UCSF

Title

Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling.

Permalink

https://escholarship.org/uc/item/2r3967v2

Journal

Oncogene, 29(6)

Authors

Beck, A
Lee, C-H
Witten, D
et al.

Publication Date

2010-02-11

DOI

10.1038/onc.2009.381

Peer reviewed

Discovery of molecular subtypes in leiomyosarcoma through integrative molecular profiling

Andrew H Beck, MD[1], **Cheng-Han Lee, MD, PhD**[2], **Daniela M Witten, MS**[3], **Briana C Gleason, MD**[4], **Badreddin Edris, MS**[1], **Inigo Espinosa, MD**[1], **Shirley Zhu, MD**[1], **Rui Li, MD**[1], **Kelli D Montgomery, BA**[1], **Robert J Marinelli, PhD**[5], **Robert Tibshirani, PhD**[3], **Trevor Hastie, PhD**[3], **David M Jablons, MD**[6], **Brian P Rubin, MD, PhD**[7], **Christopher D Fletcher, MD**[4], **Robert B West, MD, PhD**[1,8], and **Matt van de Rijn, MD, PhD**[1]

[1]Department of Pathology, Stanford University Medical Center, Stanford, CA 94305

[3]Department of Statistics, Stanford University Medical Center, Stanford, CA 94305

[5]Department of Biochemistry, Stanford University Medical Center, Stanford, CA 94305

[2]Department of Pathology, Genetic Pathology Evaluation Centre, Vancouver General Hospital and British Columbia Cancer Agency, Vancouver, BC, Canada

[4]Department of Pathology, Brigham and Women's Hospital, Boston, MA, 02115

[6]Department of Surgery, The University of California, San Francisco, CA, 94143

[7]Department of Anatomic Pathology, Pathology and Laboratory Medicine Institute, The Cleveland Clinic and the Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, 44195

[8]Department of Pathology and Laboratory Service, Palo Alto Veterans Affairs Health Care System, Palo Alto, CA 94304

Abstract

Leiomyosarcoma (LMS) is a soft tissue tumor with a significant degree of morphologic and molecular heterogeneity. We employed integrative molecular profiling to discover and characterize molecular subtypes of LMS. Gene expression profiling was performed on 51 LMS samples. Unsupervised clustering demonstrated 3 reproducible LMS clusters. Array comparative genomic hybridization (aCGH) was performed on 20 LMS samples and demonstrated that the molecular subtypes defined by gene-expression showed distinct genomic changes. Tumors from the "muscle-enriched" cluster showed significantly increased copy number changes (p=0.04). Most muscle-enriched cases showed loss at 16q24 which contains *FANCA*, known to play an important role in DNA repair, and loss at 1p36 which contains *PRDM16*, whose loss promotes muscle differentiation. Immunohistochemistry was performed on LMS tissue microarrays (n=377) for five markers with high levels of mRNA in the muscle-enriched cluster (ACTG2, CASQ2,

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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

SLMAP,CFL2, MYLK) and demonstrated significantly correlated expression of the 5 proteins (all pairwise $p < 0.005$). Expression of the 5 markers was associated with improved disease-specific survival (DSS) in a multivariate Cox regression analysis ($p < 0.04$). In this analysis that combined gene expression profiling, aCGH and immunohistochemistry, we characterized distinct molecular LMS subtypes, provided insight into their pathogenesis, and identified prognostic biomarkers.

Keywords

sarcoma; leiomyosarcoma; integrative genomics; gene expression profiling; array comparative genomic hybridization; tissue microarrays

Introduction

Cancer can be broadly divided into 3 main classes: leukemias/lymphomas (derived from cells of the hematopoetic system), carcinomas (derived from epithelial cells), and sarcomas (derived from mesenchymal tissues, including bone, muscle, and cartilage)(Abbas *et al.*, 2005). Within each of these broad categories, tumors have traditionally been further subdivided into specific diagnostic subtypes based primarily on their clinical and histopathological features (Ackerman and Rosai, 2004; Lakhani and Ashworth, 2001). In the past decade, gene expression profiling has been used to discover novel cancer subtypes in a variety of hematological (Alizadeh *et al.*, 2000; Bullinger *et al.*, 2004) and epithelial (Lapointe *et al.*, 2004; Sorlie *et al.*, 2001) malignancies. It is hoped that the further subclassification of cancer based on molecular features will facilitate the identification of prognostic and predictive biomarkers (Beck *et al.*, 2009b; Sidransky, 2002), the development of therapies targeted at oncogenic pathways altered in particular subtypes (Potti *et al.*, 2006), and ultimately, the application of a more personalized form of medicine to improve the lives of cancer patients (Sotiriou and Piccart, 2007).

Soft tissue sarcomas account for approximately 1% of all malignancies diagnosed annually, and there are ~100 recognized sarcoma diagnostic subtypes (Weiss and Goldblum, 2008). Sarcomas can be subdivided in 2 groups, one where each tumor type is characterized by a unique simple recurrent genetic abnormality such as a chromosomal translocation and the other where highly complex genetic abnormalities are present (Helman and Meltzer, 2003). Leiomyosarcomas (LMS) belong to the latter group and are malignant neoplasms of smooth muscle, which most frequently occur in the uterus or retroperitoneum but can occur throughout the body (Fletcher *et al.*, 2002). They account for ~24% of soft tissue sarcomas, making LMS the most common soft tissue sarcoma subtype (Toro *et al.*, 2006). While significant advances have been made in the molecular understanding of sarcoma subtypes with simple recurrent genetic abnormalities (Helman and Meltzer, 2003), the molecular pathogenesis and heterogeneity of LMS are poorly understood. Currently, the diagnosis of LMS is made based on the demonstration of smooth muscle differentiation in a histologically malignant neoplasm (Fletcher *et al.*, 2002). Clinical management typically consists of surgery with adjuvant doxorubicin-based chemotherapy with consideration for the addition of ifosfamide and radiotherapy in selected cases (Borden *et al.*, 2003). Doxorubicin-based chemotherapy has shown a marginal association with improved overall

survival (Cochrane Database Syst Rev, (2000)) and the addition of ifosfamide has been shown to strengthen the association with improved survival (Pervaiz *et al.*, 2008). The overall prognosis for soft tissue sarcomas is poor, with reported rates of 12 year disease-specific survival of 64% (Kattan *et al.*, 2002). There are currently no effective targeted therapies available for LMS that are directed at molecular aberrations in specific LMS subtypes. It has been shown that gene expression signatures may be able to predict metastasis in LMS (Lee *et al.*, 2004). Our laboratory has previously shown that macrophage infiltration correlates with poor outcome in LMS (Espinosa *et al.*, 2009a; Lee *et al.*, 2008). There are currently no molecular biomarkers utilized in the routine prognostication or determination of treatments in LMS.

In this study, we performed global gene expression analysis of 51 LMS samples to identify 3 distinct LMS subtypes. We then performed array comparative genomic hybridization (aCGH) to characterize the genomic changes seen in 2 of the LMS subtypes. Based on our gene expression findings, we identified biomarkers for the most distinct LMS subtype and evaluated their pattern of expression and association with clinicopathologic variables by performing immunohistochemistry (IHC) on tissue microarrays (TMAs) containing 377 LMS samples. In this integrative analysis, we characterized distinct molecular LMS subtypes, provided insight into their pathogenesis, and identified prognostic biomarkers.

MATERIALS AND METHODS

The 51 tumor samples were obtained from 46 patients (5 patients each contributed 2 samples). Clinicopathologic features of these tumors are provided in Supporting Information Table 1. The studies were performed with the approval by the Institutional Review Board at Stanford University Medical Center.

Briefly, gene expression profiling was performed on 51 LMS samples using 44K spotted cDNA microarrays. To identify molecular subtypes, unsupervised hierarchical clustering and principal component analysis (PCA) were performed. To assess the reproducibility of the clusters in an independent dataset, the clusterRepro algorithm (Kapp and Tibshirani, 2007) was utilized with the LMS samples from the GSE2553 dataset (Baird *et al.*, 2005) used as the testing dataset. For 20 of the LMS samples containing gene expression data, aCGH was performed on 44k Agilent arrays. The fused lasso algorithm was applied to identify regions of copy number gain and loss (Tibshirani and Wang, 2008). To determine whether the gene-expression defined molecular subtypes could be accurately predicted based solely on aCGH changes, 3 classification techniques were used: Prediction Analysis of Microarrays (PAM) (Tibshirani *et al.*, 2002), Prediction Analysis of Microarrays-FusedLasso (PAM-FL) (See Supporting Information Materials and Methods for explanation of PAM-FL), and K nearest neighbor (KNN). IHC was performed on LMS TMAs using antibodies for ACTG2, CASQ2, SLMAP, CFL2, and MYLK, and the stains were scored by 2 surgical pathologists (AHB and RBW). Additional information on the methods for gene expression profiling, aCGH, IHC, and statistical analysis are provided in the Supporting Information Materials and Methods. All IHC images used in this study are accesible from the accompanying website: http://tma.stanford.edu/tma_portal/LMS_IMP. In addition, gene

expression and aCGH data have been deposited in the Gene Expression Omnibus (Edgar *et al.*, 2002) with accession number GSE17555.

Results

A. Gene Expression Profiling

Unsupervised hierarchical clustering was performed on median centered gene expression data using the 3038 gene spots that had a standard deviation (SD) of at least 1 across the 51 LMS samples and passed quality-based filtering criteria (described in Supporting Information Materials and Methods). This analysis demonstrated 3 predominant clusters of LMS samples (Fig. 1A). Cluster 1 contained 13 samples derived from 11 patients, cluster 2 contained 12 samples derived from 12 patients, and cluster 3 contained 26 samples derived from 23 patients (Supporting Information Table 1). For the 5 patients with paired primary, recurrent, and/or metastatic samples in the analysis, the matched pairs clustered into the same group (Fig. 1A), suggesting that the molecular subtype is preserved during metastasis, as has been observed in other malignancies (Bernards and Weinberg, 2002; Ramaswamy *et al.*, 2003; Weigelt *et al.*, 2005; Weigelt and van't Veer, 2004). To assess the similarity of each LMS sample to its cluster's centroid, we defined a centroid for the 3 predominant LMS clusters and assessed the correlation of each LMS sample with each LMS group centroid. 48/49 samples showed the highest correlation with the sample's LMS group centroid, which provides further support of the clusters' robustness (Fig. S1). Neither of the 2 "outlier" cases from unsupervised hierarchical cluster analysis showed strong correlation with any of the LMS group centroids (correlation < 0.35).

To visualize the variability within LMS by an additional unsupervised technique, we performed PCA, which demonstrated that the molecular subtypes observed in hierarchical clustering were largely captured in the first 2 principal components (Fig. 1B). To identify a sparse set of genes composing the first 2 principal components, we performed sparse PCA (Hui *et al.*, 2006) and identified 45 genes in component 1 and 40 genes in component 2. Plotting the LMS samples along these 2 sparse principal components largely recreates the clustering structure observed in the hierarchical clustering analysis (Fig. 1C), and the classifications into three subtypes made using a simple nearest neighbor classifier based on these 2 principal components is concordant with the classifications made by unsupervised hierarchical clustering for 48/49 LMS samples (Fig. 1C).

Significance analysis of microarrays (SAM) (Tusher *et al.*, 2001) was performed to identify sets of genes highly differentially expressed between the 3 LMS subtypes. The set of genes most highly and differentially expressed in LMS Group I was significantly enriched for functional annotation terms relating to muscle contraction and the actin cytoskeleton (including:CALD1, SLMAP, DMD, ACTG2, CASQ2,CFL2, MYLK, LPP) (Supporting Information Table 1). Since this gene set is highly enriched for genes encoding proteins involved in muscle differentiation and function, we refer to this LMS group as the "Group I/ muscle-enriched" LMS molecular subtype. This gene set also showed significant enrichment for phosphoproteins, protein kinases, and kinase binding proteins (Supporting Information Table 1). The set of genes most highly and differentially expressed in LMS Group II was significantly enriched for functional annotation terms relating to protein metabolism,

regulation of cell proliferation, and organ development (Supporting Information Table 1). The set of genes most highly and differentially expressed in LMS Group III was significantly enriched for annotation terms relating to organ and system development, metal binding, extracellular proteins, proteins involved in the wound response, and ribosomal proteins involved in protein synthesis (Supporting Information Table 1). Of the 3 LMS subtypes, the Group III gene set contained the most genes (23) and highest proportion of genes from the CSF1 response gene expression signature (Beck *et al.*, 2009a). In a prior study, the CSF1 response gene expression signature was shown to be present in a subset of LMS, and the expression of 4 CSF1 response signature associated proteins was associated with poor prognosis in LMS (Espinosa *et al.*, 2009b).

To compare the pattern of gene expression seen in these LMS subtypes with other sarcomas, we performed unsupervised clustering of a large and diverse set of soft tissue tumors (STTs, n = 291, spanning 25 diagnostic subtypes) analyzed in our laboratory with the 51 LMS samples using the same gene list that was used above to cluster only the LMS samples. In this analysis, samples from the Group I/muscle-enriched cluster continue to cluster together, while cases from the other LMS clusters were interspersed in the dendrogram with other soft tissue tumor types (Fig. S2A). We performed sparse PCA on the same set of LMS and STT tumors and similarly found that the Group I/muscle-enriched LMS samples formed the most distinct cluster of LMS cases, while the other subtypes of LMS were intermixed with other STTs (Fig. S2B). These results suggest that of the three LMS subtypes, Group I/muscle-enriched shows the most distinct and specific gene expression profile.

To evaluate the reproducibility of the LMS molecular subtypes in an independent dataset, we searched the Gene Expression Omnibus for publically available sarcoma gene expression datasets and identified 1 dataset containing greater than 15 LMS samples (Baird *et al.* (GSE2553) contains 17 LMS samples) (Baird *et al.*, 2005). We utilized the clusterRepro algorithm (Kapp and Tibshirani, 2007) by training it on our LMS dataset and testing it on the Baird set of 17 LMS samples. In this analysis, 12 of the 17 Baird LMS samples were classified as Group I/muscle-enriched LMS, and the Group I LMS achieved statistically significant cluster reproducibility (IGP=1, p=0.03). Only 1 of 17 cases was classified as Group II and 4 of 17 cases were classified as Group III. Neither LMS Group II nor III achieved statistically significant reproducibility on this dataset (IGP < 0.5, p > 0.5). These findings suggest that the Group I/muscle-enriched LMS subtype is significantly reproducible in the Baird dataset, and most LMS cases from this dataset are best characterized as Group I/muscle-enriched LMS (Fig. S3). Although the Baird dataset provides no definite support for group II or III LMS, this may be partially explained by the dataset's relatively small number of LMS samples and the relatively low level of variability observed in the LMS samples in this dataset (Baird *et al.*, 2005).

Information on patient age, tumor grade, tumor histological subtype, primary tumor site, tumor status (primary, recurrence, metastasis), and patient treatment were available for the LMS cases in the gene expression analysis (Supporting Information Table 1). There was no significant difference in patient age or tumor grade between the 3 molecular subtypes (p > 0.15). Group I LMS tended to be conventional LMS histological subtype (8/11, 73%), Group II LMS were relatively evenly distributed between conventional (5/12, 42%) and

pleomorphic/mixed (7/12, 58%), and Group III LMS tended to be pleomorphic/mixed histological subtype (16/21, 76%) [p=0.03]. Group I LMS (10/11, 91% extra-uterine) and Group II LMS (9/12, 75% extra-uterine) tended to have an extra-uterine primary tumor site, while Group III LMS was relatively evenly distributed between uterine and extra-uterine primary sites (9/19, 47% extra-uterine) [p=0.04]. 9/13 Group I samples and 10/12 Group II samples analyzed by gene expression profiling were from the primary tumor site, while only 6/23 Group III tumors were from the primary site [p=0.002] and the remainder came from a recurrence (7/23) or a metastasis (10/23). [Although there are a total of 24 Group III samples, the site for one of the samples (STT4401) was not known to be primary or metastasis/recurrence, and this sample was excluded from this analysis]. Information on patient treatment with radiotherapy and/or chemotherapy was available for 45 of the Group I, II, and III samples. There was no significant difference in proportion of treatment with radiotherapy and/or chemotherapy in the molecular subtypes: 11/13 Group I samples, 11/12 Group II samples, and 14/20 Group III samples were resected from patients with no history of prior treatment with radiotherapy or chemotherapy (p = 0.3). 3 patients contributed tumor samples removed pre- and post-treatment with chemotherapy and/or radiotherapy, and in all 3 cases the pre- and post-treatment samples clustered into the same LMS molecular subtype.

B. Array Comparative Genomic Hybridization (aCGH)

aCGH was performed to characterize genomic changes in 20 LMS samples and 4 normal smooth muscle samples: 12 of the LMS samples were from Group I/muscle-enriched (derived from 10 patients), 7 from Group III (derived from 6 patients), and 1 from Group II. To identify regions of genomic gain and loss, the fused lasso technique was performed (Tibshirani and Wang, 2008) with an FDR threshold of 10% for calling regions of gain/loss (Supporting Information Table 1).

The Group I/muscle-enriched and Group III LMS samples showed distinct patterns of genomic gain and loss. The Group I samples showed significantly increased genomic gains and losses compared to the Group III samples (mean proportion of genome involved by gain/loss = 17% in Group I samples vs. 2% in Group III samples, p=0.04) (Fig. 2). Analyzing the 7 Group III samples together showed no statistically significant shared regions of gain or loss (all consensus FDR > 5%). In contrast, analyzing the 12 Group I/muscle-enriched samples together showed 691 spots that each had a consensus FDR < 5% (Supporting Information Table 1). Taken together, these findings suggest distinct pathways of oncogenesis in the 2 LMS subtypes, with decreased genomic stability in Group I/muscle-enriched LMS.

Interestingly, loss of a 291 KB region on 16q24 was seen in 7 of 12 Group I/muscle-enriched samples (all of which showed at least 7% genomic changes) and none of the Group II (0/1) or Group III (0/7) samples. This genomic region contains several cancer-associated genes, including the Fanconi anemia, complementation group A (*FANCA*) gene, which is a core Fanconi anemia protein that functions as a signal transducer and DNA-processing molecule in a DNA-damage repair network (Wang, 2007). In acute myeloid leukemia, it has been shown that acquired *FANCA* dysfunction promotes cytogenetic instability and clonal progression (Lensch *et al.*, 2003). The loss of *FANCA* may be an important event that is

specific for the molecular pathogenesis of the Group I/muscle-enriched LMS subtype and suggests an etiology for the increased genomic complexity observed in this LMS subtype. In addition, this region contains CBFA2T3, which is known to be involved in a translocation with *RUNX1* (*AML1*) in a subset of therapy related acute myeloid leukemia (Ottone *et al.*, 2009) and was identified as a putative breast tumor suppressor gene (Kochetkova *et al.*, 2002). The shared deleted region on 16q24 spans a total of 36 genes and includes *CDK10*, *TCF25*, *FOXF1*, and *IRF8* in addition to *FANCA*, *CBFA2T3* and others. The full list of genes can be found in the Supplemental Workbook.

The most commonly shared region of gain or loss in the Group I/muscle-enriched cases was a 2.5 MB region on 1p36.32, which spans PRDM16, TNFRSF14, C1orf93, and MMEL1. This region was lost in 8 of 12 Group I/muscle-enriched samples (consensus FDR = 0.01). This change was specific to this LMS subtype and there was no loss at 1p36.32 observed in the 1 Group II sample or in the 7 Group III samples. The *PRDM16* gene has recently been shown to control a brown fat/skeletal muscle switch. Loss of PRDM16 from brown fat precursors promotes skeletal muscle differentiation and leads to elevated expression of muscle specific genes (Seale *et al.*, 2008). Group I LMS showed high expression of a diverse set of muscle-associated genes, including genes expressed in smooth, cardiac, and skeletal muscle (including *ACTG2*, *MYLK*, *PDLIM5*), genes expressed primarily in cardiac and skeletal muscle (including *CFL2*, *SLMAP*), and genes expressed primarily in cardiac muscle (*CASQ2*). The loss of *PRDM16* gene in most Group I cases suggests a potential etiology of the "muscle-enriched" pattern of gene expression observed in Group I tumors, which includes both genes expressed in smooth and skeletal/cardiac muscle, suggesting either that PRDM16 may have a role in expression of genes involved in skeletal, smooth, and cardiac muscle or a separate transcriptional regulatory factor may account for the increased expression of smooth muscle and cardiac muscle associated genes in Group I/muscle-enriched LMS. *MYOCD* amplification has recently been shown to play an important role in LMS pathogenesis (Perot *et al.*, 2009). In our study, a region on 17p11 containing MYOCD (as wellas MAP2K4) was amplified in 3 Group I/muscle-enriched tumors and the 1 Group II sample.

Other cancer-associated genes that showed copy number gains in at least 6 of the Group I samples included: TCF12 (15q21), ABL2 (1q24), and the MET oncogene (7q31). MET overexpression has been previously reported in a variety of sarcomas, including alveolar soft part sarcoma (Jun *et al.*, 2009), osteosarcoma, chondrosarcomas, and leiomyosarcoma (Rong *et al.*, 1993). Copy number gain suggests a possible mechanism of MET overexpression in LMS. Additional cancer-associated genes that were lost in at least 6 Group I samples include the alveolar soft part sarcoma chromosome region, candidate 1 (ASPSCR; 17q25), BCL3 (19q13), ERCC2 (19q13), FSTL3 (19p13), RB1 (13q14), STK11 (19p13), and TCF3 (19p13) (An expanded table of cancer-associated genes with the copy number changes observed in our study is provided in the Supplemental Workbook).

There were no recurrent genomic changes seen in greater than 3 of the 7 Group III samples, and all of the changes seen in multiple Group III samples were also observed in multiple Group I/muscle-enriched samples. The only change involving a known gene that was shared in at least 3 of the 7 Group III cases was gain at 7q31.2, which includes the *CAV-1* gene.

This region was also gained in 7 of 12 Group I/muscle-enriched cases and the 1 Group II case. Caveolin-1 is known to be expressed on smooth muscle and has been shown to activate the Akt pathway in an *in vitro* prostate cancer model (Li *et al.*, 2003). The Akt pathway has been shown to play an important role in LMS (Hernando *et al.*, 2007).

To determine whether the gene-expression defined molecular subtypes could be accurately predicted based solely on aCGH changes, we used 3 classification techniques: PAM, PAM-FL, and KNN. The aCGH-based KNN classifier obtained a cross validation (CV) error rate of 4/19 (21%, corresponding to a permutation-based p value of 0.05), the PAM classifier obtained a CV error rate of 2/19 (11%), and the PAM-FL obtained a CV error rate of 1/19 (5%). These findings demonstrate that the gene expression defined Groups I and III can be predicted with significant accuracy by aCGH changes alone. The PAM-FL centroid, which summarizes genomic changes in Group I relative to Group III is presented as Fig. S4.

C. Tissue Microarray Analysis

Based on the findings from gene expression profiling (which showed that Group I/muscle-enriched LMS represented the most distinct molecular subtype) and aCGH (which showed that Group I/muscle-enriched LMS had the most recurrent regions of genomic gain and loss), we chose to focus our TMA analysis on evaluating the protein expression of genes with high levels of mRNA expression in the Group I/muscle-enriched LMS molecular subtype.

The protein expression of 5 markers highly and differentially expressed in Group I/muscle-enriched LMS was examined on LMS TMAs. For the 377 LMS cases represented on the TMAs, there were evaluable results for all 5 stains for 275 cases. For the purposes of clinicopathologic analysis, we determined the sum total of positive staining markers for each case. Clinicopathologic data (including FNLCC histologic grade, mitotic count, necrosis, presence of CSF1 response protein expression signature (Beck *et al.*, 2009a; Espinosa *et al.*, 2009b), and disease specific survival (DSS) were available for 124 of the 275 cases. Information on anatomic site (gynecological vs. non- gynecological) was available for 273 samples.

The stains showed significant correlation with each other (all pairwise Spearman's rho p < 0.005) with a minimum correlation of 0.170 between ACTG2 and CASQ2 and a maximum correlation of 0.658 between ACTG2 and SLMAP (Fig. 3). 19% (51/275) of the cases showed coordinate expression of all 5 evaluable markers, similar to the 25% of cases present in the muscle enriched cluster by gene expression arrays.

The number of positive markers showed no association with site (mean positive markers = 3.1 in both gynecological and non-gynecological LMS), no significant association with grade (mean positive markers = 3.1 in grade 1, 3.3 in grade 2, and 2.9 in grade 3; p=0.56), no significant association with mitotic figure count (p=0.335), and there was a trend for a negative association with the presence of necrosis (p=0.07). In a multivariate model incorporating the CSF1 response protein expression signature, the number of positive Group I/muscle-enriched markers, grade, tumor site (uterine vs. extra-uterine), necrosis and mitotic figures, the CSF1 response signature (summarizing the expression of 4 CSF1 response

signature-associated proteins) and the number of positive Group I/muscle-enriched markers were the only 2 significant predictors of survival, with the CSF1 response protein expression signature showing an association with poor outcome while expression of the Group I/muscle enriched markers was associated with a more favorable outcome (Table 1).

The findings from our TMA analysis demonstrate that the Group I/muscle-enriched markers show correlated protein expression, and the expression of Group I/muscle-enriched markers is associated with improved DSS independent of grade, mitotic figures, necrosis, site, and the CSF1 response signature.

DISCUSSION

LMS is an aggressive malignant neoplasm, and its molecular pathogenesis is poorly understood. Treatment options are limited, and there is a major clinical interest in gaining a better understanding of LMS pathogenesis to facilitate the development of targeted therapies.

Several prior studies have performed gene expression profiling on relatively small numbers (n=3–13) of LMS samples (Baird *et al.*, 2005; Henderson *et al.*, 2005; Nakayama *et al.*, 2007; Nielsen *et al.*, 2002; Quade *et al.*, 2004; Ren *et al.*, 2003; Segal *et al.*, 2003; Shmulevich *et al.*, 2002; Skubitz and Skubitz, 2003). Due to the small number of cases in each study it is difficult to draw conclusions on the heterogeneity within LMS based on these data. Francis et al. performed gene expression profiling on 177 soft tissue tumors, including 40 LMS samples. They identified a distinct cluster of 11 LMSs that clustered together, while the remaining 29 LMS samples showed more heterogeneous patterns of gene expression (Francis *et al.*, 2007). The distinct cluster of 11 LMS cases from this dataset were reported to show high expression levels of a group of muscle-associated genes, many of which were also identified as highly expressed in Group I/muscle-enriched LMS in our study (including CALD1, SLMAP, ACTG2, CFL2, MYLK, ACTA2, MBNL1, TPM1, PPP1R12A, DTNA, FZD6, PPP1R12A, CLIC4, CDC42EP3, BARD1, TPM1, RAB27A, MAP1B, EDIL). We find a similar muscle-enriched LMS cluster in our dataset and in the Baird dataset (Baird *et al.*, 2005). Our findings and those from the literature suggest that multiple molecular subtypes of LMS exist and that the "muscle-enriched" subtype has been reproducibly identified in at least 2 of the largest datasets.

Several prior reports have looked at CGH changes in LMS. Meza-Zepeda et al. performed aCGH on 12 LMS samples and 7 gastrointestinal stromal tumors and observed that LMS showed more genomic losses than gains with the most frequent minimal regions of loss at 10q21.3 and 13q14.2-q14.3, each detected in 9 of 12 LMS samples in their study (Meza-Zepeda *et al.*, 2006). In our study, we identified loss at 10q21.3 in 5 of 12 Group I/muscle-enriched samples and in none of the Group II or III samples. We identified loss at 13q14.2 in 6 of 12 Group I/muscle-enriched samples, 0 of 1 Group II samples, and 2 of 7 Group III samples. The common region of 13q14.2 that was lost in all 8 samples includes the *RB1* gene, a well-characterized tumor-suppressor whose loss has been shown to contribute to sarcomagenesis (Landis-Piwowar *et al.*, 2008). Meza-Zepeda *et al* also noted loss at 16q21.2-q22.1 in 6 of 12 samples and 1p36.32-p36.21 in 4 of 12 samples, which are both

changes we find in our study, specifically in Group I/muscle-enriched LMS. Larramendy *et al* evaluated 102 malignant fibrous histiocytomas (MFH) and 82 LMS cases by conventional comparative genomic hybridization (Larramendy *et al.*, 2008) and identified 11 regions with significantly increased losses in LMS compared to MFH, including 1p36.1~pter (10% of LMS vs 1% of MFH), and 16qter (34% of LMS vs. 3% of MFH), both of which were identified as lost in most Group I/muscle-enriched LMS cases in our analysis. We note that the 1p36 region contains *PRDM16* and the 16q24.3 region contains *FANCA*. To our knowledge, our study is the first to integrate aCGH data with gene expression analysis.

Prognosis in LMS is currently predicted using a combination of traditional clinicopathologic features (Kattan *et al.*, 2002). There are currently no molecular biomarkers utilized in prognostication in LMS in clinical practice. Gene expression microarrays have been used to identify signatures to predict metastasis in LMS (Lee *et al.*, 2004). Our group has previously identified macrophage infiltration (Lee *et al.*, 2008) and the CSF1 response signature (Espinosa *et al.*, 2009b) as predictors of poor prognosis in LMS. In the current study, we have identified protein markers from the Group I/muscle-enriched LMS subtype and demonstrated that their expression correlates with improved DSS. These findings suggest that despite showing increased genomic complexity, Group I/muscle-enriched LMS may be intrinsically less aggressive and more differentiated than other LMS subtypes. In a multivariate model incorporating traditional clinicopathologic features (size, grade, necrosis, site) as well as the CSF1 response signature and the Group I/muscle-enriched markers, we find that only the CSF1 response signature and the number of positive muscle-enriched markers emerged as significant predictors of survival, with the CSF1 response signature correlating with poor prognosis and the expression of Group I/muscle-enriched markers correlating with improved prognosis. These prognostic biomarkers, which can be measured with immunohistochemistry on paraffin embedded formalin fixed tissue, may prove useful for the clinical management of LMS. Ultimately, we hope that the characterization of distinct molecular subtypes in LMS will lead not only to the identification of clinically useful prognostic markers, but also to the development of treatments to target specific molecular aberrations observed in the subtypes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

REFERENCES

Adjuvant chemotherapy for localised resectable soft tissue sarcoma in adults. Cochrane Database Syst Rev: CD001419. 2000

Abbas, AK.; Cotran, RS.; Fausto, N.; Kumar, V.; Perkins, JA.; Robbins, SL. Robbins and Cotran Pathologic basis of disease [print/digital]. 7th edn.. Philadelphia: Elsevier/Saunders; 2005. xv, 1525 p.pp.

Ackerman, LV.; Rosai, J. Rosai and Ackerman's Surgical pathology [print]. 9th edn.. Edinburgh; New York: Mosby; 2004.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]

Baird K, Davis S, Antonescu CR, Harper UL, Walker RL, Chen Y, et al. Gene expression profiling of human sarcomas: insights into sarcoma biology. Cancer Res. 2005; 65:9226–9235. [PubMed: 16230383]

Beck AH, Espinosa I, Edris B, Li R, Montgomery K, Zhu S, et al. The macrophage colony-stimulating factor 1 response signature in breast carcinoma. Clin Cancer Res. 2009a; 15:778–787. [PubMed: 19188147]

Beck AH, West RB, van de Rijn M. Gene expression profiling for the investigation of soft tissue sarcoma pathogenesis and the identification of diagnostic, prognostic, and predictive biomarkers. Virchows Arch. 2009b

Bernards R, Weinberg RA. A progression puzzle. Nature. 2002; 418:823. [PubMed: 12192390]

Borden EC, Baker LH, Bell RS, Bramwell V, Demetri GD, Eisenberg BL, et al. Soft tissue sarcomas of adults: state of the translational science. Clin Cancer Res. 2003; 9:1941–1956. [PubMed: 12796356]

Bullinger L, Dohner K, Bair E, Frohling S, Schlenk RF, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med. 2004; 350:1605–1616. [PubMed: 15084693]

Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:207–210. [PubMed: 11752295]

Espinosa I, Beck AH, Lee CH, Zhu S, Montgomery KD, Marinelli RJ, et al. Coordinate expression of colony-stimulating factor-1 and colony-stimulating factor-1-related proteins is associated with poor prognosis in gynecological and nongynecological leiomyosarcoma. Am J Pathol. 2009a; 174:2347–2356. [PubMed: 19443701]

Espinosa I, Beck AH, Lee CH, Zhu S, Montgomery KD, Marinelli RJ, et al. Coordinate Expression of Colony Stimulating Factor-1 (CSF1) and CSF1 Related Proteins is Associated with Poor Prognosis in Gynecologic and Non-Gynecologic Leiomyosarcoma. Am J Pathol. 2009b Accepted for publication (February 2009).

Fletcher, CDM.; Unni, KK.; Mertens, F. Pathology and genetics of tumours of soft tissue and bone. Lyon: IARC Press; 2002. World Health Organization., International Agency for Research on Cancer; p. 427

Francis P, Namlos HM, Muller C, Eden P, Fernebro J, Berner JM, 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. [PubMed: 17359542]

Helman LJ, Meltzer P. Mechanisms of sarcoma development. Nat Rev Cancer. 2003; 3:685–694. [PubMed: 12951587]

Henderson SR, Guiliano D, Presneau N, McLean S, Frow R, Vujovic S, et al. A molecular map of mesenchymal tumors. Genome Biol. 2005; 6:R76. [PubMed: 16168083]

Hernando E, Charytonowicz E, Dudas ME, Menendez S, Matushansky I, Mills J, et al. The AKT-mTOR pathway plays a critical role in the development of leiomyosarcomas. Nat Med. 2007; 13:748–753. [PubMed: 17496901]

Hui Z, Hastie T, Tibshirani R. Sparse Principal Component Analysis. Journal of Computational and Graphical Statistics. 2006; 15:265–286.

Jun HJ, Lee J, Lim DH, Park JO, Ahn G, Seo SW, et al. Expression of MET in alveolar soft part sarcoma. Med Oncol. 2009

Kapp AV, Tibshirani R. Are clusters found in one dataset present in another dataset? Biostatistics. 2007; 8:9–31. [PubMed: 16613834]

Kattan MW, Leung DH, Brennan MF. Postoperative nomogram for 12-year sarcoma-specific death. J Clin Oncol. 2002; 20:791–796. [PubMed: 11821462]

Kochetkova M, McKenzie OL, Bais AJ, Martin JM, Secker GA, Seshadri R, et al. CBFA2T3 (MTG16) is a putative breast tumor suppressor gene from the breast cancer loss of heterozygosity region at 16q24.3. Cancer Res. 2002; 62:4599–4604. [PubMed: 12183414]

Lakhani SR, Ashworth A. Microarray and histopathological analysis of tumours: the future and the past? Nat Rev Cancer. 2001; 1:151–157. [PubMed: 11905806]

Landis-Piwowar KR, Milacic V, Dou QP. Relationship between the methylation status of dietary flavonoids and their growth-inhibitory and apoptosis-inducing activities in human cancer cells. J Cell Biochem. 2008; 105:514–523. [PubMed: 18636546]

Lapointe J, Li C, Higgins JP, van de Rijn M, Bair E, Montgomery K, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci U S A. 2004; 101:811–816. [PubMed: 14711987]

Larramendy ML, Gentile M, Soloneski S, Knuutila S, Bohling T. Does comparative genomic hybridization reveal distinct differences in DNA copy number sequence patterns between leiomyosarcoma and malignant fibrous histiocytoma? Cancer Genet Cytogenet. 2008; 187:1–11. [PubMed: 18992634]

Lee CH, Espinosa I, Vrijaldenhoven S, Subramanian S, Montgomery KD, Zhu S, et al. Prognostic significance of macrophage infiltration in leiomyosarcomas. Clin Cancer Res. 2008; 14:1423–1430. [PubMed: 18316565]

Lee YF, John M, Falconer A, Edwards S, Clark J, Flohr P, et al. A gene expression signature associated with metastatic outcome in human leiomyosarcomas. Cancer Res. 2004; 64:7201–7204. [PubMed: 15492233]

Lensch MW, Tischkowitz M, Christianson TA, Reifsteck CA, Speckhart SA, Jakobs PM, et al. Acquired FANCA dysfunction and cytogenetic instability in adult acute myelogenous leukemia. Blood. 2003; 102:7–16. [PubMed: 12637330]

Li L, Ren CH, Tahir SA, Ren C, Thompson TC. Caveolin-1 maintains activated Akt in prostate cancer cells through scaffolding domain binding site interactions with and inhibition of serine/threonine protein phosphatases PP1 and PP2A. Mol Cell Biol. 2003; 23:9389–9404. [PubMed: 14645548]

Meza-Zepeda LA, Kresse SH, Barragan-Polania AH, Bjerkehagen B, Ohnstad HO, Namlos HM, et al. Array comparative genomic hybridization reveals distinct DNA copy number differences between gastrointestinal stromal tumors and leiomyosarcomas. Cancer Res. 2006; 66:8984–8993. [PubMed: 16982739]

Nakayama R, Nemoto T, Takahashi H, Ohta T, Kawai A, Seki K, et al. Gene expression analysis of soft tissue sarcomas: characterization and reclassification of malignant fibrous histiocytoma. Mod Pathol. 2007; 20:749–759. [PubMed: 17464315]

Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, et al. Molecular characterisation of soft tissue tumours: a gene expression study. Lancet. 2002; 359:1301–1307. [PubMed: 11965276]

Ottone T, Hasan SK, Montefusco E, Curzi P, Mays AN, Chessa L, et al. Identification of a potential "hotspot" DNA region in the RUNX1 gene targeted by mitoxantrone in therapy-related acute myeloid leukemia with t(16;21) translocation. Genes Chromosomes Cancer. 2009; 48:213–221. [PubMed: 19023877]

Perot G, Derre J, Coindre JM, Tirode F, Lucchesi C, Mariani O, et al. Strong smooth muscle differentiation is dependent on myocardin gene amplification in most human retroperitoneal leiomyosarcomas. Cancer Res. 2009; 69:2269–2278. [PubMed: 19276386]

Pervaiz N, Colterjohn N, Farrokhyar F, Tozer R, Figueredo A, Ghert M. A systematic meta-analysis of randomized controlled trials of adjuvant chemotherapy for localized resectable soft-tissue sarcoma. Cancer. 2008; 113:573–581. [PubMed: 18521899]

Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Genomic signatures to guide the use of chemotherapeutics. Nat Med. 2006; 12:1294–1300. [PubMed: 17057710]

Quade BJ, Wang TY, Sornberger K, Dal Cin P, Mutter GL, Morton CC. Molecular pathogenesis of uterine smooth muscle tumors from transcriptional profiling. Genes Chromosomes Cancer. 2004; 40:97–108. [PubMed: 15101043]

Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet. 2003; 33:49–54. [PubMed: 12469122]

Ren B, Yu YP, Jing L, Liu L, Michalopoulos GK, Luo JH, et al. Gene expression analysis of human soft tissue leiomyosarcomas. Hum Pathol. 2003; 34:549–558. [PubMed: 12827608]

Rong S, Jeffers M, Resau JH, Tsarfaty I, Oskarsson M, Vande Woude GF. Met expression and sarcoma tumorigenicity. Cancer Res. 1993; 53:5355–5360. [PubMed: 7693339]

Seale P, Bjork B, Yang W, Kajimura S, Chin S, Kuang S, et al. PRDM16 controls a brown fat/skeletal muscle switch. Nature. 2008; 454:961–967. [PubMed: 18719582]

Segal NH, Pavlidis P, Antonescu CR, Maki RG, Noble WS, DeSantis D, et al. Classification and subtype prediction of adult soft tissue sarcoma by functional genomics. Am J Pathol. 2003; 163:691–700. [PubMed: 12875988]

Shmulevich I, Hunt K, El-Naggar A, Taylor E, Ramdas L, Laborde P, et al. Tumor specific gene expression profiles in human leiomyosarcoma: an evaluation of intratumor heterogeneity. Cancer. 2002; 94:2069–2075. [PubMed: 11932911]

Sidransky D. Emerging molecular markers of cancer. Nat Rev Cancer. 2002; 2:210–219. [PubMed: 11990857]

Skubitz KM, Skubitz AP. Differential gene expression in leiomyosarcoma. Cancer. 2003; 98:1029–1038. [PubMed: 12942572]

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001; 98:10869–10874. [PubMed: 11553815]

Sotiriou C, Piccart MJ. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? Nat Rev Cancer. 2007; 7:545–553. [PubMed: 17585334]

Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A. 2002; 99:6567–6572. [PubMed: 12011421]

Tibshirani R, Wang P. Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics. 2008; 9:18–29. [PubMed: 17513312]

Toro JR, Travis LB, Wu HJ, Zhu K, Fletcher CD, Devesa SS. Incidence patterns of soft tissue sarcomas, regardless of primary site, in the surveillance, epidemiology and end results program, 1978–2001: An analysis of 26,758 cases. Int J Cancer. 2006; 119:2922–2930. [PubMed: 17013893]

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A. 2001; 98:5116–5121. [PubMed: 11309499]

Wang W. Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nat Rev Genet. 2007; 8:735–748. [PubMed: 17768402]

Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, et al. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res. 2005; 65:9155–9158. [PubMed: 16230372]

Weigelt B, van't Veer LJ. Hard-wired genotype in metastatic breast cancer. Cell Cycle. 2004; 3:756–757. [PubMed: 15153810]

Weiss, SW.; Goldblum, JR. Enzinger and Weiss's Soft tissue tumors. 5th Edition. 5th edn.. Philadelphia: Mosby/Elsevier; 2008.
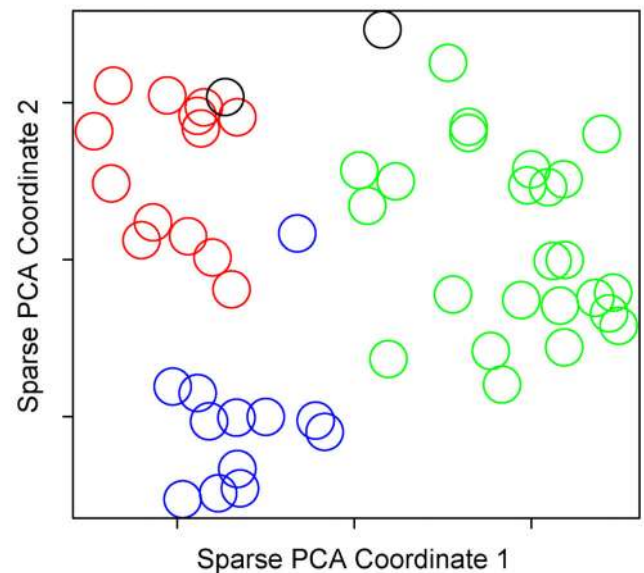
A. Unsupervised Hierarchical Clustering
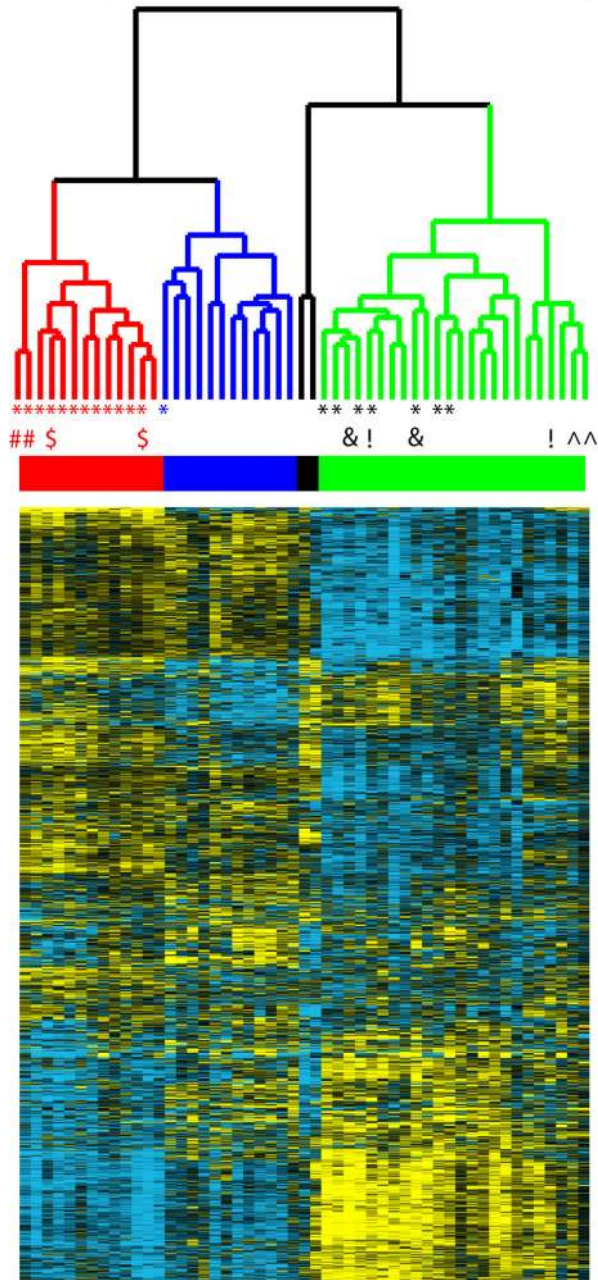
B. Principal Component Analysis
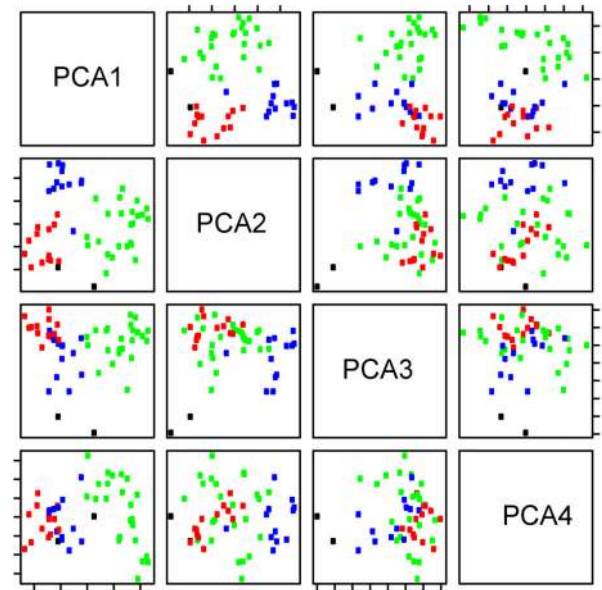
C. Sparse Principal Component Analysis

Figure 1.
Unsupervised clustering of 51 LMS samples reveals 3 reproducible molecular subtypes. A) Unsupervised hierarchical clustering was performed on 51 LMS samples with 3038 genes that showed at least 1 standard deviation across the samples. The 20 samples that were also profiled for DNA copy number changes with aCGH are indicated by an asterisk. The 5 paired primary-metastasis samples are indicated by a paired symbol (#,$,&,!,^). On the sample dendrogram, the Group I cases are heighted in red, Group II blue, and Group III green. The 2 cases that did not cluster into a group are indicated in black. Within the

heatmap, yellow indicates relatively increased expression, black indicates median expression, and blue indicates relatively decreased expression. B) Principal Component Analysis of the 51 LMS samples with 3038 genes. Each sample is represented in the figure by a colored box. The color indicates the clustering designation made by hierarchical clustering: red = Group I, blue = Group II, and Green = Group III. Most of the variance between the 3 groups is captured in the first two principal components. C) Sparse Principal Component Analysis. The 51 LMS samples were plotted against the sparse PCA coordinate 1 (containing 45 genes) and sparse PCA coordinate 2 (containing 40 genes). Each sample is represented by a colored circle, and the color indicates the clustering designation made by hierarchical clustering: red = Group I, blue = Group II, green = Group III. Most of the variance between the 3 LMS molecular subtypes is explained by these 2 sparse principal components.
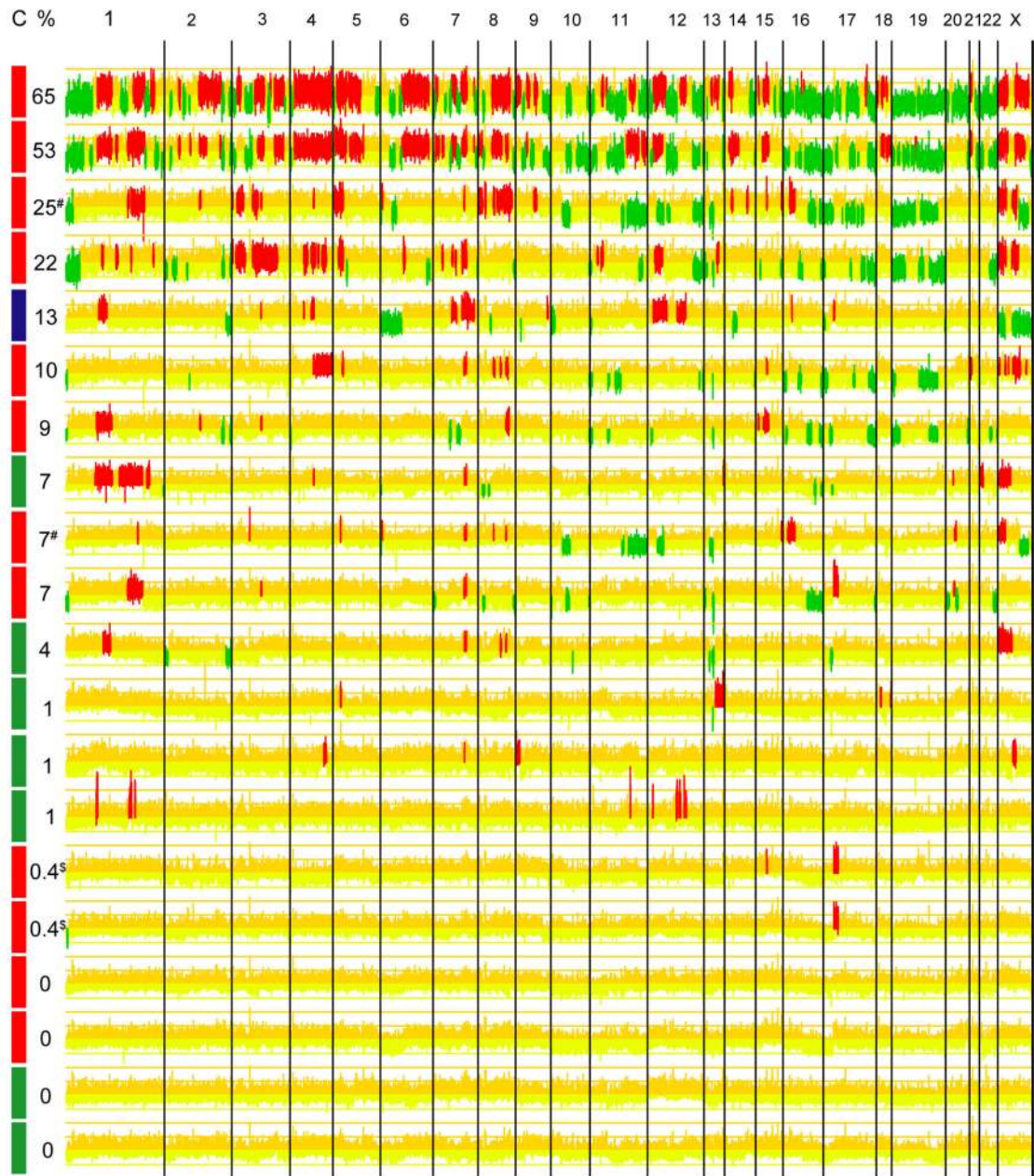
Figure 2.
Array Comparative Genomic Hybridization of 20 Leiomyosarcoma Samples. The 20 samples are arranged along the y axis and ordered according to amount of DNA copy number changes. Chromosomal locations are indicated along the x axis. Copy number changes were called using the cghFLasso algorithm with an overall false discovery rate of 0.10. Regions of genomic gain are indicated in red and loss in green. The proportion of genome showing gain or loss is indicated to the left of each row. The gene-expression defined molecular subtype is indicated on the colorbar on the left: red = Group I, blue =

Group II, green = Group III. The Group I cases show significantly increased regions of genomic gain/loss compared to the Group III cases (p=0.04).
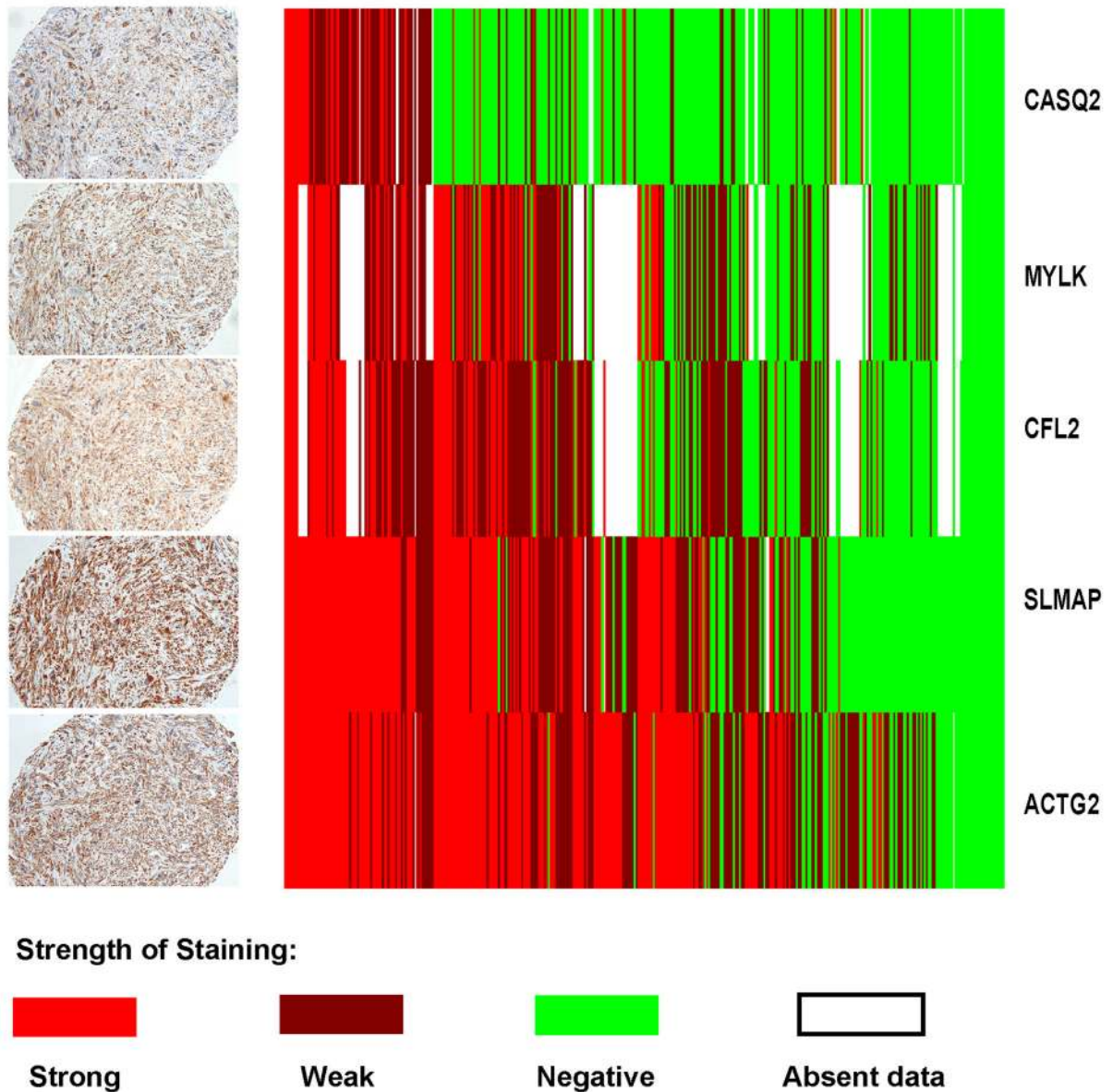
Figure 3.

Protein Expression of Group I Markers on Leiomyosarcoma Tissue Microarray. We performed IHC for 5 markers that showed high levels of expression in Group I LMS in the gene expression analysis (CASQ2, MYLK, CFL2, SLMAP, ACTG2). The LMS TMAs contained a total of 377 samples that were scored as strong positive (bright red in the heatmap), weak positive (dull red), or negative (green). The antibodies are listed along the y axis and the 377 samples along the x axis. Missing data is indicated by white in the heatmap. Pictures of an LMS sample showing strong expression of all 5 stains is shown to the left of

the heatmap (magnification=200x). The five stains showed significantly correlated expression (all pairwise Spearman's rho p < 0.005, with a minimum correlation of 0.17 between ACTG2 and CASQ2 and a maximum correlation of 0.66 between ACTG2 and SLMAP).

Table 1

Multivariate Cox Regression Analysis of Disease Specific Survival from LMS Tissue Microarray (n=124).

Multivariate Cox Regression Analysis of Disease Specific Survival			
	Wald	HR [95% CI]	P value
CSF1 response protein signature (Espinosa *et al.*, 2009)	13.0	4.7 [2.0–10.9]	< 0.001
Number of Positive Group I/muscle-enriched Markers	**4.5**	**0.77 [0.6–0.98]**	**0.035**
Site	1.8	0.35[0.1–1.6]	0.178
Mitotic Figures	0.8	1.3 [0.7–2.4]	0.371
Grade	0.32	1.3[0.6–2.5]	0.570
Necrosis	0.01	1.0 [0.5 – 2.2]	0.910

In the multivariate model, only the CSF1-response protein expression signature and the expression of Group-I/muscle-associated proteins showed a significant association with survival. In the table, the first column lists each variable included in the model, the second column lists the variable's Wald test statistic, the third column lists the variable's hazard ratio (HR) and 95% confidence interval (CI), and the fourth column lists the variable's Wald test p value.