NEOPLASIA
www.neoplasia.com

Volume 17 Number 2 February 2015 pp. 183–189 183

Molecular Subtypes of Uterine Leiomyosarcoma and Correlation with Clinical Outcome[1,2]

Joyce N. Barlin*, Qin C. Zhou†, Mario M. Leitao*,‡, Maria Bisogna*, Narciso Olvera*, Karin K. Shih*, Anders Jacobsen§, Nikolaus Schultz§, William D. Tap¶, Martee L. Hensley‡,#, Gary K. Schwartz¶, Jeff Boyd, Li-Xuan Qin† and Douglas A. Levine*,‡**

*Gynecology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, USA; †Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA; ‡Weill Cornell Medical College, New York, NY, USA; §Computational Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA; ¶Sarcoma Medical Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA; #Gynecologic Medical Oncology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA; **Division of Molecular Pathology, Fox Chase Cancer Center, Philadelphia, PA, USA

Abstract

The molecular etiology of uterine leiomyosarcoma (ULMS) is poorly understood, which accounts for the wide disparity in outcomes among women with this disease. We examined and compared the molecular profiles of ULMS and normal myometrium (NL) to identify clinically relevant molecular subtypes. Discovery cases included 29 NL and 23 ULMS specimens. RNA was hybridized to Affymetrix U133A 2.0 transcription microarrays. Differentially expressed genes and pathways were identified using standard methods. Fourteen NL and 44 ULMS independent archival samples were used for external validation. Molecular subgroups were correlated with clinical outcome. Pathway analyses of differentially expressed genes between ULMS and NL samples identified overrepresentation of cell cycle regulation, DNA repair, and genomic integrity. External validation confirmed differential expression in 31 genes ($P < 4.4 \times 10^{-4}$, Bonferroni corrected), with 84% of the overexpressed genes, including *CDC7*, *CDC20*, *GTSE1*, *CCNA2*, *CCNB1*, and *CCNB2*, participating in cell cycle regulation. Unsupervised clustering of ULMS identified two clades that were reproducibly associated with progression-free (median, 4.0 *vs* 26.0 months; $P = .02$; HR, 0.33) and overall (median, 18.2 *vs* 77.2 months; $P = .04$; HR, 0.33) survival. Cell cycle genes play a key role in ULMS sarcomagenesis, providing opportunities for therapeutic targeting. Reproducible molecular subtypes associated with clinical outcome may permit individualized adjuvant treatment after clinical trial validation.

Neoplasia (2015) 17, 183–189

Abbreviations: FDR, false discovery rate; FFPE, formalin-fixed and paraffin-embedded; GSEA, gene set enrichment analysis; LMS, leiomyosarcoma; MSKCC, Memorial Sloan Kettering Cancer Center; NL, normal myometrium; OS, overall survival; PFS, progression-free survival; ULMS, uterine leiomyosarcoma

Address all correspondence to: Douglas A. Levine, MD, Gynecology Service, Department of Surgery, Memorial Sloan Kettering Cancer Center, 1275 York Avenue, New York, NY 10065, USA.
E-mail: gynbreast@mskcc.org

[1] Presented in part at the 43rd Annual Meeting of the Society of Gynecologic Oncology, Austin, TX, March 2012. Funding: Carolyn Rosen Miller Family Foundation, The Laura Chang and Arnold Chavkin Charitable Fund, Cycle for Survival of Memorial Sloan Kettering Cancer Center, NIH R01 CA151947, and NIH P30 CA008748.
[2] This article refers to supplementary materials, which are designated by Tables S1 to S9 and Figures S1 to S4 and are available online at www.neoplasia.com.
Received 1 October 2014; Revised 12 December 2014; Accepted 17 December 2014

1476-5586/15
http://dx.doi.org/10.1016/j.neo.2014.12.007

Introduction

Uterine leiomyosarcoma (ULMS), the most common subtype of uterine sarcoma, is a rare tumor, with an annual incidence of 0.64 per 100,000 women [1]. ULMS is aggressive, with a propensity for hematogenous spread leading to local and distant recurrence [2–5]. Surgery is the primary treatment modality, and tumors are often resistant to both chemotherapy and radiation therapy [6,7]. To date, adjuvant therapy has not demonstrated a significant survival advantage. Although surgical staging and nomograms can help predict clinical outcome, the 5-year survival rate for uterus-confined disease remains less than 50% [2,8]. It is difficult to predict the clinical course of ULMS, even when considering clinical and pathologic factors beyond surgical staging. Understanding the molecular biology of ULMS may provide further prognostic and therapeutic insights.

In attempts to understand the pathobiology of ULMS, comparisons have been made to both normal myometrium (NL) and benign uterine leiomyomata. Genome-wide profiling has clustered ULMS, leiomyomas, and NL, demonstrating differences in expression profiles [9]. Immunohistochemistry has also shown differential expression of apoptotic and cell cycle regulatory proteins in ULMS compared to benign smooth muscle tumors [10]. We have previously reported that microRNA profiling supports divergent transformation pathways from normal to benign leiomyoma or ULMS, with ULMS phylogenetically more similar to mesenchymal stem cells than established leiomyomata [11]. The variable and unpredictable behavior of morphologically similar ULMS confined to the uterus supports the need for clinically relevant molecular subtyping. Previous reports have often included uterine and non-ULMS tumors comparing malignant and benign tissues or searching for clinical associations with a variety of study designs [12–14].

Given the wide disparity in outcomes among women with ULMS and the lack of benefit from adjuvant therapy, we performed gene expression profiling on a homogeneous, single-institution set of ULMS without inclusion of tumors from non-gynecologic sites of origin. We tested the hypothesis that ULMS has distinct molecular subtypes that are associated with clinical outcome and may identify therapeutic targets.

Methods

Patient Samples

Frozen and archival formalin-fixed and paraffin-embedded (FFPE) biospecimens were obtained from the tissue repositories at Memorial Sloan Kettering Cancer Center (MSKCC) after Institutional Review Board approval. A gynecologic oncology specialty pathologist reviewed all primary surgical resection specimens. NL and leiomyomata from patients undergoing hysterectomy for benign indications were used for comparison. A discovery cohort of fresh frozen tissues included 29 NL, 25 leiomyomata, and 23 consecutive ULMS specimens collected between 1998 and 2006.

External validation was performed with 46 additional ULMS samples and 14 NL samples. The ULMS samples were consecutive FFPE samples collected between 1998 and 2006. The 14 NL samples were a convenience subset of patients undergoing surgery for benign indications during the same time period. A subset of 29 ULMS cases was used initially to replicate differential expression between ULMS and NL. Two of the ULMS cases were expired during this analysis, leaving 44 ULMS cases to confirm the reproducibility of ULMS clades.

Gene Expression Profiling

RNA was extracted from frozen NL, leiomyoma, and ULMS biospecimens using Ambion *mir*Vana miRNA Isolation Kit (Life Technologies, Grand Island, NY). RNA was quantified and quality assessed using an Agilent Bioanalyzer 2100 at MSKCC and then hybridized to Affymetrix U133A 2.0 human genome microarrays (Affymetrix, Santa Clara, CA) for global mRNA gene expression profiling (Gene Expression Omnibus Series accession number GSE64763). All samples had RIN values >7.0.

RNA for the external validation cohorts was extracted from FFPE tissues using the Ambion RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies). Fragment size was assessed using the Agilent Bioanalyzer 2100 at MSKCC to ensure adequate lengths >300 nucleotides sufficient for hybridization. The NanoString nCounter gene expression system was used for external validation from FFPE biospecimens (NanoString Technologies, Seattle, WA). NanoString technology captures and counts individual mRNA transcripts without the need for enzymatic amplification [15]. Briefly, two probes were designed for each gene of interest complementary to a 100-base region of the target mRNA. Each sample was hybridized in triplicate. Fluorescent barcodes were counted using the nCounter Digital Analyzer. All genes and controls were assayed simultaneously in a multiplex reaction.

To externally validate the differentially expressed genes between ULMS and NL, RNA was hybridized to a NanoString code set of 90 genes with more than four-fold differential expression in the discovery cohort. A separate NanoString nCounter gene expression custom code set of 73 genes differentially expressed between the ULMS clades was selected on the basis of fold change and known function to cluster the 44 external validation ULMS samples.

Microarray Data Analysis

Microarray data from the discovery cohort was normalized with robust multi-array average [16].

Unsupervised hierarchical clustering using Euclidean distance and Ward linkage was performed to identify potential subgroups among the samples. ULMS subgroups identified from unsupervised clustering were compared using a modified *t* test to identify subgroup signature genes. Signature genes distinguishing the two sample clusters were selected using the *t* test *P* values as a ranking criterion and a cutoff of 0.001 [17].

Supervised class comparison generated differentially expressed genes between ULMS and NL samples using a modified *t* test [17]. False discovery rate (FDR) was calculated to adjust for multiple comparisons among the ~22,000 markers on the U133 2.0 array, and an FDR cutoff of 0.0001 was used to select genes that were significantly differentially expressed between ULMS and NL [18]. Ingenuity Pathway Analysis identified overrepresented pathways and networks from the differentially expressed genes. We used GSEA to evaluate differentially expressed genes between tumor and normal samples and between the identified ULMS molecular subtypes [19]. All curated gene sets (MSigDB c2 collection) of size 15 to 300 genes (N = 2294 gene sets) were evaluated. To account for gene-gene correlations in the enrichment analysis, GSEA gene set enrichment *P* values were computed with respect to a null distribution obtained from 100,000 randomizations of the patient-phenotype labels.

NanoString Data Analysis

The nCounter Digital Analyzer quantified RNA molecules of interest, and raw data were normalized to account for differences in hybridization and purification efficiency using 10 control genes in each custom code set.