

Indistinguishable genomic profiles and shared prognostic markers in undifferentiated pleomorphic sarcoma and leiomyosarcoma: different sides of a single coin?

Ana Carneiro¹, Princy Francis¹, Pär-Ola Bendahl¹, Josefin Fernebro¹, Måns Åkerman², Christopher Fletcher³, Anders Rydholm⁴, Åke Borg¹ and Mef Nilbert^{1,5}

Soft tissue sarcoma (STS) diagnostics and prognostics are challenging, particularly in highly malignant and pleomorphic subtypes such as undifferentiated pleomorphic sarcoma (UPS) and leiomyosarcoma (LMS). We applied 32K BAC arrays and gene expression profiling to 18 extremity soft tissue LMS and 31 extremity soft tissue UPS with the aim of identifying molecular subtype signatures and genomic prognostic markers. Both the gains/losses and gene expression signatures revealed striking similarities between UPS and LMS, which were indistinguishable using unsupervised hierarchical cluster analysis and significance analysis for microarrays. Gene expression analysis revealed just nine genes, among them tropomyosin beta, which were differentially expressed. Loss of 4q31 (encompassing the *SMAD1* locus), loss of 18q22, and tumor necrosis were identified as independent predictors of metastasis in multivariate stepwise Cox regression analysis. Combined analysis applying loss of 4q31 and 18q22 and the presence of necrosis improved the area under receiver operating characteristic curve for metastasis prediction from 0.64 to 0.86. The extensive genetic similarities between extremity soft tissue UPS and LMS suggest a shared lineage of these STS subtypes and the new and independent genetic prognosticators identified hold promise for refined prognostic determination in high-grade, genetically complex STS.

Laboratory Investigation (2009) 89, 668–675; doi:10.1038/labinvest.2009.18; published online 16 March 2009

KEYWORDS: soft tissue sarcoma; array CGH; gene expression; pleomorphic sarcomas; prognosis

Malignant fibrous histiocytoma (MFH) was the most frequent pleomorphic soft tissue sarcoma (STS) histotype, until thorough investigation showed a line of differentiation in two thirds of the cases, thereby reclassifying many of the tumors into pleomorphic liposarcomas and leiomyosarcomas (LMS).¹ Today, pleomorphic STS, without any histologically identifiable line of differentiation, are designated as undifferentiated pleomorphic sarcomas (UPS).² Gene expression studies in MFH/UPS initially suggested that the majority of tumors cluster with LMS and liposarcomas,^{3–5} but subsequent studies have identified distinct genomic and expression profiles in liposarcomas.^{6,7} Conventional comparative genomic hybridization (CGH) analyses have identified a number of frequently occurring gains and losses shared among MFH/UPS, LMS, and pleomorphic lipo-

sarcoma.^{8,9} Application of high-resolution array CGH in STS with complex karyotypes has recently suggested that genomic profiling may contribute to a refined classification and prognostication in genetically complex STS.^{10–12} Despite the presence of pleomorphic areas resembling storiform-pleomorphic MFH in LMS and myofibroblastic features in MFH, detailed genetic comparison between these two STS types has not been performed.^{13–16}

Recognition of recurrent genetic alterations constitutes a basis for reclassification in several soft tissue neoplasms and represents a diagnostic adjunct in, for example, synovial sarcoma, well-differentiated liposarcoma/de-differentiated liposarcoma, and clear-cell sarcoma.^{17,18} The genetic complexity that characterizes UPS and LMS has, however, precluded such developments, and histopathological

¹Department of Oncology, Institute of Clinical Sciences, Lund University, Lund, Sweden; ²Department of Pathology, Institute of Clinical Sciences, Lund University, Lund, Sweden; ³Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA; ⁴Department of Orthopedics, Institute of Clinical Sciences, Lund University, Lund, Sweden and ⁵Clinical Research Centre, Copenhagen University, Hvidovre Hospital, Hvidovre, Denmark
Correspondence: Dr A Carneiro, MD, Department of Oncology, Institute of Clinical Sciences, Lund University Hospital, Barngatan 2B, Lund, Lund 221 85, Sweden.
E-mail: ana.carneiro@med.lu.se

Received 14 October 2008; revised 27 November 2008; accepted 3 December 2008

characterization with complementary immunostaining remains the standard diagnostic method, albeit with sub-optimal reproducibility. Histological grade remains the most important prognostic factor in STS, but metastasis rates and survival vary among patients with histologically similar tumors, suggesting the existence of inherent prognostic features not detected by pathologic evaluation. Objective diagnostic adjuncts that may improve histotype recognition and prognostication are thus clinically required and would contribute to improved selection of high-risk patients for adjuvant treatment. Therefore, we applied 32K BAC arrays with tiling coverage of the human genome and gene expression profiling to identify molecular subtype signatures and prognostic markers in high-grade extremity UPS and LMS.

MATERIALS AND METHODS

Patients and Tumor Material

The study used a retrospective case-control design, in which primary, high-grade (grades III and IV on a four-tiered scale) UPS and LMS of the extremities and the trunk wall, operated between 1987 and 2003, were randomly selected from the Lund Sarcoma Center. Ethical permission for the study was granted by the Lund University research ethics committee. Patients with metastases at diagnosis or treated with pre-operative chemo or radiotherapy were excluded. The sample set was selected to reflect the ~50% metastasis rate in high-grade STS. All tumors were reviewed by an experienced sarcoma pathologist (M.Å.), according to the WHO classification.² In short, LMS was a tumor with the presence of eosinophilic spindle cells with vesicular, blunt-ended, intended, or lobulated nuclei arranged in a fascicular pattern at least focally, accounting for 5–10% of the surface area examined. The tumors with these characteristics were also required to show unequivocal positivity for smooth-muscle actin (SMA) as well as for desmin and/or h-caldesmon. UPS was a pleomorphic spindle-cell sarcoma that did not show evidence of specific differentiation. These tumors were also negative for melanocytic and hematopoietic markers. Adequate tissue samples and clinical data were available from all cases. As part of routine, pathology blocks were obtained at 1-cm intervals from the entire tumor and all available blocks were analyzed using 4- μ m sections. All tumors were stained at the Department of Pathology, Lund University Hospital. In 15 cases, a blinded review (including all sections available) was also performed by another experienced sarcoma pathologist (C.F.). Histopathological grade, necrosis, vascular invasion, depth, and size were obtained from the registry of Lund Sarcoma Center. All patients were followed up with clinical examinations and chest X-rays and/or CT scans every third month for the first 3 years and biannually thereafter for at least 6 years or until death. Initially, 60 tumors were selected to represent 30 patients who developed metastasis and 30 patients who remained free of disease after a minimum

Table 1 Clinical and pathological data

Clinicopathological data	UPS (n = 31)		LMS (n = 18)	
	Met. ⁺	Met. ⁻	Met. ⁺	Met. ⁻
n (frequency)	13 (0.42)	18 (0.58)	10 (0.56)	8 (0.44)
Median time to metastasis (months)	9 (3–57) ^a	—	7 (3–20) ^a	—
Age ^b				
40–60 years	1 (0.08)	4 (0.22)	1 (0.10)	0
>60 years	12 (0.92)	14 (0.78)	9 (0.90)	8 (0.10)
Vascular invasion ^b	3 (0.23)	4 (0.22)	3 (0.30)	2 (0.25)
Necrosis ^b	12 (0.92)	9 (0.50)	9 (0.90)	5 (0.62)
Depth ^b	11 (0.84)	13 (0.72)	8 (0.80)	4 (0.50)
Size ^b				
<8 cm	4 (0.31)	11 (0.61)	3 (0.30)	5 (0.63)
≥8 cm	9 (0.69)	7 (0.39)	7 (0.70)	3 (0.37)

^aRange.

^bn (frequency).

follow-up of 2 years. Eleven samples were lost because of poor quality aCGH data. A total of 49 samples, 31 UPSs and 18 LMSs, were available for aCGH analysis and represented 23 tumors from patients who developed metastases and 26 tumors from those who remained free of disease (Table 1). Parallel analysis of gene expression data could be performed in 45 tumors (28 UPSs and 17 LMSs).

DNA Extraction, Labeling, and Hybridization

DNA was extracted from freshly frozen tumor tissue using the Wizard[®] Genomic DNA Purification Kit (Promega, Madison, WI, USA), followed by a phenol-chloroform purification. Commercial genomic DNA derived from a pool of healthy male individuals was used as reference in all hybridizations (Promega, Madison, WI, USA). Labeling, slide preparation, and hybridization were performed as previously described.¹⁹ High-resolution tiling BAC arrays were produced at the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University, Sweden, using the BAC Re-Array set Ver. 1.0 (32433 BAC clones), described earlier by Krzywinski *et al*,²⁰ Gene Expression Omnibus (GEO) platform repository accession GPL4723. BAC clones were mapped to the hg17 genome build. Arrays were washed as described earlier,²¹ and fluorescence was recorded using an Agilent G2565AA microarray scanner (Agilent Technologies).

RNA Extraction, Labeling, and Hybridization

Total RNA was extracted from 80 to 120 mg frozen tissue using Trizol reagent (Invitrogen, Carlsbad, CA, USA), followed by the RNeasy Midi kit (Qiagen, Valencia, CA, USA). RNA quality was assessed using the RNA 6000 NanoLabChip Kit for Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) and concentration was determined using a NanoDrop Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). As reference RNA, we used the Universal Human Reference RNA (Stratagene, La Jolla, CA, USA). cDNA synthesis and CyDye coupling were carried out using the CyScribe cDNA Post-labeling Kit (Amersham Biosciences, UK) according to the manufacturer's recommendations and as reported earlier.⁵ The cDNA microarray slides used were produced at the Swegene DNA Microarray Resource Center, Department of Oncology, Lund University and contained 27 649 spots with sequence-verified IMAGE clones from the Research Genetics IMAGE clone library. The clone information was linked to gene names using build 180 of the Unigene database (<http://www.ncbi.nlm.nih.gov/UniGene>), and ~16 000 unique Unigene clusters were represented on the array.²²

Image Processing

On the scanned CGH arrays, the Gene Pix Pro 4.0 software (Axon Instruments Inc., Foster City, CA, USA) was used for identification of individual spots; thereafter, the quantified data matrix was loaded into the BioArray Software Environment (BASE).²³ Positive and nonsaturated spots were background-corrected using the median foreground minus the median background signal intensity for each channel, and log₂ ratios were calculated from the background-corrected intensities. Data were filtered for flagged features and signal to noise ratio >3 for each spot in both intensity channels. A lowess smooth factor of 0.33, delta of 0.1, and 4 iterations were used for popLowess normalization.¹⁹ A BASE adapted R (<http://www.r-project.org/>) implementation of CGH-Plotter²⁴ was used for automatic break point analysis. We defined gains and losses as log₂ ratios > +0.2 and < -0.2, respectively, and high-level amplification (HLA) as log₂ ratios > 1.5. BAC clones with > 11 missing values (79% presence required) across the 50 tumors were excluded from further analysis. Gained and lost regions were defined as described earlier²⁵: if two or more adjacent clones were separated by < 500 kb, the entire region spanned by the segments was considered as a single profile. For gene expression profiling, data spots, background correction, filtering, transformations, and analyses were performed using a local installation of the web-based BASE.²³ A preliminary filtering step eliminated all spots of poor quality flagged by GenePix™ Pro 4.1.1.4 version software (Axon instruments Inc., Foster City, CA, USA), and spots with diameter lesser than or equal to 60 μm, spots with > 10% pixel saturation and signal-to-noise ratio < 1.5 in either channel. The background-corrected intensity values were then normalized using the pin-based LOWESS method

to compensate for dye bias and local background effects.²⁶ Here, intensity-dependent adjustments (LOWESS fits) were performed within groups of 16 blocks to correct for spatial bias. Multiple print batches of slides were used, with 41 samples hybridized in replicates on different batches. All repeats clustered next to the first sample run irrespective of the differences in the print batch (data not shown), hence replicate assays were merged in a weighted manner, as described earlier.²² Within each slide, expression values for spots associated to the same gene symbol were merged in a similar weighted manner. After centering, the data were then transformed using an error model, as described earlier, to reduce the importance of poor-quality spots in later analysis steps.²² Filters for variation and presence of expression across hybridizations were set to reject all spots with a standard deviation of modified expression value lesser than 0.2 and with a presence in < 70% of the samples.

Data Analysis

A two-sample unpaired modified *t*-test was performed—significant analysis for microarrays (SAM)—using continuous log₂ ratios. Genomic clones showing significant copy number differences between the two tumor types were generated using 1000 permutations and a false discovery rate (FDR) ≤ 1% (TMeV version 4.0 is freely available at the TM4 microarray software suite website: <http://www.tm4.org/mev.html>). For ternary data, a two-sided Mann–Whitney *U* test was used to identify significant differences in copy number between groups. Unsupervised hierarchical cluster analysis was carried out in a TM4 microarray software suite using the average linkage-clustering algorithm with a Pearson correlation distance metric. The distance between samples was measured using the Pearson correlation distance metric. The Pearson correlation (centered, unsquared) coefficient *r* was measured for all pairs of samples, and the corresponding pair-wise distance was calculated, using BASE, as $d = 1 - r$. A rank-based statistical test²⁷ was used to identify reporters that discriminate between small/large (> 8 vs ≤ 8 cm), superficial/deep tumors, and the presence/absence of vascular invasion and necrosis.

The Cox proportional hazards model was used to assess survival in relation to altered reporters/clones (only log₂ values < -0.2 or > +0.2 were considered). Time to metastasis was analyzed in days. To avoid mass significance when fitting a large number of univariate proportional hazards models, we have considered a *P*-value of 0.01 or less as statistically significant. Only reporters that showed significant differences in the univariate analysis entered multivariate analysis. For metastasis-free survival correlations, segments with two or more adjacent reporters and two reporters or segments separated by < 500 kb were merged into one region for the multivariate analysis. For each region, the average log₂ ratio was calculated, and the data were binned into discrete log₂-transformed intervals to represent the gene copy number changes (binning criteria: > 0.2 = gain,

$-0.2 + 0.2 = \text{no change}$, $< -0.2 = \text{loss}$) as discrete events. This ternary scale was used for further multivariate analysis. The significant ($P < 0.05$) clinicopathologic variables and the genomic variables were forwarded in a stepwise selection procedure (entry only if $P \leq 0.05$ and removal only if $P \geq 0.10$). Proportional hazards assumptions were checked using Schoenfeld's test.²⁸ The final multivariate Cox model was evaluated using post-estimation analysis/prediction and the area under the receiver operating curve (ROC).

Gene Expression Analysis

The discriminatory gene lists, generated by SAM analysis, were analyzed using EASE software to functionally classify the genes and facilitate biological interpretations. The top-ranked genes were classified into groups within the categorical systems of the Gene Ontology (GO) Consortium (GO Biological Process and GO Molecular Function), the KEGG pathway, biochemical process, cellular role, and chromosomal regions. The EASE analysis used the top genes with an FDR $\leq 1\%$.

If not otherwise specified, all statistical computations were carried out using Stata SE software, version 9 (Stata Corporation, 2007, College Station, TX, USA).

aCGH and expression data for tumors will be available from NCBI's gene expression omnibus (GEO).

RESULTS

Copy Number Changes and HLAs

The aCGH profiles revealed multiple gains and losses in most samples (Figure 1a). The majority of the tumors had $>30\%$ of the genome altered, and in 16% (8/49) of the tumors, the gains and losses affected $>50\%$ of the genome. Highly complex genetic profiles, suggesting breakage-fusion-bridge cycles, were observed in six samples and sometimes affected more than one region.

Copy number changes involved all chromosomes with recurrent gains most frequently affecting 1p, 1q, 4p, 5p, 7p, 7q, 9q, 14q, 16p, 17p, 19p, 19q, 20q, and 21q and recurrent losses encompassing 1qter, 2pter, 2qter, 8p, 9p, 10p, 10q, 11qter, 13q, and 16q (Supplementary Table 1). Gains/losses of whole chromosome arms involved 1q, 2p, 3q, 5p, 7p, 7q, 8p, 8q, 10p, 13p, 14q, 15p, 15q, 16p, 16q, 17q, 18q, 19p, 19q, 20p, 21q, 22p, and 22q.

HLA were observed in 49 regions and homozygous deletions in 17 regions, without significant differences between UPS and LMS. Recurrent HLAs involved loci in 1p32.1, 1q24.2, 3p12.1, 3q23, 4q11, 12q13.3–12q14.1, and 12q15 (Supplementary Table 2). The most frequent HLA encompassed *GLI1*, *SAS*, and *CDK4* in 12q14.1 and *MDM2* on 12q14.3–q15, which were observed in 8–10% of the tumors.

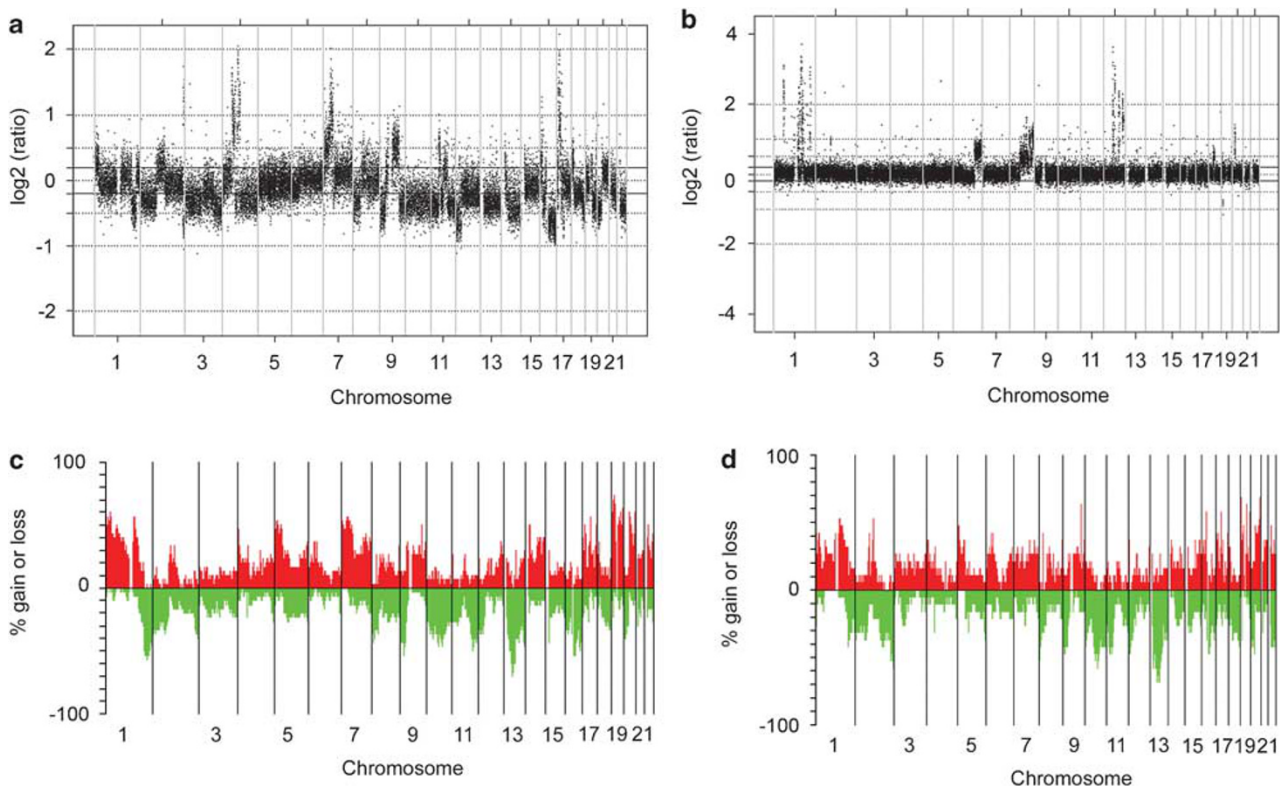


Figure 1 aCGH copy number changes profiles of (a) an LMS (sample 2201) with a complex profile containing multiple gains and losses and (b) a UPS with a simple genomic profile with the characteristic 12q HLA. Genome-wide frequency plots summarizing all gains (red) and losses (green) in (c) UPS and (d) LMS.

Five tumors (three UPS and two LMS) showed simple genetic profiles with few imbalances and a characteristic HLA in 12q15 corresponding to *MDM2* (Figure 1b). Of these tumors, four also showed concomitant HLA in 12q14 corresponding to *CDK4*. Multiple other HLAs were observed in this subset of tumors and affected 1q21, 1q23, 3p12–3p11, 3q22, 3q23, 6q22, 6q23, and 6q24–6q25.

The most common homozygous deletion, which was present in 16% of the tumors, encompassed the *RB1* locus in 13q14.2–14.3. Homozygous deletions of 9p21.3, harboring *CDKN2A/CDKN2B*, were detected in two tumors (Supplementary Table 2).

Comparison Between UPS and LMS

The gains/losses profiles in LMS and UPS showed strong similarity (Figure 1c and d), and the two tumor types were not separated by unsupervised hierarchical clustering (Figure 2a). The intratumor Pearson distance was 0.85 in LMS and 0.81 in UPS, compared with an intertumor distance of 0.83, which further supports their genetic similarity. A two-sided Mann–Whitney test (applied to ternary data) did not identify any reporters that significantly discriminated between these histotypes, whereas SAM (applied to continuous log₂ ratios) identified six reporters (mapping to 1q42.2, 2p11.1, two clones in 16p13.3, 13p13.13, and 16p13.11) that were differentially gained/lost.

The gene expression patterns of UPS and LMS showed strong similarity. Unsupervised hierarchical cluster analysis identified four major clusters, all containing both UPS and LMS (Figure 2b). SAM analysis identified nine genes, that is, *TAGLN3*, *D4S234E*, *KIAA1729*, *PDLIM5*, *TEAD3*, *TPM2*, *ALDH1B1*, *TRDMT1*, and *DHODH*, which discriminated between UPS and LMS, but EASE analysis of the latter genes did not reveal any biologically distinct groups therein. From the nine differentially expressed genes, only one (tropomyosin 2 beta) was related to muscle function, and five had clearly distinct expression levels, whereas the remaining four showed variable expression among the samples. SVM and leave one-out cross-validation were used to test the validity of such a nine-gene classifier, but as expected, the classifier performed poorly with a high misclassification rate.

Prognostic Correlations

Given that genomic profiles of the two histotypes were indistinguishable, further analysis was carried out using the entire dataset irrespective of histotype classification. To rule out a general prognostic impact of genetic complexity, the tumors were grouped in quartiles according to the number of clones with copy-number alterations. Metastasis-free survival was similar in the different groups (univariate Cox analysis, data not shown), which argues against such an effect. In univariate Cox analysis, necrosis was associated with metastasis ($P < 0.01$) (Supplementary Table 3A). Using the Cox proportional hazards model, we identified 87 BAC clones

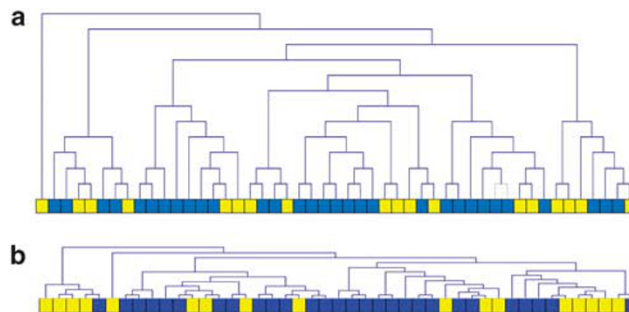


Figure 2 Unsupervised hierarchical clustering dendrogram based on log₂ ratios of (a) aCGH data and (b) gene expression data. LMS are shown in yellow and UPS are shown in blue.

Table 2 Significant variables in the multivariate stepwise Cox analysis

Variable	HR ^a	95% CI	P
Necrosis	13	(2.6–61)	0.002
Loss of 4q31	11	(3.2–38)	<0.0001
Loss of 18q22	0.06	(0.01–0.3)	0.001

^aPresence vs absence.

associated with metastasis-free survival ($P < 0.001$). The significant reporters were grouped into seven regions and herein univariate Cox analysis identified six regions that were significantly associated with metastasis-free survival (Supplementary Table 3B). In multivariate analysis, necrosis, and loss of 4q31 remained significant predictors of metastasis and loss of 18q22 a favorable prognostic factor. The 4q31 region encompasses five genes, including *SMAD1*, and the 18q22 region includes the *CDH7* locus (Table 2; Supplementary Figure 1). Combining the independent prognosticators (necrosis, loss of 4q31, and loss of 18q22) improved the area under the ROC curve for metastasis prediction, as calculated by post-estimation prediction, from 0.64 to 0.86 (Figure 3).

DISCUSSION

aCGH profiling in UPS and LMS reveals extensive genetic complexity in most tumors with median 34% of the genome altered, which is considerably more than the 12–20% reported in other malignancies.²⁹ Despite the complex genomic profiles, few aberrations were recurrent, with the most common recurrent aberration, gain of 19q13.11, present in 60% of the tumors. Gains/losses were identified in several regions, with the most common copy number gains in both UPS and LMS affecting 1p36.33–p31.3; 1q21.2–q24.3; 4p16.3; 5p15.33–p13.1; 7p22.3; 7p15.2–7p11.2; 7q32.1–q32.2; 9q34.3; 14q11.2; 14q32.33; 16p13.3; 17q12; 17q21.33;

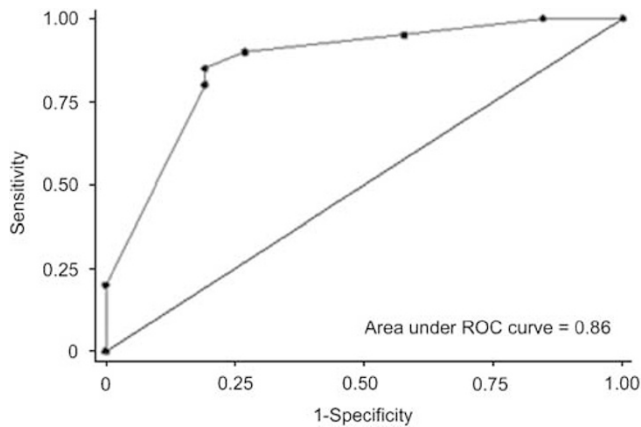


Figure 3 Area under the ROC curve calculated by post-estimation prediction. Combining the significant factors, tumor necrosis, loss of 4q31, and loss of 18q22, revealed an area under the ROC curve of 0.86 (s.e., 0.05; 95% confidence interval, 0.76–0.96).

17q23.3; 19p13.3; 19q13.11–q13.2; 19q13.42; 20q11.21–q13.33; and 21q22.3, and the most common losses affecting 1q32.1; 2p25.3; 2q36.1–q37.3; 8p23.3; 9p24.2–9p22.3; 9p21.3–p21.1; 10q21.1–q23.2; 11q22.3; 13q12.11–q31.1; 13q33.3; 16q11.2; and 16q23.1. The overall gains/losses are in accordance with studies that have applied cytogenetics and fluorescence *in situ* hybridization and showed extensive genetic alterations with recurrent numerical as well as structural aberrations, including losses of 1p36, 1q42–qter, 2p15–pter, 3p21–p23, 8p21–pter, 10q23–qter, 11q23–qter, 13q12–q13, 13q32–qter, and 18q11, and gains of 1q12–q31, in LMS.^{30–32} CGH has also been applied to LMS³³ with the most frequent losses affecting 2p, 10q, 11q, and 13q, and the most common gains encompassing 1q, 5p, 8q, 17p, and Xp.^{8,34} Application of aCGH additionally identified gains at 7p, 7q, 9q, 14q, 17q, 19p, and 20q, and allowed fine mapping of 1q21.2–24.3 (gain), 2p25.3 (loss), 4p16.3 (gain), 5p15.33–p13.1 (gain), and 11q22.3 (loss).

Deletion of 13q, affecting the entire 13q12.11–q31.1 chromosomal segment, was the most common deletion observed. Earlier studies have identified loss of 13q14–21 as the most frequent copy-number alteration in MFH/UPS and LMS, often with two separate clusters in 13q14 and 13q21.^{35–37} A large fraction of samples showed alterations within the RB1/CDK4/INK4A/INK4B and/or the TP53/ARF/MDM2 pathways with homozygous deletions of the *RB1* and *CDKN2A* (*INK4A/INK4B/ARF*) loci observed in eight and two cases, respectively. HLA of *MDM2* was identified in five cases with co-amplification of *GLI1*, *SAS*, and *CDK4* in four cases. Tumors harboring *CDK4/MDM2* amplifications had simple genomic profiles with fewer copy number changes than the average (<10 versus 30%). These profiles are reminiscent of the 12q14–15 amplicon associated with supernumerary ring chromosomes and giant rod markers,

described in well-differentiated liposarcoma/dedifferentiated liposarcoma. A blinded review of these and of five additional randomly selected cases was performed by an independent sarcoma pathologist (C.F.), but none of the cases were reclassified into lipomatous tumors (data not shown). Cytogenetic analysis revealed a ring chromosome 12 in one of the tumors, a UPS with myogenic differentiation (F Mertens, personal communication). Our data thus corroborate that the 12q14–15 amplicon is also found in a subset of pleomorphic STS.^{9,12,38,39} Despite the variations in size, STS with the 12q amplicon showed concomitant HLAs encompassing 1p32 (*JUN*) in one case, 1q23 and 1q24 in two cases, and 6q23–q24 (*MAP3KIP2*) in one case, suggesting a possibly shared oncogenic pathway linked to 12q14–15 amplification. Gene expression analysis showed significant over expression of four genes in the 12q amplicon: the angiogenic *RAP1B*, the p53 inactivating oncogene *YEATS4*, the vitamin D metabolism gene *CYP27B1*, and the nucleoporin *NUP107*, which participate in cell division. *MDM2* was not over expressed, which suggests that post-transcriptional mechanisms might be involved in its upregulation.^{40–42} *CDK4* was over expressed in all samples with *CDK4* HLA, but concomitant over expression of Cyclin D1, Cyclin D2, or of *MYC* was not observed. The recognition of a group of tumors with simple genetic changes and a 12q HLA likely reflects a biologically relevant subset evolving through a distinct pathway that could potentially be targeted with *CDK4/6* inhibitors.

The similarities between the genomic profiles of UPS and LMS are striking without significant differences when SAM and Mann–Whitney tests are applied. Unsupervised hierarchical clustering analysis did not differentiate UPS from LMS. SAM analysis identified nine differentially expressed genes, but application of these in EASE and in SVM did not allow distinction between the subtypes.

To validate the performance of the tests used to discriminate highly pleomorphic tumors, we included gene expression data from seven pleomorphic liposarcomas⁵ and separately analyzed these with the UPS and LMS samples. Unsupervised hierarchical clustering analysis clearly separated the liposarcomas from UPS as well as from LMS (data not shown). SAM analysis identified 251 discriminatory genes between liposarcomas and UPS and 908 discriminatory genes between LMSs and liposarcomas (data not shown).

The shift in diagnostic criteria for UPS complicates the interpretation/comparison with earlier studies because tumors formerly diagnosed as MFH could, with current diagnostic guidelines, be reclassified. Moreover, cytogenetic data indicate that differences in genomic aberrations may be related to tumor localization rather than to tumor morphology.⁴³ Despite these constraints, UPS and LMS have been shown to cluster together in earlier gene expression studies^{3–6,44} and to be indistinguishable in studies that have applied proteomic profiling.^{45,46} Histologic pleomorphism has been suggested to play a role with closer similarity

between pleomorphic LMS and UPS than within the LMS subtype in which nonpleomorphic tumors form a subcluster characterized by expression of tropomyosin isoforms.⁴⁶ Interestingly, tropomyosin 2 beta was one of the genes differentially expressed between UPS and LMS in our series. Hence, our results are in line with earlier reports, but specifically show that, in extremity STS, UPS, and LMS are genetically indistinguishable, suggesting a shared lineage of these two histotypes.

Data on the prognostic importance of genetic alterations in STS are limited. Cytogenetic predictors, including breakpoints in 1p1, 1q4, 14q1, and 17q2, and gain of regions 6p1 and 6p2, have been reported in a mixed series of STS and validated with the 1p1 breakpoint and with gain of 6p1, which are suggested to represent independent prognostic markers. In our series, loss of 4q31 correlated to poor prognosis, whereas loss of 18q22 predicted a favorable prognosis, both with an independent prognostic value (Table 2). Use of these independent prognostic markers in combination with necrosis improved prognostication from an ROC area of 0.64–0.86 (Figure 3). The 4q31 region encompasses four genes (*ANAPC10*, *ABCE1*, *HSH1N1*, and *SMAD1*), with the transcription factor *SMAD1* being an interesting target because of its involvement in the TGF-beta pathway. *SMAD1*, *SMAD5*, and *SMAD8* serve as substrates for BMP receptors, which have been suggested to play a role in tumor progression, although a possible role in STS remains to be clarified.⁴⁷ The 18q22 region associated with favorable prognosis harbors *CDH7*, which has been implicated in tumor progression in human melanoma, and the loci for cadherin-19 and cadherin-20 are located nearby.⁴⁸ One could speculate that deletion of the 18q22 region could compromise cadherin-mediated tumor growth and invasion.

These results encourage work applying genomic markers as adjunct prognosticators for the identification of high-risk tumors in patients who should be recommended adjuvant therapy.

The similarities in genomic and gene expression profiles shown herein suggest that, in the context of extremity STS, UPS may in fact correspond to highly pleomorphic LMS. Although specific diagnostic markers remain difficult to distinguish in these high-grade and genetically complex tumors, the identification of novel prognosticators, which can be linked to the currently used markers (necrosis), is encouraging and shows that molecular parameters can be used as a complement to histopathological risk estimates.

Supplementary Information accompanies the paper on the Laboratory Investigation website (<http://www.laboratoryinvestigation.org>)

ACKNOWLEDGEMENTS

The authors thank the orthopedic surgeons at the University Hospital in Lund, the Scandinavian Sarcoma Group registry for histopathological data (necrosis, vascular invasion, depth, and size), C1 Anna Isinger for valuable discussion, and Johan Staaf for running the poplowess algorithm in BASE.

AC has a fellowship from the Portuguese Foundation for Science and Technology (SFRH/BD/27561/2006).

- 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.
- Fletcher CU, Krishnan K, Mertens F (eds). WHO Classification of Tumours; Pathology and Genetics: Tumours of Soft Tissue and Bone. IARC Press: Lyon, France, 2002.
- Nielsen TO, West RB, Linn SC, *et al*. Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 2002;359:1301–1307.
- 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.
- Francis P, Namlos HM, Muller C, *et al*. Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. *BMC Genomics* 2007;8:73.
- Nakayama R, Nemoto T, Takahashi H, *et al*. Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. *Mod Pathol* 2007;20:749–759.
- Skubitz KM, Cheng EY, Clohisey DR, *et al*. Differential gene expression in liposarcoma, lipoma, and adipose tissue. *Cancer Invest* 2005;23:105–118.
- 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;81:211–215.
- Chibon F, Mariani O, Derre J, *et al*. A subgroup of malignant fibrous histiocytomas is associated with genetic changes similar to those of well-differentiated liposarcomas. *Cancer Genet Cytogenet* 2002;139:24–29.
- Fritz B, Schubert F, Wrobel G, *et al*. Microarray-based copy number and expression profiling in dedifferentiated and pleomorphic liposarcoma. *Cancer Res* 2002;62:2993–2998.
- Ohguri T, Hisaoka M, Kawachi S, *et al*. Cytogenetic analysis of myxoid liposarcoma and myxofibrosarcoma by array-based comparative genomic hybridisation. *J Clin Pathol* 2006;59:978–983.
- Heidenblad M, Hallor KH, Staaf J, *et al*. Genomic profiling of bone and soft tissue tumors with supernumerary ring chromosomes using tiling resolution bacterial artificial chromosome microarrays. *Oncogene* 2006;25:7106–7116.
- 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.
- Schurich W, Begin LR, Seemayer TA, *et al*. Pleomorphic soft tissue myogenic sarcomas of adulthood. A reappraisal in the mid-1990s. *Am J Surg Pathol* 1996;20:131–147.
- Oda Y, Miyajima K, Kawaguchi K, *et al*. Pleomorphic leiomyosarcoma: clinicopathologic and immunohistochemical study with special emphasis on its distinction from ordinary leiomyosarcoma and malignant fibrous histiocytoma. *Am J Surg Pathol* 2001;25:1030–1038.
- 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.
- Hill DA, O'Sullivan MJ, Zhu X, *et al*. Practical application of molecular genetic testing as an aid to the surgical pathologic diagnosis of sarcomas: a prospective study. *Am J Surg Pathol* 2002;26:965–977.
- Bridge JA. Contribution of cytogenetics to the management of poorly differentiated sarcomas. *Ultrastruct Pathol* 2008;32:63–71.
- Staaf J, Jonsson G, Ringner M, *et al*. Normalization of array-CGH data: influence of copy number imbalances. *BMC Genomics* 2007;8:382.
- Krzywinski M, Bosdet I, Smailus D, *et al*. A set of BAC clones spanning the human genome. *Nucleic Acids Res* 2004;32:3651–3660.
- Snijders AM, Nowak N, Segreaves R, *et al*. Assembly of microarrays for genome-wide measurement of DNA copy number. *Nat Genet* 2001;29:263–264.
- Fernebro J, Francis P, Eden P, *et al*. Gene expression profiles relate to SS18/SSX fusion type in synovial sarcoma. *Int J Cancer* 2006;118:1165–1172.

23. Saal LH, Troein C, Vallon-Christersson J, *et al*. BioArray Software Environment (BASE): a platform for comprehensive management and analysis of microarray data. *Genome Biol* 2002;3:SOFTWARE0003.
24. Autio R, Hautaniemi S, Kauraniemi P, *et al*. CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics* (Oxford, England) 2003;19:1714–1715.
25. Aguirre AJ, Brennan C, Bailey G, *et al*. High-resolution characterization of the pancreatic adenocarcinoma genome. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:9067–9072.
26. Yang YH, Dudoit S, Luu P, *et al*. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nucleic Acids Res* 2002;30:e15.
27. Liu Y, Ringner M. Multiclass discovery in array data. *BMC Bioinformatics* 2004;5:70.
28. Schoenfeld DA, Richter JR. Nomograms for calculating the number of patients needed for a clinical trial with survival as an endpoint. *Biometrics* 1982;38:163–170.
29. Chin SF, Wang Y, Thorne NP, *et al*. Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene* 2007;26:1959–1970.
30. Mandahl N, Fletcher CD, Dal Cin P, *et al*. Comparative cytogenetic study of spindle cell and pleomorphic leiomyosarcomas of soft tissues: a report from the CHAMP Study Group. *Cancer Genet Cytogenet* 2000;116:66–73.
31. Mertens F, Fletcher CD, Dal Cin P, *et al*. Cytogenetic analysis of 46 pleomorphic soft tissue sarcomas and correlation with morphologic and clinical features: a report of the CHAMP Study Group. *Chromosomes and morphology. Genes Chromosomes Cancer* 1998;22:16–25.
32. Wang R, Lu YJ, Fisher C, *et al*. Characterization of chromosome aberrations associated with soft-tissue leiomyosarcomas by twenty-four-color karyotyping and comparative genomic hybridization analysis. *Genes Chromosomes Cancer* 2001;31:54–64.
33. Packenham JP, du Manoir S, Schrock E, *et al*. Analysis of genetic alterations in uterine leiomyomas and leiomyosarcomas by comparative genomic hybridization. *Mol Carcinog* 1997;19:273–279.
34. Levy B, Mukherjee T, Hirschhorn K. Molecular cytogenetic analysis of uterine leiomyoma and leiomyosarcoma by comparative genomic hybridization. *Cancer Genet Cytogenet* 2000;121:1–8.
35. Mairal A, Terrier P, Chibon F, *et al*. Loss of chromosome 13 is the most frequent genomic imbalance in malignant fibrous histiocytomas. A comparative genomic hybridization analysis of a series of 30 cases. *Cancer Genet Cytogenet* 1999;111:134–138.
36. Weng WH, Lerner M, Grandner D, *et al*. Loss of chromosome 13q is a frequently acquired event in genetic progression of soft tissue sarcomas in the abdominal cavity. *Int J Oncol* 2005;26:5–16.
37. Chibon F, Mairal A, Freneaux P, *et al*. The RB1 gene is the target of chromosome 13 deletions in malignant fibrous histiocytoma. *Cancer Res* 2000;60:6339–6345.
38. Weaver J, Downs-Kelly E, Goldblum JR, *et al*. Fluorescence *in situ* hybridization for MDM2 gene amplification as a diagnostic tool in lipomatous neoplasms. *Mod Pathol* 2008;21:943–949.
39. 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.
40. Florenes VA, Maelandsmo GM, Forus A, *et al*. MDM2 gene amplification and transcript levels in human sarcomas: relationship to TP53 gene status. *J Natl Canc Inst* 1994;86:1297–1302.
41. Pollock R, Gilman M. Transcriptional activation: is it rocket science? *Proceedings of the National Academy of Sciences of the United States of America* 1997;94:13388–13389.
42. Bartel F, Taubert H, Harris LC. Alternative and aberrant splicing of MDM2 mRNA in human cancer. *Cancer Cell* 2002;2:9–15.
43. Sandberg AA. Updates on the cytogenetics and molecular genetics of bone and soft tissue tumors: leiomyosarcoma. *Cancer Genet Cytogenet* 2005;161:1–19.
44. 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.
45. Kawai A, Kondo T, Suehara Y, *et al*. Global protein-expression analysis of bone and soft tissue sarcomas. *Clin Orthop Relat Res* 2008;466:2099–2106.
46. Suehara Y, Kondo T, Fujii K, *et al*. Proteomic signatures corresponding to histological classification and grading of soft-tissue sarcomas. *Proteomics* 2006;6:4402–4409.
47. Kodach LL, Bleuming SA, Musler AR, *et al*. The bone morphogenetic protein pathway is active in human colon adenomas and inactivated in colorectal cancer. *Cancer* 2008;112:300–306.
48. Moore R, Champeval D, Denat L, *et al*. Involvement of cadherins 7 and 20 in mouse embryogenesis and melanocyte transformation. *Oncogene* 2004;23:6726–6735.