the latest edition of the WHO classification, myxoid type MFH was classified as myxofibrosarcoma in the fibroblastic category, and other subtypes of MFH without any evidence of differentiation were classified as undifferentiated high grade pleomorphic sarcoma.[6] WHO classification also suggested that the term 'MFH' might disappear when criteria for the diagnosis of pleomorphic sarcomas showing a distinct differentiation state can be reproducibly defined.[6] In this work, we used the term 'MFH' to identify tumors diagnosed as storiform and pleomorphic type MFH, and the term 'myxofibrosarcoma' for so-called MFH with predominant (>50%) myxoid features conventionally diagnosed as myxoid type MFH.

Recently, several studies report gene expression profiling of soft tissue tumors using microarray technologies to provide new insights into the tumor characterization. They described distinct patterns of gene expression in respective tumors with single, recurrent genetic aberrations, such as synovial sarcoma, myxoid/round cell liposarcoma, clear cell sarcoma or gastrointestinal stromal tumors, and heterogeneous patterns in spindle cell and pleomorphic sarcomas which are generally characterized by complex chromosomal aberrations.[7–12] No further detailed analysis of gene expression in spindle cell and pleomorphic sarcomas have been reported so far.

In this study, we analyzed gene expression profile of total 105 cases representing 10 types of soft tissue tumors to identify their molecular characteristics. We observed similarity in gene expression among spindle cell and pleomorphic sarcomas, forming a relatively loose cluster, which is separated from the distinct clusters of synovial sarcoma, myxoid/round cell liposarcoma and lipoma + well-differentiated liposarcoma. Next, we primarily analyzed 64 samples of spindle cell and pleomorphic sarcomas and showed heterogeneity of MFH in terms of gene expression. We selected genes that could clearly distinguish between dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma,

malignant peripheral nerve sheath tumor (MPNST) and fibrosarcoma and quantified similarities as distances between MFH samples and the five sarcoma types.

Materials and methods

Patients and Tumor Samples

Characteristics of 105 soft tissue tumors used in this study are shown in Supplementary data 1. Among them, 35 samples were previously analyzed in a different method.[13] All patients received histological diagnosis of primary soft tissue tumor at National Cancer Center Hospital, Tokyo, from 1996 to 2002. In this paper, we use the term 'MFH' to describe samples diagnosed as storiform and pleomorphic type MFH showing predominant pleomorphic features without immunohistochemical phenotypes characteristic of specific differentiation, and the term 'myxofibrosarcoma' to describe MFH with predominant (>50%) myxoid features conventionally diagnosed as myxoid type MFH. Before the gene expression analysis pathologists confirmed the diagnosis of MFH was appropriate at the time of primary diagnosis. Tumor samples were collected from the part with macroscopically high tumor content by pathologists immediately after surgical excision and cryopreserved in liquid nitrogen until use. This study was approved by the ethics committee of National Cancer Center and conducted according to tenets of the Declaration of Helsinki.

Gene Expression Profiling

Total RNA was isolated using TRIzol (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's instruction. Samples were analyzed with a GeneChip Human Genome U133A array (Affymetrix, Santa Clara, CA, USA) containing 22 283 probe sets. Target cRNA preparation from total RNA, hybridization to the microarray, washing and staining with an antibody amplification procedure and scanning were all carried out according to the manufacturer's instructions. The expression value (Signal) of each probe set was calculated using GeneChip Operating Software (GCOS) ver. 1.3 (Affymetrix), so that the mean of expression values in each experiment was set at 100 to adjust for minor differences between experiments.

Statistical Analysis

Gene expression data were subsequently imported into GeneSpring GX7.2 software (Agilent Technologies, Santa Clara, CA, USA) and normalized to the median of all samples enrolled in the analysis and log-transformed for each gene. Hierarchical clustering analysis was performed using Pearson's correlation. To select appropriate probe sets defining five

types of spindle cell and pleomorphic sarcomas (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma), we performed Student's *t*-tests between one and the other four sarcoma types. The top 50 probe sets with low *P*-values in each *t*-test were summed. The centroid of each sarcoma type was determined by calculating the average of the selected probe sets. The distance (D) from a centroid to a sample was defined as $D = 1 - r$, using Pearson's correlation coefficient (r, $-1 \leq r \leq 1$). Inter-centroid distances were also calculated using Pearson's correlation coefficient.

Histological Analysis

Histological sections of the tumors were stained with hematoxylin and eosin and reviewed for all samples, and representative sections were examined immunohistochemically using the labeled streptavidin–biotin method. Sections were dewaxed, rehydrated and moistened with phosphate-buffered saline (pH 7.4), autoclaved at $121°C$ for 10 min in 10 mM citrate buffer (pH 6.0) and incubated with antibodies to the following molecules on an automated immunostaining system i6000 (BioGenex, San Ramon, CA, USA) for 30 min, as described previously:[14] vimentin, desmin, α-smooth muscle actin (αSMA), muscle-specific actin, h-caldesmon, CD34, S-100 protein, epithelial membrane antigen, cytokeratin and neurofilament. Heat-induced epitope retrieval was not undertaken when sections were stained with antibodies to S-100 protein and epithelial membrane antigen.

Quantitative RT-PCR

Real-time quantitative reverse transcription (RT)-PCR was carried out using the 7500 Fast Real-Time PCR System (Applied Biosystems, Foster City, CA, USA) with FastStart TaqMan Probe Master (Rox) and Universal ProbeLibrary (Roche Applied Science, Mannheim, Germany). One microgram of total RNA from 17 tumor samples (myxofibrosarcoma ($n = 5$), MFH ($n = 7$), leiomyosarcoma ($n = 2$) and MPNST ($n = 3$)) was reverse-transcribed to synthesize single-stranded cDNAs using SuperScript III (Invitrogen), and 1/100 of the cDNA was used for each PCR. Probes and primers were designed using Probe Finder software (Roche Applied Science) (Supplementary data 2). Transcript levels were normalized to that of the *ACTB* transcript.

Results

Overview of Gene Expression in Soft Tissue Sarcomas

Gene expression data of 105 soft tissue tumor samples consisting of synovial sarcoma ($n = 16$), myxoid/round cell liposarcoma ($n = 19$), lipoma ($n = 3$), well-differentiated liposarcoma ($n = 3$), dedifferentiated liposarcoma ($n = 15$), myxofibrosarcoma ($n = 15$), leiomyosarcoma ($n = 6$), MPNST ($n = 3$), fibrosarcoma ($n = 4$) and MFH ($n = 21$) were obtained using an oligonucleotide microarray containing 22 283 probe sets. Among them, 12 599 probe sets whose expression values were not less than 100 in at least 3 of 105 samples were analyzed. To overview the transcriptome of sarcomas in our data set, we first performed principal component analysis with 12 599 probe sets (Figure 1a), which is a decomposition technique to reduce multidimensional data into several specialized dimensions. The *x* and *y* axes in Figure 1a indicate the first and second principal components, respectively, representing the top and second largest fractions of the overall variability. In this analysis, 105 samples were roughly classified into four groups based on their position relative to the first and second principal components. Both synovial sarcoma and myxoid/round cell liposarcoma samples were located on the negative side of the first principal component, while well-differentiated liposarcoma, dedifferentiated liposarcoma and other spindle cell and pleomorphic sarcoma samples were on the positive side. On the negative side of the second principal component were myxoid/round cell liposarcoma, well-differentiated liposarcoma and lipoma samples, all of which are adipocytic tumors. Interestingly, some dedifferentiated liposarcoma samples were distributed close to well-differentiated liposarcoma samples, while others were midway between well-differentiated liposarcoma and other spindle cell and pleomorphic sarcoma samples. These results suggest that the first principal component was associated with the difference between synovial sarcoma + myxoid/round cell liposarcoma and spindle cell and pleomorphic sarcomas, and that the second principal component was associated with adipocytic differentiation. Probe sets contributing significantly to the first and second principal components are listed in Supplementary data 3.

To identify genes whose expression differed in a statistically significant manner among all sarcoma types, we performed an analysis of variance (ANOVA) among 10 tumor types and selected 2590 probe sets with *P*-values of less than 1.0×10^{-5}. Two-dimensional hierarchical clustering analysis using those 2590 probe sets showed that synovial sarcoma and myxoid/round cell liposarcoma samples displayed distinct gene expression profiles and formed robust clusters (Figure 1b). On the other hand, myxofibrosarcoma, leiomyosarcoma, MPNST, fibrosarcoma and MFH samples did not show distinct gene expression profiles, but rather formed a single loose cluster and shared a similar expression profile. We also found that lipoma and well-differentiated liposarcoma samples and some of the dedifferentiated liposarcoma samples displayed similar gene expression profiles and formed a cluster, whereas
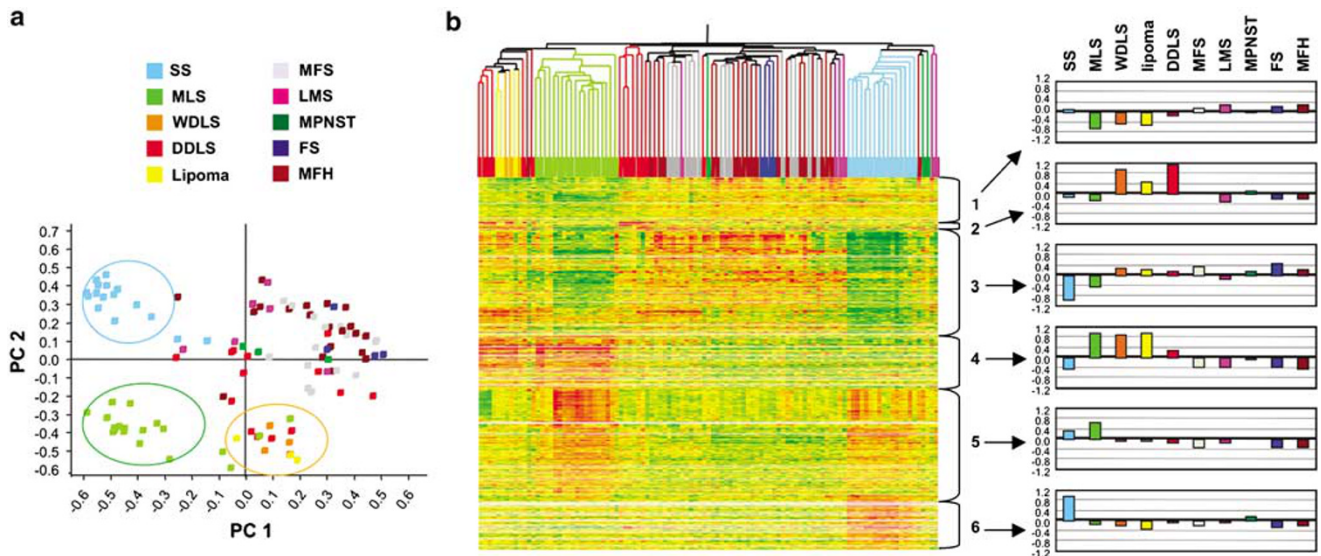
Figure 1 Gene expression overview of 105 soft tissue tumors. (**a**) Principal component analysis. A total of 12 599 probe sets with expression values not less than 100 in at least three samples were used in this analysis. *x* and *y* axes represent the first and second principal components (PC1 and PC2), respectively. Each dot represents a sample colored by its histological type. (**b**) Two-dimensional hierarchical clustering analysis. A total of 2590 probe sets differentially expressed among histological types ($P < 1.0 \times 10^{-5}$ by ANOVA) were used. Columns represent samples and rows represent probe sets. Red and green indicate high and low expression, respectively. The 2590 probe sets were roughly divided into six clusters (clusters 1–6). The six graphs on the right show averages of normalized expression values of those clusters for each histological type. Note that spindle cell and pleomorphic sarcomas, such as dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST, fibrosarcoma and MFH, form a loose cluster and share a similar expression profile compared with synovial sarcoma, myxoid/round cell liposarcoma, well-differentiated liposarcoma and lipoma. SS, synovial sarcoma; MLS, myxoid/round cell liposarcoma; WDLS, well-differentiated liposarcoma; DDLS, dedifferentiated liposarcoma; MFS, myxofibrosarcoma; LMS, leiomyosarcoma and FS, fibrosarcoma.

the other dedifferentiated liposarcoma samples did not share that profile but instead formed a loose cluster with fibrosarcoma, myxofibrosarcoma and MFH samples.

The 2590 probe sets were classified into six clusters according to their expression patterns (Figure 1b and Supplementary data 4). Interestingly, we found two major clusters (clusters 3 and 5) whose expression patterns were similar between synovial sarcoma and myxoid/round cell liposarcoma samples. Cluster 3, whose expression was low in synovial sarcoma and myxoid/round cell liposarcoma, contained many HLA genes, and cluster 5, whose expression was high in both synovial sarcoma and myxoid/round cell liposarcoma, contained many genes encoding ribosomal proteins and cancer testis antigens, such as *CTAG1B*, *CTAG2* and *PRAME*. Of note, these genes contributed largely to the first principal component (see Supplementary data 3). On the other hand, cluster 1, whose expression was low in myxoid/round cell liposarcoma, well-differentiated liposarcoma and lipoma samples, included cell cycle associated genes such as *CCNB1*, *CDKN3,* and *CDC20*, while cluster 4, whose expression was high in myxoid/round cell liposarcoma, well-differentiated liposarcoma and lipoma samples, included adipocytic differentiation-associated genes such as *LPL*, *ACACB* and *PLIN*. These genes contributed largely to the second principal component (see Supplementary data 3).

Cluster 6, whose expression was high in synovial sarcoma, included *COL2A1*, *COL9A3*, *SSX1* and *SSX2*. The small but robust cluster, cluster 2, consisted of *MDM2*, *CDK4* and other genes located in 12q13-15, which are known to be amplified in both well-differentiated liposarcoma and dedifferentiated liposarcoma.

Heterogeneity of MFH in Gene Expression and Classification of Spindle Cell and Pleomorphic Sarcomas

Spindle cell and pleomorphic sarcomas frequently display overlapping histological appearance and immunohistochemical phenotypes. Samples from these types of sarcoma did not separate into distinct histological types in the analysis using whole samples (Figure 1). To determine whether they could be grouped by gene expression, we analyzed 64 samples of spindle cell and pleomorphic sarcomas (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST, fibrosarcoma and MFH). We performed principal component analysis with 11 300 probe sets whose expression values were not less than 100 in at least three of 64 samples, and two-dimensional hierarchical clustering analysis using 1671 probe sets selected by ANOVA among six sarcoma types ($P < 0.01$) (Supplementary data 5). In the clustering analysis,
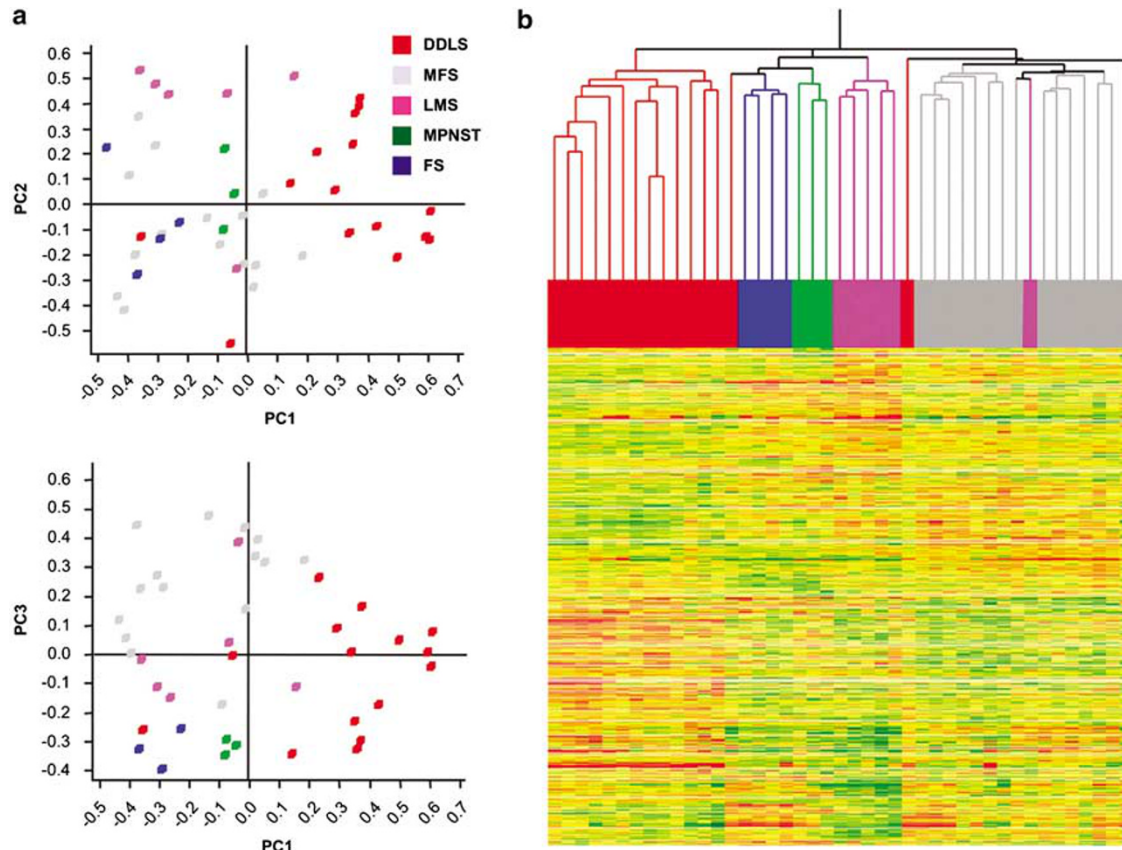
Figure 2 Classification of spindle cell and pleomorphic sarcomas without MFH. (**a**) Principal component analysis. A total of 11 300 probe sets with expression values not less than 100 in at least three of 64 spindle cell and pleomorphic sarcoma samples including MFH were used in this analysis. x and y axes in the upper panel represent the first and second principal components (PC1 and PC2), and x and y axes in the lower panel represent the first and third principal components (PC1 and PC3), respectively. (**b**) Two-dimensional hierarchical clustering analysis. A total of 1457 probe sets differentially expressed among five types of spindle cell and pleomorphic sarcomas ($P < 0.01$ by ANOVA) were used. Columns represent samples and rows represent probe sets. Red and green indicate high and low expression, respectively. Note that most samples formed clusters corresponding to their histology. DDLS, dedifferentiated liposarcoma; MFS, myxofibrosarcoma; LMS, leiomyosarcoma and FS, fibrosarcoma.

dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma samples appeared to form their own clusters, whereas those of MFH partitioned into several groups, some close to clusters of other sarcomas. These results suggest that MFH is heterogeneous in terms of gene expression as observed histologically.

Next, we analyzed 43 samples of five spindle cell and pleomorphic sarcomas (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma) and excluded MFH samples. In principal component analysis with 11 300 probe sets, samples of the same tumor type appeared to cluster (Figure 2a). We then performed two-dimensional hierarchical clustering analysis with 1457 probe sets selected from the 11 300 probe sets by ANOVA among five sarcoma types ($P < 0.01$) (Figure 2b). Although we found three exceptions (one leiomyosarcoma and two dedifferentiated liposarcoma samples), almost all dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma samples formed their

own respective clusters suggesting that each type of spindle cell and pleomorphic sarcoma formed a homogeneous group in terms of gene expression by excluding MFH samples.

Distances of MFH Samples from Other Spindle Cell and Pleomorphic Sarcomas

Since MFH samples did not form a clearly distinctive cluster, we next addressed a question whether MFH could be reclassified into other types of spindle cell and pleomorphic sarcomas by gene expression and quantified similarities between MFH samples and those sarcoma types using differentially expressed genes. To select appropriate probe sets defining spindle cell and pleomorphic sarcomas, we performed the Student's t-test between one and the other four of the five sarcoma types, namely, dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma. In this analysis, we excluded three exceptional samples
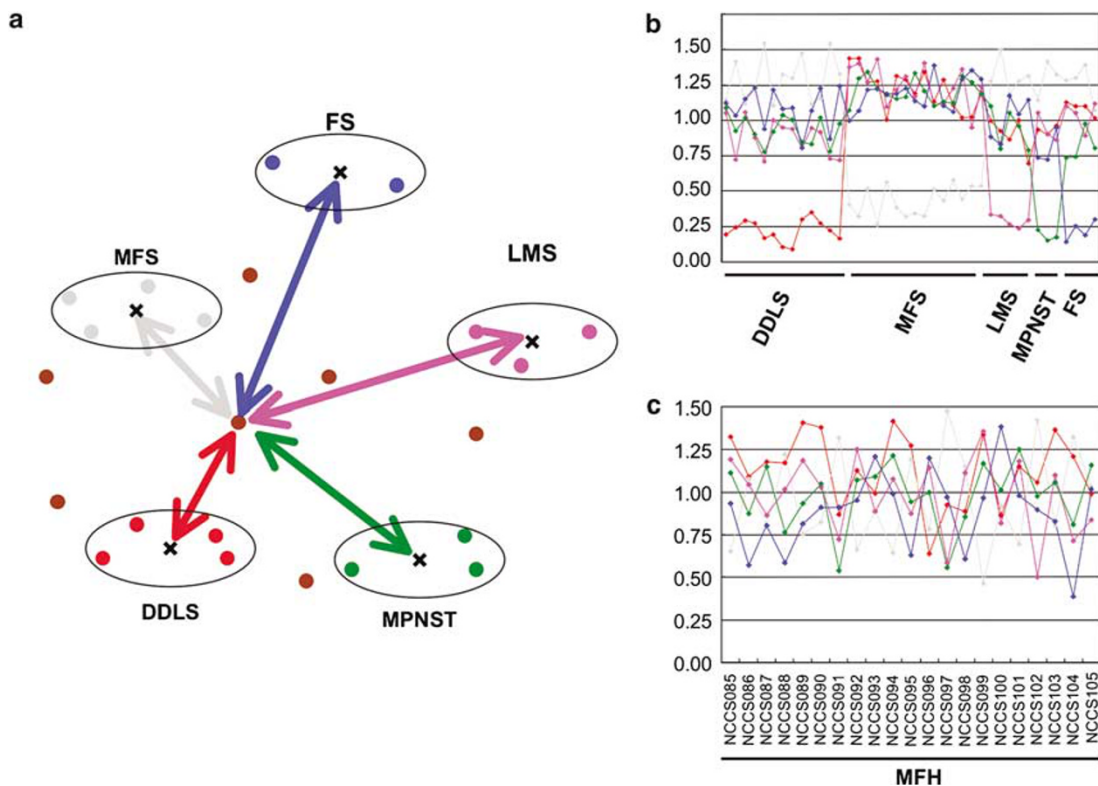
Figure 3 Distance evaluation of spindle cell and pleomorphic sarcoma samples from five sarcoma types. (**a**) Scheme of distance calculation. Each dot represents a sample colored according to its histology. Each ×-mark represents the centroid of each histological type of sarcoma. Each arrow indicates the distance from a sample to a centroid colored by the histology. (**b**) Distances of 40 control samples from the five centroids. Note that the closest centroids matched their histology. (**c**) Distances of 21 MFH samples from the five centroids. DDLS, dedifferentiated liposarcoma; MFS, myxofibrosarcoma; LMS, leiomyosarcoma and FS, fibrosarcoma.

that did not fall into the appropriate cluster (Figure 2b). The top 50 probe sets with low *P*-values in each *t*-test were summed to obtain 248 probe sets (Supplementary data 6). On the basis of the expression of these 248 probe sets, the centroids of those five sarcoma types were calculated in advance, and inter-centroid distances and distances from five centroids to each control sample ($n = 40$) were evaluated (Supplementary data 7 and Figure 3a and b). All inter-centroid distances were greater than 0.77 and the closest centroids for 40 control samples matched their histological types (Figure 3b), indicating that the evaluated distances were good indicators of sarcoma classification. We then evaluated the distances of each MFH sample from the five centroids (Figure 3c) and focused on determining the minimum (D_{min}) of the five distances. Small D_{min} values indicate high similarity to one of the five histological types in terms of gene expression. We used two cutoff values of 0.5 and 0.75 to evaluate similarity, because the majority of D_{min} values in control samples were less than 0.5 and most of the remaining four distances in each control sample were greater than 0.75. Among 21 samples, 3 showed marked similarity ($D_{min} \leq 0.5$), 12 showed moderate similarity ($0.5 < D_{min} \leq 0.75$) and the remaining 6 showed little similarity ($D_{min} > 0.75$).

Among 15 MFH samples showing high or moderate similarity ($D_{min} \leq 0.75$), 6 were similar to myxofibrosarcoma, 5 to fibrosarcoma, 2 to MPNST and 1 each to dedifferentiated liposarcoma and leiomyosarcoma.

Histological Reviews

We re-examined the histology of 21 MFH samples with the knowledge of similarity to other types of spindle cell and pleomorphic sarcomas based on gene expression. Three MFH samples that showed high gene expression similarity ($D_{min} \leq 0.5$) displayed marked pleomorphism, indicating that a diagnosis of MFH was appropriate at the time of diagnosis. However, these samples also showed histological signatures of relevant subtypes. The NCCS099 sample, which was significantly close to the myxofibrosarcoma centroid ($D_{min} = 0.46$), showed prominent myxoid features very similar to myxofibrosarcoma in one third of the tumor (Figure 4a). The NCCS102 sample which was very close to the leiomyosarcoma centroid ($D_{min} = 0.50$) was positive for desmin and αSMA (Figure 4b–d). The NCCS104 sample, which was very close to the fibrosarcoma centroid ($D_{min} = 0.39$), showed focal
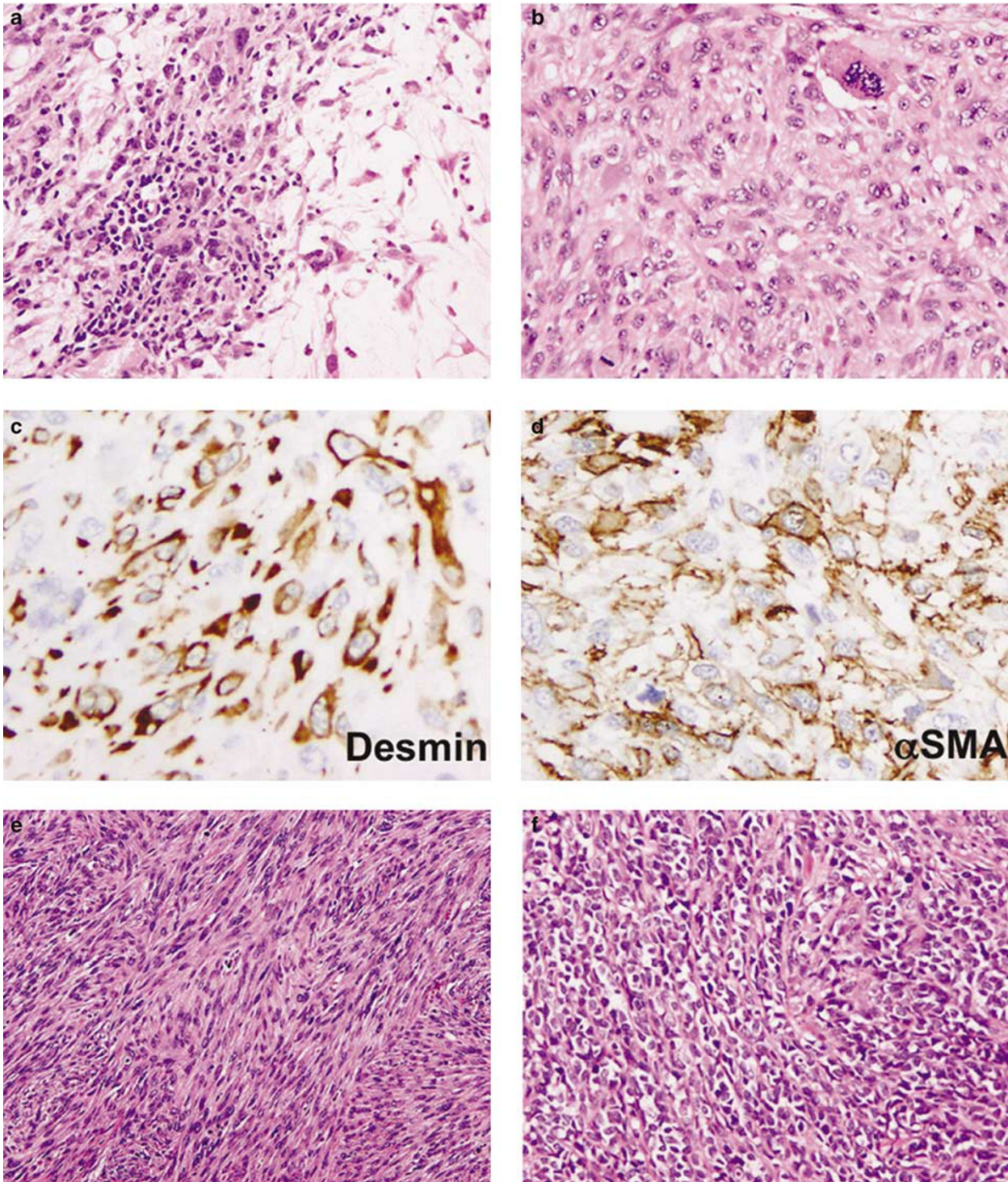
Figure 4 Histological review. (**a**) The border between pleomorphic area and myxoid area observed in the NCCS099 sample ($D_{min} = 0.46$ to myxofibrosarcoma) (hematoxylin and eosin stain). (**b–d**) Histology of the NCCS102 sample ($D_{min} = 0.50$ to leiomyosarcoma). This tumor showed marked pleomorphism (**b**) hematoxylin and eosin stain, but tumor cells were positive for desmin (**c**) and αSMA (**d**). (**e**) Fibrosarcomatous fascicular area seen in the NCCS104 sample ($D_{min} = 0.39$ to fibrosarcoma) (hematoxylin and eosin stain). (**f**) Epithelioid structure observed in the NCCS097 sample ($D_{min} = 0.56$ to MPNST) (hematoxylin and eosin stain).

fibrosarcoma-like herringbone and fascicular patterns by microscopic analysis (Figure 4e).

Among twelve samples showing moderate similarity to other types of sarcomas ($0.50 < D_{min} \leq 0.75$),

the NCCS096 sample close to the dedifferentiated liposarcoma centroid ($D_{min} = 0.64$) was obtained from a recurrent sarcoma in the retroperitoneum. Although microscopic findings did not show

evidence of adipocytic differentiation or features of well-differentiated liposarcoma in regions adjacent to the tumor, the site of involvement suggested the possibility that the tumor originated from dedifferentiated liposarcoma. All five samples close to myxofibrosarcoma $(0.50 < D_{min} \leq 0.75)$ showed scattered myxoid areas, but these findings were not sufficient to reclassify them as myxofibrosarcoma histologically. The NCCS097 sample, another pleomorphic sarcoma close to the MPNST centroid $(D_{min} = 0.56)$, exhibited scattered whorled and epithelioid structures (Figure 4f) as well as tumor cells positive for cytokeratin, neurofilament and αSMA, indicating that this tumor had neuroectodermal differentiation. Its similarity to leiomyosarcoma $(D = 0.58)$ would be reflected in αSMA positivity. For the NCCS091 sample close to MPNST and the other four close to fibrosarcoma, we did not observe any significant histological similarity to MPNST or fibrosarcoma, respectively. In summary, although more than half of the MFH samples $(n = 12)$ were moderately similar in terms of gene expression to other sarcomas $(0.50 < D_{min} \leq 0.75)$, only little resemblance was detectable by histological examination. Finally, the remaining six samples with high D_{min} values $(D_{min} > 0.75)$ showed no identifiable histological similarity to the five sarcoma types (dedifferentiated liposarcoma, myxoid/round cell liposarcoma, leiomyosarcoma, MPNST and fibrosarcoma).

Genes Overexpressed in Myxofibrosarcoma

Diagnostically useful markers for myxofibrosarcoma are not well known. To search for candidate markers that genetically characterize myxofibrosarcoma, we selected upregulated genes by comparing myxofibrosarcoma samples $(n = 15)$ with other spindle cell and pleomorphic sarcoma samples $(n = 25)$. Three samples excluded from the previous analysis and 21 samples of MFH were not used for the marker search. From 11 300 probe sets, we selected 10 probe sets (five genes) with P-values < 0.001 based on the Student's t-test and more than five-fold greater expression (Table 1). Among them, expression of four probe sets (four genes) in respective spindle cell and pleomorphic sarcomas are shown in Figure 5a by the box-and-whisker plots. Since $ANK1$ expression in MFH was much higher than that seen in myxofibrosarcoma (data not shown), its upregulation was not considered to be specific to myxofibrosarcoma. We performed quantitative RT-PCR with three other genes, $WISP2$, $GPR64$ and $TNXB$, to verify the microarray findings (Figure 5b). Quantitative RT-PCR data confirmed consistent high expression of $GPR64$ and $TNXB$ in myxofibrosarcoma samples and in some MFH samples showing similarity to myxofibrosarcoma in terms of gene expression.

Discussion

An important aim of this study was to obtain new insights to classify a diverse group of soft tissue sarcomas. Our data showed that soft tissue sarcomas examined roughly fell into four groups (Figure 1a) (1) synovial sarcoma; (2) myxoid/round cell liposarcoma; (3) lipoma, well-differentiated liposarcoma with part of dedifferentiated liposarcoma and (4) spindle cell and pleomorphic sarcomas. Six histological types of spindle cell and pleomorphic sarcomas (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST, fibrosarcoma and MFH) did not display distinct profiles but they shared a similar gene expression profile, forming a loose cluster in the hierarchical clustering analysis (Figure 1b). These results were broadly consistent with previous reports,[7,10] and histological similarity among spindle cell and pleomorphic sarcomas could be explained by similarities in gene expression. We could find some MPNST samples were located adjacent to the robust synovial sarcoma cluster in the hierarchical clustering analysis (Figure 1b), indicating that those MPNST samples shared similar expression patterns with synovial sarcoma as reported by Nagayama et al.[8] Our data also showed a common gene expression signature in synovial sarcoma and myxoid/round cell liposarco-

Table 1 Genes highly expressed in myxofibrosarcoma

Gene symbol	Fold change	P-value	Description	Probe set ID
WISP2	10.99	1.6×10^{-5}	WNT1 inducible signaling pathway protein 2	205792_at
GPR64	10.56	7.7×10^{-5}	G protein-coupled receptor 64	206002_at
TNXB	8.30	3.7×10^{-5}	Tenascin XB	208609_s_at
ANK1	7.03	1.8×10^{-5}	Ankyrin 1, erythrocytic	208352_x_at
S100A3	6.41	3.0×10^{-7}	S100 calcium binding protein A3	206027_at
ANK1	5.99	5.5×10^{-5}	Ankyrin 1, erythrocytic	205391_x_at
TNXB	5.96	2.9×10^{-5}	Tenascin XB	213451_x_at
TNXB	5.86	6.5×10^{-4}	Tenascin XB	216339_s_at
TNXB	5.74	1.3×10^{-4}	Tenascin XB	206093_x_at
TNXB	5.74	4.7×10^{-5}	Tenascin XB	216333_x_at

The top 10 probe sets with high fold changes were selected from 321 probe sets differentially expressed ($P < 0.001$ by Student's t-test) between myxofibrosarcoma samples ($n = 15$) and other spindle cell and pleomorphic sarcoma samples ($n = 25$) analyzed in Figure 3b.
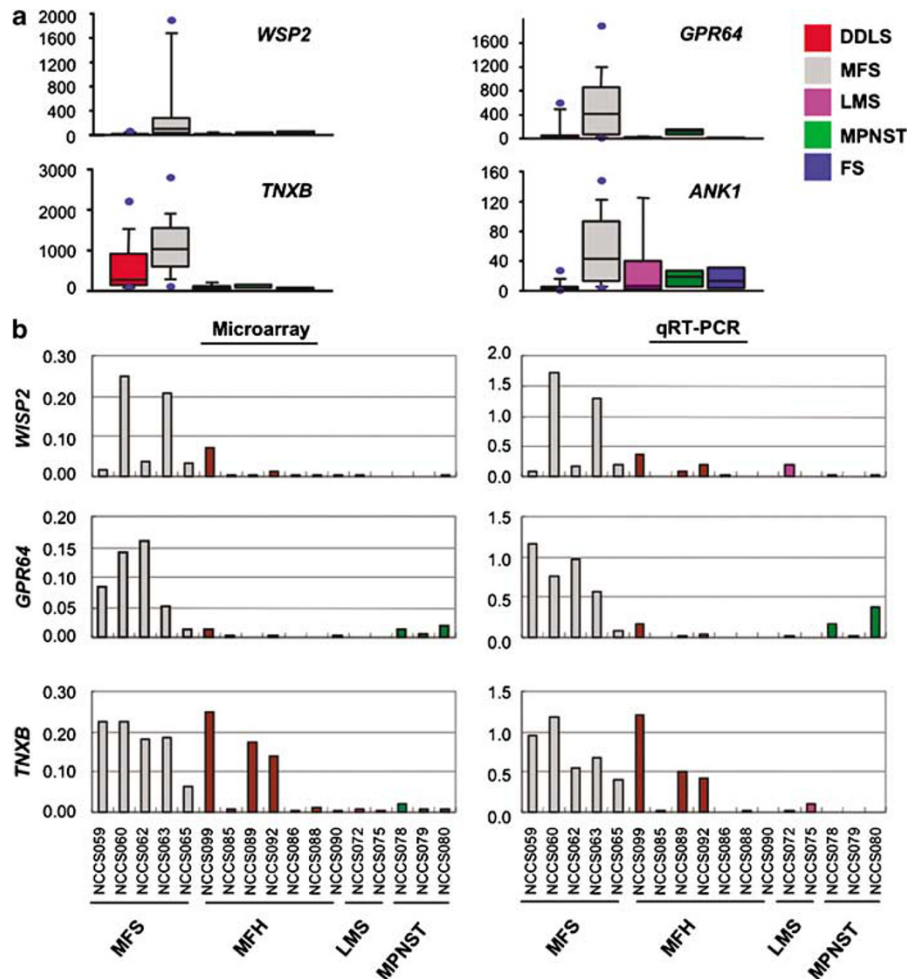
Figure 5 Genes highly expressed in myxofibrosarcoma. (**a**) Box-and-whisker plots indicating expression values for each histological type of spindle cell and pleomorphic sarcomas. A total of 40 control samples were analyzed. (**b**) Comparison between microarray analysis and quantitative RT-PCR of *WISP2*, *GPR64* and *TNXB* expression. Expression levels were normalized to that of *ACTB* in both microarray and RT-PCR data. NCCS099 ($D_{min} = 0.46$), NCCS085 ($D_{min} = 0.66$), NCCS089 ($D_{min} = 0.75$) and NCCS092 ($D_{min} = 0.66$) were similar to myxofibrosarcoma in terms of gene expression (see Figure 3c). DDLS, dedifferentiated liposarcoma; MFS, myxofibrosarcoma; LMS, leiomyosarcoma and FS, fibrosarcoma.

ma samples, distinguishing them from other tumors. Overexpression of genes encoding ribosomal proteins in myxoid/round cell liposarcoma was reported previously.[11] Another report showed that *SOX11*, *CTAG1*, *CTAG2* and *PRAME* were over-expressed in liposarcomas and absent or minimally expressed in all other tumors examined.[15] Among those genes, *CTAG1* and *PRAME* are both categorized as cancer testis antigens, and their expression in synovial sarcoma has also been reported.[16] Consistent with those reports, we found that *SOX11*, *CTAG1*, *CTAG2* and *PRAME* are highly expressed in both synovial sarcoma and myxoid/round cell liposarcoma. These similarities in gene expression may correlate with biological characteristics of synovial sarcoma and myxoid/round cell liposarcoma and suggest that these two sarcomas may share a common oncogenic pathway.

The so-called MFH was thought to be the most common soft tissue sarcoma in adults, and cur-

rently, it is widely accepted as a common morphological manifestation of a variety of poorly differentiated sarcomas. Re-evaluation of 'MFH' by different methods has been undertaken. Fletcher *et al*[5] reclassified 100 tumors primarily diagnosed as 'MFH' by histological methods and showed that the most common diagnosis was myxofibrosarcoma ($n = 29$), followed by leiomyosarcoma ($n = 20$). Hasegawa *et al*[14] examined immunoreactivity for smooth muscle markers from 100 samples of 'MFH' and reported that a large subset showed poorly differentiated smooth muscle or myofibroblastic features and should be regarded as pleomorphic leiomyosarcoma or pleomorphic myofibrosarcomas. Using comparative genomic hybridization, Derre *et al*[17] showed similar recurrent genomic imbalances in 'MFH' and leiomyosarcoma, and Coindre *et al*[18] reported that most inflammatory types of MFH developing in the retroperitoneum are identical to dedifferentiated liposarcoma. Here, we discussed

the possibility that 21 MFH samples could be reclassified into other types of spindle cell and pleomorphic sarcomas based on similarities in gene expression. For convenience of evaluation, we separated MFH samples into three groups according to the level of similarity to other sarcoma types. MFH with marked similarity ($D_{min} \leq 0.5$), MFH with moderate similarity ($0.50 < D_{min} \leq 0.75$) and MFH with no similarity ($D_{min} > 0.75$). Three samples very similar in gene expression to other sarcoma types ($D_{min} \leq 0.5$) resembled the corresponding histological types of spindle cell and pleomorphic sarcomas, and we concluded that these samples could probably be diagnosed as pleomorphic subtypes of those respective sarcomas based on current histological criteria. We then found that despite only marginal histological resemblance, more than half of the MFH samples (12/21) showed gene expression profiles similar to other sarcoma types ($0.50 < D_{min} \leq 0.75$). We considered that these moderate similarities in gene expression could correspond with pleomorphic change in each sarcoma type. Thus, although the samples cannot be diagnosed based on current histological criteria, it is possible to reclassify them as a pleomorphic subtype of those sarcomas based on gene expression. In this study, 40% (6/15) of reclassified MFH samples ($D_{min} \leq 0.75$) were similar to myxofibrosarcoma and 33% (5/15) were similar to fibrosarcoma, suggesting that a large subset of 'MFH' represents pleomorphic subtypes of fibroblastic sarcomas. Among the six cases of MFH similar to myxofibrosarcoma, five other than NCCS089 had deep-seated lesions, four (NCCS085, NCCS092, NCCS094 and NCCS101) had distant metastasis, and one (NCCS094) suffered local recurrence after surgery. Although the local recurrence rate (1/6) was unexpectedly low and distant metastasis rate (4/6) was high compared to canonical myxofibrosarcoma, these data could be consistent with the report showing that deep-seated lesions of myxofibrosarcoma were higher-grade, pleomorphic and large and increased the incidence of distant metastases.[19] About 30% of the MFH samples (6/21) did not show similarities to other sarcoma types ($D_{min} > 0.75$). One possibility is that 'de novo undifferentiated pleomorphic sarcomas' truly exist. It is also possible that these samples represent advanced stage of dedifferentiation, which is beyond the analytical power of our study design. Another possibility is that the samples were derived from sarcomas of other differentiation not examined in this study. Extraskeletal osteosarcoma, rhabdomyosarcoma and other sarcomas could be the candidate. Reclassification accuracy should be improved by examining additional histological types of spindle cell and pleomorphic sarcomas.

Given that almost one third of MFH samples shared similar gene expression patterns ($D_{min} \leq 0.75$) with myxofibrosarcoma, we hypothesize that a large subset of 'MFH' may be pleomorphic subtype of myxofibrosarcoma. Myxofibrosarcoma is one of the most frequent sarcomas seen in late adults. However, little is known about its normal tissue counterparts, or factors underlying its extremely high local recurrence rate,[19] nor are there any good markers available for histological diagnosis. Identification of genes highly expressed in myxofibrosarcoma would offer an important clue to address these problems. Here, we found *WISP2*, *GPR64* and *TNXB* were upregulated in myxofibrosarcoma compared with other spindle cell and pleomorphic sarcomas. WISP2 is a member of the WNT1 inducible signaling pathway (WISP) protein subfamily, which belongs to the connective tissue growth factor family. WISP family members are secreted, cell- and matrix-associated proteins that play critical roles in cell differentiation and survival, wound repair, vascular disease, fibrosis and progression of certain cancers.[20–22] GPR64 is a highly conserved, tissue-specific heptahelical receptor of the human epididymis,[23–25] and there are no reports on the relationship of GPR64 to any type of cancer. TNXB is the largest member of the tenascin family of extracellular matrix proteins, which have anti-adhesive effects as opposed to the adhesion activity of fibronectin. It is expressed in musculoskeletal, cardiac and dermis tissue, and its deficiency is associated with the connective tissue disorder Ehlers-Danlos syndrome.[26–28] Although it is not clear if these genes play a role in myxofibrosarcoma, they may serve as novel diagnostic markers.

In this study, we primarily analyzed gene expression of MFH and other types of spindle cell and pleomorphic sarcomas (dedifferentiated liposarcoma, myxofibrosarcoma, leiomyosarcoma, MPNST and fibrosarcoma). Although these sarcomas showed a similar gene expression pattern and formed a relatively loose cluster, samples from five types of spindle cell and pleomorphic sarcomas were classified into respective histological types by excluding MFH samples. We identified genes that were differentially expressed among the five sarcoma types and could reclassify more than 70% of MFH samples into the five sarcoma types based on their similarities in gene expression using a combination of simple statistical analysis. These results suggest that gene expression profiling will be a useful tool to reclassify MFH and to aid histological diagnosis of a diverse group of soft tissue sarcomas. Although we cannot currently predict differences in clinical behavior of reclassified MFH due to the limited number of samples analyzed, accumulation of gene expression data should improve prediction of clinically important events, such as local recurrence, metastasis or therapeutic responses.

Acknowledgement

Disclosure/conflict of interest

References

1 Helman LJ, Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer 2003;3:685–694.
2 Hasegawa T, Yamamoto S, Nojima T, *et al.* Validity and reproducibility of histologic diagnosis and grading for adult soft-tissue sarcomas. Hum Pathol 2002;33: 111–115.
3 Fletcher CD. Pleomorphic malignant fibrous histiocytoma: fact or fiction? A critical reappraisal based on 159 tumors diagnosed as pleomorphic sarcoma. Am J Surg Pathol 1992;16:213–228.
4 Hollowood K, Fletcher CD. Malignant fibrous histiocytoma: morphologic pattern or pathologic entity? Semin Diagn Pathol 1995;12:210–220.
5 Fletcher CD, Gustafson P, Rydholm A, *et al.* Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. J Clin Oncol 2001;19:3045–3050.
6 Fletcher CD, van den Berg E, Molenaar WM. Pleomorphic malignant fibrous histiocytoma/undifferentiated high grade pleomorphic sarcoma. In: Fletcher CD, Unni KK, Mertens F (eds). World Health Organization Classification of Tumours. Pathology and Genetics of Tumours of Soft Tissue and Bone. IARC Press: Washington, DC, USA, 2002, pp 120–122.
7 Nielsen TO, West RB, Linn SC, *et al.* Molecular characterisation of soft tissue tumours: a gene expression study. Lancet 2002;359:1301–1307.
8 Nagayama S, Katagiri T, Tsunoda T, *et al.* Genome-wide analysis of gene expression in synovial sarcomas using a cDNA microarray. Cancer Res 2002;62:5859–5866.
9 Lee YF, John M, Edwards S, *et al.* Molecular classification of synovial sarcomas, leiomyosarcomas and malignant fibrous histiocytomas by gene expression profiling. Br J Cancer 2003;88:510–515.
10 Segal NH, Pavlidis P, Antonescu CR, *et al.* Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. Am J Pathol 2003; 163:691–700.
11 Skubitz KM, Skubitz AP. Characterization of sarcomas by means of gene expression. J Lab Clin Med 2004; 144:78–91.
12 Baird K, Davis S, Antonescu CR, *et al.* Gene expression profiling of human sarcomas: insights into sarcoma biology. Cancer Res 2005;65:9226–9235.
13 Takahashi H, Nemoto T, Yoshida T, *et al.* Cancer diagnosis marker extraction for soft tissue sarcomas based on gene expression profiling data by using projective adaptive resonance theory (PART) filtering method. BMC Bioinform 2006;7:399.
14 Hasegawa T, Hasegawa F, Hirose T, *et al.* Expression of smooth muscle markers in so called malignant fibrous histiocytomas. J Clin Pathol 2003;56:666–671.
15 Skubitz KM, Cheng EY, Clohisy DR, *et al.* Differential gene expression in liposarcoma, lipoma, and adipose tissue. Cancer Invest 2005;23:105–118.
16 Segal NH, Blachere NE, Guevara-Patino JA, *et al.* Identification of cancer-testis genes expressed by melanoma and soft tissue sarcoma using bioinformatics. Cancer Immun 2005;5:2.
17 Derre J, Lagace R, Nicolas A, *et al.* Leiomyosarcomas and most malignant fibrous histiocytomas share very similar comparative genomic hybridization imbalances: an analysis of a series of 27 leiomyosarcomas. Lab Invest 2001;81:211–221.
18 Coindre JM, Hostein I, Maire G, *et al.* Inflammatory malignant fibrous histiocytomas and dedifferentiated liposarcomas: histological review, genomic profile, and MDM2 and CDK4 status favour a single entity. J Pathol 2004;203:822–830.
19 Mentzel T, Calonje E, Wadden C, *et al.* Myxofibrosarcoma: clinicopathologic analysis of 75 cases with emphasis on the low-grade variant. Am J Surg Pathol 1996;20:391–405.
20 Pennica D, Swanson TA, Welsh JW, *et al.* WISP genes are members of the connective tissue growth factor family that are up-regulated in wnt-1-transformed cells and aberrantly expressed in human colon tumors. Proc Natl Acad Sci USA 1998;95:14717–14722.
21 Brigstock DR. The CCN family: a new stimulus package. J Endocrinol 2003;178:169–175.
22 Perbal B. CCN proteins: multifunctional signalling regulators. Lancet 2004;363:62–64.
23 Kierszenbaum AL. Epididymal G protein-coupled receptor (GPCR): two hats and a two-piece suit tailored at the GPS motif. Mol Reprod Dev 2003;64:1–3.
24 Obermann H, Samalecos A, Osterhoff C, *et al.* HE6, a two-subunit heptahelical receptor associated with apical membranes of efferent and epididymal duct epithelia. Mol Reprod Dev 2003;64:13–26.
25 Kirchhoff C, Obermann H, Behnen M, *et al.* Role of epididymal receptor HE6 in the regulation of sperm microenvironment. Mol Cell Endocrinol 2006;250: 43–48.
26 Burch GH, Gong Y, Liu W, *et al.* Tenascin-X deficiency is associated with Ehlers-Danlos syndrome. Nat Genet 1997;17:104–108.
27 Schalkwijk J, Zweers MC, Steijlen PM, *et al.* A recessive form of the Ehlers-Danlos syndrome caused by tenascin-X deficiency. N Engl J Med 2001; 345:1167–1175.
28 Mao JR, Taylor G, Dean WB, *et al.* Tenascin-X deficiency mimics Ehlers-Danlos syndrome in mice through alteration of collagen deposition. Nat Genet 2002;30:421–425.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)