# Microarray-based Copy Number and Expression Profiling in Dedifferentiated and Pleomorphic Liposarcoma<sup>1</sup>

## Björn Fritz, Falk Schubert, Gunnar Wrobel, Carsten Schwaenen, Swen Wessendorf, Michelle Nessling, Christian Korz, Ralf J. Rieker, Kate Montgomery, Raju Kucherlapati, Gunhild Mechtersheimer, Roland Eils, Stefan Joos, and Peter Lichter<sup>2</sup>

Deutsches Krebsforschungszentrum, Abteilung Molekulare Genetik (H0700), D-69120 Heidelberg, Germany [B. F., G. W., C. S., M. N., C. K., S. J., P. L.]; Deutsches Krebsforschungszentrum, Arbeitsgruppe Intelligente Bioinformatiksysteme (H0900), D-69120 Heidelberg, Germany [F. S., R. E.]; Pathologisches Institut, Universität Heidelberg, D-69120 Heidelberg, Germany [R. J. R., G. M.]; Harvard-Partners Center for Genetics and Genomics and Harvard Medical School, Boston, Massachusetts 02115 [K. M., R. K.]; and Abteilung Innere Medizin III, D-89081 Ulm, Germany [S. W.]

### Abstract

Sixteen dedifferentiated and pleomorphic liposarcomas were analyzed by comparative genomic hybridization (CGH) to genomic microarrays (matrix-CGH), cDNA-derived microarrays for expression profiling, and by quantitative PCR. Matrix-CGH revealed copy number gains of numerous oncogenes, *i.e.*, *CCND1*, *MDM2*, *GL1*, *CDK4*, *MYB*, *ESR1*, and *AIB1*, several of which correlate with a high level of transcripts from the respective gene. In addition, a number of genes were found differentially expressed in dedifferentiated and pleomorphic liposarcomas. Application of dedicated clustering algorithms revealed that both tumor subtypes are clearly separated by the genomic profiles but only with a lesser power by the expression profiles. Using a support vector machine, a subset of five clones was identified as "class discriminators." Thus, for the distinction of these types of liposarcomas, genomic profiling appears to be more advantageous than RNA expression analysis.

### Introduction

LSs<sup>3</sup> represent the most common soft tissue tumors in adults and account for  $\sim 20\%$  of all mesenchymal malignancies. The tumors are characterized by a high morphological diversity and a great variation in biological behavior. According to the WHO classification, five morphological liposarcoma subtypes are distinguished, *i.e.*, WLS, myxoid, round cell, PL, and DL. Progression ranges from nonmetastasizing neoplasms such as WLSs to LSs with overt metastatic potential, such as round cell LS, DL, and PL (1). On the basis of recent molecular studies, DL is discussed as a variant of WLS, characterized by a stepwise morphological progression and by ring or large marker chromosomes (2, 3). Only little is known about recurrent cytogenetic aberrations in PL, and therefore it remains unclear whether PL is contained in the genetic WLS/DL continuum or whether it represents an independent pathogenetic entity (4).

Characterization of chromosomal copy number changes as well as gene expression analysis provides evidence for pathogenic mechanisms involved in tumor etiology or development. Furthermore, they have the potential to identify novel diagnostic and prognostic molecular markers. In the present study, we applied two high-throughput screening methods, *i.e.*, matrix-CGH allowing detection of genomic imbalances with high resolution and DNA chip-based gene expression profiling, as well as quantitative PCR for the analysis of eight DLs and eight PLs. This led to the identification of new candidate genes involved in the pathogenesis of these tumors. Furthermore, cluster and classification algorithms allowed us to discriminate between both LS subtypes by the pattern of genomic aberrations and revealed a minimal set of genes required for this distinction.

## Materials and Methods

**Tumor Samples, Genomic DNA, and RNA.** Diagnosis of the tumors was based on standard histopathological criteria in conjunction with immunochemical analysis according to the WHO classification (1). All tumors were classified as grade III except cases DL41 and DL45, which were grade II. For matrix-CGH, genomic DNA was isolated using a standard phenol/chloroform extraction protocol. Genomic DNA from blood of healthy donors served as control. For cDNA microarray analysis, total RNA was isolated from homogenized samples using Trizol reagent (Life Technologies, Inc.). mRNA isolation was performed using a poly(A)+ mRNA isolation kit (Qiagen) according to the manufacturer's instructions. A pool of mRNAs derived from different tissues (human total RNA panel I; Clontech) was used as a reference. To identify tumor-associated changes in gene expression, RNA isolated from healthy adipocytic tissue was used.

**Matrix-CGH.** Our recently designed DNA chip consisting of 188 DNA fragments (5) was extended by a set of 112 sequence-verified and fluorescence *in situ* hybridization-mapped BAC fragments localized on chromosome 12q. Arraying, probe labeling, and hybridization, as well as digital imaging and data analysis, were performed as described recently (5).

**cDNA Microarray.** Microarrays containing 1600 nonredundant, sequencevalidated human cDNAs, selected on their putative involvement in tumorigenesis, were processed essentially as described (6, 7). Approximately 1  $\mu$ g of tumor and 1  $\mu$ g of reference poly(A)+ mRNA were Cy3 or Cy5 labeled using the Omniscript Reverse Transcriptase kit (Qiagen) and hybridized with 10  $\mu$ g of *Cot*I DNA, 30  $\mu$ g of tRNA, and 10  $\mu$ g of poly(T) oligonucleotides in an automated hybridization chamber (GeneTac; Genomic Solutions). For all samples, we performed color switch experiments, where test and control DNA were labeled via Cy3 and Cy5, respectively, and *vice versa*. Image acquisition was done using the Axon 4000 reader (Axon Instruments). For image analysis, GenePix imaging software 3.0 was used. The normalization was done following Beissbarth *et al.* (8), but low-intensity signals were kept (9). Ratio values were computed, and ratios from low-intensity signals (15% quantile) or missing values were treated as normally expressed. The geometric mean was calculated from color shift experiments.

**RQ-PCR.** First-strand cDNA synthesis was performed using random hexamer primers (RNA PCR System; PE Applied Biosystems), and RQ-PCR was carried out using TaqMan Universal Master Mix and SYBR Green PCR Core Reagents according to the protocols provided by the manufacturer (PE Applied Biosystems). Each cDNA sample was analyzed in triplicate using the ABI PRISM 7700 Sequence Detector. To obtain a calibration graph, cDNA from Universal RNA (Stratagene) was serially diluted eight times in H<sub>2</sub>O at a ratio of 2:1 and measured in every single RQ-PCR run. To standardize the amount of sample cDNA, two endogenous control amplicons were used, the house-

Received 3/12/02; accepted 4/18/02.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

<sup>&</sup>lt;sup>1</sup> Supported by the Bundesministerium für Bildung und Forschung (BMBF, FKZ 01 KW 9937; NGFN KB PO6T7, Tumorzentrum Heidelberg/Mannheim (No. I./I.2.), Deutsche Krebshilfe (No. 70-2312 Be2) and IZKF C12.

<sup>&</sup>lt;sup>2</sup> To whom requests for reprints should be addressed, at Deutsches Krebsforschungszentrum, H0700, D-69120 Heidelberg, Im Neuenheimer Feld 280, Germany. Phone: 49-6221-424619; Fax: 49-6221-424639; E-mail: p.lichter@dkfz-heidelberg.de.

<sup>&</sup>lt;sup>3</sup> The abbreviations used are: LS, liposarcoma; WLS, well-differentiated LS; PL, pleomorphic LS; DL, dedifferentiated LS; RQ-PCR, real-time quantitative reverse transcription-PCR; BAC, bacterial artificial chromosome; HSP90, heat shock protein 90.

Fig. 1. Results from matrix-CGH analysis. A, seudocolor image of a hybridization of case DL10 (tumor labeling, Cy3; control labeling, Cy5) onto the matrix-CGH microarray consisting of an oncogene subarray of 188 BAC or P1-derived artificial chromosome clones and a chromosome 12q subarray of 112 BAC clones. B, copy number changes of oncogenes and tumor suppressor genes in 8 DL and 8 PL. HLA, high level amplification; LLA, low level amplification. Max.amp., maximum amplitude, C. normalized copy-number ratios (ordinate, log2 space) of genomic DNA of sample DL10 and normal reference DNA Clones in the oncogene and chromosome 12q subarray are ordered by position in the genome from 1p to Xq and from 12q cen to 12q ter, respectively.



keeping gene coding for phosphoglycerate kinase 1 (*PGK1*) and lamin b1 gene (*LMNB1*; SYBR Green). The average value of *PGK1* and *LMNB1* amplicons served as calibrator for a first normalization. Thereafter, the patient data were normalized using control data (human total RNA panel I, Clontech; and skeletal muscle, Stratagene) yielding the relative expression of each gene.

**Cluster Analysis and Classifying Algorithms.** Matrix-CGH clustering was performed using an implementation of the decision tree algorithm C5.0 in Clementine<sup>4</sup> and a predefined number of four clusters (Fig. 3). The principal component analysis was done using J-Express 2.01 d.<sup>5</sup> Detection of genes with predictive value from matrix-CGH (data not shown) and cDNA measurements (Fig. 4*B*) was performed using an implementation of the algorithm C5.0 in Clementine and the support vector machine SVM<sup>light.6</sup> The *P* for the detected combinations of genes was calculated using a permutation test (null hypothesis: measurements distributed at random between the tumors; test statistic: sum of the absolute values of the measurement from the decision hyperplane). For tumor type assignment of the two initially unclassified samples PL39 and DL48 (Fig. 4*B*), we used a support vector machine.

The classification error was calculated as follows. Because of the sparse data in the full 228-dimensional space (all clones), we used the distance of the classified samples to the separating hyperplane for density probability estimation of the two classes and assumed normal distribution. According to the distance of the unclassified samples to the separating hyperplane and the

## Results

**Genomic Profile.** Copy number changes of oncogenes and tumor suppressor genes in the eight DL and eight PL analyzed are indicated in Fig. 1, *A* and *B*.<sup>7</sup> High-level amplifications (intensity ratio >2) were detected for *CCND1*, *CCND2*, *MYB*, *MDM2*, *GLI*, and *CDK4*. The highest ratio values were observed for *MDM2*, *GLI*, and *CDK4* localized on chromosomal subregion 12q13–q15 and were present in all DL tumors. Copy number variations along the long arm of chromosome 12 were determined using a set of 112 BACs with a genomic spacing of roughly 1 MB (Fig. 1*C*). Interestingly, highly amplified clones (*i.e.*, RP11-1007B5 and RP11-1143G9) are interspersed by clones that are balanced, revealing a noncontiguous amplicon structure. Within chromosome band 20q13, low-level gains were detected in four of eight PLs and two of eight DLs including *BCAS1* (breast cancer amplified sequence 1), *ZNF217* (zinc finger gene 217), and *AIB1* (amplified in breast cancer 1).

density probability estimations of the two classes, we calculated the error probabilities using Bayes's rule.

<sup>&</sup>lt;sup>4</sup> Internet address: www.spss.com.

<sup>&</sup>lt;sup>5</sup> Internet address: www.molmine.com

<sup>&</sup>lt;sup>6</sup> Internet address: http://svmlight.joachims.org.

<sup>&</sup>lt;sup>7</sup> A complete list of the matrix-CGH results will be seen at http://www.dkfz-heidelberg.de/kompl\_genome/Other/mCGH-analysis.zip.



DL47

Fig. 2. Results from cDNA expression analysis. *A*, pseudocolor image of a cDNA microarray hybridization of sample DL42. *Inset*, enlarged portion of the array containing cyclin-dependent kinase 4 (*CDK4*) and HRK cDNAs. *B*, list of cDNAs that are differentially expressed (>10-fold change) in PL and DL. *C*, expression level of CDK4, CCND1, HRK, and DLK according to RQ-PCR (*ordinate*, log<sub>10</sub> space).



**RQ-PCR.** To assess the microarray data, a subset of amplified or up-regulated genes was analyzed in further detail using RQ-PCR (Fig. 2*C*). In general, chip-based and RQ-PCR-based expression data correlated well. The mean expression rate of DL as compared with PL of *CDK4* was 38-fold versus 3-fold (P = 0.01 according to Student's *t* test), of *CCND1* 3-fold versus 10-fold (P = 0.07), and of the dualleucine zipper kinase gene *DLK* 3-fold versus 9-fold (P = 0.04). A strongly reduced expression level of the putative tumor suppressor *HRK* representing an apoptosis activator was observed, *i.e.*, 68-fold in DL and 54-fold in DL by cDNA profiling and 12-fold by RQ-PCR (10-fold in DLs and 14-fold in PLs). Within the set of analyzed genes, no obvious correlation between DNA copy number and RQ-PCR data

PL33 PL32 PL11

PL6

DL42 DL29 DL28 DL10 PL40 DL47

PL33 PL32 PL40

DL10

DL42 DL29 DL28

PL11

PL6



Fig. 3. Cluster analysis of eight PL and eight DL cases using the TwoStep algorithm. Note that only samples are clustered (X axis), whereas the clones are ordered by their genomic position (Y axis). Dark green, high-level amplification ( $\log_2 1$ ); light green, low-level amplification ( $\log_2 0.32$ ); red, underrepresentation ( $\log_2 -0.41$ ); black, balanced ( $\log_2 0$ ); white, unknown.

was noticed. Candidate genes for breast carcinoma *BCAS1*, *ZNF217* and *AIB1* located on chromosome 20q13, frequently amplified in PL and DL, showed no significantly elevated transcript level.

**Clustering and Classification Analysis.** To test whether DL and PL can be subdivided into different groups based on the genomic aberration pattern, two-step cluster analysis and principal component analysis were performed. This revealed an unequivocal separation of both subtypes, *i.e.*, DL and PL strongly correlate with a different aberration pattern (Figs. 3 and 4A). For classification of the tumors, support vector machine algorithms were used to deduce the rules, which a distinct clone or a combination of clones have to meet, to act as a classifier. Fig. 4*B* depicts a classification scheme in which tumors with positive decision values are classified as PL and those with negative decision values as DL. Using a class predictor algorithm established by the identification of such discriminating ratio values, we were able to correctly classify two LS tumors of unknown pathological classification at the time of the analysis (tumors PL39 and

DL48). PL39 was scored as PL, which corresponds to the previous clinicopathological classification. DL48 was assessed as DL with a decision value close to 0, *i.e.*, close to the classification border between DL and PL. Detailed pathological investigation revealed that this tumor represented a DL but was mixed with cells of pleomorphic appearance. The error probability for wrong classification was <5% via Bayes's rule.

An analogous cluster analysis based on the expression profile of 1600 cDNAs revealed that the pleomorphic liposarcomas PL6, PL11, PL32, PL33, and PL40 were grouped in one cluster, whereas all dedifferentiated liposarcomas as well as PL27, PL30, and PL31 were grouped in a second cluster. Thus, the expression data did not correlate well with the pathological diagnosis of the tumors. To identify individual markers with diagnostic potential, the gene expression profiles of DL and PL were screened for genes that are typically differentially expressed in both subtypes. Best discrimination between PL and DL was achieved by combination of the expression data of *SCAP* and *HSP90* (P = 0.02; Fig. 4*C*) and of *DAPK* (death-associated protein kinase 1) and *calgranulin B* (P = 0.05).

## Discussion

In the present study, numerical alterations of genomic loci and the expression of genes with oncogenic potential were assessed in DL and PL. Using DNA-chip technology and quantitative PCR, a number of genes possibly relevant for the development of these tumors were identified. Cluster analysis revealed an unequivocal discrimination between DL and PL based on their pattern of genomic imbalances. Several "class predictors" were identified including, in particular, genes located on chromosomal subregion 12q13-q15. This region is frequently amplified in soft tissue tumors including WLS (10) and was found in the present study amplified in all DLs but not in PLs. Applying matrix-CGH with >100 DNA fragments spanning the entire long arm of chromosome 12, the structure of the 12q amplicon was delineated at high resolution. A characteristic discontinuous amplified region was discovered, which was interrupted by two defined balanced regions in all DL. It is possible that this specific architecture corresponds to a distinct mechanism of 12q amplifications, which might be involved in the formation of ring chromosome 12 known to occur in this tumor type (11).

Within the 12q amplicon, MDM2, GLI, and CDK4 were found coamplified in DL. This corresponds to results obtained from welldifferentiated LS and therefore suggests the molecular relationship of both subtypes (2, 10). Besides these well-known candidate genes, there is evidence for another relevant gene from this amplified region, previously not considered to play a role in LPs, the dual-leucine zipper kinase gene DLK (12q13). This gene, recurrently found overexpressed, is involved in the control of cell growth and differentiation, particularly during the process of tissue remodeling (12). MUK, the rat homologue of DLK, functions as an activator of the JNK pathway, and results in the accumulation of hyperphosphorylated c-Jun oncogene product (13). Expression analyses also revealed an interesting candidate tumor suppressor gene, which might play a role in LS tumorigenesis, i.e., the activator of apoptosis harakiri (HRK). This gene selectively interacts with the survival-promoting proteins BCL2 and BCLXL (14) and was found strongly down-regulated in all PLs and DLs analyzed.

Two other chromosomal regions, 6q25 and 20q13, were frequently involved in genomic overrepresentations. On 6q25 additional copy numbers of the *MYB* oncogene and of the estrogen receptor gene *ESR1* were found to be amplified. Gain of the long arm of chromosome 6 was shown previously in different types of soft tissue sarcomas (15). Interestingly, *ESR1* on 6q25 and genes for steroid receptor

## A





coactivators *AIB1* and *BCAS1* on 20q13 are coamplified or overexpressed in breast carcinoma. Recent data indicate that estrogendependent mitogenic stimulation of breast cancer cells are conferred by cooperation of *AIB1* with *ESR1* (16). This enhances the functional interaction of *ESR1* with the *CCND1* promoter, resulting in increased transcription of *CCND1* (17). In the present study, overexpression of *CCND1* was observed in 80% of the DLs and PLs. Thus, in breast carcinoma and LS, similar biochemical pathways could be activated.

Whereas cluster and classification analysis using expression data has come of age and is broadly used, we here show the potential of matrix-CGH data to better discriminate between DL and PL. The best discriminators were derived from DNA fragments mapping on chromosome 12q including *CDK4*, *MDM2*, and *GLI* and RP11-1143G9. The successful performance of this algorithm was demonstrated by two blind samples, which were correctly classified with a confidence interval of >95%. Amplification of these clones together with the characteristic amplicon structure found in DL not only allows a clear discrimination from PL but strongly argues for different pathogenic mechanisms leading to these LS subtypes.

Clustering on the basis of the expression of 1600 genes allowed a separation in a pleomorphic and dedifferentiated group, however, with few exceptions. This might be partly attributable to a relatively low proportion of disease-specific genes within the cDNA chip described. However, expression scores provided useful discriminators including the combination of heat shock protein *HSP90* and the adaptor protein gene *SCAP*, which were both observed to be more highly expressed in PL *versus* DL. Interestingly, these genes are located on the same chromosomal subband, *i.e.*, 1q22, which was shown recently to be significantly more often overrepresented in PL as compared with DL by chromosomal CGH (4). In addition, the presence of 1q22–q24 overrepresentation was shown to be associated with the acquisition of higher malignancy in soft tissue tumors (18).

#### Acknowledgments

We thank Daniel Göttel for his technical assistance.

## References

- Weiss, S. W., and Sobin, L. H. Histological Typing of Soft Tissue Tumours. Berlin: Springer Verlag, 1994.
- Dei Tos, A. P. Liposarcoma: new entities and evolving concepts. Ann. Diagn. Pathol., 4: 252–266, 2000.
- Mentzel, T. Biological continuum of benign, atypical, and malignant mesenchymal neoplasms—does it exist? J. Pathol., 190: 523–525, 2000.
- Rieker, R., Joos, S., Bartsch, C., Willeke, F., Schwarzbach, M., Otano-Joos, M., Ohl, S., Högel, J., Lehnert, T., Lichter, P., Otto, H. F., and Mechtersheimer, G. Distinct chromosomal imbalances in pleomorphic and in high grade dedifferentiated liposarcomas. Int. J. Cancer, 99: 68–73, 2002.
- Wessendorf, S., Fritz, B., Wrobel, G., Nessling, M., Lampel, S., Goettel, D., Kuepper, M., Joos, S., Hopman, T., Kokocinski, F., Doehner, H., Bentz, M., Schwaenen, C., and Lichter, P. Automated screening for genomic imbalances using matrix-based comparative genomic hybridization (matrix-CGH). Lab. Investig., 82: 47–60, 2002.
- Shalon, D., Smith, S. J., and Brown, P. O. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res., 6: 639–645, 1996.
- DeRisi, J. L., Iyer, V. R., and Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. Science (Wash. DC), 278: 680–686, 1997.
- Beissbarth, T., Fellenberg, K., Brors, B., Arribas-Prat, R., Boer, J. M., Hauser, N. C., Scheideler, M., Hoheisel, J. D., Schütz, G., Poustka, A., and Vingron, M. Processing and quality control of DNA array hybridization data. Bioinformatics, *16*: 1014–1022, 2000.
- Fellenberg, K., Hauser, N. C., Brors, B., Neutzner, A., Hoheisel, J. D., and Vingron, M. Correspondence analysis applied to microarray data. Proc. Natl. Acad. Sci. USA, 98: 10781–10786, 2001.
- Pilotti, S., Della Torre, G., Lavarino, C., Sozzi, G., Minoletti, F., Vergani, B., Azzarelli, A., Rilke, F., and Pierotti, M. A. Molecular abnormalities in liposarcoma: role of MDM2 and CDK4-containing amplicons at 12q13–22. J. Pathol., *185:* 188–190, 1998.
- Pedeutour, F., Forus, A., Coindre, J. M., Berner, J. M., Nicolo, G., Michielis, J. F., Terrier, P., Ranchere-Vince, D., Collin, F., Myklebost, O., and Turc-Carel, C. Structure of the supernumerary ring and giant rod chromosomes in adipose tissue tumors. Genes Chromosomes Cancer, 24: 30–41, 1999.

- Hebert, S. S., Daviau, A., Grondin, G., Latreille, M., Aubin, R. A., and Blouin, R. The mixed lineage kinase DLK is oligomerized by tissue transglutaminase during apoptosis. J. Biol. Chem., 275: 32482–32490, 2000.
- Hirai, S., Izawa, M., Osada, S., Spyrou, G., and Ohno, S. Activation of the JNK pathway by distantly related protein kinases, MEKK and MUK. Oncogene, *12*: 641–650, 1996.
- Inohara, N., Ding, L., Chen, S., and Nunez, G. Harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-X(L). EMBO J., 16: 1686–1694, 1997.
- El-Rifai, W., Sarlomo-Rikala, M., Knuutila, S., and Miettinen, M. DNA copy number changes in development and progression in leiomyosarcomas of soft tissues. Am. J. Pathol., 153: 985–990, 1998.
- Reiter, R., Wellstein, A., and Riegel, A. T. An isoform of the coactivator aib1 that increases hormone and growth factor sensitivity is overexpressed in breast cancer. J. Biol. Chem., 276: 39736–39741, 2001.
- Planas-Silva, M. D., Shang, Y., Donaher, J. L., Brown, M., and Weinberg, R. A. AIB1 enhances estrogen-dependent induction of cyclin D1 expression. Cancer Res., 61: 3858–3862, 2001.
- Forus, A., Larramendy, M. L., Meza-Zepeda, L. A., Bjerkehagen, B., Godager, L. H., Dahlberg, A. B., Saeter, G., Knuutila, S., and Myklebost, O. Dedifferentiation of a well-differentiated liposarcoma to a highly malignant metastatic osteosarcoma: amplification of 12q14 at all stages and gain of 1q22–q24 associated with metastases. Cancer Genet. Cytogenet., 125: 100–111, 2001.