NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Identification of Recurrent *NAB2-STAT6* Gene Fusions in Solitary Fibrous Tumor by Integrative Sequencing

Dan R. Robinson[1,2,12], **Yi-Mi Wu**[1,2,12], **Shanker Kalyana-Sundaram**[1,2,3], **Xuhong Cao**[1,4], **Robert J. Lonigro**[1,5], **Yun-Shao Sung**[6], **Chun-Liang Chen**[6], **Lei Zhang**[6], **Rui Wang**[1,2], **Fengyun Su**[1,2], **Matthew K. Iyer**[1,7], **Sameek Roychowdhury**[1,8], **Javed Siddiqui**[1,2], **Kenneth J. Pienta**[1,5,8,9], **Lakshmi P. Kunju**[1,2], **Moshe Talpaz**[5,8], **Juan Miguel Mosquera**[10], **Samuel Singer**[11], **Scott M. Schuetze**[5,8], **Cristina R. Antonescu**[6], and **Arul M. Chinnaiyan**[1,2,4,5,7,9]

[1]Michigan Center for Translational Pathology, University of Michigan Medical School, Ann Arbor, MI 48109 USA

[2]Department of Pathology, University of Michigan Medical School, Ann Arbor, MI 48109 USA

[3]Department of Environmental Biotechnology, Bharathidasan University, Tiruchirappalli, India

[4]Howard Hughes Medical Institute, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[5]Comprehensive Cancer Center, University of Michigan Medical School, Ann Arbor, MI 48109, USA

[6]Department of Pathology, Memorial Sloan-Kettering Cancer Center, New York, NY 10021 USA

[7]Center for Computational Medicine and Biology, University of Michigan, Ann Arbor, MI 48109, USA

[8]Department of Internal Medicine, Division of Hematology/Oncology, University of Michigan Medical School, Ann Arbor, MI 48109 USA

[9]Department of Urology, University of Michigan Medical School, Ann Arbor, MI 48109 USA

[10]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, NY 10065 USA

[11]Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10021 USA

Abstract

A 44-year old woman with recurrent solitary fibrous tumor (SFT)/hemangiopericytoma was enrolled in a clinical sequencing program including whole exome and transcriptome sequencing. A gene fusion of the transcriptional repressor *NAB2* with the transcriptional activator *STAT6* was detected. Transcriptome sequencing of 27 additional SFTs all revealed the presence of a *NAB2-STAT6* gene fusion. Using RT-PCR and sequencing, we detected this fusion in 51 of 51 SFTs, indicating high levels of recurrence. Expression of NAB2-STAT6 fusion proteins was confirmed in SFT, and the predicted fusion products harbor the early growth response (EGR)-binding domain of NAB2 fused to the activation domain of STAT6. Overexpression of the *NAB2-STAT6* gene fusion induced proliferation in cultured cells and activated EGR-responsive genes. These studies establish *NAB2-STAT6* as the defining driver mutation of SFT and provide an example of how neoplasia can be initiated by converting a transcriptional repressor of mitogenic pathways into a transcriptional activator.

Comprehensive clinical sequencing programs for cancer patients have been initiated at several medical centers including our own[1–3]. In addition to the potential for identifying actionable therapeutic targets in cancer patients, these clinical sequencing efforts may lead to the identification of novel "driver" mutations that might be relatively rare in a common cancer type or be newly found in relatively rare cancer types. In this study, a patient with cellular solitary fibrous tumor/hemangiopericytoma (SFT/HPC) was enrolled in our clinical sequencing project called MI-ONCOSEQ (the Michigan Oncology Sequencing Program). SFT represents a wide spectrum of tumor types of mesenchymal origin that can affect virtually any region of the body[4]. SFT is composed of CD34-positive fibroblastic-appearing cells, arranged in a distinctive patternless growth of alternating cellularity and collagenous stroma. Whereas most SFTs are benign and can be cured with surgery, 15–20% of patients progress with either local recurrence or distant metastases, which can be difficult to treat[4,5]. It is unclear whether SFTs originating at diverse sites such as the meninges, lung, and breast share a common pathogenesis.

The index patient was a 44 year-old woman who had surgery and post-operative radiation for an anaplastic meningioma in 2002. Later review reclassified this tumor as a meningeal malignant SFT. In 2009, magnetic resonance imaging (MRI) showed a new brain mass and also a paraspinal mass. Laminectomy was performed, and review of the tissue showed metastatic SFT, which was strongly immunoreactive for CD34. In 2011, the patient was enrolled in the MI-ONCOSEQ integrated cancer sequencing program (Supplementary Fig. 1a) after progression of sarcoma on chemotherapy. Computed tomography (CT)-guided core needle biopsies were obtained from a metastatic site in the liver (Fig. 1a). The specimen showed the typical morphologic features of SFT with HPC-like vessels, collagenous stroma, and patternless architecture of spindled-to-ovoid tumor cells (Figs. 1b,c). Immunostaining for CD34 was positive in the tumor cells and in the endothelial cells, highlighting branching vessels (Fig. 1d). The biopsy cores used for molecular analysis had over 70% tumor cell content based on morphologic analysis.

High quality DNA and RNA was isolated from the core needle biopsies and subjected to next generation sequencing. Whole-exome sequencing of the tumor and matched normal from MO_1005 identified 14 nonsynonymous point mutations (Supplementary Table 1). No significant germline aberrations or somatic point mutations were identified in genes

frequently mutated in cancer such as *TP53*, *KRAS*, *BRAF*, or *PIK3CA* among others. The exome data coupled with single-nucleotide variant (SNV) candidate modeling was used to estimate tumor content of the biopsy specimen at 70% corroborating the histologic assessment (Supplementary Fig. 2). A global landscape of somatic copy number alterations was generated from exome sequencing data (Fig. 1e), and there were only a few regions of substantial copy-number gain or loss (Supplementary Tables 2 and 3). A focal 56-kb one-copy deletion observed in the *STAT6* locus (Fig. 1e and Supplementary Fig. 3). Notably, paired-end transcriptome sequencing of RNA identified an intrachromosomal fusion between *NAB2* and *STAT6* (Fig. 1f). The *NAB2-STAT6* fusion was represented by 1,104 paired-end reads either spanning or encompassing the fusion junction of exon 6 of *NAB2* to exon 18 of *STAT6*. In the normal genome, *NAB2* and *STAT6* are adjacent genes on chromosome 12q13 that are transcribed in opposite directions.

Using primers within exon 6 of *NAB2* and exon 19 of *STAT6*, we confirmed the *NAB2-STAT6* fusion in the index case by RT-PCR followed by Sanger sequencing of the amplified product (Fig. 2a). To confirm that the fusion resulted from a DNA-level rearrangement, we carried out long-range PCR of genomic DNA. A 1.3-kb product was obtained specifically in the tumor from the index subject and not in the matched normal tissue (Fig. 2b, *left*). This allowed us to specifically map the genomic breakpoint of the *NAB2-STAT6* fusion (Fig. 2b, *right*) and confirmed that a genomic inversion occurs at the Chr12q13 locus, fusing *NAB2* and *STAT6* in a common direction of transcription.

To determine whether the *NAB2-STAT6* fusion was recurrent, we interrogated 51 cases of SFT, malignant and benign, from a range of anatomical sites (Table 1 and Supplementary Fig. 1b). We analyzed a total of 27 cases of SFT by transcriptome sequencing. All 27 cases displayed high levels of a *NAB2-STAT6* gene fusion, with some variation in the precise exon structure of the fusions (Table 1 and Supplementary Table 4). The number of paired-end reads spanning or encompassing the fusion ranged from 25 to 4,483 per case. In each case, exon 2, 4, 6, or 7 of *NAB2* was fused in frame to exon 2, 3, 5, 6,17, or 18 of *STAT6*. Most of the transcript fusions were at defined exon boundaries, but a small set had fusion junctions within exons (Table 1). Representative fusions are shown in Fig. 2c. SFT-3 is an example of a complex fusion, with a 72 bp fragment derived from the 3′ UTR OF *NAB2* and intron 16 of *STAT6* at the fusion junction, retaining the reading frame. RT-PCR combined with capillary sequencing was carried out on an additional 24 cases of SFT from the Memorial Sloan-Kettering Cancer Center (MSKCC), and all cases were also positive for a *NAB2-STAT6* fusion (Table 1). The presence of the fusion in selected cases was further confirmed by RT-PCR (qRT-PCR) analysis (Supplementary Fig. 4). Thus, regardless of anatomic site of origin or malignant versus benign status, all 51 cases of SFT harbored a *NAB2-STAT6* gene fusion. As all of the *NAB2-STAT6* gene fusions identified contained 3′ exons of *STAT6*, we took advantage of an Affymetrix gene expression data set of soft-tissue sarcomas that included a 3′ probe to STAT6 (U133A, probe set 201331_s_at) to assess expression. Importantly, 100% of the SFTs (24 out of 24) expressed the 3′ exons of *STAT6* as compared to 28 other sarcomas all with very low STAT6 abundance (Supplementary Fig. 5). RNA sequencing (RNA-seq) and RT-PCR experiments showed the presence of reciprocal *STAT6-NAB2* fusion transcripts in only 25 of 51 cases. The absence of a reciprocal transcript in the index case is consistent with the deletion of the 5′ exons of STAT6 (Supplementary Fig. 3). The reciprocal transcripts varied by case, ranging from exact *STAT6-NAB2* reciprocals to shorter variants, missing one to several *STAT6* exons, both in and out of frame.

NAB2 is comprised of an N-terminal EGR1 binding domain (EBD), a NAB conserved region 2 (NCD2) and a C-terminal transcriptional repressor domain (RD). STAT6 is comprised of a DNA-binding domain (DBD), SH2 domain, and a C-terminal transcriptional

activation domain (TAD). The domain structures of the wild-type NAB2 and STAT6 proteins as well as of six representative NAB2-STAT6 fusion proteins identified by transcriptome sequencing in this study are shown (Fig. 3a). A common feature of all of the NAB2-STAT6 fusion proteins was variable truncation in the RD motif of NAB2, which was then minimally fused to the TAD motif of STAT6.

Expression of the predicted NAB2-STAT6 fusion protein products was confirmed by immunoblot analysis of three cases of SFT (Fig. 3b). An antibody to the C-terminus of STAT6, present in all fusions, detected the respective fusion proteins only in the tumor samples, whereas matched normal tissues expressed only wild-type STAT6. SFT-10 and SFT-44 exhibit a common fusion variant, whereas SFT-14 had a larger variant similar to that in SFT-31 (Fig. 3a). Similarly, immunofluorescence using this antibody to the C terminus of STAT6 showed strong nuclear staining (Fig. 3c). Immunofluorescence using an antibody to STAT6 phosphorylated at Tyr641 showed a complete absence of phosphorylation in the eight SFTs tested, whereas nuclear staining was evident in the entrapped endothelial cells or adjacent lung parenchyma (Supplementary Fig. 6). This pattern of C-terminal STAT6 nuclear staining and absence of nuclear phosphorylated STAT6 in SFTs is consistent with the NAB2-STAT6 fusion protein rather than activated wild-type STAT6.

The *NAB2-STAT6* fusion allele from the index case (MO_1005) was cloned into a lentiviral vector with a FLAG epitope tag. Benign prostate RWPE-1 cells were infected with control virus (empty vector) or with virus encoding *NAB2-STAT6*, and pooled stable cell lines were generated. Stable cell lines expressing high and low levels of NAB2-STAT6 were characterized (Fig. 3d). Notably, the cell line with high NAB2-STAT6 expression showed markedly proliferation compared to the cells transduced with control virus, whereas the cell line with low NAB2-STAT6 expression showed an intermediate level of proliferation (Fig. 3e). The proliferation of both NAB2-STAT6-expressing cell lines could be inhibited by small interfering RNA (siRNA) knockdown of *EGR1* expression, whereas the cell line transduced with control virus was unaffected (Supplementary Fig. 7). As wild-type NAB2 is a well-characterized repressor of EGR1 transcriptional activity[6,7], we measured the expression of established EGR1 target genes in the cell lines stably expressing NAB2-STAT6 (Fig. 3f)[8]. In contrast to the known activity of NAB2, the NAB2-STAT6 fusion induced expression of EGR1 target genes. This was further verified using an EGR1 response element reporter assay (Fig. 4a). Expression of EGR1 induced expression of the reporter gene by over 200-fold compared to vector control. This induction could be repressed over 90% by coexpression of wild type NAB2 repressor. Expression of the NAB2-STAT6 fusion resulted in the opposite effect. Expression was elevated tenfold over EGR1 alone, confirming that the fusion protein functions as an activator of EGR1 targeted transcription. No significant effect was seen in parallel experiments using a STAT6 response element reporter, showing that the NAB2-STAT6 fusion functions through EGR1 target genes rather than STAT6 target genes. The interaction of the NAB2-STAT6 fusion with EGR1 was further confirmed by chromatin immunoprecipitation (ChIP)-PCR of known EGR1 target gene promoters, in cells transiently expressing of Flag-tagged NAB2-STAT6 (Fig. 4b). Cells receiving vector control showeded no specific enrichment, whereas the NAB2-STAT6 fusion showed specific enrichment for the binding of known EGR1-responsive sites and not for negative control sites. Furthermore, the physical interaction of the NAB2-STAT6 fusion protein with EGR1 could be demonstrated by coimmunoprecipitation (Supplementary Fig. 8).

To investigate the role of the NAB2-STAT6 fusion in determining the transcriptional pattern of SFT, we curated a robust set of EGR1 target genes from an EGR1 ChIP-Seq experiment in K562 cells (GSM803414)[9]. We extracted the sets of highly enriched peaks from this dataset (at 5%, 10%, and 25% percentiles) and performed initial analysis of these peaks

using the GREAT bioinformatics tool[10]. For example, EGR1 peaks with scores in the top 5% (918 peaks) were mapped to 1,222 genes (some peaks mapped to more than 1 gene) and these 1,222 genes were highly enriched for the known early growth response DNA binding motifs (Supplementary Fig. 9). To assess whether SFT gene expression profiles were enriched for this set of EGR1 target genes, we used Affymetrix U133A microarray data from 23 SFTs and 34 non-SFT soft tissue sarcomas spanning seven subtypes[11]. We then compared the gene expression profiles of SFT versus non-SFT sarcomas across our EGR1 target gene lists using gene set enrichment analysis (GSEA)[12]. Indeed, we observed a positive and highly significant enrichment of EGR1 target genes in SFT compared to non-SFT sarcomas (Supplementary Fig. 10). The list of EGR1 target genes in the top 5% was the most highly enriched in SFT tumors (enrichment score (ES) = 0.35, P value < 0.001, false discovery rate (FDR) q value = 0.025, family-wise error rate (FWER) P value = 0.006). Overall, the set of genes differentially expressed between SFT and other sarcomas was significantly enriched for EGR1 target genes. RNA-seq analysis of 7 SFTs compared with 282 other tumor samples also showed high-level expression of EGR1 target genes. We found that EGR1 target genes, including *NAB2, NAB1, IGF2, FGF2, PDGFD,* and receptor tyrosine kinases like *FGFR1* and *NTRK1,* all had outlier expression levels in SFTs relative to other tumor types (Fig. 4c, Supplementary Tables 5 and 6).

Recurrence analysis on an independent set of tumor samples suggested that nearly all SFTs (100% in this study) harbor a *NAB2-STAT6* fusion. This would suggest that the *NAB2-STAT6* gene fusion is pathognomonic for SFT and that the spectra of SFT characteristics and morphology have a common genetic origin. Assessment of *NAB2-STAT6* fusion status could be used as a genetic marker in sarcoma cases that are not unambiguously classified as SFT (*e.g.* cases of CD34-negative SFT, and malignant and de-differentiated SFT)[13]. Although the structure of the fusion proteins varies in individual SFT cases, all fusion proteins have a truncation of the transcriptional repressor domain of NAB2 with an in-frame fusion to the transcriptional activation domain of STAT6 (although additional STAT6 domains may be included). The truncation of the repressor domain attenuates its repressive activity, while addition of a strong, intact activation domain engenders transcriptional activation potential.

How does the *NAB2-STAT6* fusion potentially explain neoplastic progression in SFT? NAB2 is a well-known co-regulator of the EGR transcription factors[7], and all of the fusion proteins identified in SFT maintain an intact N-terminal EBD. EGR1 is a zinc-finger transcription factor that couples growth factor signaling with the induction of nuclear programs of differentiation and proliferation, which are mediated by EGR1 target genes (Fig. 4d)[14]. As part of a homeostatic loop, NAB2 is induced by EGR family members and functions in a negative feedback manner to repress their activity[15,16]. In the context of SFT, the NAB2 fusion inherits an activation domain from the signaling molecule STAT6, which converts a transcriptional repressor (NAB2) into a potent transcriptional activator (i.e., NAB2-STAT6) of EGR1. This leads to constitutive activation of EGR mediated transcription, culminating in a feedforward loop that drives neoplastic progression. We found that EGR target genes including *NAB2, NAB1, IGF2, FGF2, PDGFD,* and receptor tyrosine kinases like *FGFR1* and *NTRK1,* all exhibited outlier levels in SFTs relative to other tumor types (Fig. 4c). Furthermore, a number of kinases including FGFR1 are targets of EGR1 and are also overexpressed in SFT and may be explained by the feedforward loop potentiated by the *NAB2-STAT6* fusion. Although it will be a challenge to target the *NAB2-STAT6* fusion protein therapeutically, some of the downstream kinases it induces may serve as attractive drug targets that should be evaluated in clinical trials for SFT. A clinical study that tested the kinase inhibitor sunitinib and the figitumimab antibody to IGF1R in SFT showed some positive responses[17].

Taken together, this study implicates aberrations in *NAB2* and *STAT6* in neoplastic pathways of virtually all SFTs, both benign and malignant, and may predict a role for these genes in other more common tumor types, as cancer genome sequencing efforts continue. Integrative sequencing provides a molecular definition of cancers to supplement and clarify histopathological characterizations[18]. In addition to suggesting actionable mutations, unbiased clinical sequencing efforts may shed light into the biology of rare cancers or individual cases of more common cancer types. To our knowledge, this is one of the first examples of how a gene fusion can convert a transcriptional repressor (NAB2) into a transcriptional activator (NAB2-STAT6) of mitogenic pathways that can be subverted during neoplastic progression.

METHODS

Clinical Study

Research was performed under institutional review board (IRB)–approved studies. Patients are enrolled and consented through a University of Michigan IRB-approved protocol for integrative tumor sequencing, MI-ONCOSEQ (Michigan Oncology Sequencing Protocol, IRB# HUM00046018)[1]. Medically fit patients 18 years or older with advanced or refractory cancer were eligible for the study. Informed consent detailed the risks of integrative sequencing and included up-front genetic counseling. Biopsies are arranged for safely accessible tumor sites. Needle biopsies were snap frozen in OCT and a longitudinal section was cut. Hematoxylin and eosin (H&E) stained frozen sections were reviewed by study pathologist (L.P.K.) to identify cores with highest tumor content. Remaining portions of each needle biopsy core were retained for nucleic acid extraction.

Thirty SFTs with available frozen tissue material from MSKCC files were included for analysis. Seventeen of the SFT were previously analyzed as part of a prior gene expression profiling study and CEL files have been made publicly available[11]. There were 16 females and 14 males with a wide age range at diagnosis (12–79 years; mean 52 years). The retrospective validation study was approved by IRB (IRB# 02-060).

DNA/RNA isolation and cDNA synthesis

Genomic DNA from frozen needle biopsies and blood was isolated using the Qiagen DNeasy Blood & Tissue Kit, according to the manufacturer's instructions. Total RNA was extracted from frozen needle biopsies (for RNA-Seq libraries, gene expression analysis and RT-PCR) using the Qiazol reagent with disruption using a 5mm bead on a Tissuelyser II (Qiagen). RNA was purified using a miRNeasy kit (Qiagen) according to the manufacturer's instructions. cDNA was synthesized from total RNA using Superscript III (Invitrogen) and random primers (Invitrogen). For the MSKCC samples, total RNA was extracted from frozen tumor tissue using the Trizol reagent according to the manufacturer's instructions (Invitrogen).

Next generation sequencing library preparation

Exome libraries of matched pairs of tumor/normal genomic DNAs were generated using the Illumina TruSeq DNA Sample Prep Kit, following the manufacturer's instructions. RNA-Seq transcriptome libraries were prepared following Illumina's TruSeq RNA protocol, using 2μg of total RNA. PolyA+ RNA was isolated using Sera-Mag oligo(dT) beads (Thermo Scientific) and fragmented with the Ambion Fragmentation Reagents kit (Ambion, Austin, TX). cDNA synthesis, end-repair, A-base addition, and ligation of the Illumina indexed adapters were performed according to Illumina's protocol. Libraries were then size-selected for 250–300 bp cDNA fragments on a 3% Nusieve 3:1 (Lonza) agarose gel, recovered using QIAEX II gel extraction reagents (Qiagen), and PCR-amplified using Phusion DNA

polymerase (New England Biolabs) for 14 PCR cycles. Paired-end libraries were sequenced with the Illumina HiSeq 2000, (2 X100 nucleotide read length). Reads that passed the chastity filter of Illumina BaseCall software were used for subsequent analysis. Summary sequencing statistics are presented in Supplementary Table 7.

We used the publicly available software FastQC to assess sequence quality. For each lane, we examine per-base quality scores across the length of the reads. Lanes were deemed passing if the per-base quality score boxplot indicated that >75% of the reads had >Q20 for bases 1–80. All lanes passed this threshold. In addition to the raw sequence quality, we also assess alignment quality using the Picard package.

Mutation Analyses

We annotated the resulting somatic mutations using RefSeq transcripts. HUGO gene names were used. For *NAB2* mRNA and protein, positions and annotations are derived from RefSeq accessions NM_005967 and NP_005958 respectively. For *STAT6* mRNA and protein, positions and annotations are derived from RefSeq accessions NM_001178078 and NP_001171549 respectively. The impact of coding non-synonymous amino acid substitutions on the structure and function of a protein was assessed using PolyPhen-2[67]. We also assessed whether the somatic variant was previously reported in dbSNP135 or COSMIC v56[68].

Copy number aberrations were quantified and reported for each gene as the segmented normalized log2-transformed exon coverage ratios between each tumor sample and matched normal sample[19]. To account for observed associations between coverage ratios and variation in GC content across the genome, lowess normalization was used to correct per-exon coverage ratios prior to segmentation analysis. Specifically, mean GC percentage was computed for each targeted region, and a lowess curve was fit to the scatterplot of log2-coverage ratios vs. mean GC content across the targeted exome using the lowess function in R (version 2.13.1) with smoothing parameter f=0.05.

Somatic point mutations were identified in the tumor exome sequence data using the matched normal exome data to eliminate germline polymorphisms. Parameters and computational methods were as previously described[20].

To identify gene fusions, paired-end transcriptome reads passing filter were mapped to the human reference genome (hg19) and UCSC genes, allowing up to two mismatches, with Illumina ELAND software and Bowtie[21]. Sequence alignments were subsequently processed to nominate gene fusions using the methods described earlier[22,23]. In brief, paired end reads were processed to identify any that either contained or spanned a fusion junction. Encompassing paired reads refer to those in which each read aligns to an independent transcript, thereby encompassing the fusion junction. Spanning mate pairs refer to those in which one sequence read aligns to a gene and its paired-end spans the fusion junction. Both categories undergo a series of filtering steps to remove false positives before being merged together to generate the final chimera nominations. Reads supporting each fusion were realigned using BLAT (UCSC Genome Browser) to reconfirm the fusion breakpoint.

For RNA-Seq gene expression analysis, transcriptome data was processed as previously described[1].

RT-PCR, qRT-PCR and long-range PCR

For validation of fusion transcripts, RT-PCR and quantitative RT-PCR assays were performed. One microgram of total RNA from **30** SFT was used for RT-PCR using SuperScript III First-Strand System (Invitrogen), according to the manufacturer's

instructions. The primers used were: NAB2ex5 Forward and STAT6ex20 Reverse (Supplementary Table 8). The PCR products were analyzed by agarose gel electrophoresis. The amplified PCR products were purified then sequenced using the Sanger method. Quantitative RT-PCR assay was performed using SYBR Green Master Mix (Applied Biosystems) and was carried out with the StepOne Real-Time PCR System (Applied Biosystems). Relative mRNA levels of the fusion transcripts were normalized to the expression of the housekeeping gene *GAPDH*. Oligonucleotide primers were obtained from Integrated DNA Technologies (IDT) and the sequences given in Supplementary Table 8. To detect the genomic fusion junction between the *NAB2* and *STAT6* genes in the MO_1005 tumor DNA, primers were designed flanking the predicted genomic junction and PCR reactions were carried out to amplify the fusion fragments. PCR products were purified from agarose gels using the QIAEX II system (QIAGEN) and sequenced by Sanger sequencing.

Using similar conditions as above, RT-PCR for detection of STAT6-NAB2 reciprocal transcripts was performed on all 30 SFT cases and depending on the expected fusion transcript two primer pairs were used: STAT6 Ex13F and NAB2 Ex7R or STAT6 Ex1F and NAB2 Ex7R (Supplementary Table 8).

Immunoblot and immunofluorescence assays

Total protein lysates were extracted from frozen tissue from 8 SFT tumors. In three of the cases, adequate quality frozen normal tissues were available for protein extraction for comparison. Electrophoresis and immunoblotting were performed using 30 μg of total protein extract, following the standard protocol. Total STAT6 and β-actin were detected by rabbit polyclonal anti-STAT6 (Cell Signaling Technology, Cat #9362S; 1:1500 dilution) and rabbit monoclonal anti-β-actin (Cell Signaling Technology, Cat #4970; 1:1500 dilution). The secondary antibodies used were goat anti-rabbit (Santa Cruz Biotechnology, Cat #SC-2034) with 1:20000 dilution. The same total STAT6 antibody was used for immunofluorescence (IF) for detecting the cellular localization of the protein. IF for detecting Phospho-STAT6 was performed using the P-STAT6 Y641 primary antibody from Cell Signaling (#9361; 1:100 dilution) and a secondary antibody Alexa Fluor 594 goat anti-rabbit IgG from Invitrogen, (#A11037; 4ug/ml).

NAB2-STAT6 cloning, expression, and stable cell line analyses

The *NAB2-STAT6* fusion allele was PCR amplified from cDNA of the index case (MO_1005) using the primers listed in Supplementary Table 8 and the Expand High Fidelity protocol (Roche). The PCR product was digested with restriction endonuclease *Cpo* I (Fermentas) and ligated into the pCDH510B lentiviral vector (System Biosciences), which had been modified to contain an N-terminal FLAG epitope tag. Lentiviruses were produced by cotransfecting the *NAB2-STAT6* construct or vector with the ViraPower packaging mix (Invitrogen) into 293T cells using FuGene HD transfection reagent (Roche). Thirty-six hours post-transfection the viral supernatants were harvested, centrifuged at 5,000xg for 30 minutes and then filtered through a 0.45 micron Steriflip filter unit (Millipore). Benign RWPE-1 cells at 30% confluence were infected at an MOI of 20 with the addition of polybrene at 8 μg/ml. Forty-eight hours post-infection, the cells were split and placed into selective media containing 10 μg/ml puromycin. Two stable pools of resistant cells were obtained and analyzed for expression of the FLAG-*NAB2-STAT6* fusion allele by western blot analysis with monoclonal anti-FLAG M2 antibody (Sigma-Aldrich). Expression was confirmed by qPCR for the *NAB2-STAT6* fusion allele.

For the cell proliferation assay, vector control, NAB2-STAT6 high, and NAB2-STAT6 low level over-expressing cells were plated in quadruplicate at 8,000 cells per well in 24 well plates. The plates were incubated at 37°C and 5% CO_2 atmosphere using the IncuCyte live-

cell imaging system (Essen Biosciences). Cell proliferation was assessed by kinetic imaging confluence measurements at 3-hour time intervals.

Chromatin immunoprecipitation-PCR (ChIP-PCR)

We grew 293T cells to 70–80% confluence in DMEM supplemented with 10% fetal bovine serum, followed by transfection with EGR1-myc alone, or co-transfection with EGR1-myc and Flag-NAB2-STAT6 using FuGene6 reagent (Roche). Twenty-four hours after transfection, cells were cross-linked with 1% formaldehyde, neutralized with 0.125M glycine, rinsed twice with ice-cold PBS buffer, and cell pellets were collected. ChIP-seq was performed as previously described [24].

Co-immunoprecipitation of EGR1 and NAB2-STAT6 fusion proteins

293T cells were grown to ~70% confluence in DMEM supplemented with 10% fetal bovine serum, followed by transfection with myc-tagged *EGR1* alone, or co-transfection with *EGR1*-myc and Flag-tagged *NAB2-STAT6* using FuGene6 reagent (Roche). Twenty-four hours after transfection, cells were cross-linked with 1% formaldehyde, neutralized with 0.125M glycine, rinsed twice with ice-cold PBS buffer, and pelleted by centrifugation. Cell pellets were lysed in lysis buffer followed by immunoprecipitation with anti-MYC tag antibody (Sigma) and protein-G Dynabeads (Invitrogen). Precipitates were washed three times with IP Wash buffer resuspended in 3X SDS-PAGE loading buffer, and cooked at 95°C for 20 min to reverse cross-linking by formaldehyde. Flag-NAB2-STAT6 fusion protein was detected by Western blotting with anti-Flag antibody (Sigma).

siRNA knockdown of EGR1

RWPE-1 stable lines (vector control, NAB2-STAT6 high, and NAB2-STAT6 low) were transfected twice with EGR1-targeting siRNA or non-targeting siRNA (Thermo Scintific Dharmacon) using Oligofectamine (Invitrogen). The siRNAs used were as follows: ON-TARGETplus EGR1 LQ-006526-00-0002 and ON-TARGETplus Non-targeting pool. Twenty-four hours after transfection, cells were trypsinized and plated in quadruplicate at 4,000 cells per well in 24 well plates. The plates were incubated at 37°C with 5% CO_2 atmosphere in the IncuCyte live-cell imaging system (Essen Biosciences). Cell proliferation rate was assessed by kinetic imaging confluence measurements at 3-hour time intervals.

Luciferase assay

293T cells were seeded into 12-well dishes in triplicate and allowed to attach overnight. Cells were serum-starved for 12 hours and transfected with different combinations of *EGR1* (Origene), wild-type *NAB2* (Origene), and *NAB2-STAT6* fusion constructs along with a mixture of EGR1-responsive firefly luciferase construct and constitutively expressing Renilla luciferase construct (SABiosciences). Following incubation for 24 hours, cell lysates were prepared and measured for EGR1 activity using Promega Dual Luciferase reagents and Passive Lysis Buffer. Firefly luciferase levels were normalized using corresponding Renilla luciferase levels for each condition. To test if the NAB2-STAT6 fusion protein functions through the STAT6 signaling pathway, a STAT6-responsive firefly luciferase reporter was constructed in pGL4.10 (Promega) and luciferase activity assays were performed as described above. The sequences of oligos used for the preparation of the STAT6-responsive firefly luciferase construct are listed in Supplementary Table 8.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Roychowdhury S, et al. Personalized oncology through integrative high-throughput sequencing: a pilot study. Sci Transl Med. 2011; 3:111ra121.

2. Ruiz C, et al. Advancing a clinically relevant perspective of the clonal nature of cancer. Proc Natl Acad Sci U S A. 2011; 108:12054–9. [PubMed: 21730190]

3. Welch JS, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011; 305:1577–84. [PubMed: 21505136]

4. Park MS, Araujo DM. New insights into the hemangiopericytoma/solitary fibrous tumor spectrum of tumors. Curr Opin Oncol. 2009; 21:327–31. [PubMed: 19444101]

5. Gold JS, et al. Clinicopathologic correlates of solitary fibrous tumors. Cancer. 2002; 94:1057–68. [PubMed: 11920476]

6. Srinivasan R, Mager GM, Ward RM, Mayer J, Svaren J. NAB2 represses transcription by interacting with the CHD4 subunit of the nucleosome remodeling and deacetylase (NuRD) complex. J Biol Chem. 2006; 281:15129–37. [PubMed: 16574654]

7. Svaren J, et al. NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. Mol Cell Biol. 1996; 16:3545–53. [PubMed: 8668170]

8. Svaren J, et al. EGR1 Target Genes in Prostate Carcinoma Cells Identified by Microarray Analysis. J Biol Chem. 2000; 275:38524–38531. [PubMed: 10984481]

9. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS Biol. 2011; 9:e1001046. [PubMed: 21526222]

10. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]

11. Hajdu M, et al. IGF2 over-expression in solitary fibrous tumours is independent of anatomical location and is related to loss of imprinting. J Pathol. 2010; 221:300–7. [PubMed: 20527023]

12. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]

13. Mosquera JM, Fletcher CD. Expanding the spectrum of malignant progression in solitary fibrous tumors: a study of 8 cases with a discrete anaplastic component--is this dedifferentiated SFT? Am J Surg Pathol. 2009; 33:1314–21. [PubMed: 19718788]

14. Thiel G, Cibelli G. Regulation of life and death by the zinc finger transcription factor Egr-1. J Cell Physiol. 2002; 193:287–92. [PubMed: 12384981]

15. Kumbrink J, Gerlinger M, Johnson JP. Egr-1 induces the expression of its corepressor nab2 by activation of the nab2 promoter thereby establishing a negative feedback loop. J Biol Chem. 2005; 280:42785–93. [PubMed: 16260776]

16. Kumbrink J, Kirsch KH, Johnson JP. EGR1, EGR2, and EGR3 activate the expression of their coregulator NAB2 establishing a negative feedback loop in cells of neuroectodermal and epithelial origin. J Cell Biochem. 2010; 111:207–17. [PubMed: 20506119]

17. Stacchiotti S, et al. Sunitinib malate and figitumumab in solitary fibrous tumor: patterns and molecular bases of tumor response. Mol Cancer Ther. 2010; 9:1286–97. [PubMed: 20457621]

18. Aparicio SA, Huntsman DG. Does massively parallel DNA resequencing signify the end of histopathology as we know it? J Pathol. 2010; 220:307–15. [PubMed: 19921711]

19. Lonigro RJ, et al. Detection of somatic copy number alterations in cancer using targeted exome capture sequencing. Neoplasia. 2011; 13:1019–25. [PubMed: 22131877]

20. Grasso CS, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. advance online publication(2012).

21. Langmead B. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinformatics. 2010; Chapter 11(Unit 11):7. [PubMed: 21154709]

22. Iyer MK, Chinnaiyan AM, Maher CA. ChimeraScan: a tool for identifying chimeric transcription in sequencing data. Bioinformatics. 2011; 27:2903–4. [PubMed: 21840877]

23. Robinson DR, et al. Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. Nat Med. 2011; 17:1646–51. [PubMed: 22101766]

24. Yu J, et al. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. Cancer Cell. 2010; 17:443–54. [PubMed: 20478527]
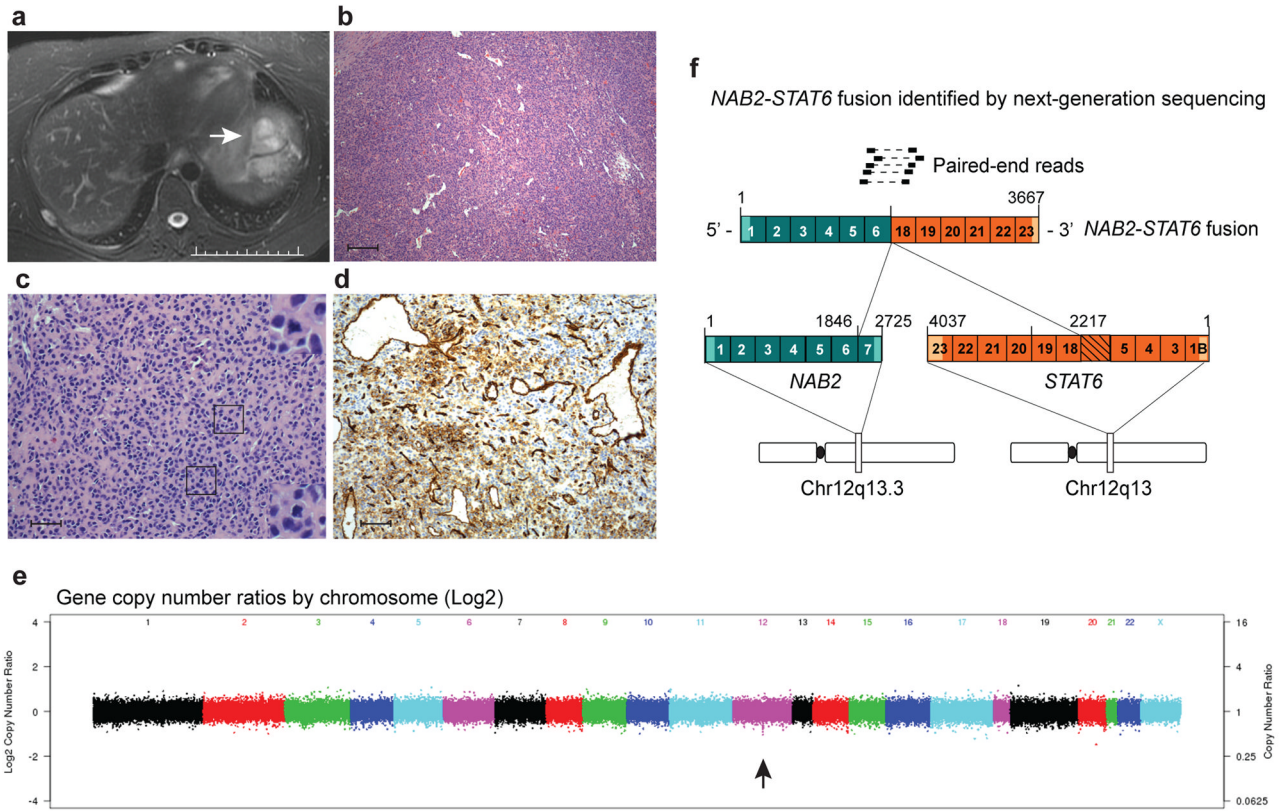
Figure 1.
Integrative sequencing and mutational analysis of patient MO_1005 (SFT index case). (**a**) CT image of the biopsied liver metastasis (arrow). Arrow indicates metastasis that was biopsied. Scale bar equals 10 cm. (**b**) Hematoxylin and eosin staining of index case. Scale bar, 200 µm. (**c**) Hematoxylin and eosin staining of index case. Scale bar, 100 µm. Please note, **b** and **c** differ only in magnification. (**d**) Immunoreactivity for CD34. Scale bar, 100 µm. Images in **a–d** were all produced from a liver biopsy sample from the index case. (**e**) Gene copy number landscape of the index case as assessed by whole-exome sequencing of the tumor to germline sequence. The arrow indicates copy loss at the *STAT6* locus. (**f**) Schematic of the *NAB2-STAT6* gene fusion detected in the index case by paired-end transcriptome sequencing. The hashed region indicates exons not shown.
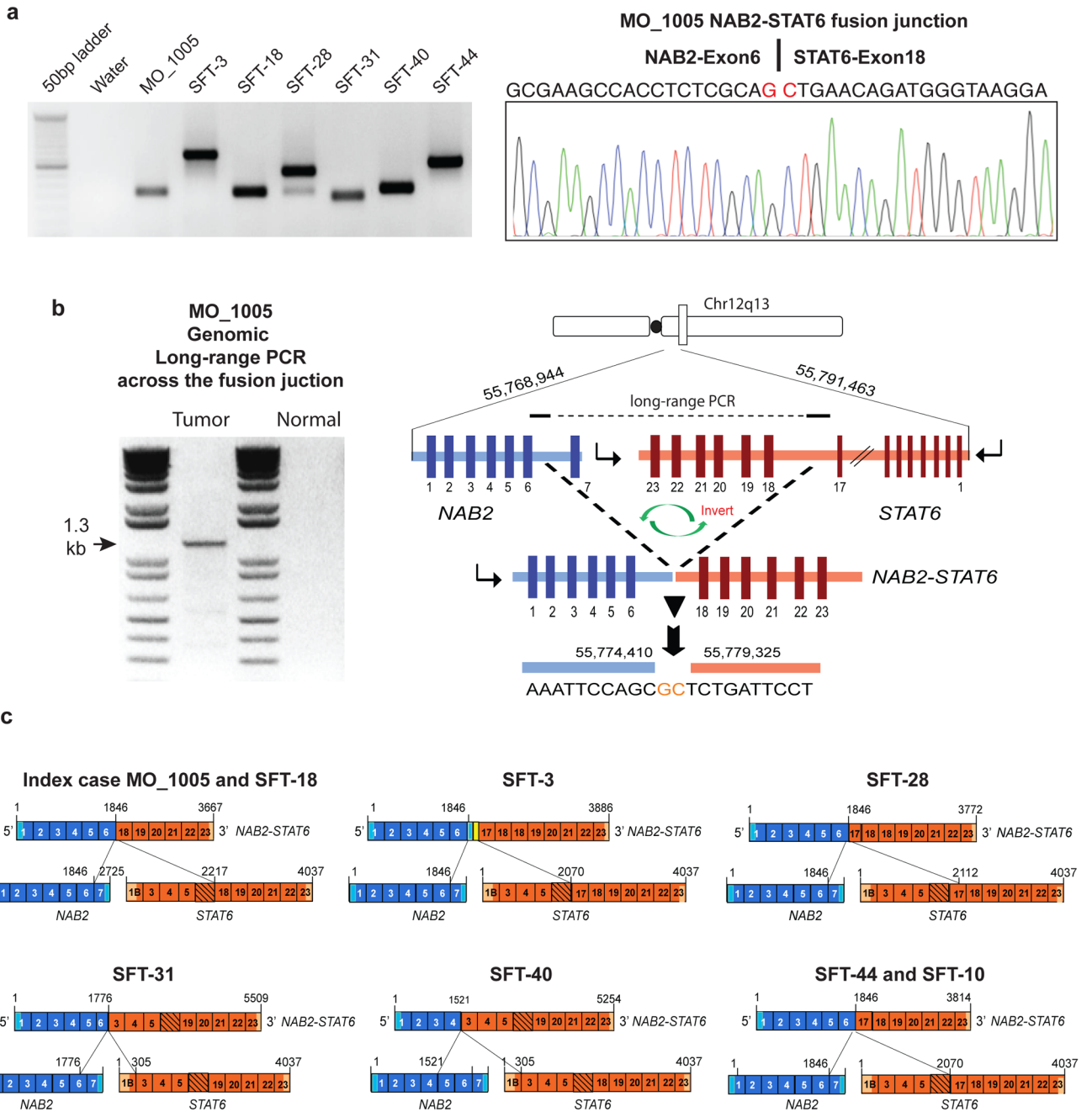
Figure 2.
Validation and recurrence of *NAB2-STAT6* gene fusions in SFT. (**a**) RT-PCR and capillary sequencing of the index case and additional SFT cases using primers for *NAB2* exon 6 and *STAT6* exon 19. The sequencing trace of the index case (right) shows the chimeric junction between *NAB2* and *STAT6*. (**b**) Genomic long-range PCR of the index case confirming the existence of the *NAB2-STAT6* gene fusion at the DNA level. Gel electrophoresis of the amplified product (left, arrow) and schematic of exon-intron structure of the index *NAB2-STAT6* gene fusion (right) are shown. (**c**) Schematics of additional *NAB2-STAT6* gene fusions identified by transcriptome sequencing of 6 SFT samples.
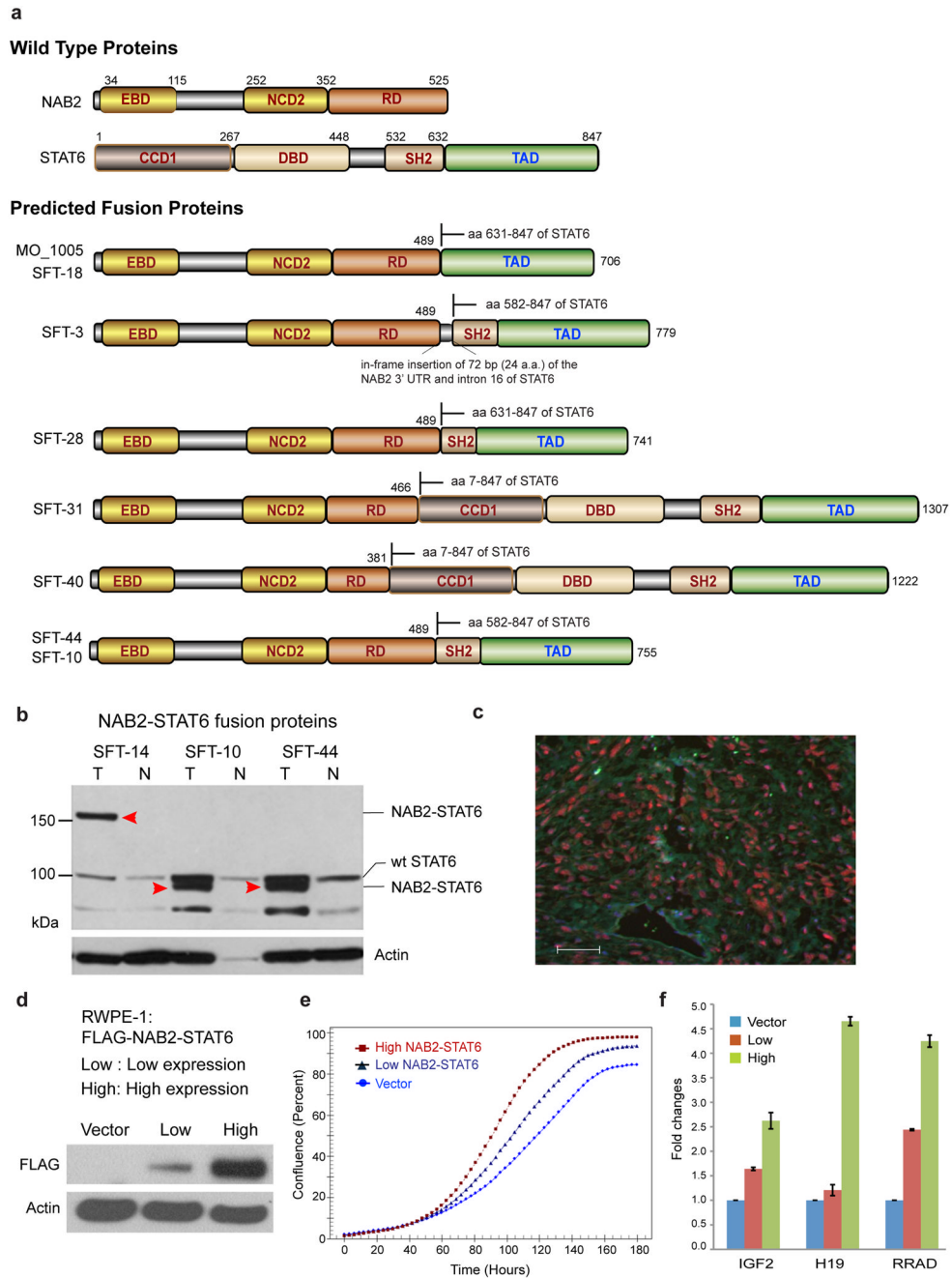
Figure 3.

Characterization and functional analysis of the NAB2-STAT6 fusion protein. (**a**) Schematics of the predicted NAB2-STAT6 fusion protein products identified in this study. EBD, EGR1-binding domain; NCD2, NAB2 conserved domain; RD, transcriptional repressor domain; SH2, Src homology 2; CCD1, coiled-coil domain 1; DBD, DNA-binding domain; TAD, transcriptional activator domain. (**b**) Immunoblot analysis of three SFT tumors (T) and matched normal tissue (N) employing an antibody against a C-terminal epitope of STAT6, which is found in all the NAB2-STAT6 fusions thus far identified. SFT-14 tumor expresses the larger 150-kDa NAB2-STAT6 fusion protein. SFT-10 and SFT-44 tumors both express the smaller, more common 90-kDa NAB2-STAT6 fusion. MW, molecular weight; WT, wild

type. Arrowheads indicate NAB2-STAT6 fusion proteins. (**c**) Immunofluorescence using the antibody in **b**, suggesting nuclear localization (red) of the NAB2-STAT6 protein in a representative SFT case. Scale bar, 50 μm. (**d**–**f**) Stable RWPE-1 cell line pools expressing low and high levels of Flag-tagged NAB2-STAT6 (MO_1005 fusion structure) were generated. (**d**) Immunoblot analysis with an antibody to Flag. (**e**) Cell proliferation assays carried out with live-cell imaging. Data are shown as cell confluence versus time at 3-h intervals. Each data point is the mean of quadruplicate measurements. (**f**) qRT-PCR for EGR1 target genes *IGF2*, *H19* and *RRAD*. Results were normalized to *GAPDH* levels. Means of triplicates with the s.d. are shown.
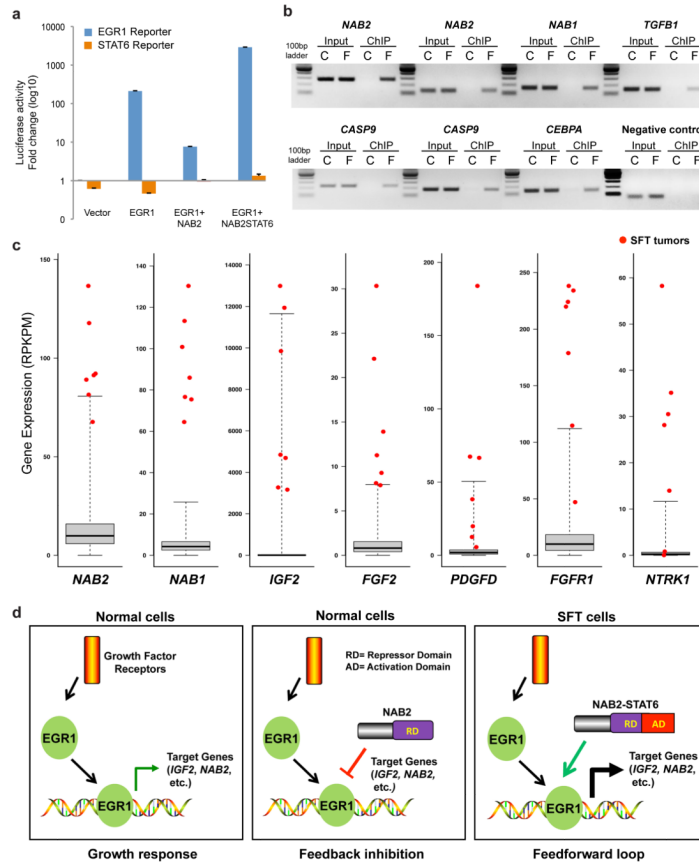
Figure 4.

A proposed model for the function of the *NAB2-STAT6* gene fusion in SFT. (**a**) NAB2-STAT6 fusion protein enhances EGR1-induced promoter activity. We transfected 293T cells with vectors expressing the indicated proteins along with a constitutively expressed *Renilla* luciferase construct and one of the two reporter constructs EGR1-responsive firefly luciferase reporter or STAT6-responsive firefly luciferase reporter. Cell lysates were prepared 24 h after transfection and measured for EGR1 and STAT6 activity. Means of quadruplicates with the s.d. are shown. (**b**) ChIP identified the recruitment of NAB2-STAT6 fusion protein to the promoters of EGR1-responsive genes. We transfected 293T cells with only vector expressing EGR1 (C) or cotransfected with vectors expressing EGR1 and Flag-NAB2-STAT6 (F). Crosslinked DNA-protein lysates were immunoprecipitated with antibody to Flag. The promoter fragments containing consensus EGR1-binding sites were PCR amplified from both samples. Input samples were loaded as a control. IGX1A primers that detect a specific genomic DNA sequence within an ORF-free intergenic region were used as a negative control. Two distinct EGR1-binding sites in *NAB2* and *CASP9* are shown. (**c**) Outlier gene expression in SFT predicted to be the result of NAB2-STAT6–driven constitutive activation of EGR1-mediated pathways. Data from RNA-seq analysis of 7 SFTs relative to a compendium of 282 previously sequenced tumor tissues are shown (Supplementary Table 6). RPKPM (reads per kilobase per million reads) values for the compendium are summarized by box-and-whisker plots showing the minimum, quartile 1, median, quartile 3 and maximum observed expression levels across the compendium. RPKPM values for the seven SFT samples are shown in red. (**d**) Model of the NAB2 and EGR1 signaling loops. RD, repressor domain; AD, activation domain.

Table 1

Summary of the SFT samples analyzed in this study.

Case	Age/Gende	Location	Malignant	CD34	RNA-Seq	Q RT-PCR	PCR screenin	Affy U133A	NAB2 exon	STAT6 exon
MO_1005	44/F	Liver	+	+	+	+			6	18
SFT-1	50/F	ST, pelvis	−	+			+	+	6	17
SFT-2	79/M	ST, thigh	−	+			+		7	3
SFT-3	58/F	ST, buttock	+	+	+	+	+	+	6 complex	17
SFT-4	67/M	Pleura	−	+			+		7	3
SFT-5	67/M	Pleura, recurrence	+	+			+	+	6	17
SFT-6	70/F	ST, pelvis	+	+			+	+	6	17
SFT-7	57/M	Pleura	−	+			+	+	6	17
SFT-8	62/F	Meningeal	−	+			+	+	6	18
SFT-10	78/F	ST, thigh	+	+			+	+	6	17
SFT-11	35/F	ST, shoulder	−	+			+		7	3
SFT-13	54/F	ST, periscapular	−	+			+	+	6	17
SFT-14	72/M	Pleura	+	+			+	+	7	3
SFT-15	54/F	ST, pelvis	−	+			+		6	17
SFT-18	67/F	ST, pelvis	−	+	+	+	+	+	6	18
SFT-22	40/M	Meningeal, lung met	+	+			+	+	6	17
SFT-23	49/F	Meningeal, orbital recurrence	+	+			+	+	6	17
SFT-24	53/M	Meningeal, T11 spinal cord	−	+			+		6	17
SFT-28	33/F	Meningeal, kidney met	+	+	+	+	+	+	6	17
SFT-31	67/F	Meningeal, brain recurrence	+	+	+	+	+	+	6 internal	17 internal
SFT-33	41/F	ST, thigh	−	+			+		6	3
SFT-35	63/F	Pleura, recurrence	+	+			+	+	6	17
SFT-38	31/M	ST, pelvis	+	+			+	+	6	17
SFT-40	67/M	Pleura, recurrence	+	+	+	+	+	+	4	3
SFT-42	40/F	ST, pelvis	−	+			+		6	17
SFT-43	12/M	ST, hand	−	+			+		6	17
SFT-44	29/M	Meningeal origin, pancreatic met	+	+	+	+	+		6	17

Case	Age/Gende	Location	Malignant	CD34	RNA-Seq	Q RT-PCR	PCR screenin	Affy U133A	NAB2 exon	STAT6 exon
SFT-46	59/F	ST, retroperitoneum	+	+			+		6	17
SFT-47	32/M	Meningeal origin, small bowel me	+	+			+		6	17
SFT-49	71/M	ST, pelvis	−	+			+		6	17
SFT-50	27/M	ST, pelvis	−	+			+		6	18
UM-SFT-1	68/F	Pancreas	+	+	+				6	17
UM-SFT-2	64/F	Liver	+	+	+				6	2 complex
UM-SFT-3	46/F	Presacral mass	−	+	+				6	17
UM-SFT-4	50/F	Left flank	+	+	+				6	17
UM-SFT-5	55/M	Left lung	+	+	+				4	5
UM-SFT-6	41/F	Retroperitoneum	−	+	+				6	17
UM-SFT-7	75/M	Dura	−	+	+				4	3
UM-SFT-8	53/M	Prostate	−	+	+				6	18
UM-SFT-9	33/M	Right thigh	+	+	+				6	17
UM-SFT-10	17/M	Left cheek	−	+	+				6	17
UM-SFT-11	42/M	C7 paraspinal mass	−	+	+				6	17
UM-SFT-12	57/F	Perirectal mass	+	+	+				2	6
UM-SFT-13	88/F	Left parapharyngeal	−	+	+				2 internal	1
UM-SFT-14	58/M	Posterior neck	−	+	+				6	17
UM-SFT-15	28/F	Right frontal tumor	−	+	+				6	17
UM-SFT-16	22/M	Right frontal tumor	−	+	+				6	18
UM-SFT-17	41/M	Retroperitoneum	−	+	+				7 internal	2 internal
UM-SFT-18	71/M	Right lung	−	+	+				4	3
UM-SFT-19	51/F	Brain Posterior fossa	−	+	+				6	17
UM-SFT-20	30/M	Left frontal tumor	−	+	+				6	17