In-depth genetic analysis of sclerosing epithelioid fibrosarcoma reveals recurrent genomic alterations and potential treatment targets


1Division of Clinical Genetics, Department of Laboratory Medicine, Lund University, SE-221 84 Lund, Sweden;
2Department of Pathology and Clinical Genetics, Sahlgrenska University Hospital, Gothenburg, Sweden;
3Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA;
4Department of Histopathology, Royal National Orthopaedic Hospital, Stanmore, UK;
5Department of Pathology, KU Leuven and University Hospitals, Leuven, Belgium;
6Center for Human Genetics, KU Leuven and University Hospitals, Leuven, Belgium;
7Department of Musculoskeletal Pathology, Royal Orthopaedic Hospital NHS Foundation Trust, Birmingham, UK;
8Department of Clinical Genetics, Office for Medical Services, Division of Laboratory Medicine, 221 85 Lund, Sweden

Running Title: Genetic characterization of SEF.

Keywords: SEF, LGFMS, CD24, fusion genes, DMD.

Abbreviations: desmoid tumors (DFM), extraskeletal myxoid chondrosarcomas (EMC), formalin-fixed paraffin-embedded (FFPE), global gene expression (GGE), low grade fibromyxoid sarcoma (LGFMS), myxofibrosarcomas (MFS), ossifying fibromyxoid tumor (OFMT), principal component analysis (PCA), RNA sequencing (RNA-seq), sclerosing epithelioid fibrosarcoma (SEF), single nucleotide polymorphism (SNP), single nucleotide variant (SNV), solitary fibrous tumors (SFT), whole exome sequencing (WES).
Financial support: the Swedish Cancer Society, the Swedish Childhood Cancer Foundation, and Region Skåne (FM); P50 CA140146-01 and P30-CA008748 (CRA).

Corresponding author: Elsa Arbajian; mailing address: BMC C13, Klinikksgatan 26, 22184 Lund, Sweden; email: elsa.arbajian@med.lu.se
ABSTRACT

Purpose: Sclerosing epithelioid fibrosarcoma (SEF) is a highly aggressive soft tissue sarcoma closely related to low-grade fibromyxoid sarcoma (LGFMS). Some tumors display morphological characteristics of both SEF and LGFMS, so called hybrid SEF/LGFMS. Despite the overlap of gene fusion variants between these two tumor types, SEF is much more aggressive. The present study aimed to further characterize SEF and hybrid SEF/LGFMS genetically in order to better understand the role of the characteristic fusion genes and possible additional genetic alterations in tumorigenesis.

Experimental Design: We performed whole exome sequencing, single nucleotide polymorphism (SNP) array analysis, RNA-sequencing (RNA-seq), global gene expression analyses and/or IHC on a series of 13 SEFs and 6 hybrid SEF/LGFMS. We also expressed the \textit{FUS-CREB3L2} and \textit{EWSR1-CREB3L1} fusion genes conditionally in a fibroblast cell line; these cells were subsequently analyzed by RNA-seq and expression of the CD24 protein was assessed by FACS analysis.

Results: The SNP array analysis detected a large number of structural aberrations in SEF and SEF/LGFMS, many of which were recurrent, notably \textit{DMD} microdeletions. RNA-seq identified \textit{FUS-CREM} and \textit{PAX5-CREB3L1} as alternative fusion genes in one SEF each. \textit{CD24} was strongly upregulated, presumably a direct target of the fusion proteins. This was further confirmed by the gene expression analysis and FACS analysis on Tet-On 3G cells expressing \textit{EWSR1-CREB3L1}.

Conclusions: While gene fusions are the primary tumorigenic events in both SEF and LGFMS, additional genomic changes explain the differences in aggressiveness and clinical outcome between the two types. CD24 and DMD constitute potential therapeutic targets.
TRANSLATIONAL RELEVANCE

Sclerosing epithelioid fibrosarcoma (SEF) is a highly malignant soft tissue sarcoma; while patients with localized tumors may be cured by surgery, the prognosis for patients with disseminated disease is dismal. Previous studies have shown that most SEF display gene fusions that to some extent overlap with those found in low-grade fibromyxoid sarcoma (LGFMS), a tumor type associated with much better outcome than SEF. We here show that the clinical differences between SEF and LGFMS can largely be explained by complex genomic rearrangements, including recurrent intragenic deletions of the DMD gene encoding dystrophin. Furthermore, both in vivo and in vitro, the EWSR1-CREB3L1 fusion, which predominates in SEF, leads to higher expression of the CD24 gene than does the FUS-CREB3L2 fusion, which predominates in LGFMS. Both DMD and CD24 constitute promising treatment targets.
INTRODUCTION

Sclerosing epithelioid fibrosarcoma (SEF) is a malignant soft tissue tumor that preferentially affects middle-aged and elderly individuals. It is highly aggressive, with high rates (>50%) of local recurrence and distant spreading; apart from tumor size and location, no strong predictive markers of metastasis have been identified (1-5). MUC4 is a useful immunohistochemical marker for diagnosis (4).

In some cases, more or less extensive areas reminiscent of low-grade fibromyxoid sarcoma (LGFMS) can be seen, either in synchronous or metachronous combination with SEF areas; such tumors are known as hybrid SEF/LGFMS (1,5). LGFMS is a morphologically low-grade fibroblastic tumor that typically affects younger adults and children. Also LGFMS cells are positive for MUC4, in accordance with the fact that it was the most strongly differentially expressed gene compared to various histologically similar tumors (6,7). Although it has a much lower potential for recurrence (10%) and metastasis (5%) than SEF within the first 5 years, it is notorious for late occurring metastases (8,9).

The intriguing morphological and clinical overlaps and differences between SEF and LGFMS could perhaps be explained by shared and distinct genetic features, respectively. LGFMS is by far the more extensively analyzed subtype of the two. Around 95% of the cases show a FUS-CREB3L2 fusion gene, typically on the basis of a balanced translocation t(7;16)(q33;p11) (10,11). The remaining cases display either a t(11;16)(p11;p11), resulting in a FUS-CREB3L1 fusion (12) or, in rare cases, an EWSR1-CREB3L1 fusion (13,14). Cytogenetic and genomic array data have identified few additional changes, suggesting that the fusion gene is both necessary and sufficient for tumorigenesis, and global gene expression analysis has identified a distinct profile...
(7,10). The genetic information on SEF is limited to data on fusion gene status, revealing a predominance (80-90% of the cases) of EWSR1-CREB3L1 (15-17), but occasionally showing recurrent alternative EWSR1-CREB3L2 and FUS-CREB3L2 fusions (17,18). Furthermore, a similar FUS-CREB3L2 fusion as seen in LGFMS has been also found in most hybrid SEF/LGFMS cases (4,15,17), but some cases may display rearrangements of EWSR1 and/or CREB3L1 (4).

To understand better why SEF and LGFMS show different clinical phenotypes and why some LGFMS recur as SEF, we analyzed pure SEFs and hybrid SEF/LGFMS, using high-resolution single nucleotide polymorphism (SNP) arrays, global gene expression (GGE) profiling, transcriptome sequencing (RNA-seq), and whole exome sequencing (WES). In addition, the impact of the FUS-CREB3L2 and EWSR1-CREB3L1 chimeras was further investigated in vitro using the Tet-On 3G inducible gene expression system.

**MATERIALS AND METHODS**

**Patients and Tumors**

The study included 15 tumors from 13 patients classified as pure SEF and six tumors from four patients with hybrid SEF/LGFMS; of the latter six tumors, five showed SEF morphology and one LGFMS morphology. The histological appearance of case 17, with synchronous SEF and LGFMS areas, can be seen in Fig 1. In addition, tumors from 18 patients with LGFMS were included for comparison of gene fusion status, GGE profiles, genomic imbalances, and/or whole
exome mutation profiles. Finally, 40 other soft tissue tumors – 11 myxofibrosarcomas (MFS), six desmoid tumors (DFM), 13 solitary fibrous tumors (SFT), and six extraskeletal myxoid chondrosarcomas (EMC) – were used for comparisons of GGE profiles. For filtering of mutations at WES, mutation data on five simultaneously analyzed cases of ossifying fibromyxoid tumor (OFMT) were used. All tumors were diagnosed according to established criteria (19). Clinical data and information on the analyses performed in each case are summarized in Supplementary Table 1.

The studies were conducted in accordance with the Declaration of Helsinki. All studies were performed after approval by the institutional review board (IRB) and after obtaining written consent from the subjects.

**Immunohistochemistry**

CD24 immunohistochemistry on formalin-fixed paraffin-embedded (FFPE) tissue was performed following pressure cooker epitope retrieval (Novocastra Epitope Retrieval Solution pH 6, Newcastle Upon Tyne, UK), using a rabbit polyclonal antibody raised against KLH-conjugated synthetic peptide derived from human CD24 (dilution 1:100, overnight incubation at 4°C, Biorbyt Ltd. Cambridge, UK). The EnVision Dual Link system (Dako) was used for visualization.

**RNA sequencing**

RNA was extracted from 12 fresh frozen tumor samples and from 25 cell lines. mRNA libraries were prepared for sequencing using the Truseq RNA sample preparation kit v2 (Illumina, San
Diego, USA), as described (20). Paired-end 151 bp reads were generated from the mRNA libraries on a NextSeq 500 (Illumina). From another five tumors mRNA with DV$_{200}$ values $\geq 40$ could be extracted from FFPE blocks using Qiagen’s RNeasy FFPE Kit (Qiagen, Valencia, CA, USA), as described (21). mRNA libraries were prepared from 20-50 ng of RNA, depending on the DV$_{200}$ value, using the capturing chemistry of the TruSeq RNA Access Library Prep Kit (Illumina). Paired-end 85 nt reads were generated from the mRNA libraries on a NextSeq 500 (Illumina). Sectioning, RNA extraction, library preparation and sequencing were performed as described (21).

ChimeraScan and FusionCatcher, using default settings, were used to identify candidate fusion transcripts from the sequence data (22,23). The GRCh37/hg19 build was used as the human reference genome.

For gene expression studies, data were normalized using Cufflinks with default settings (24). Correlation-based principal component analysis (PCA) and hierarchical clustering analysis were performed using the Qlucore Omics Explorer version 3.2 (Qlucore AB, Lund, Sweden). Differences between cell clones in log2 transformed expression data were calculated using a t-test, and corrections for multiple testing were based on the Benjamini-Hochberg method (Qlucore AB). Genes with $P < 0.01$ and a false discovery rate (FDR) $< 0.07$ were considered significantly altered.

**RT-PCR analyses to confirm gene fusions**

The same RNA that had been used for RNA-seq was reverse transcribed and PCR amplified as described (7,25). Primers specific for $FUS$, $EWSR1$, $PAX5$, $CREB3L1$, $CREB3L2$, and $CREM$ were designed to detect fusion transcripts (Supplementary Table 2). Transcripts were amplified
using an initial denaturation for 5 min at 94°C, followed by 30 cycles of 1 min at 94°C, 1 min at 58°C, and 2 min at 72°C, and a final extension for 5 min at 72°C. Amplified fragments were purified from agarose gels and directly sequenced using the Big Dye v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA) on an ABI-3130 genetic analyzer (Applied Biosystems). The BLASTN software (http://www.ncbi.nlm.nih.gov/blast) was used for the analysis of sequence data.

**SNP array analysis and WES**

DNA was extracted from eight fresh frozen tumor samples (six pure SEF, two hybrid SEF/LGFMS) and prepared for SNP array analysis using the Affymetrix Cytoscan HD array (Affymetrix, Santa Clara, CA, USA), as described (26). From another four tumor samples (one pure SEF, three hybrid SEF/LGFMS) DNA was extracted from FFPE blocks using QIAamp DNA FFPE Tissue Kit and prepared for SNP array analysis using the Oncoscan CNV array (Affymetrix), according to the manufacturer’s instructions. The position of the SNPs was based on the GRCh37/hg19 sequence assembly. SNP array analysis was done using the Affymetrix Chromosome Analysis Suite (ChAS) software. The SNP array data on SEF and hybrid SEF/LGFMS cases were compared to previously published, and here re-evaluated, SNP array data on 10 cases of LGFMS (7) that had been analyzed using the Illumina HumanOmni-Quad version 1.0 array (Illumina). The LGFMS data were converted to the hg19 genome assembly using UCSC LiftOver.

See Supplementary Methods for details regarding WES.
Global gene expression microarray analysis

RNA from 10 tumors (eight pure SEF, two hybrid SEF/LGFMS) was of sufficient quality for GGE analysis using the Human GeneChip Gene 1.0 ST Array (Affymetrix, Santa Clara, CA). Extraction of total RNA from frozen tumor biopsies, RNA concentration and quality measurements, and hybridization were carried out as described (27). Gene expression data were normalized, background-corrected, and summarized by using the Robust Multichip Analysis algorithm implemented in the Expression Console version 1.4 software (Affymetrix). These data were compared to previously published (7) gene expression data from 17 LGFMS, 11 MFS, six DFM, 13 SFT, and 6 EMC. Correlation-based PCA and hierarchical clustering analysis were performed using the Qlucore Omics Explorer version 3.2 (Qlucore AB, Lund, Sweden). Differences between tumor groups in log2 transformed expression data were calculated using a t-test, and corrections for multiple testing were based on the Benjamini-Hochberg method (Qlucore AB). Genes with P < 0.01 and FDR < 0.07 were considered significantly altered.

Tet-On 3G setup

An hTert immortalized Bj5ta human fibroblast cell line (cultured in 4:1 DMEM: 199 medium with 10% Tet approved FBS + 0.01 mg/ml Hygromycin B) was used throughout the experiments. Bj5ta was sequentially transduced with the pLVX-Tet3G vector encoding the regulator Tet-On 3G protein and with one of the response plasmids pLVX-TRE3G-FUS-CREB3L2, pLVX-TRE3G-EWSR1-CREB3L1, pLVX-TRE3G-EWSR1-NR4A3 or pLVX-TRE3G-EMPTY, encoding the different fusion genes of interest. Expression of the inserted gene fusion was turned on by addition of doxycycline (dox) to the culture medium. The EWSR1-NR4A3 expressing cell line was used for comparison in subsequent expression studies (See Supplementary Methods for further details).
Flow cytometry

To evaluate the expression of the CD24 protein in the fusion gene-expressing Tet-On 3G transduced cells cultured with or without dox, flow cytometry was performed (See Supplementary Methods for further details).

RESULTS

Identification of SEF-specific expression patterns

At unsupervised PCA, the data were variance filtered until the 63 samples formed clusters which corresponded to the different tumor types. In this setting (variance ratio, F= 0.4, 882 genes), the LGFMS and SEF groups were clearly different from the other groups and appeared most similar to each other and then to the MFS group. The filtered data set was then subjected to ANOVA tests, generating 424 genes with a significant (P< 0.01) differential expression pattern between SEF and the control tumors; a second ANOVA test generated 375 genes with a significant (P< 0.01) differential expression pattern between SEF and LGFMS. In the PCA, the data were then subjected to one sided t-tests to extract the most up- and downregulated genes. The cluster of the 100 most upregulated transcripts in SEF versus controls and SEF vs LGFMS are listed in Supplementary Table 3.

CD24 was the most upregulated gene (80x) in SEF (Fig.2A). Consistent with this finding, IHC for CD24 on SEF cases showed cytoplasmic and membranous staining (Fig.2C). CD24 was also
upregulated (37x) in LGFMS, in accordance with the findings by Möller et al (7). Also *MUC4* was upregulated in both LGFMS and SEF (Supplementary Fig.1).

The two hybrid SEF/LGFMS cases appeared to be most closely related to each other and then to pure SEF cases. Among pure SEF, *EWSR1-CREB3L1*- and *EWSR1-CREB3L2*-positive tumors did not cluster separately, whereas the single case with a *FUS-CREM* fusion clustered further apart from the other SEF cases.

**Fusion gene status**

The gene fusion status of the SEF and hybrid SEF/LGFMS cases was assessed through RNA-Seq, SNP array, RT-PCR, and/or FISH. Among the 13 cases of pure SEF, eight were positive for *EWSR1-CREB3L1*, two for *EWSR1-CREB3L2*, and one each for *FUS-CREM, FUS-CREB3L2, and PAX5-CREB3L1*. All six samples, including five with SEF morphology and one with LGFMS morphology, from four cases of hybrid SEF-LGFMS displayed a *FUS-CREB3L2* fusion (Supplementary Table 1). The previously undescribed *FUS-CREM* and *PAX5-CREB3L1* transcripts were confirmed by RT-PCR, confirming the presence of an in frame fusion of *FUS* ex 11 with *CREM* ex 7 and *PAX5* ex 2 with *CREB3L1* ex 6, respectively.

**SNP array**

Detailed information on all imbalances exceeding 100 kb, as well as smaller (> 10 kb) homozygous deletions and imbalances affecting genes involved in fusions, is given in Supplementary Table 4 and visualized in Fig.3. In all cases, the profiles were compatible with a near-diploid chromosome count.
Using the 100 kb cut-off, imbalances were found in 7/10 LGFMS cases (mean 1.4, range 0-3). The only recurrent (2-3 cases) aberrations were gain of material from chromosomes/arms 7q, 12 and 15q. Three and four cases, respectively, displayed microdeletions involving the FUS and CREB3L2 loci.

All seven samples from pure SEF showed multiple imbalances (mean 28, range 6-78). Most imbalances were due to structural rearrangements; only 11/197 aberrations affected whole chromosomes. Several imbalances were recurrent: loss of several regions in 22q12-qter (minimal overlapping regions 38.12-40.93 Mb, 41.78-42.56 Mb, 44.97-49.46 Mb, 49.92-50.26 Mb, and 50.85-51.15 Mb) in six cases, loss of 11p (1-45.96 Mb) and proximal 22q (16.89-23.05 Mb) in five cases each, and gain of 1q31-qter (188.32-248.84 Mb), loss of 10q21 (56.18-56.65 Mb), 13q21 (71.31-72.01 Mb), 22q12 (29.29-29.68 Mb), and Xp21 (32.48-32.70 Mb) in three cases each. High-level (> 5 copies) gain was rare, involving three different regions in three cases. Complete loss (0 alleles) was seen in four cases, affecting the DMD gene in Xp21 in two of them. One or more copy number transitions in the EWSR1, CREB3L1, and CREB3L2 loci, typically deleting the parts not included in the fusion genes, were seen in five, four, and one case, respectively. The single case with a FUS-CREM fusion showed copy number transitions in both loci.

All five samples from three patients with hybrid SEF/LGFMS showed multiple imbalances (mean 20, range 17-22). The two samples from Case 15 (a local recurrence and a metastasis) had less than 15% of the imbalances in common, whereas the two samples from the primary tumor of Case 17 (one showing SEF morphology and one LGFMS morphology) shared >80% of the aberrations. Two imbalances were present in all three patients with hybrid SEF/LGFMS: loss and/or LOH for the entire 11p, and loss of one copy of chromosome 22. In addition, both samples
from both Cases 15 and 17 showed homozygous deletions of the \textit{DMD} gene (Fig. 4). All three cases showed copy number shifts in the \textit{FUS} and \textit{CREB3L2} loci.

In summary, the SNP array profiles of SEF and hybrid SEF/LGFMS were very similar to each other and much more complex than those in LGFMS.

\textbf{Mutation pattern in tumors}

WES on five SEF and two hybrid SEF/LGFMS achieved a mean coverage >100x, with 99\% of the targeted bases covered to a depth of >x20. Filtering of the data resulted in 260 potential mutations (Supplementary Table 5). Targeted DNA sequencing of the tumor samples, with a median depth of 204 reads per amplicon, confirmed 72/125 tested variants, of which 50 were unique mutations (Supplementary Table 5). No gene was consistently involved.

WES was also performed on five cases of LGFMS with paired blood samples, achieving a mean coverage 85x. 1-12 single nucleotide variants (SNVs) and 0-1 insertions/deletions were found per case (Supplementary Table 5). None of the mutations was recurrent. Two samples from a hybrid SEF/LGFMS case were analyzed; 17A showing LGFMS morphology and 17B showing SEF morphology. This resulted in 7 SNVs each (Supplementary Table 6).

\textbf{Cells with stable expression of \textit{FUS-CREB3L2}, \textit{EWSR1-CREB3L1}, or \textit{EWSR1-NR4A3}}

Of the 100 most highly expressed genes in SEF tumors compared to controls, 38 were also significantly upregulated (fold change>2, p<0.05) in Tet-On 3G EWSR1-CREB3L1 cells
cultured with dox (Supplementary Table 6). Among these, CD24 was upregulated 135x (Fig 2A).
CD24 was also upregulated (28x) in TetOn 3G FUS-CREB3L2 cells cultured with dox.

CD24 protein expression in the transduced cells cultured with or without dox and in Bj5ta cells was assessed by flow cytometry. TetOn 3G EWSR1-CREB3L1 cells cultured with dox showed significant upregulation of cell surface CD24 (9x) and of total CD24 (50x), as compared to the no dox counterparts (Fig. 2B). TetOn 3G FUS-CREB3L2 cells cultured with dox showed a 7x upregulation of cell surface CD24 and a 3x upregulation of total CD24 compared to the no dox counterparts (Supplementary Fig. 2).

DISCUSSION

Based on the limited genetic information previously available on SEF and LGFMS, the prognostic difference between the two tumor types was difficult to explain; although two different gene fusions predominate in these tumors – EWSR1-CREB3L1 and FUS-CREB3L2, respectively – these chimeras are so closely related that they could be predicted to have similar down-stream effects. Indeed, the comprehensive genetic analyses presented here strongly suggest that the poor outcome for SEF patients could largely be accounted for by extensive secondary genomic alterations. While LGFMS has few aberrations at the genome or nucleotide level in addition to the characteristic FUS-CREB3L2 chimera, SEF consistently displays complex genomic alterations. Strong support for the relative importance for outcome of genomic imbalances compared to gene fusion status was derived from the hybrid SEF/LGFMS cases. These tumors have the same FUS-CREB3L2 fusion as LGFMS, but display the same types of genomic alterations and show the same poor outcome as pure SEF.
Point mutations seem to have little impact on the clinical outcome, since no consistent mutations were identified, in either SEF, LGFMS or hybrid SEF/LGFMS. WES on LGFMS, performed with paired blood samples, showed no recurrent mutation and no recurrently mutated gene. The results of WES on SEF and SEF/LGFMS, which resulted in 260 mutations, were harder to interpret because no paired blood sample was available and we can therefore not rule out that some of the detected mutations were constitutional.

Some of the recurrent genomic imbalances in SEF were related to the underlying gene fusion, notably deletions of 11p (where CREB3L1 is located) and 22q (where EWSR1 is located). However, at least the 11p deletions could not simply be considered side effects of translocations giving rise to the EWSR1-CREB3L1 chimera; copy neutral LOH or deletions of 11p were found also in SEFs without CREB3L1 fusions. Loss/LOH of 11p is frequently seen in many other sarcomas, notably embryonal rhabdomyosarcomas (26). The expression pattern of imprinted genes in 11p15 (Supplementary Fig. 1B) indicates that the maternal copy of 11p is lost in SEF. However, the correlation between 11p deletion and gene expression is unclear; cases without 11p deletion displayed the same pattern of expression and only two genes in 11p, PAMRI and ADM, were among the 100 most downregulated genes in SEF vs controls (Supplementary Table 3C).

Another link with myogenic sarcomas was the frequent finding of intragenic deletions of the DMD gene, encoding the dystrophin protein that is causally involved in inherited muscular dystrophies. While no such aberrations were seen in LGFMS, 2/7 pure SEF and 2/3 hybrid SEF/LGFMS with SNP array data showed DMD deletions of variable size (Supplementary Table 4). The deletions always started after exon 1 and extended no longer than intron 74. In all cases the deletion on the X-chromosome was quite distinct, targeting only the DMD locus, implying that these deletions are not by chance, but selected for during tumor growth. From two of the
patients with hybrid SEF/LGFMS we had two samples each with *DMD* deletions. In Case 15, the deletions were different in the local recurrence (15A) and the metastasis (15B), strongly arguing for late and independent mutations in the two samples, while identical deletions were seen in the two portions from a primary hybrid SEF/LGFMS, one showing LGFMS morphology (Case 17A) and one SEF morphology (17B). In Case 17, the deletion must therefore have occurred relatively early, and cannot explain the morphological transition from LGFMS to SEF. No effect on the transcription levels of the *DMD* gene was seen at global gene expression profiling, which strongly suggests that a truncated mRNA is expressed, in keeping with the fact that none of the deletions involved the entire coding sequence.

Similar intragenic *DMD* deletions were recently described in advanced sarcomas showing myogenic differentiation (28). Based on various types of supportive data, the authors concluded that intragenic *DMD* deletions, abrogating expression of full-length dystrophin, is an important mechanism behind tumor progression in myogenic sarcomas (28). As far as we are aware, SEF does not show any myogenic differentiation, which combined with the occasional finding of *DMD* deletions also in other malignancies, such as malignant melanomas, suggests that a tumor suppressor function of DMD is not limited to myogenic cells. The *DMD* locus encodes a variety of protein isoforms, some of which are expressed by non-muscle cells. Notably, the Dp71 protein, encoded by exons in the 3’-end of *DMD*, is involved in many cellular processes including cell adhesion, mitosis, and nuclear architecture (29). This notwithstanding, the finding of *DMD* deletions in yet another highly malignant type of sarcoma provides further incentive for evaluating means to restore dystrophin expression as a treatment option. Another possible scenario is that *DMD* does not act as a tumor suppressor in non-myogenic cells and that instead, the remaining translated part of DMD should be the focus in these cases.
Also GGE identified a potential therapeutic target: CD24. Indeed, CD24 was the top-upregulated gene in SEF compared to the control tumors, showing on average 80x higher expression (Fig. 2A). We previously showed that CD24 was among the most strongly differentially expressed genes also in LGFMS, and that the high transcriptional activity of CD24 could be directly linked to the presence of chimeric CREB3L transcription factors (7). The CD24 locus harbors several CRE half-sites (TGACG), and in vitro studies showed that truncated FUS/CREB3L2 and CREB3L2 constructs in HEK 293 cells activated transcription of CD24 (7).

CD24 encodes a mucin-like protein that is heavily glycosylated, and may be found both in the cell membrane and in the cytoplasm. When CD24 is expressed at the cell surface, it is known that P-selectin can act as a ligand. This binding has been shown to affect the adherence of CD24-expressing lymphocytes to activated thrombocytes and endothelial cells. Possibly, this mechanism is also of importance for lung colonization of solid tumor cells (30,31). Indeed, according to numerous studies, several tumor types display increased expression of CD24, typically at the protein level, and increased expression is often associated with more aggressive phenotype, increased risk of metastasis development, and worse clinical outcome (31,32).

It was recently shown that CD24 is most abundant intracellularly, where it can be found both in the cytoplasm and in the nucleus (33). This is in agreement with our IHC results on SEF showing both cytoplasmic and membranous staining (Fig. 2C), and with FACS data on EWSR1-CREB3L1 expressing cells (Fig. 2B). When introducing ectopic expression of CD24 in an osteosarcoma cell line, it was found that CD24 inhibits the binding of ARF to nucleophosmin (NPM1), a key regulator of mitosis. In turn, this increased the levels of MDM2 and decreased the levels of TP53 and its target CDKN1A (33). Here, we could show that CD24 is transcriptionally activated also in the fibroblast cell lines expressing EWSR1-CREB3L1 and FUS-CREB3L2. Interestingly, the
CD24 expression was higher in EWSR1-CREB3L1 expressing cells, in line with the higher expression levels in SEF samples compared to LGFMS (Fig 2A). The high expression of CD24 in both LGFMS and SEF has promising therapeutic implications. SEF is a highly malignant tumor, as exemplified by the present study in which 14/17 patients with SEF or hybrid SEF/LGFMS developed metastases and/or died of disease. CD24 has been suggested as a potential treatment target in several other tumor types. For instance, in both in vitro and mouse models of colorectal, pancreatic, and bladder cancer, silencing of CD24 through siRNA or a monoclonal antibody resulted in reduced tumor growth and/or reduced risk for lung metastases (34,35).

In addition to CD24 and MUC4, 36 more genes among the 100 top up-regulated genes in SEF tumors were also significantly up-regulated in Tet-On 3G cells expressing EWSR1-CREB3L1 (Supplementary Table 6). Thus, in spite of the extensive genomic alterations in SEF, a significant part of the characteristic gene expression profile can be recapitulated simply by expressing the underlying gene fusion, providing strong evidence for its importance in SEF tumorigenesis.

Detection of fusion genes is important for clinical, diagnostic purposes (36). It is hence of interest that we here detected some novel fusions in SEF. One MUC4-positive SEF (Case 11) showed an in-frame fusion of exon 11 of FUS with exon 7 of CREM at RNA-Seq; the fusion was verified by RT-PCR, and SNP array analysis revealed copy number transitions in both loci (Supplementary Tables 1 and 4). Interestingly, this case had the lowest mRNA expression of MUC4 of all SEFs, in line with a scattered staining of tumor cells (Supplementary Figure 1A). The breakpoint in FUS was unusually distal, but breakpoints as distal as exon 14 have been described in other FUS fusions (37). CREM (cAMP responsive element modulator) belongs to a large family of basic leucine zipper (bZIP)-containing proteins, including also CREB3L1 and CREB3L2. The b-ZIP
domain of CREM, which is encoded by exons 7-8 and thus retained in the fusion chimera, preferentially binds to palindromic CRE (5’-TGACGTCA-3’) or half-CRE (5’-TGACG-3’ or 5’-CGTCA-3’) sites in promoters of target genes. However, studies using chromatin immunoprecipitation followed by sequencing have revealed that CREM occupies promoters of thousands of genes, that more than 15% of the binding sites do not contain CRE or half-CRE sites, and that binding sites vary in a cell-specific way (38).

Another MUC4-positive pure SEF had a PAX5-CREB3L1 fusion; the breakpoint in CREB3L1 (exon 6) was similar to that in the more common EWSR1-CREB3L1 fusions. PAX5, which has not before been reported in sarcoma-associated gene fusions, is a recurrent 5’-partner to many different 3’-genes in acute lymphoblastic leukemia, where the breakpoints in PAX5 range from exon 4-9; the cellular consequences of these translocations are not fully understood (39).

In conclusion, we provide evidence for the respective fusion genes in SEF and LGFMS being the main tumorigenic drivers. We also show that additional genomic imbalances found in SEF and SEF/LGFMS can account for the significant difference in clinical outcome between the highly aggressive SEF and SEF/LGFMS and LGFMS, morphologically low grade but notorious for late metastases. Also, CD24 and DMD were identified as potential therapeutic targets for SEF.

ACKNOWLEDGMENTS

We would like to thank Dr Philippe Taniere, University Hospitals Birmingham, for providing tumor material.

REFERENCES


17. Prieto-Granada C, Zhang L, Chen HW, Sung YS, Agaram NP, Jungbluth AA, et al. A genetic dichotomy between pure sclerosing epithelioid fibrosarcoma (SEF) and hybrid
SEF/low-grade fibromyxoid sarcoma: a pathologic and molecular study of 18 cases.


Legends to Figures

Figure 1: Histological appearances of hybrid SEF/LGFMS with *FUS-CREB3L2* fusion (case 17). A. Low magnification with abrupt transition from LGFMS (left) to SEF (right). B. High magnification appearances of increasing cellularity, atypia and decreasing extracellular matrix during progression of LGFMS to SEF (left to right) within the same lesion.

Figure 2: CD24 expression. A. By global gene expression, CD24 is found to be highly upregulated in SEF (80x) and in LGFMS (37x) as compared to a set of control tumors (left boxplot). In line with that, Tet-On 3G cells expressing the *EWSR1-CREB3L1* fusion (E-C Dox cells) and Tet-On 3G cells expressing the *FUS-CREB3L2* fusion (F-C Dox cells) also show upregulation of CD24 (125x and 28x, respectively) as compared to the No Dox counterpart and the controls (right boxplot). B. CD24 FACS on Tet-On 3G cells expressing *EWSR1-CREB3L1* (E-C Dox) and *FUS-CREB3L2* (F-C Dox) was well in line with the RNA-seq results visualized in the box plots above. Cell surface CD24 expression was 42x higher, and total CD24 expression was 50x higher in E-C dox (red line) than in E-C No Dox (blue line) in the cell clone illustrated. C. IHC showed cytoplasmic and membranous staining of CD24 in Cases 5 (left) and 6A (right).

Figure 3: SNP array results. Circos plot showing SNP array results in low-grade fibromyxoid sarcomas (LGFMS; inner 10 circles, light grey), hybrid sclerosing epithelioid fibrosarcoma (SEF)/LGFMS (middle five circles, grey), and pure SEF (outer seven circles, dark grey). Gains, losses, and copy-neutral loss of heterozygosity exceeding 100 kb, as well as smaller homozygous deletions or deletions affecting genes involved in fusions, are indicated.

Figure 4: DMD deletions in case 15. Homozygous *DMD* deletions (i.e., 0 copies), detected by SNP array analysis in Cases 15A and B (upper and lower, respectively), illustrate convergent clonal evolution; two separate events target the same gene. Case 15A harbors a 130 kb deletion; Case 15B harbors two deletions of 172 and 257 kb each.
In-depth genetic analysis of sclerosing epithelioid fibrosarcoma reveals recurrent genomic alterations and potential treatment targets

Elsa Arbajian, Florian Puls, Cristina R. Antonescu, et al.

Clin Cancer Res Published OnlineFirst September 22, 2017.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-17-1856

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2017/09/22/1078-0432.CCR-17-1856.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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, use this link http://clincancerres.aacrjournals.org/content/early/2017/09/22/1078-0432.CCR-17-1856. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.