

# Gene expression profiling for the investigation of soft tissue sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers

Andrew H. Beck · Robert B. West · Matt van de Rijn

Received: 8 December 2008 / Revised: 31 March 2009 / Accepted: 14 April 2009  
© Springer-Verlag 2009

**Abstract** Soft tissue sarcomas are malignant neoplasms derived from mesenchymal tissues. Their pathogenesis is poorly understood and there are few effective treatment options for advanced disease. In the past decade, gene expression profiling has been applied to sarcomas to facilitate understanding of sarcoma pathogenesis and to identify diagnostic, prognostic, and predictive markers. In this paper, we review this body of work and discuss how gene expression profiling has led to advancements in the understanding of sarcoma pathobiology, the identification of clinically useful biomarkers, and the refinement of sarcoma classification schemes. Lastly, we conclude with a discussion of strategies to further optimize the translation of gene expression data into a greater understanding of sarcoma pathogenesis and improved clinical outcomes for sarcoma patients.

**Keywords** Gene expression profiling · Microarrays · Genomics · Bioinformatics · Tumor biology · Soft tissue tumors · Sarcomas · Biomarkers · Molecular pathology · Surgical pathology

## Introduction

Sarcomas are malignant neoplasms derived from mesenchymal tissues. They represent approximately 1% of malignancies diagnosed annually. Sarcomas are currently diagnosed based on histopathological evaluation supplemented with the use of immunohistochemistry and molecular diagnostic techniques in selected cases [1]. There are ~100 recognized histopathological subtypes of soft tissue tumors [2]. The rarity of these tumors and the large number of diagnostic entities make sarcomas a challenging area of investigation.

Currently, treatment for sarcomas consists of surgery with consideration for chemotherapy and radiotherapy. The growing importance of comprehensive treatment regimens involving neoadjuvant chemotherapy [3] and preoperative radiotherapy [4] emphasizes the importance of pathological analysis of preoperative biopsy material. The strongest predictor of patient outcome is stage at time of diagnosis. Treatment options are limited and few effective adjuvant therapies exist for advanced disease. Due to the poor prognosis and limited effective therapeutic options available for most sarcomas, there is a major clinical need for increased understanding of sarcoma pathogenesis to facilitate the development of new diagnostic markers and therapeutic agents [5, 6].

Gene expression microarray technology was developed in the mid-1990s to permit genome-wide monitoring of gene expression [7–9]. The principal types of array platforms used for gene expression monitoring involve the robotic deposition of DNA fragments onto a glass slide [7] or the in situ synthesis of oligonucleotides on high-density arrays using photolithography [10]. Each of these technologies is able to produce genome-wide measurements of expression to produce a “snap shot” of a tumor’s gene expression at a specific point in time.

---

A. H. Beck · R. B. West · M. van de Rijn (✉)  
Pathology Department, Stanford University Medical Center,  
300 Pasteur Drive,  
Stanford, CA 94305, USA  
e-mail: mrijn@stanford.edu

R. B. West  
Pathology and Laboratory Service,  
Palo Alto Veterans Affairs Health Care System,  
Palo Alto, CA, USA

Standard molecular biology techniques (Southern blotting, Northern blotting, Western blotting, polymerase chain reaction, sequencing) and ancillary molecular techniques routinely utilized by pathologists (immunohistochemistry, fluorescence in situ hybridization), typically measure one biomarker at a time, and consequently relatively little statistical analysis is required to interpret the result. In contrast, a gene expression profiling experiment measures the expression of thousands (10,000–500,000) of transcripts in parallel. The enormous amount of quantitative data produced by a single sample measured on a gene expression microarray requires the application of statistical procedures to standardize the data and determine the significance of the observed findings [11]. In recent years, the statistical community has adopted recommendations for biomedical researchers working with these technologies [11, 12]. These recommendations hope to facilitate standardization of the primary steps of microarray data analysis (design, preprocessing, inference, classification, and validation). Following application of statistical procedures to the dataset, it is useful to apply additional analytic tools to integrate the findings with databases containing biomedical annotation to facilitate translation of the microarray data into biologically and clinically useful knowledge [13–16].

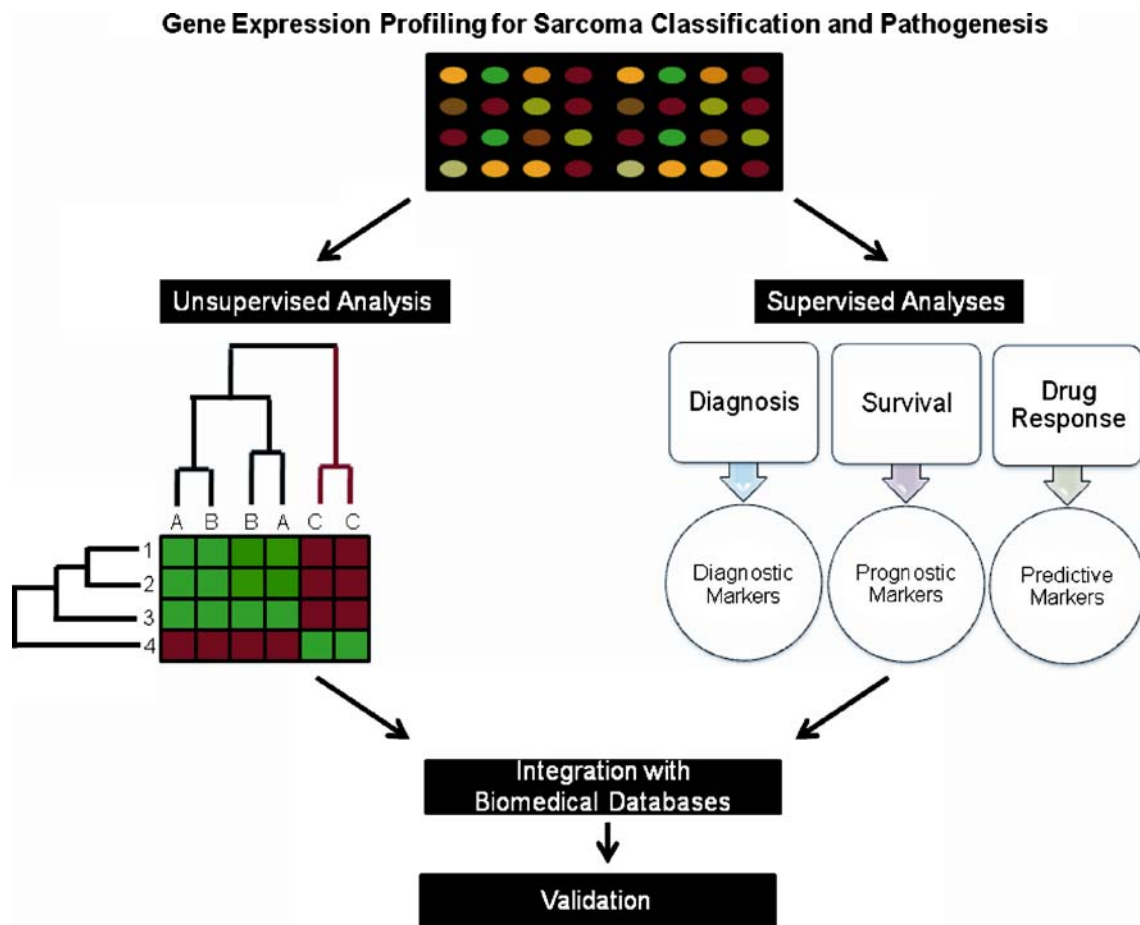
It is hoped that application of this technology to sarcomas will afford a greater understanding of their biology and facilitate the development of more effective diagnostic and treatment regimens to improve outcomes for sarcoma patients. The primary goals of gene expression profiling in sarcomas are to: (1) gain a deeper understanding of the aberrant biological pathways contributing to sarcomagenesis, (2) refine the classificatory scheme for sarcomas to better reflect the underlying biology of the tumors, and (3) identify markers to predict diagnosis, patient outcome, and response to therapy (Fig. 1) [17]. In the following sections, we review studies utilizing gene expression profiling in soft tissue sarcoma.

## Synovial sarcoma

Synovial sarcomas compose 5–10% of all soft tissue sarcomas. They are aggressive sarcomas and distant metastasis is the major cause of mortality [18]. The tumors are defined molecularly by a specific t(X;18)(p11.2;q11.2) translocation which fuses *SS18* (*SYT*) on chromosome 18 with either *SSX1*, *SSX2*, or *SSX4* on the X chromosome [19, 20]. The aberrant pathways induced by this translocation are incompletely understood [21]. Due in part to limited understanding of synovial sarcoma pathogenesis, clinical management consists primarily of surgery and radiotherapy and no effective targeted therapies are currently available [18].

Gene expression profiling has been performed to better characterize the tumor's pathogenesis. In most profiling studies, synovial sarcomas showed a distinct pattern of gene expression and clustered together (Fig. 2) [22–27]. In several other reports, some synovial sarcomas showed a similar pattern of gene expression to malignant peripheral nerve sheath tumor, resulting in a subset of synovial sarcomas clustering with malignant peripheral nerve sheath tumors (MPNSTs) [28, 29]. Taken together, these studies support that synovial sarcoma represents a relatively homogenous sarcoma subtype, whose gene expression profile shows the most similarity to malignant peripheral nerve sheath tumor.

Several groups have published lists of genes highly expressed in synovial sarcoma. Francis et al. [27] examined 32 synovial sarcomas in a study that included 177 sarcomas spanning 13 histological subtypes. Of these 13 subtypes, they found that synovial sarcoma showed the most distinct gene expression signature. They identified 4,000 genes differentially expressed in synovial sarcoma (false discovery rate=11%), including the SS chromosome X breakpoint genes *SSX1* and *SSX3* and genes involved in: neural differentiation (*FNA1*, *NCAMI*, *NEDD5*, *NPDC1*, and *OLFM1*), the retinoic acid pathway (*RARA*, *RARG*, *MDK*, *MEIS1*, and *PRAME*), and the epidermal growth factor (EGF) and fibroblast growth factor (FGF) receptor signaling pathways (*ERBB2*, *FGFR1*, *FGFR3*, *FGF18*, and *FRAG1*). Nielsen et al. [22] performed gene expression profiling on eight monophasic synovial sarcomas in a study that contained a total of 41 soft tissue tumors spanning six histologic subtypes. They identified a cluster of genes (including *SSX*, retinoic acid pathway genes, and epidermal growth factor receptor (*EGFR*)) which showed high levels of expression in synovial sarcoma. Nielsen et al. subsequently validated these gene microarray findings on a tissue microarray and showed that most synovial sarcomas showed protein expression of EGFR and SALL2 [30]. This group subsequently showed that the marker TLE1, which was identified through gene expression profiling experiments, was an effective diagnostic immunohistochemical marker for synovial sarcoma [31]. Allander et al. performed gene expression profiling on 14 synovial sarcomas, four malignant fibrous histiocytomas, and one fibrosarcoma and found that synovial sarcomas showed increased expression of *ERBB2* and *IGFBP2* and *IGF2* [26]. They validated and extended this finding by performing immunohistochemistry on a tissue microarray containing 37 synovial sarcomas. In this analysis, they noted that strong protein expression of *ERBB2* and *IGFBP2* localized to the glandular epithelial component of biphasic synovial sarcomas and to the solid epithelioid areas of some monophasic synovial sarcomas, suggesting that these proteins are involved in epithelial differentiation in synovial sarcoma [26]. Nagayama et al. [28] performed gene expression profiling on a total of 47



**Fig. 1** Schematic overview of the use of gene expression profiling for the investigation of sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers. Following collection of biological samples for analysis, gene expression microarray experiments begin with the hybridization of cDNA or cRNA to a microarray chip, producing tens of thousands of gene expression measurements per sample. Following data collection, the most commonly used data analysis techniques are: unsupervised two-way clustering of samples and genes to discover novel biologic pathways observed in sarcoma subtypes and to discover new relationships within and between sarcoma subtypes. Supervised analyses complement

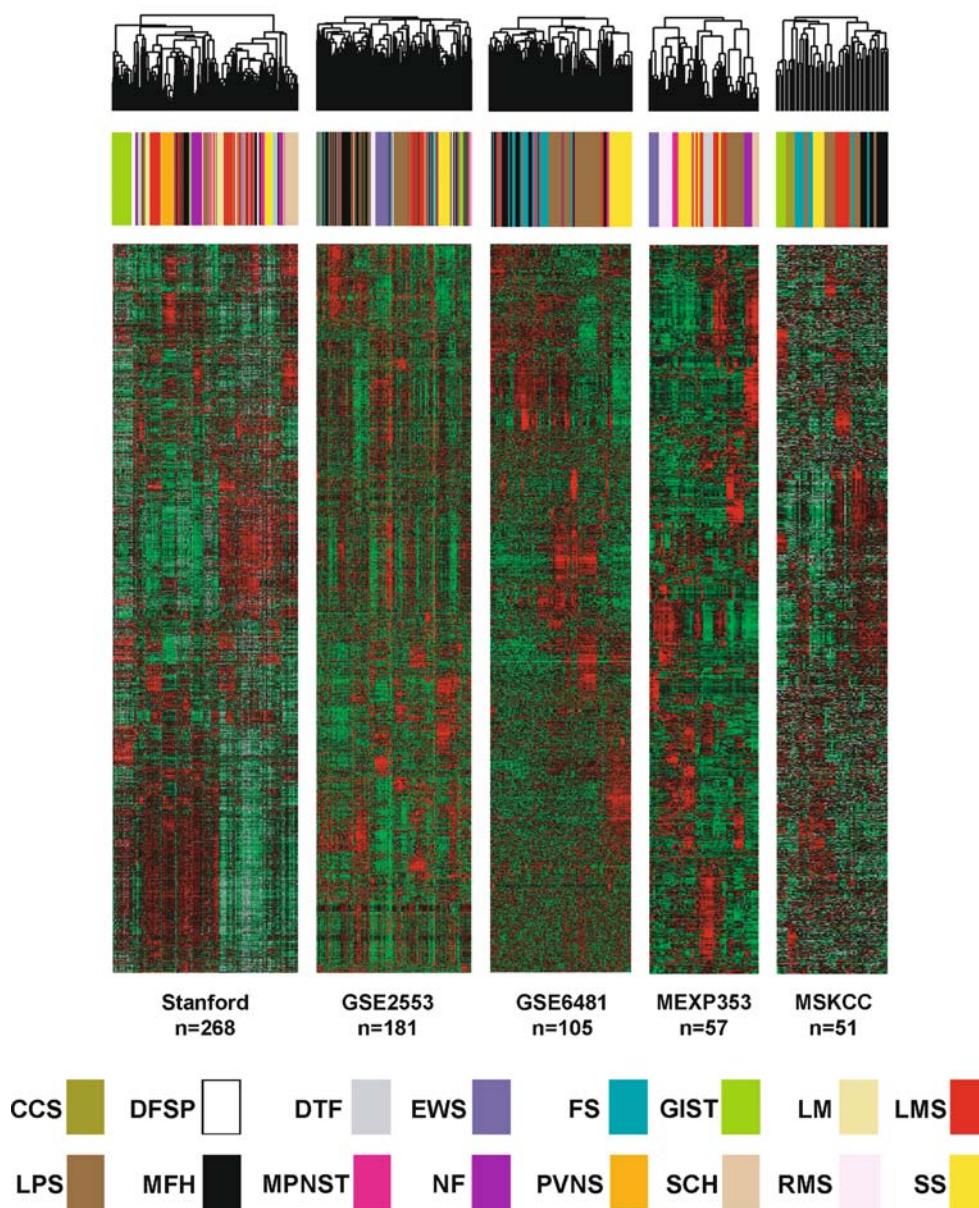
the unsupervised approaches and are used to identify gene expression signatures and biomarkers to predict particular clinical phenotypes, such as: diagnostic subtype, prognosis, and response to treatment. Due to the complexity of microarray data and the large gene lists typically produced by both unsupervised and supervised analyses, it is useful to integrate the data with repositories of biomedical annotation information to optimize the translation of the findings into biomedically useful knowledge. A final step of independent validation (ideally on an independent platform with independent samples) is essential to ensure the validity and reproducibility of the findings

sarcomas spanning six histologic subtypes and including 13 synovial sarcomas. They found that MPNST and SS clustered together and identified genes involved in neural differentiation shared by the two tumor types (including *EPHA4*, *EFNB3*). Interestingly, it has recently been shown that *SS18-SSX2* exerts part of its oncogenic effect by altering cytoskeletal architecture in an Eph-dependent manner [32]. YF Lee et al. performed gene expression profiling on nine synovial sarcomas, nine leiomyosarcomas, and nine pleomorphic sarcomas (malignant fibrous histiocytoma (MFH)) and identified 44 genes with significantly higher expression in synovial sarcoma [33]. This gene list includes genes involved in embryonic development (*FGF9*), transcriptional regulation (*SSX4*, *NCOA3*), cell

signaling (*EFNB1*, *IGF2*), and cellular adhesion (*CDH1*, *ICAM1*) [33]. Fernebro et al. [34] identified a gene expression signature that discriminated between cases with the variant *SS18-SSX1* translocation as compared with *SS18-SSX2*. These data may provide insight into the potential association of the *SS18-SSX1* variant with poorer prognosis, which has been suggested in several papers [20, 35], but not confirmed in subsequent studies [36, 37]. Taken together, the most consistent findings from these analyses are that overexpression of *ERBB2*, neural markers, and genes involved in retinoic acid and EGF and FGF signaling is seen in synovial sarcoma.

Based on the study by Nielsen et al. [22] that showed high levels of expression of *EGFR* (Her-1) in synovial

## Unsupervised Hierarchical Clustering of Sarcoma Gene Expression Datasets



**Fig. 2** Unsupervised hierarchical clustering of five sarcoma gene expression datasets. The cases are arranged along the *columns* and the genes along the *rows*. Within the heat map, *red* indicates relatively increased expression, *black* median expression, and *green* decreased expression. The *color bar* underlying the *dendrogram* indicates the diagnostic subtype of each case as indicated in the legend, which uses the following abbreviations: clear cell sarcoma (*CCS*), dermatofibrosarcoma protuberans (*DFSP*), desmoid-type fibromatosis (*DTF*), Ewing's sarcoma (*EWS*), fibrosarcoma (*FS*), gastrointestinal stromal tumor (*GIST*), leiomyoma (*LM*), leiomyosarcoma (*LMS*), liposarcoma (*LPS*), malignant fibrous histiocytoma/pleomorphic sarcoma (*MFH*), malignant peripheral nerve sheath tumor (*MPNST*), neurofibroma (*NF*), pigmented villonodular synovitis (*PVNS*), schwannoma (*SCH*), rhabdomyosarcoma (*RMS*), and synovial sarcoma (*SS*). The clustering

pattern suggests that one subset of sarcomas (including *GIST*, *SS*, *EWS*) shows relatively homogenous gene expression patterns and a second subset of sarcomas (including *LMS*, *LPS*, and *MFH*) shows more heterogeneous patterns of gene expression. The gene expression data presented in this figure were downloaded from the Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>), Stanford Microarray Database (<http://genome-www5.stanford.edu/>), ArrayExpress (<http://www.ebi.ac.uk/microarray-as/ae/>), and supplemental materials accompanying the paper by Segal et al. [25]. Prior to clustering, the genes were mean-centered. Clustering was performed with average linkage and correlation (uncentered) as the distance metric using Cluster 3.0 software (<http://bonsai.ims.u-tokyo.ac.jp/~mdehoon/software/cluster/software.htm#ctv>)

sarcoma, Ray-Coquard et al. [38] performed a phase II trial of gefitinib (an EGFR tyrosine kinase inhibitor) for the treatment of patients with synovial sarcomas expressing EGFR by immunohistochemistry that were refractory to standard chemotherapy regimens. In this trial, gefitinib failed to show significant therapeutic efficacy to warrant further evaluation in this setting. These negative results highlight the difficulty of translating findings from gene expression profiling into the oncology clinic and suggest the importance of considering other biologically important features (such as phosphorylation) in addition to messenger RNA and protein expression levels.

While most gene expression studies have focused on genes highly expressed in synovial sarcoma, Lubniecka et al. noted that a significant number of genes are down-regulated by the *SS18-SSX* fusion [39]. They investigated *EGR1*, which is a gene that shows low expression levels in synovial sarcoma, and showed that *EGR1* expression is significantly inhibited by the *SS18-SSX* fusion. They then demonstrated that *EGR1*'s expression can be reactivated in a synovial sarcoma cell model following treatment with the histone deacetylase inhibitor romidepsin, suggesting a potential mechanistic rationale for the use of histone deacetylase inhibitors in the treatment of synovial sarcoma [39]. There is a major clinical need for new targeted therapies for synovial sarcoma [18], and additional functional work will need to be performed using techniques such as animal modeling [21] to further delineate the pathways mediated by the *SS18:SSX* fusion and to identify effective targeted therapeutic agents.

### Leiomyosarcoma

Leiomyosarcomas are malignant tumors of smooth muscle, which most frequently occur in the retroperitoneum or uterus and may also arise in large blood vessels and extremities [2]. The pathogenesis of leiomyosarcoma is not well understood. Retroperitoneal site and larger size predict poorer prognosis, and local recurrence and metastasis are common in leiomyosarcoma. Clinical management typically consists of surgery, radiation, and chemotherapy. There are currently no effective targeted therapies available for leiomyosarcoma.

Gene expression studies to date have shown inconsistent results regarding the variability within leiomyosarcoma. Most experiments have found that leiomyosarcomas show diverse patterns of gene expression, with a relatively homogenous subset of leiomyosarcomas clustering together while one or multiple separate groups of leiomyosarcoma cluster with pleomorphic sarcoma (MFH) [22, 27, 33]. In contrast, to these studies, Baird et al. found relatively homogenous patterns of gene expression among 17 LMS with all cases clustering together [24].

Several groups have identified lists of genes that show high expression in leiomyosarcomas, many of which are known to be involved in smooth muscle differentiation (for example: *CALD1*, *MYL4*, *SLMAP*) [22, 27]. Our laboratory has found that elevated expression of these muscle-associated genes is seen in a distinct subset of leiomyosarcoma; however, two other distinct subsets of leiomyosarcoma exist, which each show increased expression of distinct gene sets. Taken together, these findings suggest that in contradistinction to relatively homogenous sarcoma subtypes (such as synovial sarcoma and gastrointestinal stromal tumor), a significant amount of variability exists within leiomyosarcoma. It may well be that subsets of this entity can be found in which subsets of cases have dysregulation of specific pathways that could then be targeted for therapy. Much more work needs to be done to define these subgroups and identify gene targets.

Several studies have used microarrays to identify prognostic markers in leiomyosarcoma. YF Lee et al. performed a supervised analysis and identified a list of 335 genes significantly differentially expressed between 20 primary tumors and seven metastatic tumors [40]. They then clustered a set of 30 nonmetastatic tumors with this gene set and showed that the clustering classification strongly predicted the presence of concurrent metastasis and correlated with time to metastasis. Most of the genes from the metastasis signature showed increased expression in the cluster of cases with poorer prognosis and encoded proteins involved in cell cycle, signal transduction, and apoptosis. Future studies will need to be performed to assess whether this gene expression signature predicts metastatic outcome on an independent dataset.

In a recent study, CH Lee et al. performed gene microarray analysis on a series of leiomyosarcomas and identified highly variable expression levels of macrophage-associated genes within leiomyosarcoma [41]. This finding prompted the investigators to evaluate the protein expression of macrophage markers (CD163 and CD68) on a leiomyosarcoma tissue microarray with clinical follow-up data. This analysis revealed that increased macrophages within leiomyosarcoma are associated with poor outcome in nongynecological leiomyosarcoma. Further research will be needed to validate the prognostic findings in leiomyosarcoma. In addition, further data mining of the leiomyosarcoma datasets is necessary to further identify aberrant molecular pathways to permit the development of targeted therapies for leiomyosarcoma patients.

### Liposarcoma

Liposarcomas represent ~20% of soft tissue sarcomas and are the single most prevalent subtype of soft tissue sarcoma

[2]. There are three principle histologic subtypes of liposarcoma (well differentiated/dedifferentiated, myxoid/round cell, and pleomorphic), which show distinct morphologic features and cytogenetic aberrations [42]. Several studies have performed gene expression profiling to assess the relationship between liposarcoma subsets, to provide insight into their pathogenesis, and to identify diagnostic and therapeutic markers.

Most studies have shown that myxoid/round cell liposarcoma, well-differentiated/dedifferentiated, and pleomorphic liposarcoma show distinct patterns of gene expression [24, 27, 29]. Singer et al. performed gene expression profiling to create a 142-gene classifier that distinguished between liposarcoma subtypes, lipoma, and normal fat on a training set of 80 samples and an independent test set of 49 samples [43]. They found that cell cycle and checkpoint pathways tended to be activated in well-differentiated/dedifferentiated liposarcoma and that Nutlin-3a, an antagonist of MDM2, preferentially induces apoptosis and growth arrest in dedifferentiated liposarcoma cells compared with normal adipocytes [43]. Other groups had previously shown that *MDM2* and *CDK4* are frequently amplified in atypical lipomatous tumor/well-differentiated liposarcoma and dedifferentiated liposarcoma [44]. To further characterize biological relationships between adipogenic tumors, Matushansky et al. created a developmental model of lipogenic sarcomagenesis by first defining gene expression profiles of stages of differentiation during in vitro differentiation of human mesenchymal stem cells into adipose tissues. They then created a differentiation-based classifier to correlate each of the major liposarcoma subtypes (dedifferentiated, well differentiated, round cell, and pleomorphic) to one of the stages of differentiation [45]. This novel approach provides additional insight into the relationship between liposarcoma subtypes. Taken together, these findings show that the liposarcoma subtypes show distinct gene expression profiles, which may be characteristic of particular stages of mesenchymal stem cell differentiation along an adipogenic path.

Despite the fact that liposarcoma is the most common single sarcoma subtype, few gene expression studies have been performed to identify prognostic or predictive markers for this entity, making this an important area for future investigation.

### Pleomorphic sarcoma/MFH

Pleomorphic sarcoma, previously known as MFH, remains one of the most frequently diagnosed sarcoma types, despite the fact that leading sarcoma experts question the validity and usefulness of this diagnostic category [46]. The diagnostic term was first proposed in the mid-1960s when it

was thought that cultured fibroblasts adopted phagocytic qualities resembling histiocytes [47]. The diagnostic category was seriously called into question following a careful analysis of 159 lesions identified by morphology as MFH. This analysis revealed that most cases could be recategorized as specific types of pleomorphic sarcoma and only 13% were morphologically compatible with the MFH category [48]. In a subsequent study, Fletcher et al. [49] implemented a combination of morphological, immunohistochemical, and ultrastructural criteria and showed that myogenic differentiation was associated with worse prognosis in cases that had previously been diagnosed as MFH. In the most recent World Health Organization (WHO) classification system, the term pleomorphic MFH has been replaced with undifferentiated pleomorphic sarcoma [2]. This is thought to be a rare entity that accounts for no more than 5% of soft tissue sarcomas [50].

Motivated in part by uncertainty concerning the biological basis for this diagnostic category, several gene expression studies have been performed to characterize patterns of gene expression in pleomorphic sarcoma/MFH. Unsupervised hierarchical clustering analyses have suggested that pleomorphic sarcoma/MFH may not represent a distinct biological entity but may instead represent a heterogeneous collection of high-grade sarcomas that shares patterns of gene expression with other high-grade karyotypically complex sarcomas [22–27].

While there is evidence for great heterogeneity in this group of tumors, they appear to have shared characteristics as well as evidenced by the work of Matushansky et al. [51]. To evaluate the pathogenesis and patterns of gene expression in pleomorphic sarcoma/MFH, Matushansky et al. [51] first defined a stem cell gene signature and showed that the stem cell signature was highly and specifically expressed by pleomorphic sarcoma/MFH. Using functional studies, they showed that expression of *DKK1*, which is known to be involved in embryonic development through inhibition of the WNT signaling pathway, inhibits mesenchymal stem cell differentiation via the Wnt2/ $\beta$ -catenin pathway and that mesenchymal stem cells can be transformed via inhibition of WNT signaling to form pleomorphic sarcoma-like tumors in nude mice. These findings suggest a particular cell of origin for pleomorphic sarcoma/MFH and suggest that alteration of WNT signaling pathway may provide a potential therapeutic strategy.

Taken together, most gene expression studies on pleomorphic sarcoma/MFH fail to show significant homogeneity within the tumor types; however, the novel technique employed by Matushansky et al. provides interesting insight into the pathogenesis of pleomorphic sarcoma/MFH. Future work will need to be undertaken to evaluate this hypothesis and to identify new therapeutic targets in this malignancy.

## GIST

Gastrointestinal stromal tumors (GIST) are mesenchymal neoplasms that arise in the wall of the gastrointestinal tract and that compose approximately 1–3% of gastrointestinal neoplasms [52]. The majority of GISTs show activation of the KIT or PDGFR $\alpha$  receptor tyrosine kinase proteins, and imatinib, a tyrosine kinase inhibitory with activity against GIST, shows therapeutic efficacy in these tumors [53]. Heinrich and colleagues [54] have shown that specific mutations in *KIT* or *PDGFR $\alpha$*  correlate with biologic features of the tumor and predict response to therapy. In contrast, the protein expression of CD34, KIT (CD117),  $\alpha$ -smooth muscle actin, and desmin has shown no association with survival in GIST patients treated with imatinib [55].

Several large sarcoma gene expression studies included significant numbers of GIST, and these studies all show that GISTs have a distinct and homogenous gene expression pattern and tend to all group together in unsupervised hierarchical clustering [22, 24, 27, 29]. Subramanian et al. [56] performed gene expression profiling on 26 GISTs with known *KIT* and *PDGFR $\alpha$*  status and identified gene expression profiles characteristic of different mutation types. A small percentage of GISTs do not show expression of KIT (CD117) by immunohistochemistry, and these tumors represent a diagnostic challenge. West et al. [57] identified DOG1 as a promising marker for CD117-negative GIST. This study was recently validated using a novel monoclonal antibody against DOG1 [58]. Yang et al. [59] recently showed that DOG1 (TMEM16A) is a calcium-activated chloride channel. Both DOG1 and KIT are expressed at high levels in cells of Cajal in the myenteric plexus from which GIST originates. Price et al. [60] performed gene expression profiling on a large collection of GISTs and leiomyosarcoma and created a two-gene classifier that was highly discriminant between the two diagnoses. Taken together, these studies show that GISTs show relatively homogenous patterns of gene expression with relatively little variability within this diagnostic category.

### Small round blue cell tumors

Small round blue cell tumors encompass several diagnostic categories, including: primitive neuroectodermal tumor (PNET/Ewing's), rhabdomyosarcoma, neuroblastoma, and desmoplastic small round cell tumor. In an early and widely cited gene expression profiling experiment, Khan et al. used gene expression profiling and artificial neural networks to create a highly accurate gene-expression-based classifier to differentiate between neuroblastoma,

rhabdomyosarcoma, non-Hodgkin lymphoma, and the Ewing family of tumors [61]. Gene expression studies have been performed to predict poor prognosis in the Ewing family of tumors [62, 63]. Ferreira et al. performed gene expression profiling to identify distinct genomic instability patterns associated with DNA repair and cell cycle checkpoint pathways in Ewing's sarcoma [64]. Multiple studies have looked at the oncogenic profile of the *EWS-FLII* fusion. In a recent meta-analysis, Hancock et al. identified a core *EWS-FLII* signature that supports the hypothesis that these tumors are derived from mesenchymal stem cells [65].

Most alveolar rhabdomyosarcomas are characterized by the t(2;13) translocation, which results in the generation of the PAX3-FKHR oncogenic fusion protein. Several gene expression studies have been performed to identify a PAX-FKHR gene expression signature [66–69]. Ren et al. showed that rhabdomyosarcomas may originate from mesenchymal stem cells, and the PAX3-FKHR fusion product induces skeletal muscle differentiation [70]. Several gene expression studies have been performed to identify predictive [71] and prognostic markers in rhabdomyosarcoma [66].

### Low-grade fibroblastic tumors

The diagnostic category of low-grade fibroblastic tumors includes a large number of diagnostic entities, which are comprised of tumors characterized by a proliferation of spindle cells with fibroblastic features. Gene expression profiling has been performed on several low-grade fibroblastic tumor subsets to better define their pathogenesis, to identify prognostic markers, and to discover novel patterns of fibroblastic gene expression that may provide insight into stromal reaction patterns in the carcinoma microenvironment.

Dermatofibrosarcoma protuberans (DFSP) is an aggressive spindle cell neoplasm that frequently involves subcutaneous soft tissues. The tumor is associated with the t(17;22) chromosomal translocation, which fuses the *COL1A1* and *PDGF $\beta$*  genes. DFSP have been shown to have a characteristic gene expression signature [72]. Gene expression profiling of DFSP led to the identification of APOD as an immunohistochemical marker expressed at high levels in DFSP, which may prove to be clinically useful as a diagnostic marker for differentiating DFSP from fibrous histiocytoma [73].

Desmoid-type fibromatosis (DTF) is a fibroblastic neoplasm that arises in the deep soft tissues and show infiltrative local growth with frequent recurrence, although they do not metastasize. DTFs can occur in association with familial cancer syndromes, such as familial adenomatous

polyposis and familial infiltrative fibromatosis, both of which result from germ line mutations in the *APC* gene. Beta-catenin dysregulation is a common finding in sporadic fibromatosis, and specific mutations in the beta-catenin gene (*CTNNB1*) predict recurrence risk [74]. DTF have been shown to have a distinct gene expression profile, characterized by high levels of expression of genes involved in extracellular matrix structure and function, including collagens, collagen binding, cell–cell adhesion, and cell-surface receptor-linked signal transduction regulation [75–77]. Gene signatures have been identified that differentiate DTF from nodular fasciitis [77] as well as solitary fibrous tumor [76]. West and colleagues have utilized DTF’s gene expression signature to provide insight into a stromal reaction pattern consistently observed in ~25% of breast carcinomas and which correlates with improved prognosis [75, 76]. Currently, no effective therapies exist to prevent recurrence in desmoid-type fibromatosis, although imatinib has recently been shown to be an active agent in the treatment of some advanced fibromatosis [78]. Increased molecular understanding of this neoplasm is essential to permit the further development of more efficacious therapies [79].

### Low-grade “fibrohistiocytic” tumors

Pigmented villonodular synovitis (PVNS) and tenosynovial giant cell tumor (TGCT) are related tumors, which are characterized by a proliferation of synovial-like mononuclear cells interspersed with multinucleate giant cells and inflammatory cells [2]. These tumors are classified as benign “fibrohistiocytic tumors” in the WHO classification of soft tissue tumors [2]. West et al. [80] evaluated the expression of a series of type III receptor tyrosine kinases on a large soft tissue tumor tissue microarray and identified increased expression of colony stimulating factor 1 (CSF1) and its receptor (CSF1R) in PVNS/TGCT. This finding led the investigators to identify a translocation that fuses *CSF1* on 1p13 to *COL6A3* on 2q35, resulting in the over-expression of CSF1 in a minority of tumor cells and the recruitment of CSF1R-expressing macrophages that compose the majority of the tumor mass in PVNS/TGCT [80]. The identification of the important role of the CSF1–CSF1R pathway in the pathogenesis of PVNS/TGCT suggests that inhibition of this pathway through the use of small-molecule inhibitors (such as SU11248) [81] could be effective for management of aggressive/recurrent PVNS/TGCT. It has recently been shown that the CSF1 response signature observed in PVNS/TGCT is seen in a subset (~25%) of breast carcinomas and is associated with higher grade, decreased hormone receptor expression, and increased *TP53* mutations [82].

### Future directions

Over the past decade, many sarcoma types have been profiled with gene microarrays. However, we are still in the early stages of translating these data into deeper understanding of sarcomagenesis to facilitate the identification of predictive diagnostic markers and effective therapeutic regimens.

A first step towards this goal is increased utilization of the great wealth of sarcoma gene expression data that have been placed in public repositories, such as Gene Expression Omnibus [83] and Array Express [84]. A gene expression meta-analysis utilizing these datasets in conjunction with novel bioinformatics analytic techniques [15, 85–91] may prove useful for more fully understanding the aberrant biological pathways contributing to sarcomagenesis and for identifying and designing effective drugs to treat these conditions. Several investigators have utilized large-scale meta-analyses to identify common core biological pathways that are seen across diverse cancer types [13, 14, 92, 93]. This technique has not yet been applied systematically to sarcomas. Several studies have utilized gene expression profiles to guide chemotherapy, [16, 94, 95] and increased utilization of bioinformatics tools and publically available datasets may permit the development of clinical studies to assess the efficacy of gene expression profiling to guide chemotherapy in sarcoma.

A second step towards increased utilization of sarcoma gene microarray data is to encourage integration of these data with other sources of biomedical information measured with emerging molecular biology techniques, such as ultrahigh-throughput sequencing, array comparative genomic hybridization, and micro-RNA profiling [96]. It is hoped that the integration of these technologies with clinical and functional data will facilitate a deeper understanding of sarcoma biology and the development of diagnostic and therapeutic tools to improve the survival of sarcoma patients.

The final step towards translating these data into clinical practice will be developing techniques for measuring the most useful biomarkers using tools widely available in diagnostic pathology laboratories. A promising approach to achieving this goal is by analyzing data produced by massively parallel high-throughput genome-wide technologies (such as microarrays) to identify sparse sets of potentially clinically useful biomarkers, which can then be tested on large numbers of patient samples using tissue microarrays [97]. The ultimate goal is to allow measurement of a set of validated prognostic and predictive markers in routine clinical practice with tools available in diagnostic pathology laboratories, such as immunohistochemistry, fluorescent in situ hybridization, and polymerase-chain-reaction-based DNA/RNA analyses.

**Conflict of interest statement** We declare that we have no conflict of interest.

## References

- Weiss SW, Goldblum JR (2008) *Enzinger and Weiss's soft tissue tumors*, 5th edn. Mosby/Elsevier, Philadelphia
- Fletcher CDM, Unni KK, Mertens F et al (2002) *Pathology and genetics of tumours of soft tissue and bone*. IARC Press, Lyon
- Wendtner CM, Abdel-Rahman S, Krych M et al (2002) Response to neoadjuvant chemotherapy combined with regional hyperthermia predicts long-term survival for adult patients with retroperitoneal and visceral high-risk soft tissue sarcomas. *J Clin Oncol* 20:3156–3164
- O'Sullivan B, Davis AM, Turcotte R et al (2002) Preoperative versus postoperative radiotherapy in soft-tissue sarcoma of the limbs: a randomised trial. *Lancet* 359:2235–2241
- Borden EC, Baker LH, Bell RS et al (2003) Soft tissue sarcomas of adults: state of the translational science. *Clin Cancer Res* 9:1941–1956
- Judson I (2008) State-of-the-art approach in selective curable tumours: soft tissue sarcoma. *Ann Oncol* 19(Suppl 7):vii166–vii169
- Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* 21:33–37
- Lockhart DJ, Winzler EA (2000) Genomics, gene expression and DNA arrays. *Nature* 405:827–836
- Fodor SP, Rava RP, Huang XC et al (1993) Multiplexed biochemical assays with biological chips. *Nature* 364:555–556
- Allison DB, Cui X, Page GP et al (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 7:55–65
- Butte A (2002) The use and analysis of microarray data. *Nat Rev Drug Discov* 1:951–960
- Rhodes DR, Chinnaiyan AM (2004) Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. *Ann N Y Acad Sci* 1020:32–40
- Rhodes DR, Kalyana-Sundaram S, Tomlins SA et al (2007) Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* 9:443–454
- Segal E, Friedman N, Kaminski N et al (2005) From signatures to models: understanding cancer using microarrays. *Nat Genet* 37 (Suppl):S38–S45
- Potti A, Dressman HK, Bild A et al (2006) Genomic signatures to guide the use of chemotherapeutics. *Nat Med* 12:1294–1300
- Tschoep K, Kohlmann A, Schlemmer M et al (2007) Gene expression profiling in sarcomas. *Crit Rev Oncol Hematol* 63:111–124
- Eilber FC, Dry SM (2008) Diagnosis and management of synovial sarcoma. *J Surg Oncol* 97:314–320
- Clark J, Rocques PJ, Crew AJ et al (1994) Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nat Genet* 7:502–508
- Kawai A, Woodruff J, Healey JH et al (1998) SYT-SSX gene fusion as a determinant of morphology and prognosis in synovial sarcoma. *N Engl J Med* 338:153–160
- Haldar M, Randall RL, Capocchi MR (2008) Synovial sarcoma: from genetics to genetic-based animal modeling. *Clin Orthop Relat Res* 466:2156–2167
- Nielsen TO, West RB, Linn SC et al (2002) Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* 359:1301–1307
- Nakayama R, Nemoto T, Takahashi H et al (2007) Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. *Mod Pathol* 20:749–759
- Baird K, Davis S, Antonescu CR et al (2005) Gene expression profiling of human sarcomas: insights into sarcoma biology. *Cancer Res* 65:9226–9235
- Segal NH, Pavlidis P, Antonescu CR et al (2003) Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. *Am J Pathol* 163:691–700
- Allander SV, Illei PB, Chen Y et al (2002) Expression profiling of synovial sarcoma by cDNA microarrays: association of ERBB2, IGFBP2, and ELF3 with epithelial differentiation. *Am J Pathol* 161:1587–1595
- Francis P, Namlos HM, Muller C et al (2007) Diagnostic and prognostic gene expression signatures in 177 soft tissue sarcomas: hypoxia-induced transcription profile signifies metastatic potential. *BMC Genomics* 8:73
- Nagayama S, Katagiri T, Tsunoda T et al (2002) Genome-wide analysis of gene expression in synovial sarcomas using a cDNA microarray. *Cancer Res* 62:5859–5866
- Henderson SR, Guiliano D, Presneau N et al (2005) A molecular map of mesenchymal tumors. *Genome Biol* 6:R76
- Nielsen TO, Hsu FD, O'Connell JX et al (2003) Tissue microarray validation of epidermal growth factor receptor and SALL2 in synovial sarcoma with comparison to tumors of similar histology. *Am J Pathol* 163:1449–1456
- Terry J, Saito T, Subramanian S et al (2007) TLE1 as a diagnostic immunohistochemical marker for synovial sarcoma emerging from gene expression profiling studies. *Am J Surg Pathol* 31:240–246
- Barco R, Hunt LB, Frump AL et al (2007) The synovial sarcoma SYT-SSX2 oncogene remodels the cytoskeleton through activation of the ephrin pathway. *Mol Biol Cell* 18:4003–4012
- Lee YF, John M, Edwards S et al (2003) Molecular classification of synovial sarcomas, leiomyosarcomas and malignant fibrous histiocytomas by gene expression profiling. *Br J Cancer* 88:510–515
- Fernebro J, Francis P, Eden P et al (2006) Gene expression profiles relate to SS18/SSX fusion type in synovial sarcoma. *Int J Cancer* 118:1165–1172
- Ladanyi M, Antonescu CR, Leung DH et al (2002) Impact of SYT-SSX fusion type on the clinical behavior of synovial sarcoma: a multi-institutional retrospective study of 243 patients. *Cancer Res* 62:135–140
- Guillou L, Benhattar J, Bonichon F et al (2004) Histologic grade, but not SYT-SSX fusion type, is an important prognostic factor in patients with synovial sarcoma: a multicenter, retrospective analysis. *J Clin Oncol* 22:4040–4050
- Takenaka S, Ueda T, Naka N et al (2008) Prognostic implication of SYT-SSX fusion type in synovial sarcoma: a multi-institutional retrospective analysis in Japan. *Oncol Rep* 19:467–476
- Ray-Coquard I, Le Cesne A, Whelan JS et al (2008) A phase II study of gefitinib for patients with advanced HER-1 expressing synovial sarcoma refractory to doxorubicin-containing regimens. *Oncologist* 13:467–473
- Lubieniecka JM, de Bruijn DR, Su L et al (2008) Histone deacetylase inhibitors reverse SS18-SSX-mediated polycomb silencing of the tumor suppressor early growth response 1 in synovial sarcoma. *Cancer Res* 68:4303–4310
- Lee YF, John M, Falconer A et al (2004) A gene expression signature associated with metastatic outcome in human leiomyosarcomas. *Cancer Res* 64:7201–7204
- Lee CH, Espinosa I, Vrijaldenhoven S et al (2008) Prognostic significance of macrophage infiltration in leiomyosarcomas. *Clin Cancer Res* 14:1423–1430
- Dei Tos AP (2000) Liposarcoma: new entities and evolving concepts. *Ann Diagn Pathol* 4:252–266

43. Singer S, Socci ND, Ambrosini G et al (2007) Gene expression profiling of liposarcoma identifies distinct biological types/subtypes and potential therapeutic targets in well-differentiated and dedifferentiated liposarcoma. *Cancer Res* 67:6626–6636
44. Sirvent N, Coindre JM, Maire G et al (2007) Detection of MDM2-CDK4 amplification by fluorescence in situ hybridization in 200 paraffin-embedded tumor samples: utility in diagnosing adipocytic lesions and comparison with immunohistochemistry and real-time PCR. *Am J Surg Pathol* 31:1476–1489
45. Matushansky I, Hernando E, Socci ND et al (2008) A developmental model of sarcomagenesis defines a differentiation-based classification for liposarcomas. *Am J Pathol* 172:1069–1080
46. Tos AP (2006) Classification of pleomorphic sarcomas: where are we now? *Histopathology* 48:51–62
47. Ozzello L, Stout AP, Murray MR (1963) Cultural characteristics of malignant histiocytomas and fibrous xanthomas. *Cancer* 16:331–344
48. Fletcher CD (1992) Pleomorphic malignant fibrous histiocytoma: fact or fiction? A critical reappraisal based on 159 tumors diagnosed as pleomorphic sarcoma. *Am J Surg Pathol* 16:213–228
49. Fletcher CD, Gustafson P, Rydholm A et al (2001) Clinicopathologic re-evaluation of 100 malignant fibrous histiocytomas: prognostic relevance of subclassification. *J Clin Oncol* 19:3045–3050
50. Fletcher CD (2006) The evolving classification of soft tissue tumours: an update based on the new WHO classification. *Histopathology* 48:3–12
51. Matushansky I, Hernando E, Socci ND et al (2007) Derivation of sarcomas from mesenchymal stem cells via inactivation of the Wnt pathway. *J Clin Invest* 117:3248–3257
52. Miettinen M, Lasota J (2006) Gastrointestinal stromal tumors: review on morphology, molecular pathology, prognosis, and differential diagnosis. *Arch Pathol Lab Med* 130:1466–1478
53. Joensuu H, Roberts PJ, Sarlomo-Rikala M et al (2001) Effect of the tyrosine kinase inhibitor STI571 in a patient with a metastatic gastrointestinal stromal tumor. *N Engl J Med* 344:1052–1056
54. Heinrich MC, Corless CL, Demetri GD et al (2003) Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumor. *J Clin Oncol* 21:4342–4349
55. Sciot R, Debiec-Rychter M, Daugaard S et al (2008) Distribution and prognostic value of histopathologic data and immunohistochemical markers in gastrointestinal stromal tumours (GISTs): an analysis of the EORTC phase III trial of treatment of metastatic GISTs with imatinib mesylate. *Eur J Cancer* 44:1855–1860
56. Subramanian S, West RB, Corless CL et al (2004) Gastrointestinal stromal tumors (GISTs) with KIT and PDGFRA mutations have distinct gene expression profiles. *Oncogene* 23:7780–7790
57. West RB, Corless CL, Chen X et al (2004) The novel marker, DOG1, is expressed ubiquitously in gastrointestinal stromal tumors irrespective of KIT or PDGFRA mutation status. *Am J Pathol* 165:107–113
58. Espinosa I, Lee CH, Kim MK et al (2008) A novel monoclonal antibody against DOG1 is a sensitive and specific marker for gastrointestinal stromal tumors. *Am J Surg Pathol* 32:210–218
59. Yang YD, Cho H, Koo JY et al (2008) TMEM16A confers receptor-activated calcium-dependent chloride conductance. *Nature* 455:1210–1215
60. Price ND, Trent J, El-Naggar AK et al (2007) Highly accurate two-gene classifier for differentiating gastrointestinal stromal tumors and leiomyosarcomas. *Proc Natl Acad Sci U S A* 104:3414–3419
61. Khan J, Wei JS, Ringner M et al (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7:673–679
62. Ohali A, Avigad S, Zaizov R et al (2004) Prediction of high risk Ewing's sarcoma by gene expression profiling. *Oncogene* 23:8997–9006
63. Cheung IY, Feng Y, Danis K et al (2007) Novel markers of subclinical disease for Ewing family tumors from gene expression profiling. *Clin Cancer Res* 13:6978–6983
64. Ferreira BI, Alonso J, Carrillo J et al (2008) Array CGH and gene-expression profiling reveals distinct genomic instability patterns associated with DNA repair and cell-cycle checkpoint pathways in Ewing's sarcoma. *Oncogene* 27:2084–2090
65. Hancock JD, Lessnick SL (2008) A transcriptional profiling meta-analysis reveals a core EWS-FLI gene expression signature. *Cell Cycle* 7:250–256
66. Davicioni E, Finckenstein FG, Shahbazian V et al (2006) Identification of a PAX-FKHR gene expression signature that defines molecular classes and determines the prognosis of alveolar rhabdomyosarcomas. *Cancer Res* 66:6936–6946
67. Ebauer M, Wachtel M, Niggli FK et al (2007) Comparative expression profiling identifies an in vivo target gene signature with TFAP2B as a mediator of the survival function of PAX3/FKHR. *Oncogene* 26:7267–7281
68. Lae M, Ahn EH, Mercado GE et al (2007) Global gene expression profiling of PAX-FKHR fusion-positive alveolar and PAX-FKHR fusion-negative embryonal rhabdomyosarcomas. *J Pathol* 212:143–151
69. Wachtel M, Dettling M, Koscielniak E et al (2004) Gene expression signatures identify rhabdomyosarcoma subtypes and detect a novel t(2;2)(q35;p23) translocation fusing PAX3 to NCOA1. *Cancer Res* 64:5539–5545
70. Ren YX, Finckenstein FG, Abdueva DA et al (2008) Mouse mesenchymal stem cells expressing PAX-FKHR form alveolar rhabdomyosarcomas by cooperating with secondary mutations. *Cancer Res* 68:6587–6597
71. Schaefer KL, Eisenacher M, Braun Y et al (2008) Microarray analysis of Ewing's sarcoma family of tumours reveals characteristic gene expression signatures associated with metastasis and resistance to chemotherapy. *Eur J Cancer* 44:699–709
72. Linn SC, West RB, Pollack JR et al (2003) Gene expression patterns and gene copy number changes in dermatofibrosarcoma protuberans. *Am J Pathol* 163:2383–2395
73. West RB, Harvell J, Linn SC et al (2004) Apo D in soft tissue tumors: a novel marker for dermatofibrosarcoma protuberans. *Am J Surg Pathol* 28:1063–1069
74. Lazar AJ, Tuvin D, Hajibashi S et al (2008) Specific mutations in the  $\beta$ -catenin gene (CTNNB1) correlate with local recurrence in sporadic desmoid tumors. *Am J Pathol* 173:1518–1527
75. Beck AH, Espinosa I, Gilks CB et al (2008) The fibromatosis signature defines a robust stromal response in breast carcinoma. *Lab Invest* 88:591–601
76. West RB, Nuyten DS, Subramanian S et al (2005) Determination of stromal signatures in breast carcinoma. *PLoS Biol* 3:e187
77. Bacac M, Migliavacca E, Stehle JC et al (2006) A gene expression signature that distinguishes desmoid tumours from nodular fasciitis. *J Pathol* 208:543–553
78. Heinrich MC, McArthur GA, Demetri GD et al (2006) Clinical and molecular studies of the effect of imatinib on advanced aggressive fibromatosis (desmoid tumor). *J Clin Oncol* 24:1195–1203
79. Kotilgim D, Lazar AJ, Pollock RE et al (2008) Desmoid tumor: a disease opportune for molecular insights. *Histol Histopathol* 23:117–126
80. West RB, Rubin BP, Miller MA et al (2006) A landscape effect in tenosynovial giant-cell tumor from activation of CSF1 expression by a translocation in a minority of tumor cells. *Proc Natl Acad Sci U S A* 103:690–695
81. Murray LJ, Abrams TJ, Long KR et al (2003) SU11248 inhibits tumor growth and CSF-1R-dependent osteolysis in an experimental breast cancer bone metastasis model. *Clin Exp Metastasis* 20:757–766

82. Beck AH, Espinosa I, Edris B et al (2009) The macrophage colony-stimulating factor 1 response signature in breast carcinoma. *Clin Cancer Res* 15:778–787
83. Edgar R, Domrachev M, Lash AE (2002) Gene expression omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210
84. Parkinson H, Sarkans U, Shojatalab M et al (2005) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res* 33:D553–D555
85. Bild AH, Potti A, Nevins JR (2006) Linking oncogenic pathways with therapeutic opportunities. *Nat Rev Cancer* 6:735–741
86. Hu P, Bader G, Wigle DA et al (2007) Computational prediction of cancer-gene function. *Nat Rev Cancer* 7:23–34
87. Lamb J, Crawford ED, Peck D et al (2006) The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313:1929–1935
88. Lamb J (2007) The connectivity map: a new tool for biomedical research. *Nat Rev Cancer* 7:54–60
89. Dennis G Jr, Sherman BT, Hosack DA et al (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4:P3
90. von Mering C, Jensen LJ, Kuhn M et al (2007) STRING 7—recent developments in the integration and prediction of protein interactions. *Nucleic Acids Res* 35:D358–D362
91. Subramanian A, Tamayo P, Mootha VK et al (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545–15550
92. Segal E, Friedman N, Koller D et al (2004) A module map showing conditional activity of expression modules in cancer. *Nat Genet* 36:1090–1098
93. Rhodes DR, Yu J, Shanker K et al (2004) Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 101:9309–9314
94. Bonnefoi H, Potti A, Delorenzi M et al (2007) Validation of gene signatures that predict the response of breast cancer to neo-adjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00-01 clinical trial. *Lancet Oncol* 8:1071–1078
95. Dressman HK, Berchuck A, Chan G et al (2007) An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. *J Clin Oncol* 25:517–525
96. Subramanian S, Lui WO, Lee CH et al (2008) MicroRNA expression signature of human sarcomas. *Oncogene* 27:2015–2026
97. Camp RL, Neumeister V, Rimm DL (2008) A decade of tissue microarrays: progress in the discovery and validation of cancer biomarkers. *J Clin Oncol* 26:5630–5637