Emely Möller¹, Jason L. Hornick⁴, Linda Magnusson¹, Srinivas Veerla³, Henryk A. Domanski², and Fredrik Mertens¹

Abstract

Purpose: Low-grade fibromyxoid sarcoma (LGFS) is typically characterized by the specific translocation t(7;16)(q33;p11) and the corresponding fusion gene FUS-CREB3L2. The present study aimed to extract LGFS-specific, and putatively FUS-CREB3L2-dependent, gene expression patterns to learn more about the pathogenesis of this tumor.

Experimental Design: We carried out single nucleotide polymorphism (SNP) and global gene expression array analyses, and/or immunohistochemical (IHC) analyses on 24 LGFS tumor biopsies. Tumor types that are important differential diagnoses to LGFS were included as comparison in the gene and protein expression analyses. In addition, cells that stably expressed FUS-CREB3L2 were analyzed with gene expression array and the influence of FUS-CREB3L2 on gene expression was investigated in vitro.

Results: The SNP array analysis detected recurrent microdeletions in association with the t(7;16) chromosomal breakpoints and gain of 7q in cases with ring chromosomes. Gene expression analysis clearly distinguished LGFS from morphologically similar tumors and MUC4 was identified as a potential diagnostic marker for LGFS by gene expression and IHC analysis. FOXL1 was identified as the top upregulated gene in LGFS and CD24 was upregulated in both LGFS tumors and FUS-CREB3L2 expressing cells. FUS-CREB3L2 was capable of activating transcription from CD24 regulatory sequences in luciferase assays, suggesting an important role for the upregulation of this gene in LGFS.

Conclusions: The gene expression profile of LGFS is distinct from that of soft tissue tumors with similar morphology. The data could be used to identify a potential diagnostic marker for LGFS and to identify possible FUS-CREB3L2 regulated genes. Clin Cancer Res; 17(9); 2646–56. ©2011 AACR.

Introduction

Low-grade fibromyxoid sarcoma (LGFS) typically arises in the deep, intramuscular soft tissue of the proximal extremities or trunk of young adults (1). Histologically, it is characterized by uniform, bland spindle cells of fibroblastic or myofibroblastic differentiation, growing in a whorling pattern within myxoid or collagenized areas. LGFS may be confused with benign tumors such as desmoid fibromatosis (DFM; ref. 2). However, LGFS has a potential for local recurrence and metastasis and myxofibrosarcoma (MFS), another fibroblastic/myofibroblastic myxoid tumor with a heterogeneous appearance is considered an important differential diagnosis to LGFS (3, 4).

The discovery that LGFS has a specific translocation t(7;16)(q33;p11), or in rare cases t(11;16)(p11;p11), has greatly facilitated the diagnosis (5, 6). Through these translocations, the chimeric genes FUS-CREB3L2 or FUS-CREB3L1, respectively, are created. At the molecular level, the 5’-part of FUS, encoding a transactivation domain, is fused to the 3’-part of CREB3L2 or CREB3L1, encoding a basic leucine zipper (bZIP) DNA-binding domain (5, 7). A subset of LGFS cases expresses the FUS-CREB3L2 fusion transcript but lacks the typical t(7;16) and instead harbors a supernumerary ring chromosome, which may contain the fusion gene, as the sole aberration (6, 8). Karyotypic information on LGFS reveals few other recurrent aberrations, suggesting that the chromosomal translocations are tumorigenic events.

The CREB3L2 and CREB3L1 proteins are believed to be endoplasmic reticulum (ER)-resident through their carboxy (COOH)-terminal helical transmembrane domain and activated by regulated intramembrane proteolysis in response to the accumulation of misfolded proteins in the ER (ER stress; ref. 9). On ER stress, the cleaved fragment which contains the bZIP domain is translocated to the nucleus where it may activate transcription more potently...
Translational Relevance

Low-grade fibromyxoid sarcoma (LGFMS) may be confused with other myxoid spindle cell soft tissue tumors of more benign, or malignant, character. The identification of the typical t(7;16)(q33;p11) and corresponding fusion gene FUS-CREB3L2 are the most solid diagnostic criteria for LGFMS; however, additional diagnostic markers are much needed when cytogenetic and/or molecular genetic analyses are not feasible. By comparing the gene expression profile of LGFMS tumor samples with those of morphologically similar tumors, we could in the present study extract LGFMS-specific genes and, in combination with immunohistochemical analyses, identify a potential diagnostic marker for LGFMS, MUC4. Moreover, gene expression array analysis of cells with stable FUS-CREB3L2 expression identified potential FUS-CREB3L2 regulated genes. The presence of FUS-CREB3L2 is believed to be essential for LGFMS tumorigenesis and our results are important to understand more about the role of the fusion protein in LGFMS and the cellular pathways that underlie LGFMS pathogenesis.

Tumor samples

Gene expression array (cases 1–19), IHC analyses (cases 1, 2, 10, 11, and 20–24), ICC (cases 5 and 6), and/or SNP array (cases 1–9) were carried out on 24 LGFMS cases listed in Table 1. The clinical, cytogenetic, and fusion transcript data on 20 of the cases have been published before (5–8). One tumor (case 14) had the FUS-CREB3L1 fusion transcript. The remaining cases had either the FUS-CREB3L2 fusion transcript and/or the t(7;16): 5 of these cases (2–4, 20, and 22) had 1 or more supernumerary ring chromosomes. To detect the presence of FUS-CREB3L2 fusion transcripts when required, cDNA synthesis, reverse-transcriptase PCR (RT-PCR) with the primers TLS-165F and BBF2-1435R, and sequencing were carried out as previously described (6, 14; data not shown). MFS, DFM, SFT, and EMCS were included as comparison to the LGFMS group in the gene and protein expression analyses. The MFS cases were negative for t(7;16) and FUS-CREB3L2 fusion transcript (data not shown) and were of the low-grade variant with 37 to 58 chromosomes (4). The EMCS cases were characterized by translocations t(9;22)(q22;q12) or t(9;17) (q22;q11) resulting in the expression of EWSR1-NR4A3 or TAF15-NR4A3 fusion transcripts (data not shown). Samples were obtained after informed consent and the studies were approved by the local ethics committees. Two pools of total RNA from normal skeletal muscle (Clontech Laboratories) were also included in the gene expression analysis.

Gene expression microarray analyses

RNA from 19 LGFMS cases, and from 6 cases each of MFS, DFM, SFT, and EMCS, was of sufficient quality for the global gene expression analysis. Extraction of total RNA from frozen tumor biopsies, RNA concentration and quality measurements, and hybridization of cDNA to the Human GeneChip Gene 1.0 ST Array (Affymetrix) were carried out as described (15). Background correction, normalization, and probe summarization were done using the Robust Multichip Average (RMA) Method implemented in the Expression Console software Version v 1.0 (Affymetrix).

Two LGFMS samples (cases 10 and 11) and 1 SFT sample were identified as technical outliers in the quality control analysis (using the cutoff thresholds recommended by Affymetrix) and removed from subsequent analyses. Filtering, hierarchical clustering (HCL), principal component analysis (PCA), and statistical analysis of log2 transformed expression data were conducted using the Qlucore Omics Explorer v 2.0 (Qlucore AB). For PCA based on Pearson correlation matrix, the data were normalized through the settings mean = 0 and σ = 1 and variance filtered on the basis of the ratio σ/σmax. HCL was conducted on the basis of Euclidean distance (samples) and Pearson correlation (genes). Differentially expressed genes were identified utilizing correlation-matrix based PCA in combination with ANOVA statistical tests. The Benjamini–Hochberg method was used for error correction (q-value calculation).
<table>
<thead>
<tr>
<th>Case</th>
<th>Analyses</th>
<th>Sex/ age, y</th>
<th>Size(^a)</th>
<th>Site</th>
<th>RT-PCR(^b)</th>
<th>gPCR(^b)</th>
<th>Karyotype</th>
<th>Reference(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SNP, GE, RQ, IHC</td>
<td>M/38</td>
<td>9</td>
<td>Thigh</td>
<td>ex7/ex5</td>
<td>ND</td>
<td>46,XY,t(7;16)(q32;p11)</td>
<td>Case 1838/04, Mertens et al. (5)</td>
</tr>
<tr>
<td>2</td>
<td>SNP, GE, RQ, IHC</td>
<td>M/46</td>
<td>3</td>
<td>Upper arm</td>
<td>ex7/ex5</td>
<td>ND</td>
<td>47,XY,+r</td>
<td>Case 6, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>3</td>
<td>SNP, GE, RQ</td>
<td>M/52</td>
<td>10</td>
<td>Upper arm</td>
<td>in7/ex5</td>
<td>ND</td>
<td>47,XY,+r,t(7;16)(q10)</td>
<td>Case 7, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>4</td>
<td>SNP, GE, RQ</td>
<td>M/34</td>
<td>9</td>
<td>Lower back</td>
<td>in7/ex5</td>
<td>ND</td>
<td>47,XY,+r</td>
<td>Case 8, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>5</td>
<td>SNP, GE, RQ, ICC</td>
<td>M/43</td>
<td>9</td>
<td>Lower leg</td>
<td>in7/ex5</td>
<td>ND</td>
<td>46,XY,t(7;16)(q33;p11)</td>
<td>Case 4, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>6</td>
<td>SNP, GE, RQ, ICC</td>
<td>F/17</td>
<td>8</td>
<td>Axilla</td>
<td>ex6/ex5</td>
<td>ex6/ex5</td>
<td>46,XX,der(5)(l15;16)(q7;15),der(7)(l5;7)(q33;p11),der(7)(ins17;16)(q33;p11),inv(9)(p1q13),c,der(16)(l17;16)(p11)</td>
<td>Case 3, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>7</td>
<td>SNP, GE, RQ</td>
<td>M/42</td>
<td>4</td>
<td>Thigh</td>
<td>ex6/ex5</td>
<td>ex6/ex5</td>
<td>46,XY</td>
<td>Case 10, Panagopoulos et al. (7)</td>
</tr>
<tr>
<td>8</td>
<td>SNP, GE, RQ</td>
<td>M/75</td>
<td>7</td>
<td>Axilla</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>46,XY</td>
<td>Case 3845/03, Mertens et al. (5)</td>
</tr>
<tr>
<td>9</td>
<td>SNP, GE</td>
<td>M/34</td>
<td>?</td>
<td>Foot</td>
<td>ex7/ex5</td>
<td>ex7/ex5</td>
<td>43,XY,-6,add(7)(q32),-13,del(16)(p13),add(21)(p11),-22</td>
<td>Case 1, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>10</td>
<td>GE, IHC analysis</td>
<td>M/39</td>
<td>5</td>
<td>Thigh</td>
<td>ex5/ex6</td>
<td>ND</td>
<td>46,XY,t(7;16)(q33;p11)</td>
<td>Case 3127/03, Mertens et al. (5)</td>
</tr>
<tr>
<td>11</td>
<td>GE, IHC analysis</td>
<td>F/40</td>
<td>7</td>
<td>Thigh</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>45,XX,add(3)(p25),l(7;16)(q33;p11),del(14;15)(q10;10),42-45,idem,-22</td>
<td>Case 5, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>12</td>
<td>GE, RQ</td>
<td>F/21</td>
<td>ND</td>
<td>Trunk wall</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>46,XX</td>
<td>Case 1076/04, Mertens et al. (5)</td>
</tr>
<tr>
<td>13</td>
<td>GE, RQ</td>
<td>M/38</td>
<td>10</td>
<td>Shoulder</td>
<td>ex6/ex5</td>
<td>ex6/ex5</td>
<td>46,XY,t(7;16)(q33;p11)/46,idem,r(10)</td>
<td>Case 1, Storlazzi et al. (7) Case 2, Panagopoulos et al. (6)</td>
</tr>
<tr>
<td>14</td>
<td>GE, RQ</td>
<td>M/42</td>
<td>7</td>
<td>Thoracic wall</td>
<td>ex9/ex5</td>
<td>ND</td>
<td>ND</td>
<td>Case 1453/04, Mertens et al. (5)</td>
</tr>
<tr>
<td>15</td>
<td>GE</td>
<td>F/36</td>
<td>6</td>
<td>Leg</td>
<td>ex5/ins5</td>
<td>ND</td>
<td>ND</td>
<td>Case 1162/04, Mertens et al. (6)</td>
</tr>
<tr>
<td>16</td>
<td>GE</td>
<td>F/13</td>
<td>ND</td>
<td>Thigh</td>
<td>ex6/ins/ex5</td>
<td>ND</td>
<td>ND</td>
<td>Case 1454/03, Mertens et al. (5)</td>
</tr>
<tr>
<td>17</td>
<td>GE</td>
<td>F/33</td>
<td>6</td>
<td>Buttock</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>ND</td>
<td>Case 1455/04, Mertens et al. (5)</td>
</tr>
<tr>
<td>18</td>
<td>GE</td>
<td>M/44</td>
<td>4</td>
<td>Buttock</td>
<td>ins/ex5</td>
<td>ND</td>
<td>46,XY,t(7;16)(q34;p11)/46,idem,tas(15;21)(p11)</td>
<td>Case 1820/04, Mertens et al. (5)</td>
</tr>
<tr>
<td>19</td>
<td>GE</td>
<td>F/12</td>
<td>3</td>
<td>Buttock</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>ND</td>
<td>Case 1905/03, Mertens et al. (5)</td>
</tr>
<tr>
<td>20</td>
<td>IHC analysis</td>
<td>M/77</td>
<td>5</td>
<td>Lung</td>
<td>ex7/ex5</td>
<td>ND</td>
<td>48,XY,+der(7)(l16)(p7),+mar,48,XY,+der(7)(l16),x2</td>
<td>Bartuma et al. (8)</td>
</tr>
<tr>
<td>21</td>
<td>IHC analysis</td>
<td>M/26</td>
<td>3</td>
<td>Abdominal wall</td>
<td>ex6/ex5</td>
<td>ND</td>
<td>46,XY</td>
<td>unpublished</td>
</tr>
<tr>
<td>22</td>
<td>IHC analysis</td>
<td>F/39</td>
<td>?</td>
<td>Thigh</td>
<td>ND</td>
<td>ND</td>
<td>46,XX,t(7;16)(q33;34;34;p11)/46,idem,+r</td>
<td>unpublished</td>
</tr>
<tr>
<td>23</td>
<td>IHC analysis</td>
<td>M/40</td>
<td>7</td>
<td>Thigh</td>
<td>ND</td>
<td>ND</td>
<td>46,XX,t(7;16)(q34;p11)/46,idem,</td>
<td>unpublished</td>
</tr>
<tr>
<td>24</td>
<td>IHC analysis</td>
<td>M/49</td>
<td>1.5</td>
<td>Arm</td>
<td>ND</td>
<td>ND</td>
<td>46,XY,add(1)(p36),l(7;16)(q33;p11)</td>
<td>unpublished</td>
</tr>
</tbody>
</table>

Abbreviations: GE, gene expression array; RQ, relative quantification of gene expression (real-time PCR); gPCR, genomic PCR; ex, exon; in, intron; ins, insertion; ND, not determined.

\(^a\)Largest diameter in cm.

\(^b\)FUS and CREB3L2 breakpoints. Case 14 expresses the FUS-CREB3L1 transcript.

\(^c\)References in which clinical, cytogenetic, and fusion transcript data have been published before.
**Stable transfection of FUS-CREB3L2**

The full-length coding sequence of FUS-CREB3L2 from case 13 cloned into pcR3.1 (13) was used as template in the subcloning procedures. The PCR amplification, subcloning, and sequencing procedures were carried out as described (13). The FUS-CREB3L2 cDNA was amplified with TSL11bBamH1 (5’-GGCGGATCCATGGGCTCAACG-GAATTAACC) and BBF21907Xhol (5’-GGCGGCTCGAGGTGACGGCGACCCCTTCTAGAA), with restriction sites in bold, and cloned in-frame between the BamH1 and Xhol sites of the pCMV-Tag2B vector (Stratagene) and sequenced. pCMV-Tag2B enables the constitutive expression of amino (NH2)-terminal FLAG-tagged proteins in mammalian cells and selection of stable transfectants using G418. HEK293 cells (human embryonic kidney, ICLC) were cultured as described previously (13). A total of 1.2 × 10^6 cells were seeded in 60-mm^2 petri dishes and transfected with 4 μg of the pCMV-Tag2B-FUS-CREB3L2 construct (FC-HEK) or the empty vector (pCMV), using the PolyFect transfection reagent (Qiagen) according to the manufacturer’s instructions. Forty-eight hours later, 500 μg/mL of G418 (Roche) was added to the culture medium for 2 (replicate no. 1) or 3 (replicate no. 2) weeks. RT-PCR was used to confirm the presence of FUS-CREB3L2 fusion transcripts as described above. The expression and subcellular localization of the FLAG-tagged FUS-CREB3L2 fusion proteins in control and transfected cells were detected by Western blot and ICC using the murine anti-FLAG M2 antibody (Stratagene), as described in the Supplementary Material and Methods.

For gene expression array analysis, total RNA from control and FC-HEK (2 culture replicates each) was hybridized to the Human GeneChip Gene 1.0 ST Arrays as described (23). The cloning, and sequencing procedures were carried out as described above. Background correction, normalization, and probe summarization, as well as data and statistical analyses were conducted as described above. cDNA synthesis and real-time PCR of the CD24 and wt CREB3L2 genes were carried out as described above.

**In silico analyses of transcription factor binding sites and promoter regions**

Gene lists of approximately 100 of the most upregulated genes in LGFMS and FC-HEK, as identified by the gene expression analyses, were used for the identification of transcription factor binding sites (TFBS) that were significantly enriched in the data set. The sequence from 1,500 bp upstream (~1,500) to 500 bp (~+500) downstream of each gene’s predicted transcription start site (TSS), as well as the conserved regions (between human and mouse) within the −5,000 to +1,000 sequence, was searched for TFBSs with significant enrichment (P_p) and presence (P_p) in the gene list (compared with 100,000 randomly generated gene lists) using the SMART (Systematic Motif Analysis Retrieval Tool) Software (18). The P_p value is the probability of finding the observed number of instances of a given TFBS and P_p is the probability of finding the observed number of promoters with at least one instance of a given TFBS, in random gene lists. The same software was used to identify TFBS clusters with the QTC (QT-clust) algorithm. The TFBS clusters define specific gene sets which have the same pattern of cooccurring TFBSs in their promoters.

The putative regulatory region of CD24 investigated in silico in the present study extends from −3,000 to +1,500 of the NM_013230 TSS and is part of the sequence with accession no FJ226006 (19). The alignment of FOXL1 orthologous sequences was conducted with Genome-VISTA as described (20), and the sequence which covered the most conserved regions from −3,000 to +1,000 relative to the NM_005250 TSS was chosen for further analyses. The CD24 and FOXL1 regulatory regions were investigated with PromoterScan using the default settings (21). CpG plot was used for putative CpG island identification and TFBS predictions were conducted with MatInspector and Patch public v 1.0 as described (20). The CREB (cAMP responsive element–binding protein) Target gene database was also used to identify CRE full-sites (TGACGTCGA) and half-sites (TGAC/GCGTCA) in the promoters of selected genes (22).

**Reporter gene plasmids and luciferase assays**

The CD24 upstream regulatory region from −2,080 to −1,050 and downstream intron 1 sequence from +400 to +1,300 and the FOXL1 putative promoter region from −1,400 to +1,000, both containing 2 CRE half-sites, were cloned upstream of the firefly luciferase gene in the vector pFLhRL (23). The cloning procedure is described in the Supplementary Material and Methods. HEK293 cells were seeded at a density of 7,000 cells per well in 96-well plates and 24 hours later transfected using the FuGENE HD Transfection Reagent (Roche Applied Science) according to the manufacturer’s recommendations. A total of 500 ng of pFLhRL construct were cotransfected with 100 ng or 1 μg of the pcR3.1-FUS-CREB3L2ΔTM, pcR3.1-CREB3L2ΔTM, or empty pcR3.1 expression plasmids. The plasmids pcR3.1-FUS-CREB3L2ΔTM and pcR3.1-CREB3L2ΔTM have been described before (24). The “ΔTM” proteins lack the transmembrane and COOH-terminal domains and thus correspond to the active, cleaved forms of the chimeric and wt, respectively, transcription factors.
which localize to the nucleus (13). The luciferase activity was quantified 48 hours after transfection. The cell lysis, luciferase measurements, and statistical analysis were conducted as previously described (23). The experiments were repeated twice.

**Tissue microarray–immunohistochemistry**

Paraffin-embedded material from 9 LGFMS, 11 MFS, 12 DFM, 10 SFT, and 6 EMCS cases was available for protein level analyses. Two 1-mm tissue columns were cut from selected donor block areas with a tissue arrayer (Beecher Instruments) and inserted into recipient tissue microarray (TMA) paraffin blocks with 50 cases on each block. The TMA blocks were prepared in duplicate and cut into 4-μm sections for the IHC. Epitope retrieval was achieved by heating the TMA slides in citrate buffer (pH 6.0) in a pressure cooker. The sections were stained with a mouse anti-MUC4 monoclonal antibody (8G-7/ab52263; Abcam) at 1:100 dilution or a rabbit anti-CREB3L2 polyclonal antibody (HPA015068; Atlas Antibodies AB) at 1:50 dilution and counterstained with hematoxylin. The specificity and sensitivity of the MUC4 antibody has been shown in a larger tumor series (25). The anti-CREB3L2 antibody recognizes a NH2-terminal epitope of the wt protein. Normal colonic mucosa and placenta were used as positive controls for the MUC4 and CREB3L2 antibodies, respectively. The extent of immunoreactivity and staining intensities were graded as described (25).

**Combined CD24 ICC and FISH**

Cell cultures from cases 5 and 6 were analyzed with ICC using a CD24 FITC (fluorescein isothiocyanate)-conjugated monoclonal antibody (SN3/ab30350; Abcam). Interphase FISH was then carried out on the same slides by cohybridization of BAC probes for the **FUS** (BAC RP11-388M20) and **CREB3L2** (BACs RP11-29B3 and RP11-377B19) loci (NCBI Build 37, BACPAC Resource Center). The combined FISH and CD24 ICC procedure is described in the Supplementary Material. The human prostate carcinoma cell line DU-145 (DSMZ no. ACC 261) was used as positive control for CD24 expression. HEK293 cells were used as negative control.

**SNP array analysis and data interpretation**

DNA from 9 LGFMS cases was analyzed with SNP array to detect global copy number aberrations. The DNA was extracted as described (26) and DNA concentrations were measured with a NanoDrop 1000 spectrophotometer (Saveen & Werner AB). DNA was hybridized to the HumanOmni1-Quad v1.0 array (Illumina) following standard protocols supplied by the manufacturer. Data analysis was conducted using the GenomeStudio software v2010 1.6.1 (Illumina). Imbalances were identified through combining visual inspection with segmentation analysis of normalized data (27, 28).

**Results**

**Identification of LGFMS-specific expression patterns**

To extract genes that distinguish LGFMS from histologically similar tumor types, MFS, DFM, EMCS, and SFT cases were used as comparison to the LGFMS group (Table 1) in gene expression array analysis. In unsupervised PCA, the data were variance filtered until the 40 samples formed clusters which corresponded to the different tumor types. In this setting (variance ratio, $F = 0.4, 715$ genes), the LGFMS group was clearly different from the other groups and appeared most similar to the MFS group, and then to DFM. An outlier LGFMS sample (case 8) could be identified which appeared to lie in between the LGFMS and MFS groups (Fig. 1A). The filtered data set was then subjected to ANOVA, generating 555 genes with a significant ($P \leq 0.001$) differential expression pattern across the groups, as visualized with HCL (Fig. 1B). In PCA, the data were then filtered on the basis of $P$ values until the up- and down-regulated genes became clearly separated and the most significant genes could be extracted. The cluster of the 54 most upregulated transcripts in LGFMS are listed in Supplementary Table S1 and visible by HCL in Figure 1C. The most significant genes in LGFMS were to a large extent identical regardless of the clustering method (HCL vs. PCA), statistical method (ANOVA vs. multiclass SAM), or program (Quacore vs. tMEV) used or whether the skeletal muscle samples were included or not (data not shown).

The differential expression of selected genes (**CASK**, **CD24**, **FOXL1**, **SLCA1A2**, **TANC2**, and **TMPRSS2**) was confirmed with real-time PCR (Supplementary Fig. S1). **CASK** expression was not detected in MFS, DFM, or EMCS, and **TMPRSS2** was not detected in DFM, SFT, or EMCS ($C_T \geq 35$). The expression of wt **CREB3L2** was also investigated with real-time PCR, showing that this transcript is expressed in all the samples investigated (Supplementary Fig. S1).

**TFBS enrichment in LGFMS upregulated genes**

When analyzing the promoters of the most upregulated genes in LGFMS, we found significant enrichment ($P_E < 0.001$) of binding sites for activator protein 2 (AP2α, AP2γ), the zinc finger proteins ZF5 and CHCH, E2F, and members of the forhead box (FOX) family. Of these, FOX sites were also significantly present ($P_E = 0.001$). Within the conserved promoter regions, only FOX sites were significantly enriched ($P_E < 0.001$). Three QTC clusters, that is, groups of genes with a similar binding site pattern in their promoters, were identified (Supplementary Table S1). QTC cluster 1 (QTC1, 53 genes) was defined by, in particular, the overrepresentation ($P_E < 0.001$) of binding sites for several different FOX factors. Among those, **FOXF1** and **FOXL1** were specifically ($P_E \leq 0.001, q < 1 \times 10^{-10}$) upregulated in LGFMS, and more importantly, **FOXL1** itself belongs to **QTC1** and is the top upregulated gene in LGFMS (Fig. 1E). **FOX1** and **FOXL1**, and **FOXC2** which was also upregulated in LGFMS ($P < 0.001, q < 1 \times 10^{-10}$), are located within a 70-kbp region on 16q24 and transcribed in the same
direction, an organization that is conserved in the mouse (29). The gene MTHFSD is located between the FOX genes, transcribed in the opposite direction, and not specifically upregulated in LGFMS (Fig. 1D). The QTC2 (42 genes) contained genes with coenriched (P < 0.001) binding sites for E2F, AP2 (α/γ), ZF5, CHCH, FOXXN1, EGR, ETF, SP1, and several zinc finger proteins. The 4 genes in QTC3 contained coenriched (P < 0.005) sites for several different transcription factors families, as well as FOX factors. In the QTC clusters, most of the enriched TFBSs were also significantly present (P < 0.005). A few genes from the input list could not be clustered into any of the QTC clusters [denoted (–) in Supplementary Table S1].

TMA–IHC analysis identifies LGFMS-specific expression of MUC4

As MUC4 was one of the top upregulated genes in LGFMS, we evaluated the expression of this gene on the protein level by TMA–IHC. The LGFMS samples (9/9) showed strong cytoplasmic expression of MUC4 (Fig. 2A), whereas the other tumor samples on the TMA were negative. However, one SFT sample was also positive for MUC4 (Fig. 2B). This finding led us to reevaluate the karyotype from this case and to carry out RT-PCR which disclosed a FUS-CREB3L2 fusion transcript (data not shown). This SFT sample was removed from the gene expression analysis due to poor quality but was observed to cluster among the LGFMS samples in unsupervised PCA (not shown). The other SFT cases that were used for gene expression analyses were negative for the FUS-CREB3L2 fusion transcript. The positive control for the MUC4 antibody also displayed cytoplasmic staining (not shown).

Cells with stable expression of FUS-CREB3L2 upregulate LGFMS-specific genes

FUS-CREB3L2 expression from FC-HEK was confirmed with RT-PCR (data not shown), Western blot (Fig. 3A, left),
and ICC (Fig. 3A, right). The expressed FUS-CREB3L2 protein corresponds to the active, cleaved (FUS-CREB3L2ΔTM) form (13), which is localized to the nucleus and thereby capable of influencing transcription (Fig. 3A).

On unsupervised examination of the filtered gene expression array data ($F = 0.1, 3,366$ genes), FC-HEK were clearly different from cells transfected with empty vector (pCMV) and nontransfected cells (control; Fig. 3B, top left). ANOVA ($P < 0.005$) on the filtered gene set gave $196$ differentially expressed genes (whereof $125$ upregulated), as visualized with PCA (Fig. 3B, bottom left) and HCL (Fig. 3B, right). By HCL and PCA, the transfected cells (FC-HEK and pCMV) were more similar to each other than to the controls. A subset of the $125$ upregulated genes in FC-HEK was significantly upregulated also in the tumors: ARSE, CD24, FAM159B, and PAPSS2 (among the LGFMS $54$ top genes) and TSPAN13 and MYOM2 (among the LGFMS $100$ top genes; Fig. 3B and C, top). CREB3L2 was also among these $125$ genes, although this signal likely corresponds to the FUS-CREB3L2 transcript, as wt CREB3L2 was not specifically expressed in FC-HEK by real-time PCR (Fig. 3C, bottom). The specific upregulation of $CD24$ in FC-HEK was confirmed with real-time PCR (Fig. 3C, bottom). Moreover, when examining the unfiltered data with scatter plots, we observed that also TMEM90B, EYA1, NPTX1, ROR1 (among the LGFMS $54$ top genes) and BLM/C, CD9 and VSIN1 (among the LGFMS $100$ top genes) were significantly upregulated ($P < 0.01$) in FC-HEK (not shown).

When identifying enriched TFBSs in promoters of genes that were upregulated in FC-HEK, we found the most significant coenrichment of sites for E2F, AP2 ($\alpha/\gamma$), ZF5, CHCH, EGR, and ETF ($P < 0.0001$), similar to the TFBS pattern of the LGFMS upregulated genes of QT2C.

**FUS-CREB3L2ΔTM activates transcription from the CD24 intronic sequence**

The possibility that FUS-CREB3L2 may activate transcription from the $CD24$ and $FOXL1$ regulatory regions was investigated by utilizing the reporter gene firefly luciferase. Within the $4$-kb $CD24$ regulatory region, Promoter Scan identified $2$ possible promoter regions (scores: $53.7$ and $64.7$) directly upstream of exon $1$ and another (score: $74.3$) in the downstream intron $1$ sequence (from $+820$ to $+1,080$; Fig. 3D, left). All regions coincided with the locations of predicted CpG islands. Two CRE half-sites (TGACC) were identified at $-1,200$ and $-1,101$ shortly upstream of the putative promoter region and at $+768$ and $+1,126$ in association with the downstream intronic regulatory region (Fig. 3D, left). Therefore, both these regions were investigated with the firefly luciferase assay system. Promoter Scan identified $2$ FOXL1 putative promoter regions; at $-2,485$ to $-2,235$ (score: $59.4$) and at $-1,169$ to $-919$ (score: $53.9$; Fig. 3D, right). The latter contained TATA box sequences and was more conserved across species. A CRE half-site was identified at $+871$; we therefore cloned the $-1,372$ to $+1,000$ sequence (Fig. 3D, right). The plasmids containing the cloned $CD24$ or FOXL1 regulatory sequences were cotransfected with plasmids expressing FUS-CREB3L2ΔTM or CREB3L2ΔTM into HEK293 cells. The luciferase assays showed that the FUS-CREB3L2ΔTM chimera and CREB3L2ΔTM activated transcription from the $CD24$ intronic sequence $2.75$ and $5$ times, respectively, more than the empty vector pCR3.1. Only CREB3L2ΔTM activated the $CD24$ upstream sequence, $5$ times more than pCR3.1 (Fig. 3D, left). CREB3L2ΔTM had an effect also through the FOXL1 putative promoter, whereas FUS-CREB3L2ΔTM only had a weak effect (Fig. 3D, right). Moreover, FUS-CREB3L2ΔTM had no effect (not more than empty vector) and CREB3L2ΔTM had a small effect on an active promoter fragment without CRE sites, the pE4 EWSR1 promoter fragment (ref. 20; data not shown).

**Combined FISH and ICC show CD24 expression in tumor cells**

Because our results suggest that FUS-CREB3L2 enhances the expression of $CD24$, we investigated the expression of the $CD24$ protein in tumor cells with $t(7;16)$ rearrangement. The FISH analysis of cells from cases $5$ and $6$ showed the presence of $2$ separated CREB3L2 fusion. Cytoplasmic CD24 expression was also among these $125$ genes, although this signal likely corresponds to the FUS-CREB3L2 transcript, as wt CREB3L2 was not specifically expressed in FC-HEK by real-time PCR (Fig. 3C, bottom). The specific upregulation of $CD24$ in FC-HEK was confirmed with real-time PCR (Fig. 3C, bottom). Moreover, when examining the unfiltered data with scatter plots, we observed that also TMEM90B, EYA1, NPTX1, ROR1 (among the LGFMS $54$ top genes) and BLM/C, CD9 and VSIN1 (among the LGFMS $100$ top genes) were significantly upregulated ($P < 0.01$) in FC-HEK (not shown).

When identifying enriched TFBSs in promoters of genes that were upregulated in FC-HEK, we found the most significant coenrichment of sites for E2F, AP2 ($\alpha/\gamma$), ZF5, CHCH, EGR, and ETF ($P < 0.0001$), similar to the TFBS pattern of the LGFMS upregulated genes of QT2C.
Figure 3. Stable expression of FUS-CREB3L2 in cells (A–C) and the influence of FUS-CREB3L2 on the regulatory sequences of CD24 and FOXL1 (D). A, Western blot (left) of total protein content from cells transfected with FUS-CREB3L2 constructs (FC-HEK) or empty vector (pCMV) and nontransfected cells (control), showing an approximately 55-kDa polypeptide corresponding to the cleaved FUS-CREB3L2 fragment. FUS-CREB3L2 was expressed in the nuclei of transfected cells by ICC (right). B, the gene expression data (F = 0.1, 3,366 genes) show that FC-HEK (purple) are clearly different from pCMV (orange) and controls (turquoise; top left). Lines connect the 2 k-nearest neighbors. Three component axes of the PCA plots are shown. ANOVA (P < 0.005) on the filtered gene set gave 196 differentially expressed genes, as shown by PCA (bottom left) and HCL (right). The HCL color scale ranges from log2 values +1.6 (red; upregulated genes) to −1.6 (green; downregulated genes). The upregulated genes explained most of the variance (76%). Among the significant genes, a subset was also among the 100 most upregulated genes in LGFMS (marked with white triangles in the PCA plot and black lines in the heat map). C, the specific expression of CD24 in FC-HEK shown by scatter plot (top) and real-time PCR (bottom, dark gray bars). Wt CREB3L2 was not specifically expressed in FC-HEK (bottom, light gray bars). D, transcription activation through the CD24 upstream (1) and downstream intronic (2) regulatory sequences (left) and the FOXL1 promoter sequence (right). The positions (bp) are relative the respective TSS (+1). Dark gray boxes represent untranslated exons, black boxes coding exons, lines intronic regions and white arrow heads the location of CRE half-sites (TGACG). The vertical light gray areas denote the putative promoter regions identified by Promoter Scan. The regulatory sequences, each with 1 or 2 CRE half-sites, were cloned upstream of the firefly luciferase gene and cotransfected with 100 ng or 1 μg of the FUS-CREB3L2TM (F-C) or CREB3L2TM (C) expression vectors, or empty vector (pCR3.1). The CD24 intronic sequence (2) was responsive to both F-C and C (**, P < 0.001), whereas the upstream sequence (1) was responsive only to C, compared with the influence of the empty vector (pCR3.1) and the activity of the cloned sequence itself (P). The Genome VISTA alignment of the FOXL1 promoter region (right) shows the sequence conservation (minimum 50% and maximum 100% conservation) between human and Rhesus monkey (top), horse, dog, mouse, zebrafish, and chicken (bottom). The sequence which covers the most conserved regions and contains 1 CRE half-site was responsive to both C and F-C. This effect was, however, not as pronounced as with CD24. DAPI, 4′,6-diamidino-2-phenylindole.
SNP analysis reveals translocation-associated deletions

The LGFMS cases of the present study had no cytogenetically visible aberrations in common other than translocation t(7;16)(q33;p11) and supernumerary ring chromosomes (Table 1). By SNP array analysis, the only recurrent aberrations detected were submicroscopic deletions in association with the genomic breakpoints in CREB3L2 (7q33) and FUS (16p11) and gain of 7q in cases with ring chromosomes. Four cases (cases 2, 3, 5, and 6) had deletions directly upstream of (telomeric to) the CREB3L2 exon 5 breakpoints, and case 6 had an additional 7q33 deletion centromeric to CREB3L2, involving the DGKI gene (Supplementary Fig. S3A). These deletions are most likely located on the homologue involved in the translocation, as all the LGFMS cases expressed the wt CREB3L2 section that is a part of the deletion. Cases 2, 5, and 6 also had deletions directly downstream of (centromeric to) their respective FUS genomic breakpoints (exon 7/intron 7/exon 6). In addition, cases 2 and 6 displayed 16p11 deletions telomeric to FUS involving the genes STX1B (case 2), BCKDK, and MYST1 (case 6; Supplementary Fig. S3B). The positions of the deletions are summarized in Supplementary Table S2. Two (cases 3 and 4) of three cases with ring chromosomes investigated here displayed gain of 7q and 1 (case 3) also had gain of a small 16p11 segment (28.73–28.95 Mbp), which is in agreement with previous reports (8, 30). Combined SNP and karyotypic data from case 3 suggest that the tumor cells contain the gained 7q material in the form of an isochromosome or a supernumerary ring chromosome. In case 4, the ring chromosome, which supposedly contains the gained 7q material, was the sole aberration (Supplementary Fig. S3C).

Discussion

To date, identification of the characteristic t(7;16)(q33;p11) or t(11;16)(p11;p11) and detection of the FUS-CREB3L2/L1 fusion transcripts are the most solid diagnostic criteria for LGFMS. In the present study, 4 LGFMS cases were found at SNP array analysis to have microdeletions in association with the FUS and CREB3L2 breakpoints. In fact, these were the only recurrent imbalances, together with gain of 7q in 2 cases with ring chromosomes, in the cases investigated, emphasizing the pathogenetic importance of the fusion gene in LGFMS. Moreover, the deletions effectively remove most of the FUS and CREB3L2 portions that are not retained in the fusion gene, thereby explaining the absence of reciprocal fusion gene expression in most reported LGFMS cases (6, 7).

On immunohistologic examination, LGFMS stains positive for vimentin and occasionally for CD99, EMA (focally), smooth muscle actin (SMA), and Bcl-2, whereas most cases are negative for desmin, CD34, S100, and cytokeratins (1, 31). These markers are not specific for LGFMS, making the distinction of this entity from other tumors with spindle cell and myxoid characteristics challenging. In the present study, we included MFS and DFM samples, 2 of the most important differential diagnoses to LGFMS, in the gene expression and protein analyses. The gene expression data clearly show that LGFMS is more similar to MFS, followed by DFM, than to SFT and EMCS. MUC4 was one of the top LGFMS upregulated genes and could be shown by TMA–IHC to display LGFMS-specific expression also on the protein level. It is noteworthy that all the analyzed LGFMS cases were clearly positive for MUC4 expression including one case that had been misdiagnosed as an SFT. The specificity of MUC4 as a diagnostic marker for LGFMS has recently been confirmed in a larger series of LGFMS and differential diagnostic entities (25). MUC4 is normally expressed by the epithelial cell layer of most tissues and is overexpressed in several epithelial malignancies (32, 33). It has been suggested that MUC4 interacts with ERBB2 (HER2) to enhance the proliferation, motility, and tumorigenic capacity of epithelial cancer and fibroblast cells through activating ERBB2 downstream pathways and inhibiting integrin-mediated cell adhesion (34–36). It is possible that MUC4 has a similar tumorigenic role also in LGFMS.

The role of the FUS-CREB3L2 chimera in LGFMS pathogenesis is not known but can be hypothesized to involve anomalous regulation of genes normally controlled by CREB3L2 or, alternatively, inappropriate dimerization with members of the CREB family affecting the DNA-binding properties. In line with the first notion, FUS-CREB3L2 and FUS-CREB3L2ΔTM were found to be stronger transcriptional activators than CREB3L2 and CREB3L2ΔTM, respectively, and FUS-CREB3L2 was the strongest activator through the promoter of HSPA5 (alias GRP78 or BiP) and box-B, ATF6, and CRE binding sites (13). CRE binding sites in the HSPA5 promoter are normally bound directly by the CREB3L2/L1 proteins in the unfolded protein response, which acts to rescue the cell from ER stress–induced apoptosis (9). However, FUS-CREB3L2ΔTM was found to be the weakest activator through the cloned sites (13) and had no effect on the CREB3L2 promoter, which contains a conserved CRE site, whereas CREB3L2ΔTM did (24), suggesting a discrepancy in the regulatory actions of the wt and chimeric proteins. In the present study, HSPA5 and CREB3L2 were not differentially expressed in LGFMS tumor samples or in FC-HEK, suggesting that the regulation of these genes is not FUS-CREB3L2 dependent. Potentially FUS-CREB3L2 regulated genes may be found in the subset of genes that was specifically upregulated in both LGFMS tumors and FC-HEK. Among those genes, CD24 was chosen for in vitro studies because the regulatory sequences of this gene contain several CRE half-sites. We found that CREB3L2ΔTM and FUS-CREB3L2ΔTM enhanced transcription from a CD24 downstream intronic sequence, which contains 2 CRE half-sites, whereas only CREB3L2ΔTM activated the upstream sequence. However, the upstream cloned sequence did not include the in silico predicted promoter regions or CpG island and had a weak promoter activity on its own. It is possible that FUS-CREB3L2ΔTM requires additional binding sites for full activity. In LGFMS, the CREB3L2 wt transcript was detected and the protein was...
predominantly found to be weakly expressed, a pattern that is not specific for LGFMS and reflects the normal condition. Hence, the CREB3L2 bZIP domain is overexpressed in LGFMS through the fusion protein as a gain-of-function mechanism. This might be sufficient to cause anomalous target gene expression, even though the transcription activation by FUS-CREB3L2 may be weaker than that of CREB3L2 through the CRE half-site. Our results suggest that FUS-CREB3L2 enhances the expression of CD24; however, a possible direct, or indirect, interaction between FUS-CREB3L2 and CD24 regulatory sequences remains to be proven. CD24 was upregulated in FUS-DDIT3 expressing NIH-3T3 murine fibroblasts and mesenchymal progenitor cells (37, 38), perhaps suggesting the contribution of the FUS domain to the upregulation of CD24. CD24 is a glycosylated cell surface mucin which is involved in T-cell proliferation, synaptic transmission, immune response, and cell adhesion through its interaction with P-selectin on endothelial cells or platelets (39, 40). CD24 is expressed at higher levels in carcinomas of the breast, ovary, lung, prostate, pancreas, bladder, gastrointestinal and biliary tracts, and in neuroepithelial tumors, compared with the corresponding normal tissues (40, 41). LGFMS cells with the t(7;16) were shown to have cytoplasmic expression of CD24, further suggesting a role for this mucin in LGFMS.

Few properties of CREB3L2/L1 regulated genes have been suggested. In the promoters of LGFMS upregulated genes, we found a significant overrepresentation of binding sites for the AP2, CHIC, E2F, FOX, and ZF3 factors, parting the genes into specific, potentially coregulated, subgroups. Of the FOX factors with overrepresented binding sites, FOXI1 and FOXL1 were specifically expressed in LGFMS, suggesting an important role for these FOX factors in the tumorigenesis of LGFMS. Because the FOXL1 promoter region contains a CRE half-site, we investigated the transcriptional activation from this sequence in vitro. CREB3L2ΔTM was found to activate the FOXL1 promoter, whereas FUS-CREB3L2ΔTM had a weak effect. The effect of these proteins on the CD24 regulatory sequences was much more pronounced. In agreement with these results, FOXL1 was not upregulated in FC-HEK, suggesting that the upregulation of this gene is not directly caused by FUS-CREB3L2. Foxl1 and Foxf1 have been identified as targets of the hedgehog (Hh) signaling pathway in the murine developing mesoderm; their promoters were directly bound by Gli1 and Gli2 and the Foxl1, Foxf1, and Gli1 transcript levels increased in response to Sonic hedgehog (Shh) treatment (29). Ligand-dependent activation of the Hh pathway has been associated with epithelial cancers in which Hh ligands and/or targets, such as SHH and GLI1, are overexpressed (42). Pitch1, SHH, GLI1, and GLI2 were found to be expressed at higher levels in LGFMS compared with many of the other tumor samples, although not displaying clearly distinct LGFMS-specific expression (Supplementary Fig. S4). This suggests that the Hh pathway may be involved in LGFMS tumorigenesis, although further studies are needed to investigate this aspect.

Our results show that the gene expression profile of LGFMS is distinct from that of soft tissue tumors with similar morphology. We could use the gene expression data to identify a potential diagnostic marker for LGFMS, MUC4, which displayed highly LGFMS-specific expression on both the transcript and protein levels. As the LGFMS tumors have no other aberrations in common, the fusion gene event is likely to be necessary for tumor formation. Although HEK293 cells and LGFMS have diverse genetic backgrounds and microenvironments, specific FUS-CREB3L2-dependent gene expression patterns could be extracted and used to identify putative FUS-CREB3L2 target genes.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were declared.

Acknowledgments

The authors thank the Swedge Centre for Integrative Biology at Lund University (SCIBLU) for help with the SNP and gene expression microarray laboratory work and image analysis. The authors also Margareth Isaksson and Jenny Nilsson at the Department of Clinical Genetics in Lund for help with the laboratory work and SNP data analysis.

Grant Support

This work was supported by the Swedish Children’s Cancer Foundation, the Swedish Cancer Society, the Swedish Research Council, and the Royal Physiographic Society of Lund (E. Möller and F. Mertens). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 18, 2011; revised February 25, 2011; accepted March 6, 2011; published online April 26, 2011.

References


FOXL1 and CD24 Expression Profile with Upregulation of CD24 and FOXL1

Emely Möller, Jason L. Hornick, Linda Magnusson, et al.

Clin Cancer Res 2011;17:2646-2656.

Updated version
Access the most recent version of this article at:
http://clincancerres.aacrjournals.org/content/17/9/2646

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2011/05/05/17.9.2646.DC1

Cited articles
This article cites 40 articles, 10 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/17/9/2646.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/17/9/2646.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.