NAB2-STAT6 drives an EGR1-dependent neuroendocrine program in Solitary Fibrous Tumors

Connor M Hill[1,2], Alexandra Indeglia[1,3], Francis Picone[1], Maureen E. Murphy[1], Cara Cipriano[4], Robert G. Maki[4,5] and Alessandro Gardini[1,6]

[1]The Wistar Institute, Philadelphia, PA, US

[2]Cell and Molecular Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.

[3]Graduate Group in Biochemistry and Molecular Biophysics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S.

[4]Department of Orthopedic Surgery, University of Pennsylvania, Philadelphia, PA, U.S.

[5]Abramson Cancer Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, U.S. Present address: Department of Medicine, Memorial Sloan-Kettering Cancer Center, and Weill Cornell Medical College, Cornell University, New York, NY, U.S.

[6]Corresponding author

The pathogenesis of many rare tumor types is poorly understood, preventing the design of effective treatments. Solitary fibrous tumors (SFTs) are neoplasms of mesenchymal origin that affect 1/1,000,000 individuals every year and are clinically assimilated to soft tissue sarcomas. SFTs can arise throughout the body and are usually managed surgically. However, 30-40% of SFTs will relapse local-regionally or metastasize. There are no systemic therapies with durable activity for malignant SFTs to date. The molecular hallmark of SFTs is a gene fusion between the *NAB2* and *STAT6* loci on chromosome 12, resulting in a chimeric protein of poorly characterized function called NAB2-STAT6. We use primary samples and an inducible cell model to discover that NAB2-STAT6 operates as a transcriptional coactivator for a specific set of enhancers and promoters that are normally targeted by the EGR1 transcription factor. In physiological conditions, NAB2 is primarily localized to the cytoplasm and only a small nuclear fraction is available to operate as a co-activator of EGR1 targets. NAB2-STAT6 redirects NAB1, NAB2, and additional EGR1 to the nucleus and bolster the expression of neuronal EGR1 targets. The STAT6 moiety of the fusion protein is a major driver of its nuclear localization and further contributes to NAB2's co-activating abilities. In primary tumors, NAB2-STAT6 activates a neuroendocrine gene signature that sets it apart from most sarcomas. These discoveries provide new insight into the pathogenesis of SFTs and reveal new targets with therapeutic potential.

INTRODUCTION

The pathogenesis of many rare tumor types is poorly understood, preventing the design of effective treatments. For some of these tumors, the discovery of gene fusion events and the characterization of their biological repercussions has led to the design of effective treatments and improved patient outcomes. Notable examples include the MLL fusions that cause early onset mixed-lineage leukemias, the NUT-BRD4 fusions in NUT carcinoma and the SS18 (SWI/SNF) fusions in synovial sarcoma(*1, 2*).

Solitary fibrous tumors (SFTs) are rare mesenchymal tumors with an estimated incidence rate of 1 in 1 million people per year(*3, 4*). They can develop in any location in the body, but most commonly arise in pleural, dural, or pelvic soft tissues. The majority of these sarcomas are indolent and can be surgically removed with curative intent. However, 30-40% of tumors recur local-regionally or metastasize and have no curative treatment options. Histology demonstrates that tumor masses present peculiar vascular features (hemangiopericytoma-like vessels) and are made up of patternless fibroblastic cells that are CD34+(*5*). SFTs present no recurrent mutations at known oncogenes or tumor suppressors loci. However, an intrachromosomal inversion on chromosome 12 was identified as the molecular hallmark of SFTs in 2013(*6, 7*). The inversion on the long arm of chromosome 12 results in a gene fusion between *NAB2* and *STAT6.* Immunohistochemistry using antibodies targeting the C-terminus of STAT6 reveals strong nuclear staining in the presence of the chimeric protein NAB2-STAT6, leading to STAT6 nuclear staining as a diagnostic tool for SFTs(*8*). Hemangiopericytomas were previously diagnosed as a distinct soft tissue neoplasm and upon discovery of the *NAB2-STAT6* gene fusion, are now classified as SFTs(*3, 4*).Despite the introduction of molecular diagnostic tools, many SFTs are still misdiagnosed or misclassified, owing to our fundamental lack of understanding of their etiology.

The precise role of NAB2-STAT6 in SFTs pathogenesis remains unclear. Both NAB2 and STAT6 are physiologically active as transcriptional regulators. STAT6 is a DNA-binding transcription factor operating downstream to the JAK/STAT signaling pathway in leukocytes. STAT6 is commonly activated by cytokines such as IL-4 and IL-13 and mediates key immunological processes such as macrophage polarization and T-cell/B-cell activation(*9-11*). Activation of STAT6 occurs via JAK-mediated phosphorylation at Y641, triggering homodimerization via the SH2 domain and ultimately leading to nuclear translocation(*12, 13*). Nuclear STAT6 stimulates transcription through its transactivation domain that recruits RNA Helicase A (RHA), SND1, and p300/CBP(*14-17*).

Conversely, the NGFI-A binding protein 2 (NAB2) is expressed in most tissues and was originally characterized as co-repressor of the Early Growth Response transcription factors (EGR1/2)(*18*). The EGR proteins are Immediate Early Genes (IERs), like other ubiquitously expressed transcription factors such as AP-1 (all FOS/JUN family members)(*19, 20*). Immediate Early Genes are upregulated in most cell types in response to a variety of growth stimuli(*21, 22*). In fact, EGR1/2 are considered, like AP-1, broad regulators of cell proliferation and peak during developmental processes(*23*). Their role has been best studied in the central nervous system (CNS), where they also regulate neuronal activity post mitosis(*24-26*). EGR1 can modulate the expression of growth factors, such as IGF2 and TGF-β1, as well as that of NAB2, and is believed to act either as oncogene or tumor suppressor depending on the cellular context(*27-30*). NAB2 forms homodimers and heterodimerizes with its family member NAB1, which was also described as co-repressor of EGR1-dependent transcription(*31*). We recently demonstrated that NAB2 may also function as co-activator of transcriptional enhancers during myeloid differentiation(*32*). The global role of NAB2 as a chromatin regulator in tumors, either in its wild-type form or as a fusion protein partner, remains poorly elucidated.

Several different NAB2-STAT6 isoforms have been identified. Most tumors (40-70%, especially those of pleural origin) carry a large fusion product of ~140Kda resulting from exons 1-4 of NAB2 (retaining a truncated NCD2 domain and missing the c-terminal CID domain) and exons 2-22 of STAT6 (with the entire CDS). Another frequent isoform retains a larger NAB2 moiety (exon 1-5/6, to include a truncated CID domain) and a much

smaller STAT6 moiety (exon 16/17-22, missing the SH2 and DNA binding domains but retaining the transactivation domain). This shorter isoform accounts for 10-20% of cases and is most common in SFTs originating in extra-pleural locations. Risk of malignant progression, recurrence and metastasis are not preferentially associated with either isoform of NAB2-STAT6(*33-37*). In fact, the current risk assessment model (Demicco score) does not incorporate any molecular or genetic features of the primary tumor(*38*).

The determinants of NAB2-STAT6 occupancy genome-wide have not been previously investigated, nor have the epigenetic and transcriptional consequences of the fusion protein been identified(*6, 34, 39, 40*). Here, we employ functional genomics and proteomics strategies to demonstrate that NAB2-STAT6 drives a EGR1-dependent neuroendocrine transcriptional program via epigenetic activation of a network of distal enhancers and proximal promoters that is normally operating in neuronal cells. We further show that NAB2 is primarily contained in the cytoplasm and becomes nuclear-bound when fused to STAT6. The fusion protein utilizes the full EGR1 regulatory axis, including homo- and heterodimerization with NAB1/NAB2, to drive a highly distinctive transcriptional program that sets SFTs apart from most other tumors of mesenchymal origin.

RESULTS

Solitary Fibrous Tumors express a neuronal gene signature

Developing treatments for SFTs is particularly challenging since their molecular landscape is largely undefined. Whole-exome sequencing of primary tumors has led to the discovery of *NAB2-STAT6* gene fusions. Additional transcriptomic data have been previously obtained, but were difficult to interpret in the absence of matching normal tissues(*6, 41*). We examined Formalin-Fixed Paraffin-Embedded (FFPE) sections of 8 tumors and their normal tissue counterpart. All tumors were removed from the pleura or pleural-adjacent locations such as lung, chest wall, and hilum. Three of these tumors were diagnosed as malignant, three as low risk or benign, and the other two had uncertain biological potential (Fig.1a). We extracted RNA from all FFPE sections and subjected it to exon targeted sequencing (Fig.1b). To validate SFT diagnosis for each tumor we employed Arriba, a computational tool to identify gene fusions from RNA-seq data. All tumors shared one common NAB2-STAT6 isoform, spanning the first 4 exons of NAB2 and all but the first STAT6 exon (Fig.1c, Supplementary Fig 1a). The transcriptomes of the primary SFTs well correlated with previously published SFT RNA-seq datasets (Supplementary Fig.1b). We performed differential gene expression analysis between the tumors and the adjacent normal tissues (Supplementary Fig.1c). We found 2,429 genes significantly upregulated across all SFTs. Notably, the most upregulated gene was *IGF2*, an EGR1 target that was recognized as major driver of a hypoglycemic condition occurring in <5% of SFTs (Doege-Potter syndrome). We further identified key regulators of neuronal development as upregulated, including *SHOX2, KCNA1, LHX2*, and *ROBO2*. A set of upregulated genes encoded GABA receptor subunits (*GABRD, GABRA2, GABRB2, GABRG1*). We also identified 3,769 downregulated genes, including a set of immune processes regulators such as *CCL18, IGHA1, SCGB3A1, IGHA2,* and *PIGR* (Fig.1d and Supplementary Table 1). Gene ontology and gene set enrichment analysis (GSEA) revealed that neuronal development programs are broadly upregulated in SFTs (Fig. 1e, Supplementary Fig. 1e). Conversely, immune and cell signaling pathways are significantly downregulated (Supplementary Fig. 1d-e), suggesting that SFTs may be immunologically ‘cold’. We analyzed the enrichment of transcription factor motifs within the promoter regions of all upregulated genes identified by RNA-seq. We found that binding sites for DB1, MAZ, MOVO−B, and EGR1 were the top enriched motifs by adjusted p-value and that EGR1 motifs were over-represented in the whole enrichment matrix (Fig.1f). Conversely, the top motifs enriched at the promoters of downregulated genes belonged to transcription factors active in mesoderm derived tissues, like HTF4 and MRF4 (Supplementary Fig.1f). These results suggest that an EGR1-bound neuronal transcriptional program is being activated in SFT cells.

Generation of an inducible NAB2-STAT6 cell model

SFTs are presumed to be mesenchymal in origin, therefore we used the osteosarcoma-derived U2OS cell line to generate an inducible model of NAB2-STAT6 expression and study the early transcriptional events driven by the fusion protein. Previous work on NAB2-STAT6 relied on overexpression in bulk cell populations that were primarily of non-mesenchymal origin, potentially limiting their utility(*6, 34, 39, 40*). We generated a single cell derived clone expressing the most common isoform of NAB2-STAT6 (exons 1-4 of *NAB2* and exons 2-22 of *STAT6*) under the control of a doxycycline inducible promoter (tet-ON). We observed expression of NAB2-STAT6 starting at 24 hours of doxycycline treatment, peaking at 48 hours (Fig. 2a). To profile early gene expression changes induced by the expression of NAB2-STAT6 we performed 3' mRNA QuantSeq. We observed widespread gene dysregulation that increased in magnitude over time, indicating that the fusion protein has direct and robust effects genome-wide (Fig. 2b, Supplementary Fig. 2a). We observed a smaller cluster of 299 genes exhibiting the strongest upregulation after 1 day of NAB2-STAT6 expression (cluster 1) and tapering off at 48h and 72h of doxycycline (Fig 2b), suggesting they could be indirect targets. This cluster was enriched for translation and protein biosynthesis genes that are regulated by E2F, ZF5 and HES-7 transcription factors (Supplementary Fig. 2b-c), further pointing to an immediate cellular stress response to the ectopic expression of NAB2-STAT6. Larger scale gene expression changes were apparent after 48h of doxycycline with 562 upregulated genes (73%), including IGF2 and several neuronal regulators. About 27% of differentially expressed genes were downregulated (Fig 2c and Supplementary Table 2), indicating that the primary function of NAB2-STAT6 may be transcriptional activation. Upregulated pathways were primarily involved in neuronal differentiation and development (Fig 2d). Promoters of upregulated genes were overwhelmingly enriched for EGR1 motifs, while STAT motifs were not significantly enriched (Fig. 2e). The expression of several neuronal markers such as *LHX2, ROBO2, SHOX2* increased over time of doxycycline treatment and peaked at 72h of NAB2-STAT6 expression (Fig. 2f). Interestingly, the expression of endogenous *NAB2, NAB1*, and *EGR1* followed a similar trend, suggesting that the fusion protein enables a feed forward mechanism of the EGR1 regulatory axis (Fig 2f). NAB2 has been proposed to repress its own expression and that of *NAB1* and *EGR1*. However, this trend suggests that NAB2-STAT6 does not possess any co-repressor activity and is instead functioning as a co-activator.

The GO and TF enrichment analyses in U2OS cells closely matched our findings from primary SFTs (Fig. 1), despite a limited overlap between the gene lists (15%, Supplementary Fig. 2d). Key EGR-1 targets such as *IGF2* or the neuronal development regulators *LHX2* and *ROBO2* were commonly upregulated, indicating that U2OS cells can be utilized to model the activity of the fusion protein but do not accurately represent the cell of origin of SFTs (Supplementary Fig. 2d).

EGR1 targeted promoters and enhancers are activated by NAB2-STAT6

We established that EGR1 neuronal targets are upregulated in SFTs and after NAB2-STAT6 expression in U2OS cells. To investigate the mechanistic underpinnings of NAB2-STAT6 dysregulation we performed ChIP-seq of the fusion protein through its C-terminal FLAG epitope after 48h of doxycycline. We identified 1,394 peaks that gained significant NAB2-STAT6 signal compared to day 0 and were almost equally distributed between promoters (57%) and distal cis-regulatory regions (43%, Fig. 3a, Supplementary Fig. 3a, and Supplementary Table 3). Next, we analyzed endogenous EGR1 at NAB2-STAT6 sites and found seeding binding across the vast majority of fusion protein peaks (Fig. 3a). Upon NAB2-STAT6 expression, EGR1 binding increased at least two-fold at most sites. We performed ATAC-seq to investigate chromatin accessibility status. NAB2-STAT6 peaks were moderately accessible at day 0 and their accessibility increased two-fold (similar to EGR1 binding) at 48h of doxycycline (Fig. 3a). We next profiled endogenous NAB2 and STAT6 in control conditions and found that NAB2 localized at NAB2-STAT6 peaks but not STAT6 (Supplementary Fig. 3b). Conversely, NAB2-STAT6 did not bind any STAT6 control sites (Supplementary Fig. 3c). To further assess determinants of NAB2-STAT6 binding we performed motif analysis using HOMER and identified a strong EGR1/EGR2 signature but no STAT family motifs (Fig. 3b). Gene set enrichment analysis (GSEA) showed that NAB2-STAT6 target genes, as determined by ChIP-seq, were highly enriched in NAB2-STAT6 upregulated genes found by RNA-seq analysis (Fig 3c). NAB2-STAT6 and EGR1 co-localize at promoters and enhancers of neuronal markers such as KNDC and UNCX (Fig. 3d and Supplementary Fig. 3d). Additional enhancer sites targeted by NAB2-STAT6 were found near known EGR1 target genes such as IGF2,

the most upregulated gene in SFTs (Fig. 3e). Interestingly, the fusion protein also targeted the proximal promoter of EGR1, thereby boosting its expression and providing a robust feed-forward mechanism to sustain its own epigenetic network (Supplementary Fig. 3e). Collectively, these results demonstrate that NAB2-STAT6 localizes to EGR1 targets increasing their accessibility and expression.

NAB2-STAT6 localizes to EGR1 targets in primary tumors

To validate our U2OS model of NAB2-STAT6 as a robust activator of certain EGR1 targets, we set out to determine genome-wide NAB2-STAT6 occupancy in a primary solitary fibrous tumor. We obtained a fresh primary pre-sacral SFT post-surgery (no radiation or chemotherapy had been administered to the patient). This tumor expressed the short isoform of NAB2-STAT6 containing exons 1-6 of NAB2 and exons 17-22 of STAT6 (Supplementary Fig. 4a). We isolated and fixed in single cell suspension ~50 million cells and performed ChIP-seq for NAB2, STAT6, and RNA Polymerase II (RNAPII). Since we previously showed that endogenous NAB2 and STAT6 localization do not overlap in physiological conditions (Supplementary Fig. 4b), we reasoned that we could pinpoint NAB2 and STAT6 primary binding sites as well as specific binding sites of the fusion protein, based on the convergence of NAB2 and STAT6 signals (Fig. 4a). We identified 6,284 NAB2 peaks, 1,640 STAT6 peaks, and 38,036 RNAPII peaks in primary tumor cells. Peak overlap analysis revealed 718 putative NAB2-STAT6 binding sites, 69% of which were distal to gene promoters and likely to represent transcriptional enhancers (Fig. 4b, Supplementary Fig. 4c and Supplementary Table 4). The NAB2-STAT6 peaks also showed robust RNAPII signal, further suggesting that the fusion protein elicits transcriptional activation (Fig. 4b). We also identified 5,912 unique NAB2 peaks and 1,285 unique STAT6 peaks (Fig 4b). Next, we performed motif analysis using HOMER. NAB2 only sites were enriched for EGR1/2 motifs (Fig 4c and Supplementary Fig 4d) and the same motif profile was found at NAB2-STAT6 peaks (Fig 4c and Supplementary Fig 4e). STAT6 only peaks, instead, were enriched for GRE and STAT motifs (Fig 4c and Supplementary Fig 4f). Gene set enrichment analysis of NAB2-STAT6 peaks (closet gene) showed significant correlation with upregulated genes in SFTs (Fig 4d). Collectively, these findings align with data obtained in U2OS cells and suggest that NAB2-STAT6 localization is exclusively driven by EGR1 binding. We found limited overlap between NAB2-STAT6 sites in the primary tumor and those retrieved in U2OS (Fig. 4e-f and Supplementary Fig. 4g-h), consistent with the limited overlap between their transcriptomes. It is likely that the mechanism of transcription activation by NAB2-STAT6 is conserved, whereas the targets are cell type specific. Interestingly, NAB2 appeared upregulated in both systems as either the fusion protein or endogenous NAB2 bound the proximal promoter region (Fig 4f), furthermore nucleosome accessibility and EGR1 binding increased in U2OS cells upon doxycycline treatment (Fig. 4f). The oncogenic NAB2-STAT6 fusion may thereby reinforce its own expression in SFTs.

The subcellular localization and interactome of NAB-STAT6

Previous work established that NAB2-STAT6 localizes to the nucleus. It has been proposed that NAB2 is primarily guiding the subcellular localization of the chimeric protein, barely contributing any further functional domain, whereas STAT6 endows the fusion product with activation domain(*8*).Since the interactome of NAB2-STAT6 has not been previously investigated with unbiased proteomic approaches, we performed LC-MS/MS analysis on NAB2-STAT6 eluates using a MudPIT approach. First, we affinity purified NAB2-STAT6 from nuclear fractions of U2OS cells using antibodies directed against NAB2 or STAT6. Both antibodies efficiently immunoprecipitated the chimeric protein, as well as NAB1 and EGR1 (Fig 5a). To validate EGR1 as a NAB2-STAT6 interactor we performed the reciprocal IP. Upon doxycycline induction in U2OS, EGR1 antibodies co-precipitated NAB2-STAT6. In addition, EGR1 co-purified along with CBP/P300 and RNAPII subunits suggesting that NAB2-STAT6 co-opts EGR1 co-activator functions (Supplementary Fig 5a). To further validate NAB2-STAT6's interactome we performed FLAG affinity purification in U2OS as well as in a 293T stably expressing clone (Supplementary Fig 5b). This approach pulled down the fusion product as well as near-stoichiometric amounts of NAB1, additional interactors included the co-activators SND1 (normally recruited by STAT6) and RHA (recruited by either STAT6 or EGR1, Supplementary Fig 5c-d). NAB1 is a NAB2 paralog that was loosely studied for its ability to repress EGR1-mediated activation. NAB1 is the strongest interactor of NAB2-STAT6 across multiple experiments (Supplementary Fig. 5d), suggesting that a heterodimeric form of

the oncogenic protein may be its most common form. Endogenous NAB2 can also heterodimerize with NAB1 as we have shown (Supplementary Fig. 5d). Additionally, the chimeric protein can heterodimerize with endogenous, full length NAB2 (Supplementary Fig 5e-f).

To further clarify the subcellular localization of NAB2-STAT6 we initially performed fractionation experiments in U2OS cells to find that NAB2 and EGR1 localization is primarily cytoplasmic prior to NAB2-STAT6 expression, whereas endogenous STAT6 is nuclear (Fig. 5b). The fusion protein is fully retained in the nucleus and drives endogenous NAB2 and EGR1 to the nucleus (Fig. 5b). To corroborate these findings, we employed immunocytochemistry (ICC). We first examined the primary tumor cells that we profiled by ChIP-seq (Fig. 4) and retrieved a robust nuclear signal for NAB2 and STAT6 (Fig. 5c). Next, we performed ICC in the U2OS inducible clone and confirmed that in wild type conditions, NAB2 localization is largely cytoplasmic, while STAT6 signal is mostly nuclear (Fig. 5c, Supplementary Fig. 5g). Upon NAB2-STAT6 expression, NAB2 and STAT6 signals co-localize in the nucleus, while a small amount of endogenous NAB2 remains cytoplasmic. Taken together, these data suggest that the STAT6 moiety drives nuclear localization of the fusion protein. The NAB2 paralog NAB1 heterodimerizes with the fusion protein (Fig. 5a, Supplementary Fig. 5d). Accordingly, NAB1 under physiological conditions is predominantly cytoplasmic (Fig. 5d, Supplementary Fig. 5h) and is redirected to the nucleus by NAB2-STAT6 (Fig. 5d), having a similar fate to that of endogenous NAB2 (Fig. 5b). Importantly, NAB1 antibodies do not cross-recognizen NAB2, and viceversa (see Author Response to Reviewers).

The SFT gene signature is expressed in neuroendocrine tumors

SFTs are traditionally assimilated to soft tissue sarcomas and are often positive to CD34 staining. They are broadly classified as mesenchymal tumors based on their histological patterns, albeit their cell of origin remains uncertain(*42, 43*). To unbiasedly determine which tumor transcriptomes are the closest to SFTs, we performed single sample gene set enrichment analysis (ssGSEA) probing the SFT gene signature across the entire Cancer Genome Atlas database (22,687 samples from 223 different tumors). About 30% of TCGA tumors, including most sarcoma subtypes, had a positive ssGSEA score indicating that a significant number of EGR1-driven genes are upregulated (Fig. 6a). Strikingly, a group of neuroendocrine tumors (Glioblastoma, Mixed Glioma, Neuroblastoma, and Pheochromocytoma/Paraganglioma) stood out as the most closely correlated to SFTs (Fig. 6a). Conversely, myeloid and lymphoid malignancies showed the poorest correlation (Fig 6a). Neuroendocrine tumors are broadly characterized by nerve cell traits as well as by their unique secretory features that may impact distal organs and tissues. In addition to IGF2 and IGF1 hormones, as we previously described, SFTs upregulate secreted metabolic regulators such CTRP11, neuropeptides and their activating enzymes (NPW, PCSK2), and chemokines/cytokines such as PDGFD, NTN3, DEFB136 (Supplementary Table 1). Neuroendocrine tumors are challenging to treat, presenting highly variable degrees of aggressiveness. We asked whether the SFT gene signature correlated with prognosis by performing a survival analysis across all TCGA samples. SFT-like tumors showed significantly worse outcomes, suggesting that expression of an EGR1 gene signature underlies tumor aggressiveness and/or therapy resistance (Fig 6b). We next wanted to validate the SFT signature for its ability to classify solitary fibrous tumors a priori, reasoning that ssGSEA scores should pinpoint misclassified SFTs within a pool of different tumors. Prior to NAB2-STAT6 discovery and its diagnostic use, some SFTs were misclassified as mesotheliomas due to their pleural localization (*44*). We therefore analyzed all transcriptomes from the TGCA collection of 87 mesotheliomas and identified one patient presenting an outlier ssGSEA positive score (Fig. 6c). We used Arriba to confirm that SH-A7BH was, in fact, the only tumor in the mesothelioma cohort containing a NAB2-STAT6 fusion (exons 1-6 of NAB2 and 16-22 of STAT6, Fig. 6d).

The previously unrecognized neuroendocrine signature in SFTs and the startling number of neural genes activated via NAB2-STAT6/EGR1 elicits further questions about the identity of the tumor-initiating cells. To address this question, we first investigated the correlation of SFTs to normal tissue types by calculating the Spearman correlation with all samples from the Human Protein Atlas. SFTs correlated best with several neuronal tissue types and were the least correlated with lung and gut tissues (Supplementary Fig. 6). In addition to neuronal genes, we noticed that some of the top upregulated genes were implicated in embryogenesis and early development (i.e. shh regulators such as GLI2 or wnt pathway activators such as WNT2). We resolved to examine the protein domains that were enriched within the top 400 genes of the SFT signature using the functional annotation tool InterPro. The most significantly enriched domain was the

Homeobox domain (Fig. 6e), which functions as a DNA binding domain for a large class of transcription factors that are expressed during embryonic and fetal development to drive pattern formation, axis specification, ultimately leading to proper tissue and organ morphogenesis. Notably, several homeodomain transcription factors were overexpressed in SFTs (ALX4, SHOX2, SIX1) and entire HOX clusters (HOXA, HOXC) were upregulated as further confirmed by RNAPII profiling of a primary tumor tissue (Fig. 6f). However, we did not identify a homeobox binding motif in our previous analysis of NAB2-STAT6 binding sites (Fig. 3b and 4c). We then searched for enriched TF binding sites using all RNAPII-bound enhancers (excluding NAB2-STAT6 targeted enhancers) and found an overwhelming enrichment for homeotic proteins as well as Nanog (Fig. 6g). These results suggest that a homeobox TF network, elicited by NAB2-STAT6 upregulation of homeotic genes, drives a significant part of the SFT transcriptomes.

DISCUSSION

Fusion oncoproteins frequently occur in a broad range of tumors, such as AML, sarcomas, non-small cell lung cancer, and prostate cancer(*45*). The loci encoding fusion protein products originate via chromosomal aberrations such as translocations or, as in the case of SFTs, inversions(*46*). In most cases, the aberrant protein product retains select biological activities from both protein partners and drives neoplastic transformation(*47-49*).

In this work we characterize the mechanistic role of the NAB2-STAT6 fusion in Solitary Fibrous Tumors through a combination of genomics data generated from primary SFTs, a tet-inducible model system in human cell lines, and by comparing SFT gene signatures to the Cancer Genome Atlas collection. Detection of NAB2-STAT6 protein products has become the primary diagnostic tool for this rare tumor type, however the role of the fusion protein in the etiology of SFTs has remained elusive(*8*). A variety of mechanisms have been proposed for NAB2-STAT6's function, including activation of STAT6 targets and conversion of NAB2 from a repressor to an activator(*6, 34, 39, 40, 50*). Here we show that NAB2-STAT6 activates an EGR1-driven neuroendocrine gene expression signature by translocating EGR1, NAB1, and wild type NAB2 to the nucleus (Fig. 7). Our data suggest that the NAB2 moiety of the fusion protein targets a specific set of EGR1-dependent enhancers and proximal promoter sites and uses NAB1, or endogenous NAB2, as co-activators. In fact, our data suggest that NAB2/NAB1 are co-activators of EGR-1 targets under physiological conditions, however cytoplasmic localization restrains their activity. Unlike previously proposed(*8*), we find that the STAT6 moiety of the fusion is the major driver of its nuclear localization (Fig. 7). STAT6, however, does not endow the fusion protein with the ability to recognize a subset of STAT motifs, but may recruit additional co-activators. NAB1/NAB2 were originally proposed to act as co-repressor of EGR1 targets on the basis of transactivation assays(*51*). Data from this work and our previous analysis of NAB2 activity during myeloid differentiation establish that these proteins operate as robust co-activators of EGR1 targets, at enhancers and proximal promoters(*32*). In myeloid cells, we found that full-length NAB2 recruits the Integrator protein complex. The NAB2 moiety of NAB2-STAT6 does not appear to recruit Integrator subunits, yet retains co-activator abilities by partnering with chromatin modifiers (CBP/P300) and helicases (RHA).

For the first time, we were able to establish transcriptional changes that occur in SFTs through comparison with the adjacent matching normal tissue. This led to identifying a robust EGR-1 driven neuronal gene signature, including GABA receptor subunits, synapses modulators, ion channels and modulators of axonal morphology. In the CNS, EGR1 is known to regulate a variety of neuronal-dependent processes such as memory and behavior. We also identified a distinct secretory gene signature associated with SFTs. In fact, IGF2 is the most upregulated gene, via activation of an intronic enhancer by EGR1. IGF2 was pinpointed as the cause of hypoglycemia occurring in a very small subset of SFTs (Doege–Potter syndrome)(*52*). Our data suggest that IGF2 (and IGF1) upregulation is a common feature of all SFTs. In addition to insulin-like growth factors, STFs may secrete a host of peptides with diverse functions in neuronal processes, chemotaxis, and growth stimulation. The previously unrecognized neuronal features and the putative secretory phenotype of STFs set them apart from mesenchymal malignancies and relate them to neuroendocrine malignancies such

as pheochromocytoma, oligodendroglioma and neuroblastoma. Neuroendocrine tumors originate from neuron-like cells that are able to send and receive signals from the nervous system, but also function as endocrine organs by producing hormones such as IGF2, with a systemic effect across distal tissues and organs. Similar to many neuroendocrine neoplasms, SFTs are immunologically cold and downregulate a set of immune response genes, perhaps with further support from IGF2 itself(*53*).

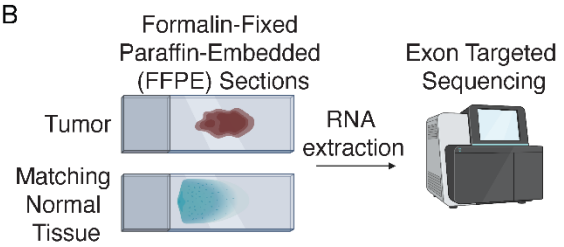
SFTs were classified as mesenchymal on the basis of their histological patterns and are believed to originate from soft tissue cell types, such as fibroblasts(*42, 43*). When comparing SFT signatures to human tissue RNA-seq datasets, we identified greater similarities to CNS tissues but not mesenchymal tissues types. Furthermore, we unveiled a set of embryonically and fetally expressed transcription factors that appear to coordinate part of SFT transcriptomes. Ectopic expression of few homeotic genes has been observed across many tumors and may contribute to their growth and plasticity(*54*). SFTs appear to upregulate an overwhelming number of homeotic transcription factors as we showed, perhaps suggesting that the tumor mass may originate from a corrupted developmental process from embryogenesis or fetal development.

Fig. 1

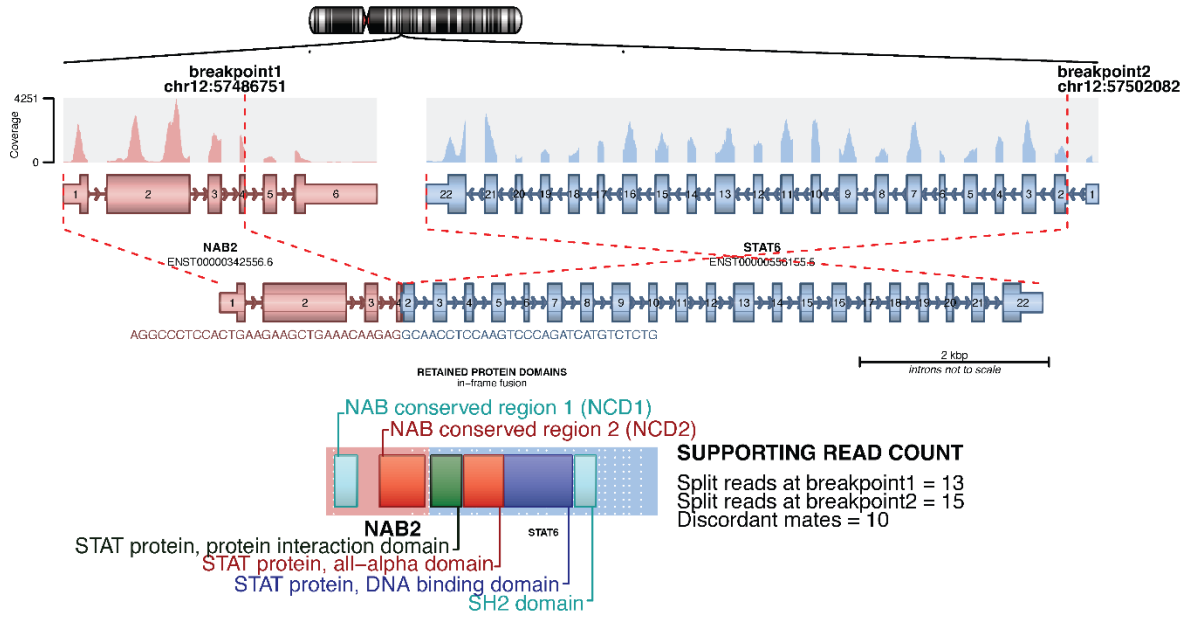
A

ID	Age at surgery	Sex	Diagnosis	Site
SFT 004	68	male	SFT uncertain biologic potential	Left lung
SFT 006	68	female	Malignant SFT	Right hilum
SFT 007	74	female	SFT	Right lung
SFT 008	72	female	Malignant SFT	Left pleura
SFT 010	67	male	SFT high risk	Left chest wall
SFT 011	68	female	Malignant SFT	Left pleura
SFT 014	55	female	SFT low risk	Right lung
SFT 015	80	male	SFT low risk	Right lung

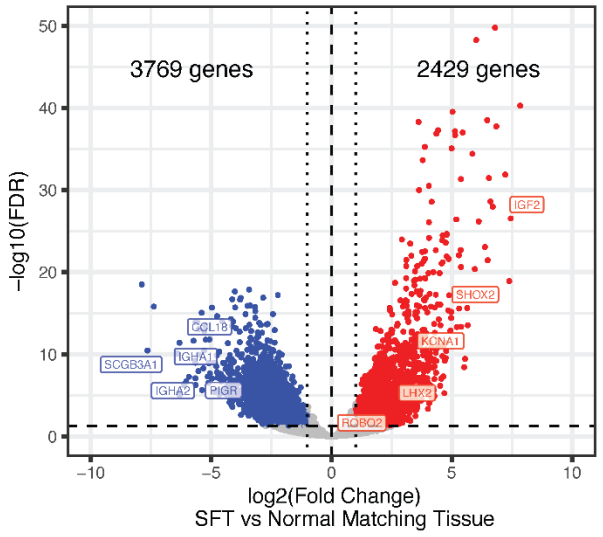
B



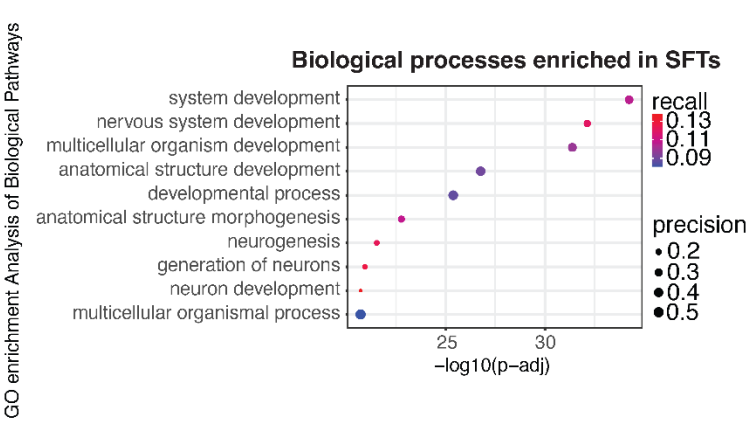
C



D



E



F

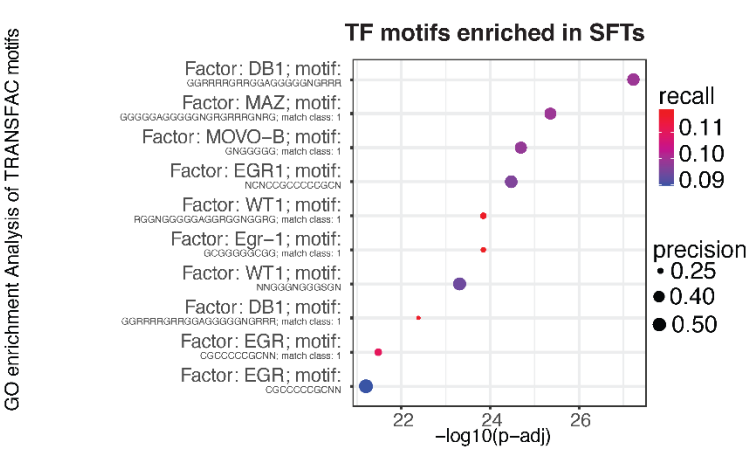


Figure legend

Figure 1. Solitary Fibrous Tumors express a neuronal gene signature.

(a) Table of the eight formalin-Fixed Paraffin-Embedded (FFPE) samples used for RNA-seq listing the sample ID, Age at surgery, Sex, Diagnosis of tumor status, and site of resection.

(b) Graphic demonstrating the workflow for FFPE RNA-seq. RNA was extracted from FFPE tumor and matching normal tissue and sequenced using exon targeted sequencing which increases quality of FFPE RNA-seq's.

(c) Most abundant gene fusion present in all SFTs in FFPE RNA-seq, NAB2-STAT6 with exons 1-4 of NAB2 and exons 2-22 of STAT6. Graphic generated with Arriba, a fusion detection algorithm.

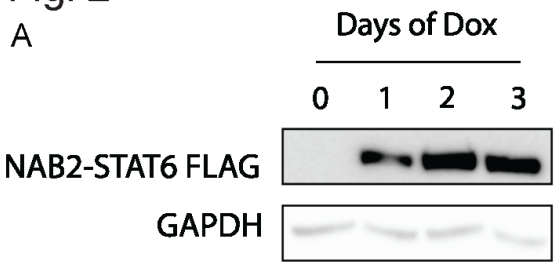
(d) Volcano plot showing differentially expressed genes in SFTs versus normal matching tissues as determined by FFPE RNA-seq (n = 8). 2,429 genes were upregulated (indicated by red dots) and 3,769 genes were downregulated (indicated by blue dots). Fold change >1, FDR <0.1.

(e) Biological pathway GO analysis of 2,429 upregulated genes in INTS10 KO cells revealed enrichment for developmental and specifically neuronal developmental pathways.

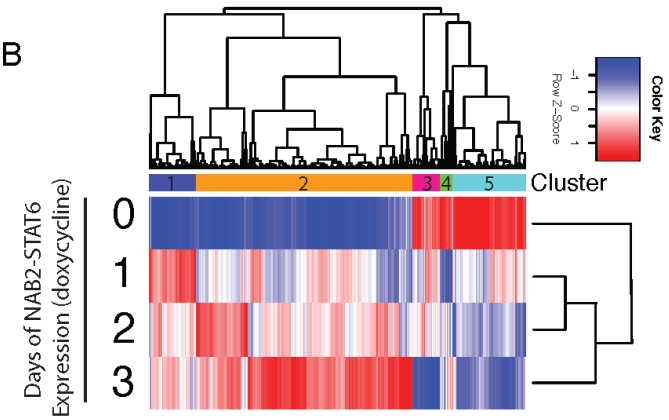
(f) TRANSFAC motif (transcription factor motifs at +/-1kb from transcription start site) GO analysis of 2,429 upregulated genes in INTS10 KO cells revealed enrichment for DB1, MAZ, MOVO-B, EGR1, and WT1 motifs.

Fig. 2

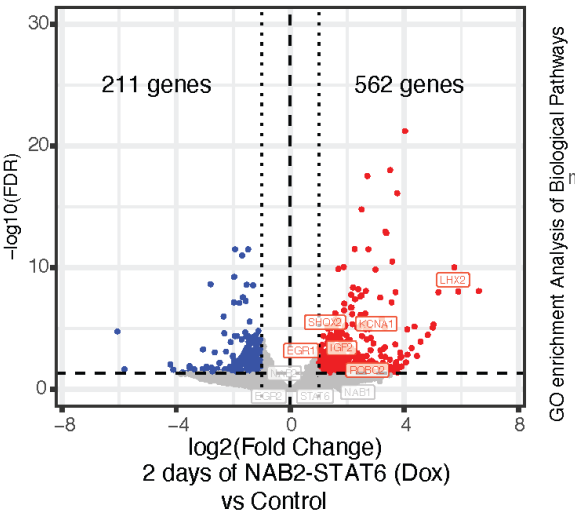
A


Days of Dox
0
1
2
3
NAB2-STAT6 FLAG
GAPDH

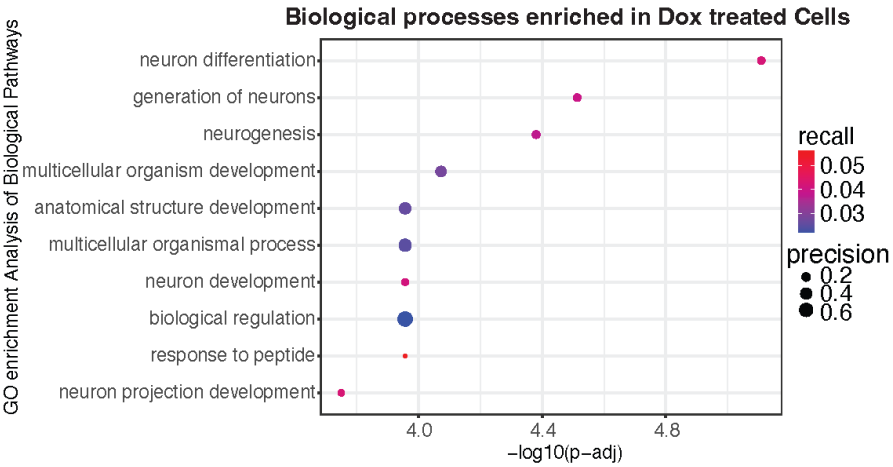
B


Color Key
Row Z-Score
Cluster
1
2
3
4
5
Days of NAB2-STAT6 Expression (doxycycline)
0
1
2
3

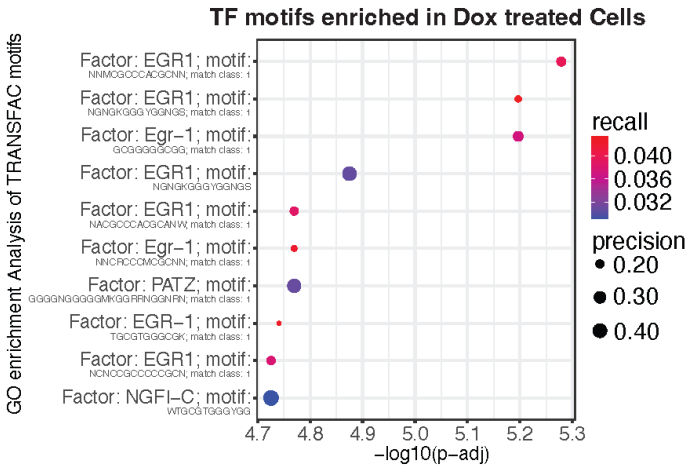
C


211 genes
562 genes
−log10(FDR)
log2(Fold Change)
2 days of NAB2-STAT6 (Dox)
vs Control
LHX2
SHOX2
KCNA1
EGR1
IGF2
ROBO2
NAB1

D


Biological processes enriched in Dox treated Cells
GO enrichment Analysis of Biological Pathways
neuron differentiation
generation of neurons
neurogenesis
multicellular organism development
anatomical structure development
multicellular organismal process
neuron development
biological regulation
response to peptide
neuron projection development
−log10(p−adj)
recall
0.05
0.04
0.03
precision
0.2
0.4
0.6

E


TF motifs enriched in Dox treated Cells
GO enrichment Analysis of TRANSFAC motifs
Factor: EGR1; motif:
Factor: EGR1; motif:
Factor: Egr−1; motif:
Factor: EGR1; motif:
Factor: EGR1; motif:
Factor: Egr−1; motif:
Factor: PATZ; motif:
Factor: EGR−1; motif:
Factor: EGR1; motif:
Factor: NGFI−C; motif:
−log10(p−adj)
recall
0.040
0.036
0.032
precision
0.20
0.30
0.40

F

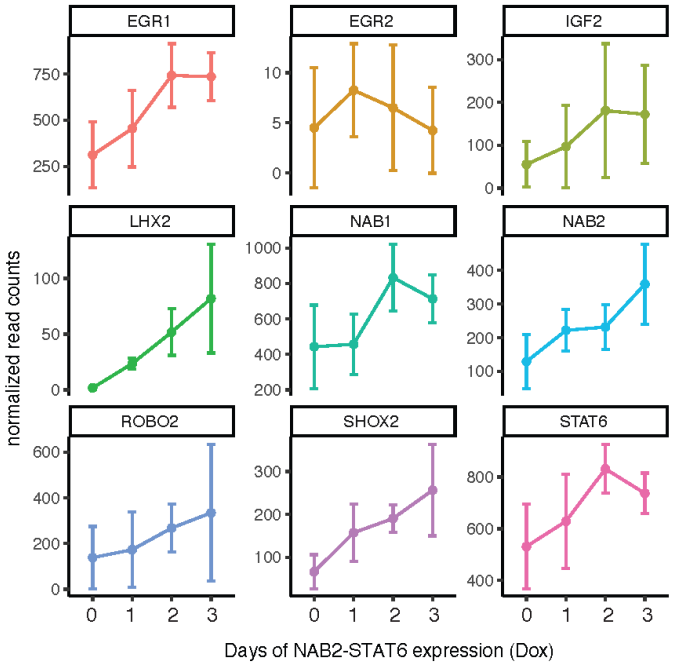
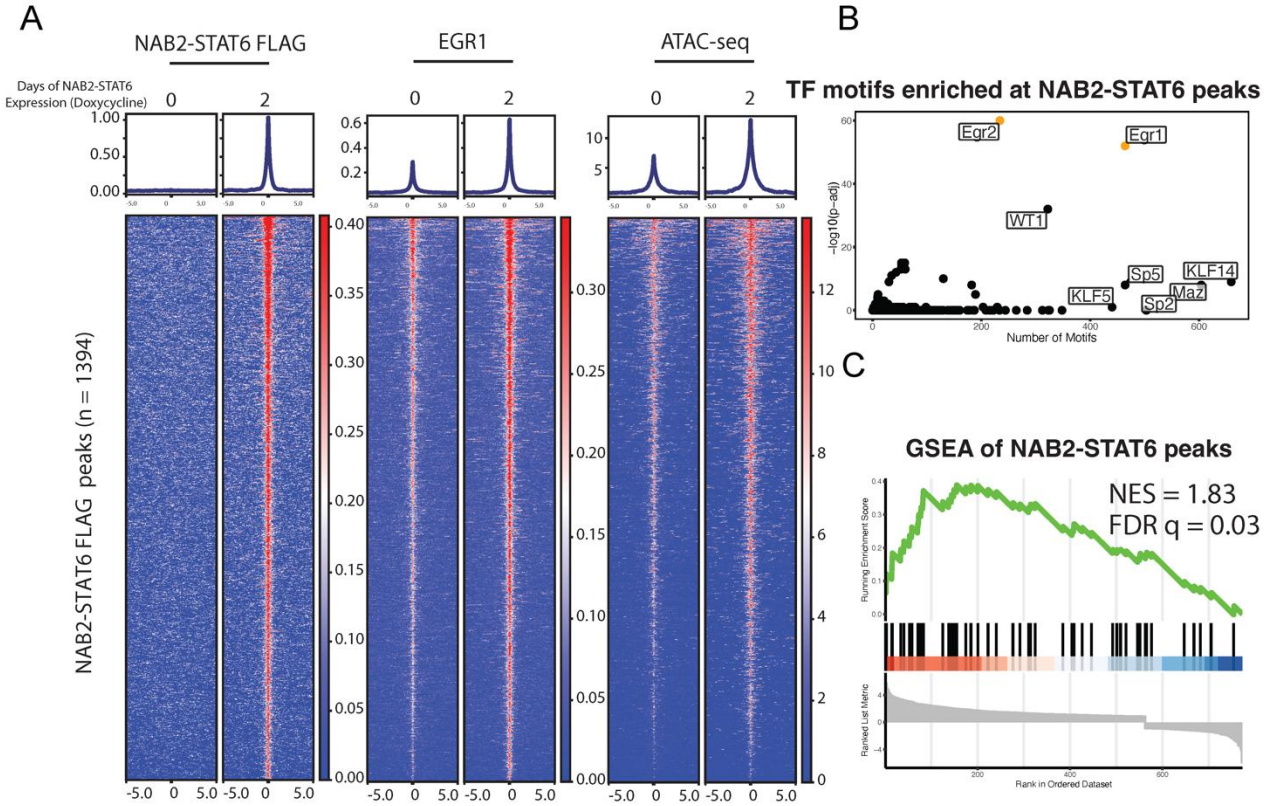

EGR1
EGR2
IGF2
LHX2
NAB1
NAB2
ROBO2
SHOX2
STAT6
normalized read counts
Days of NAB2-STAT6 expression (Dox)

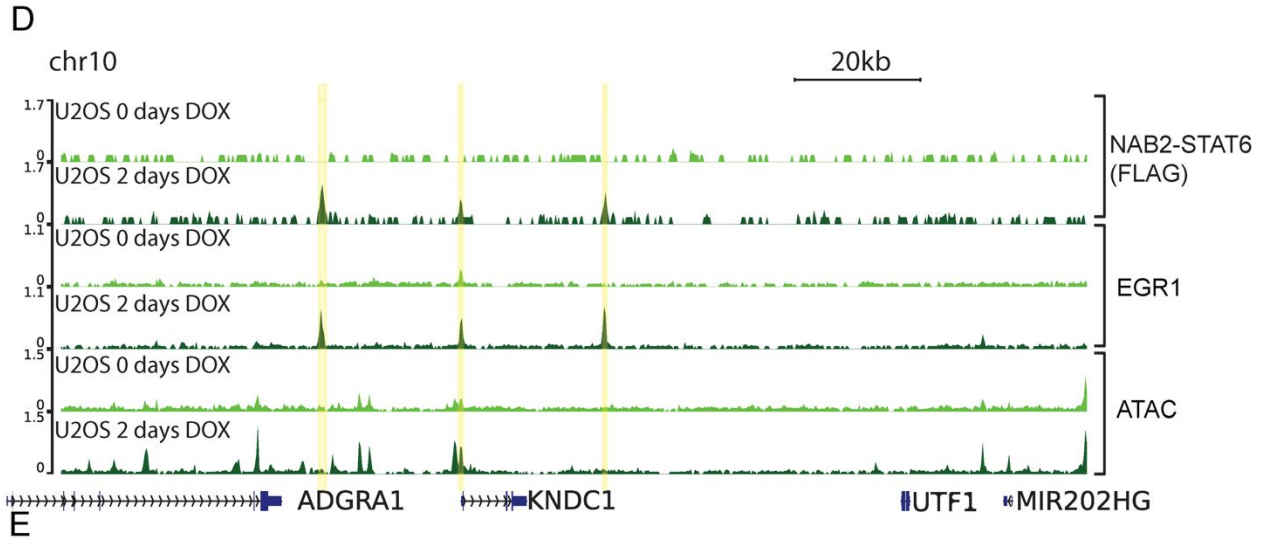
Figure 2. Generation of an inducible NAB2-STAT6 system to investigate early transcriptional changes.
(a) We generated a doxycycline inducible clone that expresses NAB2-STAT6 (NAB2 exons 1-4, STAT6 exons 2-22) with a C-terminal FLAG tag. Immunoblot analysis of whole cell extracts shows strong expression of NAB2-STAT6 after 1,2, and 3 days of doxycycline treatment using a FLAG antibody. GAPDH was used as control.
(b) Heatmap clustering analysis of 2,430 genes that are differentially expressed (fold change >1, FDR <0.1) across 1, 2, and 3 days of NAB2-STAT6 expression (Dox) as determined by 3' mRNA Quant-seq (n = 4)
(c) Volcano plot showing differentially expressed genes in cells expressing NAB2-STAT6 (Dox) for 2 days versus control cells as determined by 3' mRNA Quant-seq (n = 4). 562 genes were upregulated (indicated by red dots) and 211 genes were downregulated (indicated by blue dots). Fold change >1, FDR <0.1.
(e) Biological pathway GO analysis of 562 upregulated genes in INTS10 KO cells revealed enrichment for neuronal developmental pathways.
(e) TRANSFAC motif (transcription factor motifs at +/-1kb from transcription start site) GO analysis of 562 upregulated genes in INTS10 KO cells revealed enrichment for EGR1 and PATZ motifs.
(f) We plotted normalized read counts of NAB1, NAB2, EGR1, EGR2, STAT6, EGR target IGF2 and neuronal markers m (LHX2, ROBO2 and SHOX2) over 3 days of NAB2-STAT6 (Dox) expression. Targets of EGR1 were gradually upregulated; *NAB1, NAB2, EGR1, IGF2, LHX2, ROBO2* and *SHOX2*.

Fig. 3


A
NAB2-STAT6 FLAG
EGR1
ATAC-seq
Days of NAB2-STAT6 Expression (Doxycycline)
0
2
NAB2-STAT6 FLAG peaks (n = 1394)
-5.0
0
5.0

B
TF motifs enriched at NAB2-STAT6 peaks
Egr2
Egr1
WT1
Sp5
KLF14
KLF5
Maz
Sp2
-log10(p-adj)
Number of Motifs

C
GSEA of NAB2-STAT6 peaks
NES = 1.83
FDR q = 0.03
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset


D
chr10
20kb
U2OS 0 days DOX
U2OS 2 days DOX
NAB2-STAT6 (FLAG)
EGR1
ATAC
ADGRA1
KNDC1
UTF1
MIR202HG

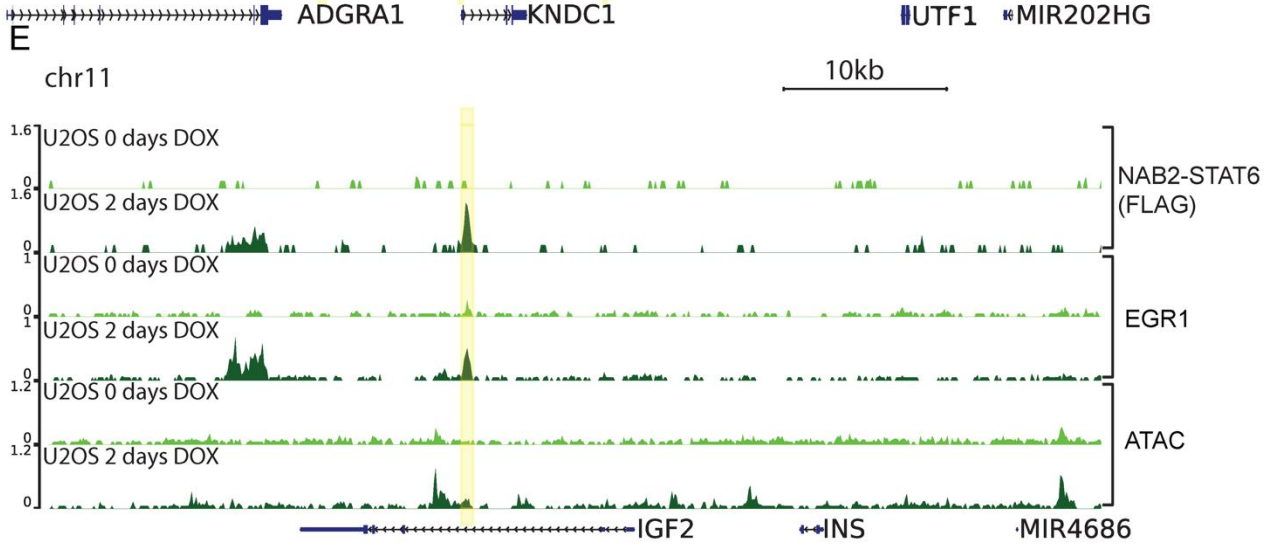

E
chr11
10kb
U2OS 0 days DOX
U2OS 2 days DOX
NAB2-STAT6 (FLAG)
EGR1
ATAC
IGF2
INS
MIR4686

Figure 3. EGR1 targeted promoters and enhancers are activated by NAB2-STAT6.
(a) Average profiles and heatmaps of NAB2-STAT6 FLAG and EGR1 ChIP-seq and ATAC-seq in both control and 2 days NAB2-STAT6 (Dox) expressing U2OS cells at 1394 NAB2-STAT6 FLAG peaks. NAB2-STAT6 FLAG becomes significantly localized to these peaks which have significant increases in EGR1 and ATAC-seq signal.
(b) Motif analysis of 1394 NAB2-STAT6 FLAG peaks using HOMER shows EGR1, EGR2, and WT1 as the most significantly enriched TF matrices.
(c) GSEA shows that genes nearest to NAB2-STAT6 FLAG peaks (n = 1394) are significantly upregulated after 2 days of NAB2-STAT6 (Dox) expression in U2OS cells when compared with control cells from Fig 2.
(d) Screenshot displays two enhancers and the promoter (highlighted in yellow) of *KNDC1* that gains NAB2-STAT6 FLAG localization and have increases in EGR1 localization and accessibility by ATAC-seq.
(e) Screenshot displays an enhancer (highlighted in yellow) of *IGF2* that gains NAB2-STAT6 FLAG localization and has increased EGR1 localization and accessibility by ATAC-seq.

Fig. 4

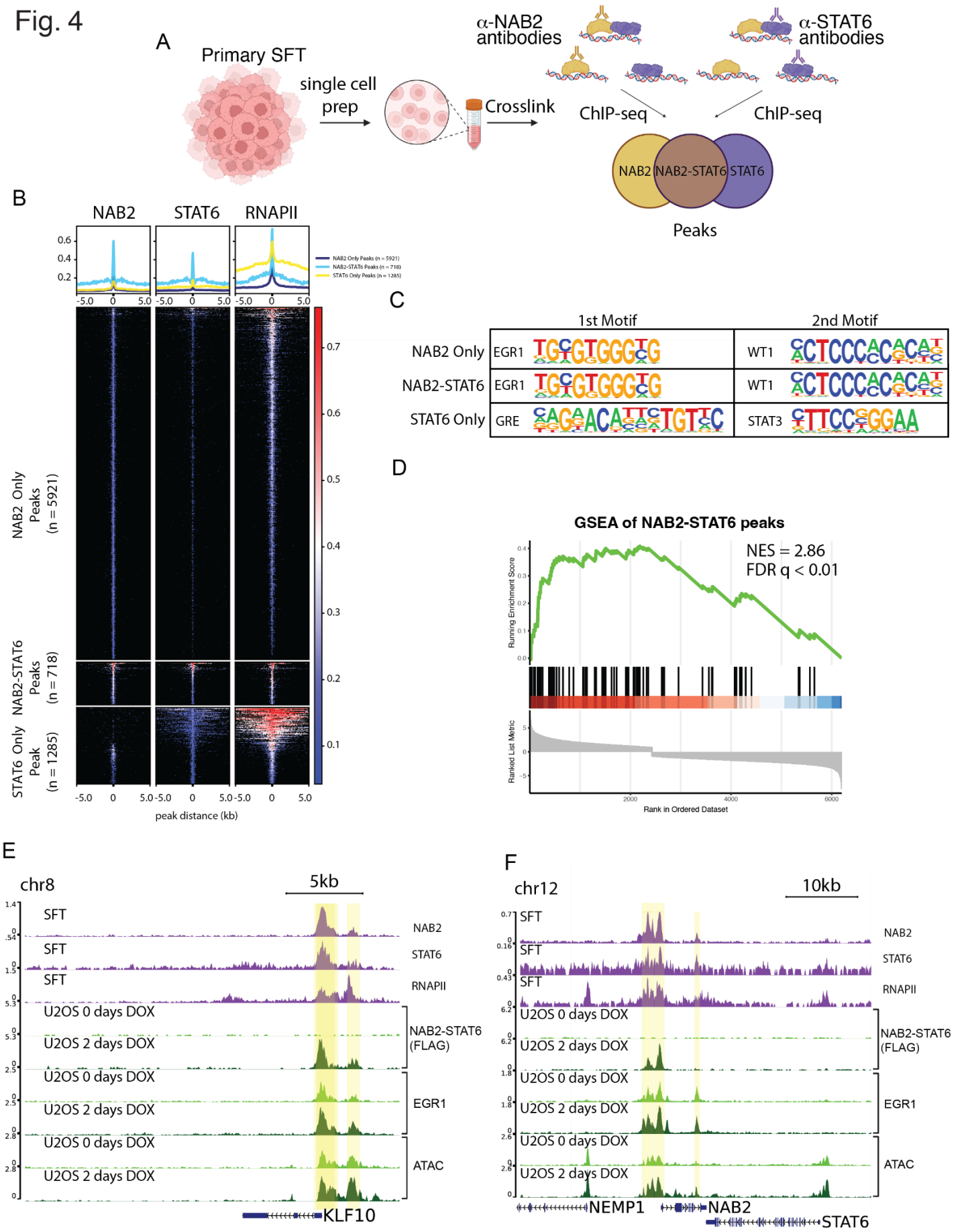
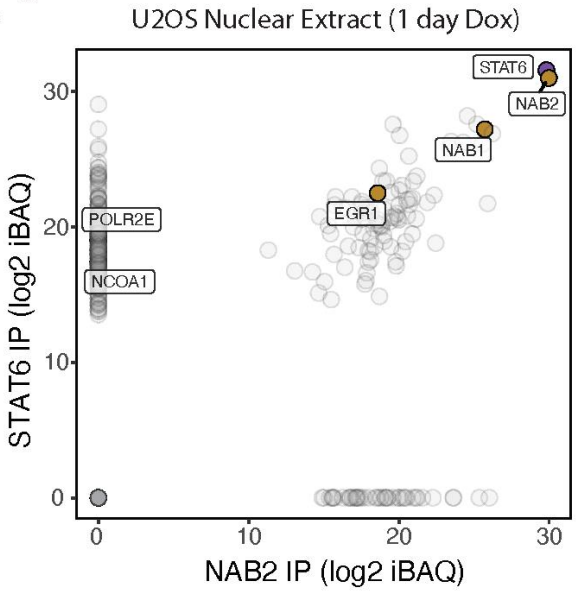

A
Primary SFT
single cell prep
Crosslink
α-NAB2 antibodies
α-STAT6 antibodies
ChIP-seq
ChIP-seq
NAB2
NAB2-STAT6
STAT6
Peaks
B
NAB2
STAT6
RNAPII
NAB2 Only Peaks (n = 5921)
NAB2-STAT6 Peaks (n = 718)
STAT6 Only Peaks (n = 1285)
NAB2 Only Peaks (n = 5921)
NAB2-STAT6 Peaks (n = 718)
STAT6 Only Peak (n = 1285)
peak distance (kb)
C
1st Motif
2nd Motif
NAB2 Only
EGR1
WT1
NAB2-STAT6
EGR1
WT1
STAT6 Only
GRE
STAT3
D
GSEA of NAB2-STAT6 peaks
NES = 2.86
FDR q < 0.01
Running Enrichment Score
Ranked List Metric
Rank in Ordered Dataset
E
chr8
5kb
SFT
NAB2
SFT
STAT6
SFT
RNAPII
U2OS 0 days DOX
U2OS 2 days DOX
NAB2-STAT6 (FLAG)
U2OS 0 days DOX
U2OS 2 days DOX
EGR1
U2OS 0 days DOX
U2OS 2 days DOX
ATAC
KLF10
F
chr12
10kb
SFT
NAB2
SFT
STAT6
SFT
RNAPII
U2OS 0 days DOX
U2OS 2 days DOX
NAB2-STAT6 (FLAG)
U2OS 0 days DOX
U2OS 2 days DOX
EGR1
U2OS 0 days DOX
U2OS 2 days DOX
ATAC
NEMP1
NAB2
STAT6

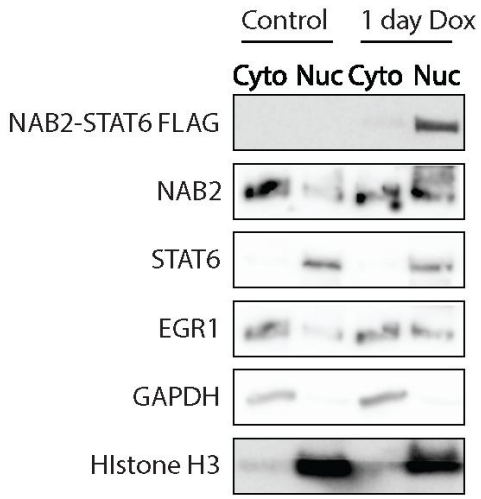
Figure 4. NAB2-STAT6 localizes to EGR1 targets in primary tumors.
(a) Summary of the strategy to profile NAB2-STAT6 binding in a primary SFT. The primary tumor was designated and single cells were isolated and fixed. Fixed cells were then used for NAB2 and STAT6 ChIP-seq. Peaks overlapping in both NAB2 and STAT6 ChIP-seq were characterized as NAB2-STAT6 peaks.
(b) Average profiles and heatmaps of NAB2, STAT6, and RNAPII ChIP-seq in a primary SFT at 5921 NAB2 only peaks, 718 NAB2-STAT6 peaks, and 1285 STAT6 only peaks. NAB2-STAT6 peaks had significant NAB2, STAT6, and RNAPII signal.
(c) Top 2 motif from motif analysis of 5921 NAB2 only peaks, 718 NAB2-STAT6 peaks, and 1285 STAT6 only peaks using HOMER shows EGR1 and WT1 as the most significantly enriched TF matrices at NAB2 and NAB2-STAT6 sites and GRE and STAT3 at STAT6 sites.
(d) GSEA shows that genes nearest to NAB2-STAT6 peaks (n = 718) are significantly upregulated in SFTs when compared with matching normal tissue from Fig 1.
(e) Screenshot displays the promoter (highlighted in yellow) of *KLF10* that has significant NAB2, STAT6, and RNAPII localization in SFTs and in U2OS gains NAB2-STAT6 FLAG localization and has increased EGR1 localization and accessibility by ATAC-seq.
(f) Screenshot displays an enhancer and promoter (highlighted in yellow) of *NAB2* that has significant NAB2 and STAT6 localization in SFTs and in U2OS gains NAB2-STAT6 FLAG localization and has increased EGR1 localization and accessibility by ATAC-seq.

Fig. 5

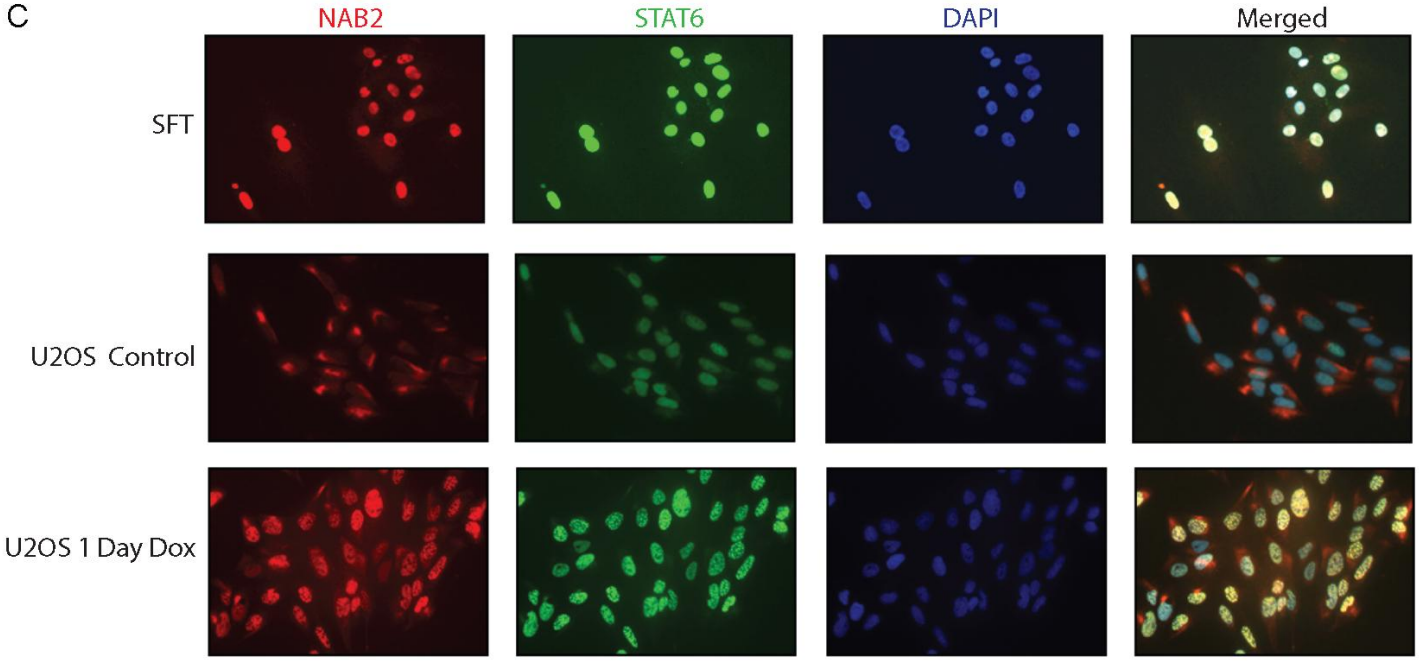
A


U2OS Nuclear Extract (1 day Dox)
STAT6
NAB2
NAB1
EGR1
POLR2E
NCOA1
STAT6 IP (log2 iBAQ)
NAB2 IP (log2 iBAQ)
0
10
20
30

B


Control
1 day Dox
Cyto Nuc Cyto Nuc
NAB2-STAT6 FLAG
NAB2
STAT6
EGR1
GAPDH
HIstone H3

C


NAB2
STAT6
DAPI
Merged
SFT
U2OS Control
U2OS 1 Day Dox

D

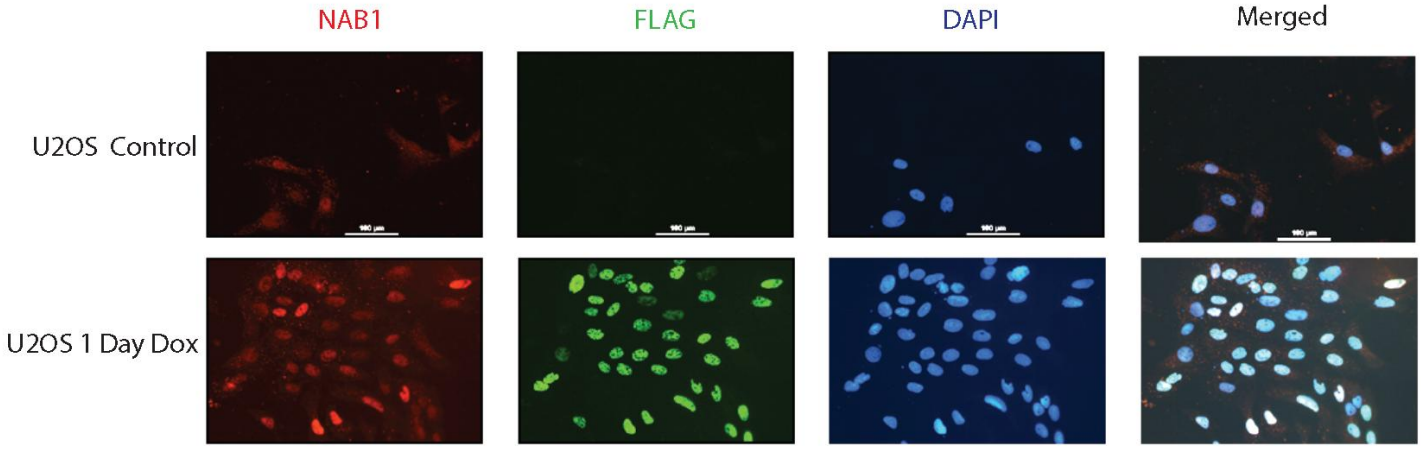

NAB1
FLAG
DAPI
Merged
U2OS Control
U2OS 1 Day Dox

Figure 5. NAB2-STAT6 interacts with EGR1 and NAB1 directs them to the nucleus.

(a) Eluates from NAB2 and STAT6 IPs from U2OS nuclear extracts expressing NAB2-STAT6 for 1 day were subjected to MudPIT LC-MS/MS analysis for unbiased identification of the top interactors. Log2 iBAQ protein scores of STAT6 IP interactors are plotted against scores of NAB IP. NAB1 and EGR1 were the top interactors.

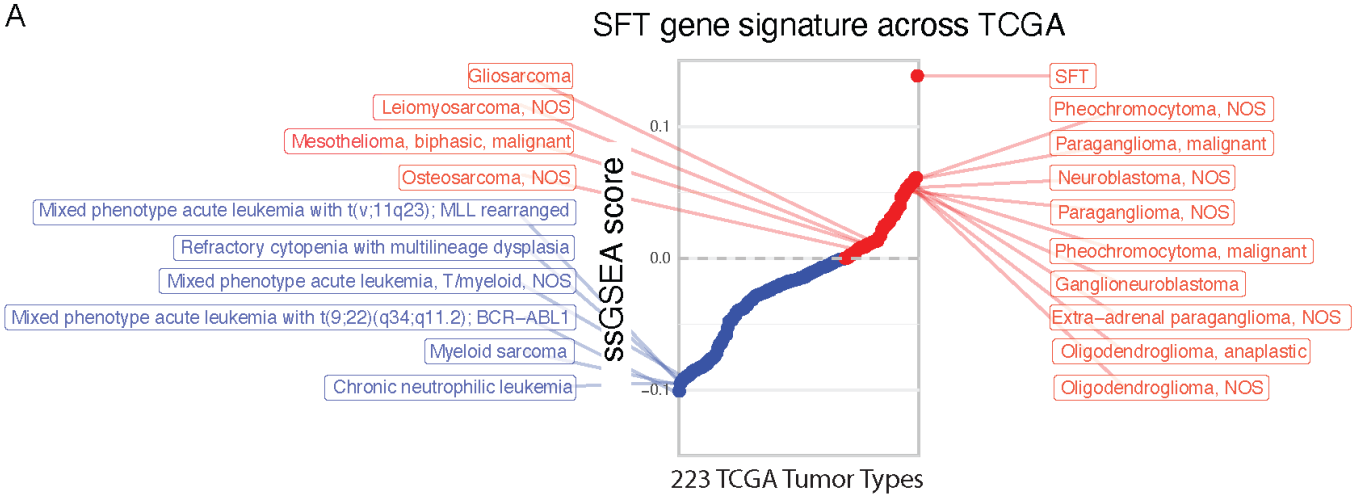
(b) Control and U2OS cells expressing NAB2-STAT6 for 1 day were sub cellularly fractionated into nuclear and cytoplasmic fractions. Immunoblots analysis shows that NAB2-STAT6 was only present in the nuclear fraction of Dox conditions. STAT6 was nuclear in both conditions. NAB2 and EGR1 were cytoplasmic in control conditions but became nuclear in Dox conditions. GAPDH was cytoplasmic control and HIstone H3 was nuclear control

(c) Immunocytochemistry (ICC) of NAB2 (red), STAT6 (green), and DAPI (blue) in SFT primary cells from Fig 4 and U2OS control and NAB2-STAT6 (Dox) expressing for one day cells. SFT and NAB2-STAT6 expressing cells show strong nuclear staining for NAB2 and STAT6. Control U2OS cells have nuclear STAT6 and cytoplasmic NAB2 staining.

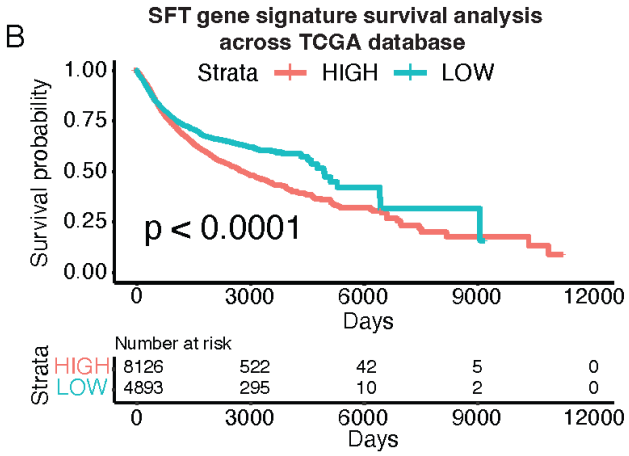
(d) Immunocytochemistry (ICC) of NAB1 (red), FLAG (green), and DAPI (blue) in U2OS control and NAB2-STAT6 (Dox) expressing for one day cells. NAB2-STAT6 expressing cells show strong nuclear staining for FLAG and NAB1. Control U2OS cells have no FLAG and cytoplasmic and nuclear NAB1 staining.

Fig. 6

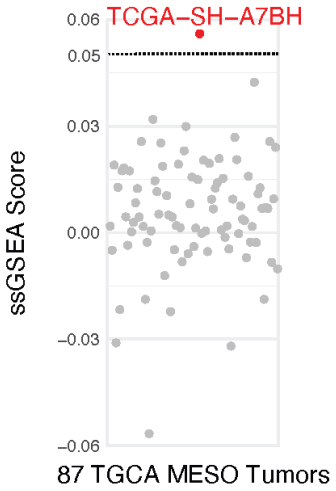
A



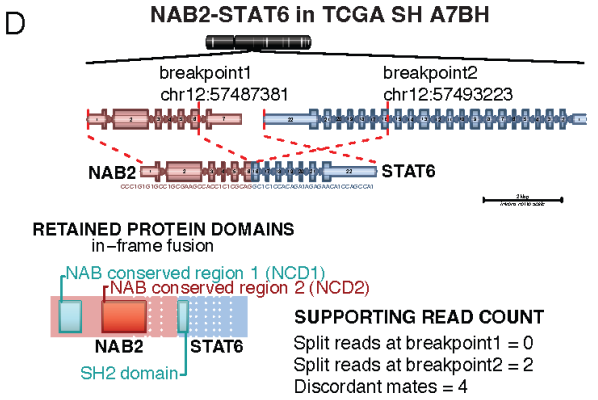
B



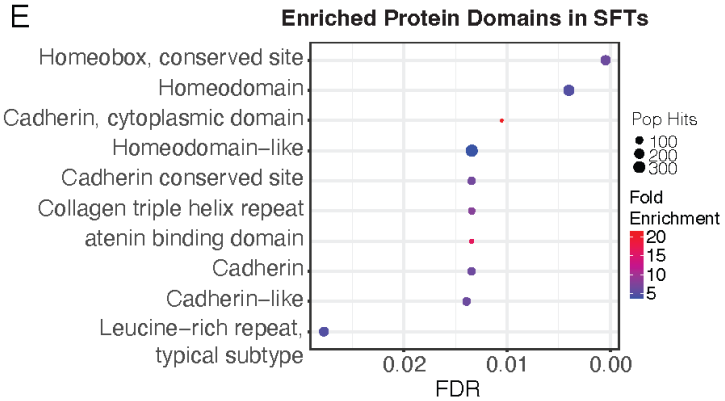
C



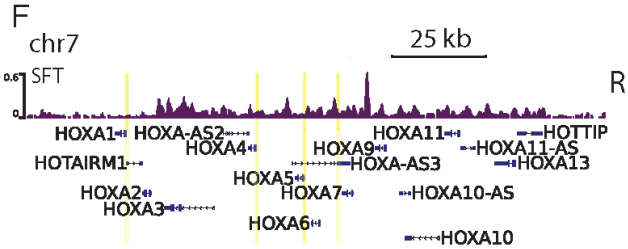
D



E



F



G

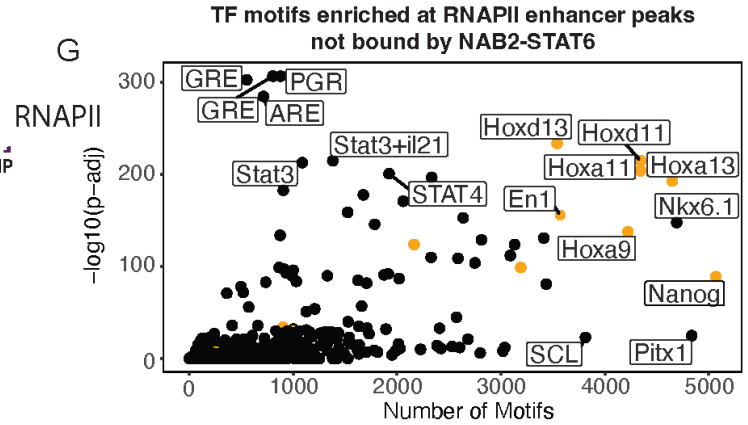


Figure 6. The SFT gene signature resembles EGR1 activated tumors.
(a) Ranking of TCGA tumors by their average single sample gene set enrichment analysis (ssGSEA) score using the SFT gene signature of upregulated genes from Fig 1. (n = 2,429). Neuroendocrine tumors highly express the signature while leukemias down regulate the signature.
(b) Kaplan meier curve showing survival analysis of tumors in the TCGA database stratified by high or low expression of SFT gene signature of upregulated genes from Fig 1 (n = 2,429).
(c) ssGSEA of 87 mesotheliomas from TCGA using SFT gene signature of upregulated genes from Fig 1 (n = 2,429). Shows significant upregulation of the SFT gene signature in the TCGA SH A7BH sample.
(d) NAB2-STAT6 gene fusion in (NAB2 exons 1-6, STAT6 exons 17-22) present in TCGA SH A7BH, originally diagnosed mesothelioma. Graphic generated with Arriba, a fusion detection algorithm.
(e) InterPro domain analysis of the top 400 most upregulated genes in SFTs shows Homeobox and Cadherin as the most significantly enriched protein domains.
(f) Screenshot displays HOXA locus which has significant RNAPII localization in SFTs indicating that the Homeobox genes are actively transcribed in SFTs.
(g) Motif analysis of 11155 distal enhancer (1>kb from nearest TSS) RNAPII peaks using HOMER shows HOX, GRE, and PGR as the most significantly enriched TF matrices.

Fig. 7

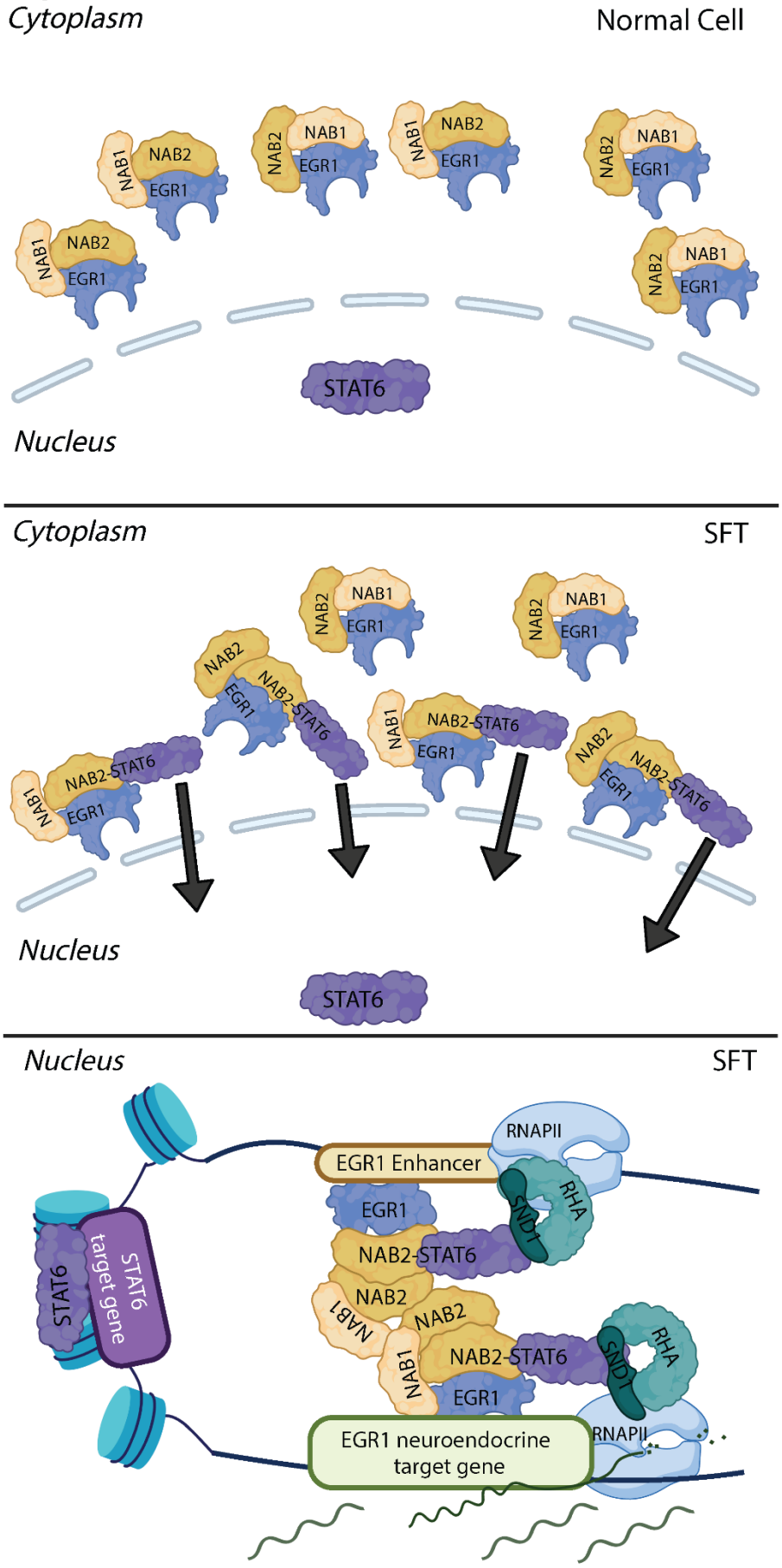

Cytoplasm
Normal Cell
NAB1
NAB2
EGR1
STAT6
Nucleus
Cytoplasm
SFT
NAB2-STAT6
Nucleus
STAT6
Nucleus
SFT
RNAPII
EGR1 Enhancer
RHA
SND1
EGR1
NAB2-STAT6
STAT6
STAT6 target gene
EGR1 neuroendocrine target gene

Figure 7. NAB2-STAT6 drives the expression of EGR1 targets by driving co-activators to the nucleus. Model for NAB2-STAT6's function; NAB2-STAT6 is directed to the nucleus by its STAT6 moiety. The fusion protein then directs co-activators NAB1, NAB2, and EGR1 to the nucleus which in turn direct NAB2-STAT6 to EGR1 target promoters and enhancers which are highly activated by the complex of co-activators recruited.

Methods

FFPE RNA-seq

FFPE sections were obtained from the the Tumor Tissue and Biospecimen Bank at the Hospital of the University of Pennsylvania. Total RNA was purified using the RNeasy DSP FFPE Kit (73604) following manufacturer's manual. Libraries were generated using the TruSeq® RNA Library Prep for Enrichment (20020189) with the TruSeq® RNA Enrichment (20020490), TruSeq RNA Single Indexes Set A (20020492), and the Illumina Exome Panel (20020183). Libraries were sequenced on Illumina NextSeq 2000 with 40 base pair paired-end reads. Reads were aligned to hg19 human reference genome using STAR v2.5. FeatureCounts was used for counting reads to the genes. Data were normalized using Voom and differential gene expression analysis was performed using DEseq2 in R (v1.38.3) unless otherwise noted. Data was visualized using ggplot2 (3.3.6). GO enrichment analysis was done using gprofiler2 package in R (v 0.2.1). Gene set enrichment analysis (GSEA) was 500 randomly selected genes from the select data sets using the clusterProfiler package in R (v4.6.2).

Cell culture

293T were purchased from American Type Culture Collection (ATCC) and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% super calf serum and Glutmax. U2OS cells were obtained from the Lakadamyali lab and maintained in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% tet free FBS or regular FBS and Glutmax. Cells were fed with fresh medium every 3-4 days until reaching 80-90% confluence and split to 1:6-1:12 during passaging.

Lentiviruses packaging

Lentiviruses were produced in HEK293T cells by co-transfection 8 ug of select plasmid with three packaging plasmids (2.5 ug of pRSV-REV, 5 ug of pMDLg/pRRE and 3 ug of pMD2.G per 10 cm cell culture dish) using calcium phosphate transfection(*55*). Lentiviruses were harvested 48 - 72 hours after transfection. Lentivirus was used fresh or pelleted by centrifugation at 16,500 rpm for 90 minutes then pellets were air dried for 30 min and resuspended in 10 ul of PBS, aliquoted and frozen in - 80 ℃.

Cloning

Gene blocks containing three fragments of the long isoform of NAB2-STAT6 (exons 1-4 of NAB2 and exons 2-22 of STAT6) with a c-terminal FLAG tag were ordered from GenScript. Using Gibson Assembly® Master Mix (NEB, cat#E2611), gene blocks were assembled with a digested pLVX-Tight-Puro (ClonTech) or pLENTI-CMV-GFP-Puro (Addgene, cat#17448) vectors. shRNA oligos were ordered form IDT and cloned into digested Tet-pLKO-puro (21915) vector by ligation. All plasmids generated were verified by Sanger sequencing.

Generating tet inducible NAB2-STAT6 system in U2OS cells

U2OS cells were plated into a 6-well plate at 30-40 % confluence in media with tet-free FBS which was used for the rest of the process. When the cells reached 60-70% confluence, 2 ml of medium fresh lentivirus media generated using pLVX-Tet-On Advanced (ClonTech) with 8 ug/ml polybrene (Thermo Fisher, cat#TR1003G)

per well was added to replace the old medium. 24 hours after induction, the virus medium was removed and replaced with fresh cell culture medium for another 48 hours. After that, cells were selected with 200 ug/ml of Neomycin (Corning, cat#MT30234CR) in fresh medium. Cells were selected in Neomycin for two weeks with fresh media being added 3-4 weeks and cells split 1:3 when reaching 80-90% confluency. Then cells were plated into a 6-well plate at 30-40 % confluence. When the cells reached 60-70% confluence, 2 ml of medium fresh lentivirus media generated using pLVX-NAB2-STAT6-FLAG-Tight-Puro with 8 ug/ml polybrene (Thermo Fisher, cat#TR1003G) per well was added to replace the old medium. 24 hours after induction, the virus medium was removed and replaced with fresh cell culture medium for another 48 hours. After that, cells were selected with 0.5 ug/ml of puromycin in fresh medium (InvivoGen, cat#ant-pr-1). 48 hours after selection with puromycin, cells were disassociated with Trypsin and plated at a low density in a 15 cm dish. Single cells were cultured with 0.5 ug/ml of puromycin for the next 2-3 weeks until colonies appeared. Individual microcolonies were moved to a 96-well plate for clonal expansion. Clones were screened after addition of 1ug/mL Doxycline by verified by Western Blot.

Generating constitutively expressing NAB2-STAT6 in HEK293T

pLenti-NAB2-STAT6-FLAG-puro plasmid transfected to HEK293T cells using Lipofectamine 2000 (ThermoFisher, cat#11668030) as the manufacturer's instructions. 24-48 hours after transfection, the old medium was replaced with fresh medium with 400 ug/ml zeocin (InvivoGen, cat#ant-zn-1) for 48 hours to remove non-transfected cells. After that, trypsinized single cells were plated in 15 cm dishes and maintained in fresh medium containing 400 ug/ml zeocin for 2-3 weeks for clonal growth. Expression of NAB2-STAT6 FLAG in microclones was verified by immunoblotting with an M2-Flag antibody

Immunoprecipitation-mass spectrometry (IP-MS)

Nuclear fractions were extracted from cells for IP experiments. Briefly, cells were collected and washed two times with cold PBS. Cell pellets were resuspended in 5 packed volumes (PCV) of Buffer A (10mM HEPES, 5mM MgCl2, 0.25M Sucrose, 0.5mM DTT, and 1mg/ml each of protease inhibitors aprotinin, leupeptin, and pepstatin) then NP-40 was added to 0.1% concentration and cells were mixed again. Resuspended cells were incubated on ice for 10 minutes. Cells were then pelleted at 8000 rpm for 10 minutes at 4 °C. Supernatant was saved as the cytoplasmic fraction. Pellets from previous spin was resuspended in 4 PCV of Buffer B (10mM HEPES, 1.5mM MgCl2, 25% glycerol, 0.1mM EDTA, 0.5mM DTT and 1mg/ml each of protease inhibitors aprotinin, leupeptin, and pepstatin) then NaCl was added to 0.5M concentration and cells were mixed again. The resuspended extract was incubated for 20 minutes on ice. Cells were then sonicated briefly for 6 seconds. The extract was then pelleted at 10000 rpm for 10 minutes at 4 °C. Supernatant was saved as the nuclear fraction. Nuclear and Cytoplasmic fractions were further cleared at 15000 rpm for 30 minutes at 4 °C. All the saved extracts were dialyzed overnight in BC80 (20mM Tris pH 8.0, 80mM KCl, 0.2mM EDTA, 10% glycerol, 1mM B-mercaptoethanol, 0.2mM phenylmethylsulfonyl fluoride (PMSF)) at 4 °C. The extracts were spinned at 20,000 g for 60 minutes at 4 °C on the next day. Supernatant was used for experiments or saved at -80 °C.

For each IP, 2 mg of nuclear extract, 4 ug of antibody, 30 ul of Dynabeads Protein A or G was mixed in co-IP buffer (20mM Tris pH 8.0, 100mM NaCl, 0.1% NP-40, 0.5mM DTT and 1mg/ml each of protease inhibitors aprotinin, leupeptin, and pepstatin) to make a 500 ul volume of reaction. IPs were incubated at 4 °C for 2 hours with rotation and then washed three times with co-IP buffer and one time with PBS with 0.05% NP-40. Proteins were eluted by Igg elution buffer and analyzed by Western blot or LC-MS/MS.

For Flag IP, 2 mg of nuclear extract was incubated with 20 ul of anti-FLAG M2 Magnetic beads and eluted with FLAG peptide.

Western blot

Cells were harvested and washed three times in 1 X PBS. Cell pellets were resuspended and lysed in cold RIPA buffer (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 1% Igepal, 0.5% sodium deoxycholate, 0.1% SDS, 500 µM DTT) supplemented with 1 µg/ml of each of the protease inhibitor aprotinin, leupeptin and pepstatin. 20 µg of cell lysate was loaded in Bolt 4% - 12% Bis-Tris gel (Invitrogen) and separated through gel electrophoresis in 1 X Bolt MES running buffer. Proteins were transferred to ImmunoBlot PVDF membranes in Tris-Glycine buffer (Bio-Rad). Membranes were blocked with 10% BSA in TBST for 35 minutes at room temperature, then incubated with primary antibodies diluted in 5% BSA in 1 X TBST for 2 h at room temperature or overnight at 4 ℃. Membranes were washed three times with TBST for 10 minutes and incubated with HRP- conjugated secondary antibodies for one hour at room temperature. Proteins were detected using Clarity Western ECL substrate (Biorad) and imaged with ImageQuant LAS 4000 (GE healthcare).

RNA extraction

RNA was extracted from cells using TRIzol and purified with the Zymo Directed RNA mimiprep kit (Zymo Research, R2050) following manufacturer's instructions. Briefly, media was removed from the cells and wells were washed with PBS once. 1 ml Accutase was added to each well of a 6 well plate for 5 min. Cell suspension was collected and centrifuged at 300 g for 5 min. 300 ul TRI reagent was added to each pellet and either frozen at - 80 ℃ or proceeded purification following the Zymo Directed protocol. RNA concentration was quantified by Nanodrop.

3' RNA Quant-Seq

Cells were lysed with TRIzol and total RNA was purfiled using Zymo Research Direct-zol RNA mini-prep kit (R2050) following the manufacturer's manual. Libraries were generated using the QuantSeq 3' - mRNA Seq Library Prep Kit for Illumina (Lexogen). 75 base-pair single-end reads were sequenced on the Illumina NextSeq 2000. Reads were aligned to hg19 human reference genome using STAR v2.5. FeatureCounts(*56*) was used for counting reads to the genes. Data were normlized using Voom and differential gene expression analysis was performed using DEseq2 in R (v1.38.3) unless otherwise noted. Data was visualized using ggplot2 (3.3.6). GO enrichment analysis was done using gprofiler2 package in R (v 0.2.1). Gene set enrichment analysis (GSEA) was done with 500 randomly selected genes from the given set of genes across the C2 collection of the human molecular signatures database or custom signatures using the GSEA function in clusterProfiler package in R (v4.6.2).

RNA-Seq

Cells were lysated with TRIzol and total RNA was purfiled using Zymo Research Direct-zol RNA mini-prep kit (R2050) following manufacturer's manual. Libraries were generated using the Illumina Stranded Total RNA Prep with Ribo-Zero Plus (20040525). 40 base-pair single-end reads were sequenced on the Illumina NextSeq 2000. Reads were aligned to hg19 human reference genome using STAR v2.5. FeatureCounts(*56*) was used for counting reads to the genes. Data were normlized using Voom and differential gene expression analysis was performed using DEseq2 in R (v1.38.3) unless otherwise noted. Data was visualized using ggplot2 (3.3.6). GO enrichment analysis was done using gprofiler2 package in R (v 0.2.1). Gene set enrichment analysis (GSEA) was done with 500 randomly selected genes from the given set of genes across the C2 collection of the human molecular signatures database or custom signatures using the GSEA function in clusterProfiler package in R (v4.6.2). Interpro domain analysis was done using DAVID (https://david.ncifcrf.gov/tools.jsp).

Chromatin Immunoprecipitation Sequencing

Cells were resuspended at a concentration of 10 million cells per 10 ml in fresh media at room temperature in 50 ml falcon tubes. For each replicate, 10 - 20 million cells were harvested for cross-linking. The tube was then rotated on a rocker for 15 min at room temperature for 5 min with 1% of formaldehyde (Sigma; Cat#252549). To quench the cross-linking reaction, 560 ul of 2.5 M Glycine per 10 ml media was added and cells were incubated at room temperature with rotation for another 10 min. After being washed twice with cold PBS and spun at 2500 rpm for 10 min at 4℃ then froze at -80 ℃. Cells were then resuspended in ChIP lysis buffer (150 mM NaCl, 1% TritonX-100, 5mM EDTA, 10mM TrisCl, 500 uM DTT, 0.4% SDS) and sonicated to an average length of 200-250 bp using a Covaris S220 Ultrasonicator. Fragmented chromatin was cleared at 13000 rpm for 10 min and diluted with SDS-free ChIP lysis buffer. For each immunoprecipitation, cleared fragmented chromatin was incubated with 5 ug of human antibody, and Protein A or Protein G Dynabeads (Invitrogen) at 4 ℃ overnight. After incubation, beads were washed twice with each of the following buffers: Mixed Micelle Buffer (150 mM NaCl, 1% Triton X-100, 0.2% SDS, 20 mM Tris-HCl (pH 8.0), 5 mM EDTA, 65% sucrose), Buffer 500 (500mM NaCl, 1% Triton X-100, 0.1% Na-deoxycholate, 25mM HEPES, 10mM Tris-HCl (ph 8.0), 1mM EDTA), LiCl/detergent wash buffer (250 mM LiCl, 0.5% Na-deoxycholate, 0.5% NP-40, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA), followed by a final wash with 1X TE. To elute samples, beads were then resuspended with 1X TE supplemented with 1% SDS and incubated at 65 ℃ for 10 min. After eluted twice, samples and the untreated input (5% of the total sheared chromatin) were incubated at 65 ℃ overnight to reverse cross-link. After reverse cross-linking, samples were treated with 0.5 mg/ml proteinase K at 65 ℃ for one hour and purified with Zymo ChIP DNA Clean Concentrator kit (Zymo Research D5205) as the manufacturer's manual and quantified by QUBIT. Barcoded libraries were made using NEB Ultra II DNA Library Prep Kit for Illumina following manufacturer's instructions and quantified by Agilent 2100 Bioanalyzer System. Libraries were sequenced on Illumina NextSeq 2000 with 40 base pair paired-end reads. Sequences were aligned to human reference hg19. Samtools (1.9.0) was used to remove the PCR duplicates (rmdup) and the reads with a mapping quality score of less than 10 from the aligned reads. Bigwig files of the data generated with deeptools (v2.4.2, bamCoverage–binSize 10–normalizeTo1× 3137161264–extendReads 150–ignoreForNormalization chrX) and visualized on the WashU Epigenome Browser (https://epigenomegateway.wustl.edu/browser/) or the USCS Genome Browser (https://genome.ucsc.edu/). For normalization of the data, each number of the filtered reads was divided by the lowest number of the filtered reads in the same set of experiments, generating a downsampling factor for each sample. Normalized BAM files were generated using samtools view -s with the above downsampling factors and further converted to normalized BAM files using bamCoverage–binSize 10–extendReads 150. Peaks were called by MACS2(*57*).

Heatmaps were generated from read depth normalized bigwig files using deeptools ComputeMatrix and visualized with plotHeatmap. Differential binding analysis was done using Diffbind package (v3.4.11). Motif analysis was done using HOMER's (4.10.1) findMotifsGenome command and known motif's were plotted

FLAG Chromatin Immunoprecipitation Sequencing

FLAG ChIP-seq was done as above with the following modifications. M2-FLAG antibody was washed in one volume of PBST (1X PBS + 0.01% tween) and then pre-bound with protein G Dynabeads for at least 4 hours. Fixed cells were then resuspended in Sonication buffer (1/80 volume of 20% Sarkosyl (FC 0.25%), 1mM DTT, and protease inhibitors into RIPA Buffer 3.0 (0.1% SDS, 1% Triton X-100, 10mM Tris-HCl (pH 7.4), 1mM EDTA, 0.1% NaDOC, 0.3M NaCl)). After incubation, beads were washed twice with each of the following buffers: Low Salt (150 mM NaCl, 1% Triton X-100, 0.1% SDS, 50 mM Tris-HCl (pH 8.0), 5 mM EDTA, 65% sucrose), High Salt (500 mM NaCl, 1% Triton X-100, 0.1% Na-deoxycholate, 25 mM HEPES, 10 mM Tris-HCl, 1 mM EDTA), LiCl/detergent wash buffer (150 mM LiCl, 0.5% Na-deoxycholate, 1% NP-40, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA, 0.1% SDS), followed by a final wash with 1X TE with 50mM NaCl. To elute samples,

beads were then resuspended in 210uL Elution Buffer (1% SDS, 50mM Tris-HCL (pH 8.0), 10mM EDTA, 200mM NaCl) and incubated at 65 ℃ for 30 min shaking at 1200rpm then spun down for 1 min at 16,000g at RT. The supernatant was then collected then samples and the untreated input (5% of the total sheared chromatin) were incubated at 65 ℃ overnight to reverse cross-link.

Omni-ATAC-seq

Omni ATAC-seq was performed according to Corces et al 2017. Briefly, 50,000 cells were washed in ATAC-resuspension buffer (RSB) with 0.1% Tween-20 and then lysed by incubating on ice for 3 minutes in RSB with 0.1% NP-40 and 0.1% Tween-20. Lysis was washed out in RSB 0.1% Tween-20 and nuclei were pelleted by centrifugation at 600g/4°C/5 minutes. Nuclei were then incubated in the transposition mixture for 30 minutes at 37°C while shaking at 1000rpm. DNA was then purified using Zymo DNA clean and concentrator-5 Kit (cat# D4014). DNA was then amplified for 5 cycles with ATAC index primers and NEBNext Ultra II Q5 Master Mix (NEB #M0544). Additional amplification cycles were then determined by qPCR using 5uL of original amplification with PerfeCTa SYBR Green FastMix Reaction Mixes (Quantabio 95072-012). After additional amplification cycles with remaining 15uL DNA was then purified using Zymo DNA clean and concentrator-5 Kit (cat# D4014) and quantified by QUBIT.

Immunocytochemistry (ICC)

Immunofluorescence experiments were performed as described[25]. Briefly, cells were fixed with 4% formaldehyde for 15 min at room temperature, washed three times with PBS for 15 min and incubated with 1% goat serum in PBST containing 0.1% Triton X-100 for 30 min at room temperature. Cells were incubated with primary antibodies at 4℃ for overnight, washed three times with PBS for 10 min at room temperature and incubated with secondary antibodies for 1 hour at room temperature. After that, cells were incubated with 1 ug/ml of DAPI for 15 min and mounted in SlowFade™ Gold Antifade Mountant (ThermoFisher Scientific, Cat#S36938) and imaged using Nikon 80i Upright Microscope.

Public data processing

Gene expression counts from TCGA were downloaded form the GDC data portal (https://portal.gdc.cancer.gov/repository?files_offset=22700&files_size=100&filters=%7B%22op%22%3A%22and%22%2C%22content%22%3A%5B%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22files.data_format%22%2C%22value%22%3A%5B%22tsv%22%5D%7D%7D%2C%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22files.data_type%22%2C%22value%22%3A%5B%22Gene%20Expression%20Quantification%22%5D%7D%7D%2C%7B%22op%22%3A%22in%22%2C%22content%22%3A%7B%22field%22%3A%22files.experimental_strategy%22%2C%22value%22%3A%5B%22RNA-Seq%22%5D%7D%7D%5D%7D). Publically available SFT RNA-seq was downloaded from the database of Genotypes and Phenotypes (dbGaP) under accession phs000567.v1.p1. Single sample gene set enrichment analysis (ssGSEA) was done using using the GSVA (1.48.1) package in R. 500 randomly selected genes from curated data sets and the TCGA, SFT FFPE RNA-seq samples, and phs000567.v1.p1 SFT RNA-seq. Survival Analysis was done using the Survival (3.5-5) package in R. Spearman correlation of SFTs and human tissues and cells was performed on the Human Protein Atlas RNA-seq datasets (https://www.proteinatlas.org/about/download#:~:text=rna_tissue_consensus.tsv.zip and https://www.proteinatlas.org/about/download#:~:text=rna_single_cell_type.tsv.zip**).**

Key Resources Table

Antibodies

Antibodies	Dilutions	Source	Identifier
NAB2	WB, 1:1000, IF, 1:500, ChIP	ThermoFisher	Cat#PA5-27925
STAT6	WB, 1:1000, IF, 1:500, ChIP	Santa Cruz Biotechnology	Cat#sc-374021
EGR1	WB, 1:1000, ChIP	Bethyl	Cat#A303-390A
Rabbit polyclonal anti-GAPDH	WB, 1:10,000	Cell Signaling	Cat#2118
Mouse monoclonal anti-FLAG	WB, 1:5000, IF 1:500, ChIP	Sigma	Cat#F1804
Rabbit polyclonal anti-H3K27ac	ChIP	Abcam	Cat#ab4279
Rabbit polyclonal anti-H3K4me1	ChIP	Abcam	Cat#ab8895
Rabbit polyclonal anti-NAB1	IF, 1:500	Protein Tech	18541-1-AP
Rabbit polyclonal anti-RNAPII	ChIP	Barbieri et at., 2018	
Anti-rabbit IgG (HRP)	WB, 1:10,000	Cell Signaling	Cat#7074
Anti-mouse IgG (HRP)	WB, 1:10,000	Cell Signaling	Cat#7076
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 594	IF, 1:500	ThermoFisher Scientific	Cat#A-11012
Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 488	IF, 1:500	ThermoFisher Scientific	Cat#A-11008

Reagents	Source	Indentifier
Dynabeads Protein A	Invitrogen	Cat#10002D
Dynabeads Protein G	Invitrogen	Cat#10004D
Igg elution buffer	G-Biosciences	Cat#786-545
Anti-FLAG M2 Magnetic Beads	Sigma-Aldrich	Cat#M8823
FLAG peptide	Sigma-Aldrich	Cat#F3299
Gibson Assembly Master Mix	New Englands Biolabs	Cat#E2611S
Phusion High-fidelity DNA polymerase kit	New Englands Biolabs	Cat#M0530S
Direct-zol RNA Miniprep kit	Zymo Research	Cat#R2051
Rvertaid first strand cDNA synthesis kit	ThermoFisher Scientific	Cat#K1622
ChIP DNA Clean&Concentrator kit	Zymo Research	Cat#D5201
NEBNext Ultra II DNA Library Prep Kit for Illumina	New Englands Biolabs	Cat#E7645S
TruSeq® RNA Library Prep for Enrichment TruSeq®	Illumina	Cat#20020189
TruSeq® RNA Enrichment	Illumina	Cat#20020490
TruSeq RNA Single Indexes Set A	Illumina	Cat#20020492
Illumina Exome Panel	Illumina	Cat#20020183
NEBNext Multiplex Oligos for Illumina (Index Primers set 1)	New Englands Biolabs	Cat#E6310S
RNeasy DSP FFPE Kit	Qiagen	Cat#73604
Illumina® Stranded Total RNA Prep, Ligation with Ribo-Zero Plus	Illumina	Cat#20040525
Antifade Mountant	ThermoFisher Scientific	Cat#S36938

Deposited Data

SuperSeries	This paper	GEO:
ChIP-seq data	This paper	GEO:
RNA-seq data	This paper	GEO:
ATAC-seq	This paper	GEO:

Cell Lines

Human:293T cells	ATCC	Cat#CRL-3216
Human: U2OS	Lakadamyali Lab	

Cell culture

Fetal bovine serum	Corning	Cat#35010CV
Fetal Bovine Serum, Tet system approved, USDA-approved regions	ThermoFisher Scientific	Cat#A4736401
L-glutamine	Corning	Cat#5031C
Doxycycline hyclate	Sigma-Aldrich	Cat#D9891

Primers

Name	Sequence	Application
pINTO_INTS13-F	AACCGGTAGCGG TAC C AAGATTTTTTCTGA ATCT	Gibson assembly cloning
pINTO_INTS13-R	ATCCGAGCTCGGT ACCTCACTGCCGG CTGGCTTT	Gibson assembly cloning
pINTO_INTS14-F	AACCGGTAGCGG TAC C CCGACAGTGGTG GTAATG	Gibson assembly cloning
pINTO_INTS14-R	ATCCGAGCTCGGT ACCTCAAATTCTTT	Gibson assembly cloning

	CAGTGCT	
pINTO_INTS10-F	AACCGGTAGCGG TAC C TCTGCCCAGGGG GACTGC	Gibson assembly cloning
pINTO_INTS10-R	ATCCGAGCTCGGT ACC TCAGGTCAGAGTC TGAAG	Gibson assembly cloning
INTS10-F	TATACCAGTGGCC TCTTGTC	PCR and verification of knockout clones
INTS10-R	CGTCTTCCTTTAT CCATCTGCC	PCR and verification of knockout clones

Recombinant DNA

pLVX-Tight-Puro	Clontech	Cat#632162
pLVX-Tet-On Advanced	Clontech	Cat#632162
pLENTI-CMV-GFP-Puro	Addgene	Cat#17448
Tet-pLKO-puro	Addgene	Cat#21915
pLVX-NAB2-STAT6-FLAG-Tight-Puro	This paper	
pLenti-NAB2-STAT6-FLAG-puro	This paper	

Software and Algorithms

STAR (2.5)	(*58*)
Samtools (0.1.19)	(*59*)
FeatureCounts	(*56*)
DESeq2 (1.38.3)	(*60*)
MACS2	(*57*)
bedtools (2.26.0)	(*61*)

ggplot2	(*62*)
deeptools (3.5.1)	(*63*)
DiffBind (3.4.11)	(*64*)
gprofiler2 (0.2.1)	(*65*)
Survival (3.5-5)	(*66*)
GSVA (1.48.1)	(*67*)
HOMER (4.10.1)	(*68*)

Acknowledgements

This study was supported by grants from the G. Harold and Leila Y. Mathers Charitable Foundation (A.G.) and the NIH (R01 HL141326 and R01 CA252223). C.M.H. was also supported by a Ruth L. Kirschstein NRSA Award F31 CA265257.

We thank the Sarma lab for reagents and useful discussions. We thank the Tumor Tissue and Biospecimen Bank and We also thank the Genomics Core, the Proteomics & Metabolomics Core, and the Imaging Facility of The Wistar Institute (P30-CA010815) for providing outstanding technical support. Primary SFT tissue was collected in the context of an IRB approved protocol (Penn IRB #843351). We thank the University of Pennsylvania's Tumor Tissue and Biospecimen Bank (TTAB) for help providing primary samples. Illustrations were generated with BioRender.com.

References

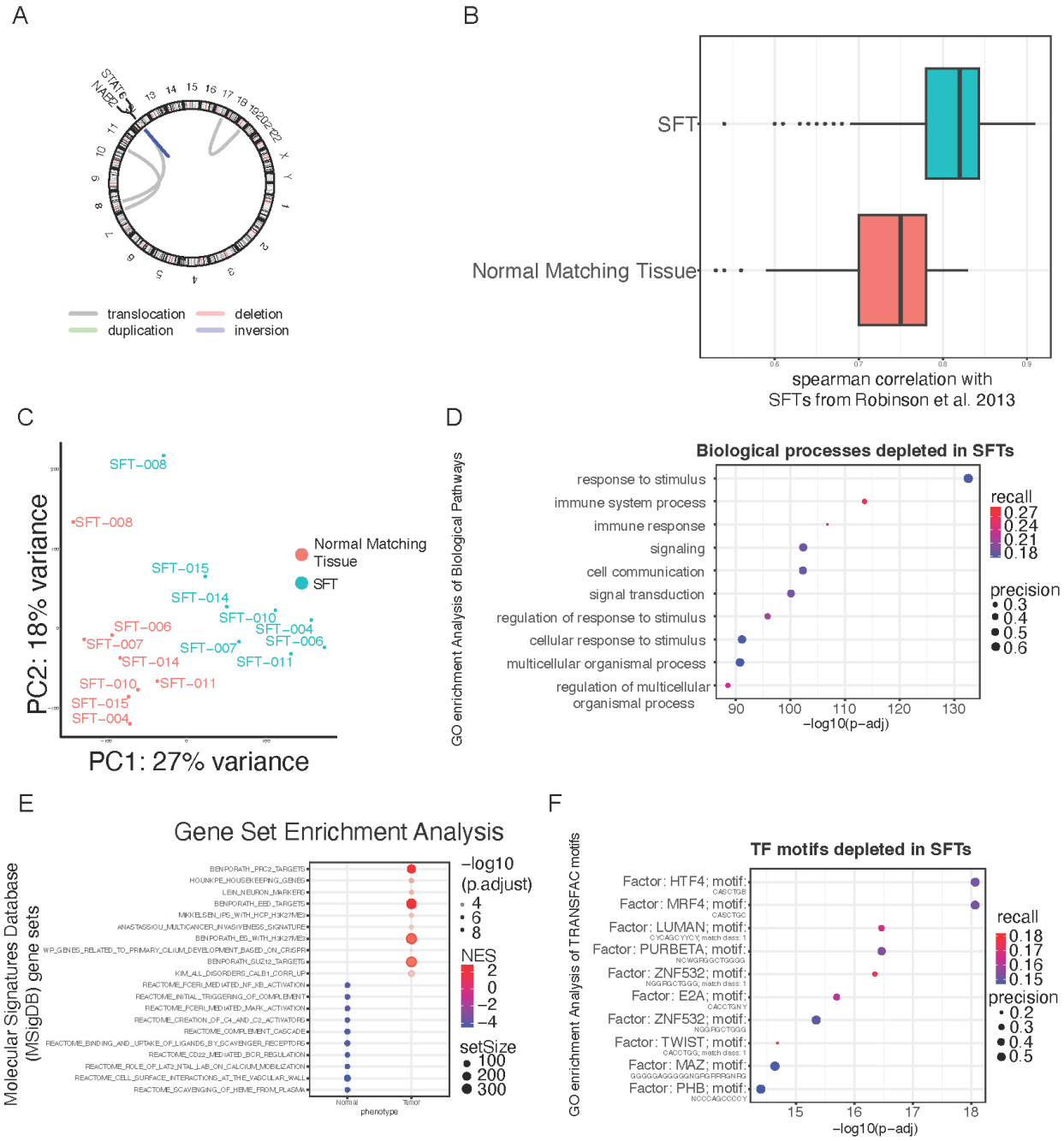
1. C. A. French *et al.*, BRD4-NUT fusion oncogene: a novel mechanism in aggressive carcinoma. *Cancer research* **63**, 304-307 (2003).
2. C. Kadoch, G. R. Crabtree, Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* **153**, 71-85 (2013).
3. A. Bansal, S. Goyal, A. Goyal, M. Jana, WHO classification of soft tissue tumours 2020: An update and simplified approach for radiologists. *European Journal of Radiology*, 109937 (2021).
4. C. Gengler, L. Guillou, Solitary fibrous tumour and haemangiopericytoma: evolution of a concept. *Histopathology* **48**, 63-74 (2006).
5. K. Kazazian, E. G. Demicco, M. de Perrot, D. Strauss, C. J. Swallow, Toward Better Understanding and Management of Solitary Fibrous Tumor. *Surg Oncol Clin N Am* **31**, 459-483 (2022).
6. D. R. Robinson *et al.*, Identification of recurrent NAB2-STAT6 gene fusions in solitary fibrous tumor by integrative sequencing. *Nature Genetics* **45**, 180-185 (2013).

7. J. Chmielecki *et al.*, Whole-exome sequencing identifies a recurrent NAB2 - STAT6 fusion in solitary fibrous tumors. *Nature Genetics* **45**, 131-132 (2013).
8. L. Schweizer *et al.*, Meningeal hemangiopericytoma and solitary fibrous tumors carry the NAB2-STAT6 fusion and can be diagnosed by nuclear expression of STAT6 protein. *Acta Neuropathol* **125**, 651-658 (2013).
9. G. Vahedi *et al.*, STATs Shape the Active Enhancer Landscape of T Cell Populations. *Cell* **151**, 981-993 (2012).
10. Z. Czimmerer *et al.*, Extensive and functional overlap of the STAT6 and RXR cistromes in the active enhancer repertoire of human CD14+ monocyte derived differentiating macrophages. *Molecular and Cellular Endocrinology* **471**, 63-74 (2018).
11. K. Shimoda *et al.*, Lack of IL-4-induced Th2 response and IgE class switching in mice with disrupted State6 gene. *Nature* **380**, 630-633 (1996).
12. J. Hou *et al.*, An interleukin-4-induced transcription factor: IL-4 Stat. *Science* **265**, 1701-1706 (1994).
13. K. Takeda, M. Kamanaka, T. Tanaka, T. Kishimoto, S. Akira, Impaired IL-13-mediated functions of macrophages in STAT6-deficient mice. *The Journal of Immunology* **157**, 3220-3222 (1996).
14. C. M. Litterst, E. Pfitzner, Transcriptional Activation by STAT6 Requires the Direct Interaction with NCoA-1. *Journal of Biological Chemistry* **276**, 45713-45721 (2001).
15. X. Wang *et al.*, Coactivator P100 protein enhances STAT6-dependent transcriptional activation but has no effect on STAT1-mediated gene transcription. *Anat Rec (Hoboken)* **293**, 1010-1016 (2010).
16. J. Yang *et al.*, Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. *The EMBO Journal* **21**, 4950-4958 (2002).
17. T. Valineva, J. Yang, R. Palovuori, O. Silvennoinen, The transcriptional co-activator protein p100 recruits histone acetyltransferase activity to STAT6 and mediates interaction between the CREB-binding protein and STAT6. *J Biol Chem* **280**, 14989-14996 (2005).
18. J. Svaren *et al.*, NAB2, a corepressor of NGFI-A (Egr-1) and Krox20, is induced by proliferative and differentiative stimuli. *Molecular and Cellular Biology* **16**, 3545-3553 (1996).
19. I. C. Weaver *et al.*, The transcription factor nerve growth factor-inducible protein a mediates epigenetic programming: altering epigenetic marks by immediate-early genes. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **27**, 1756-1768 (2007).
20. D. L. Coleman, A. H. Bartiss, V. P. Sukhatme, J. Liu, H. D. Rupprecht, Lipopolysaccharide induces Egr-1 mRNA and protein in murine peritoneal macrophages. *J Immunol* **149**, 3045-3051 (1992).
21. S. Healy, P. Khan, J. R. Davie, Immediate early response genes and cell transformation. *Pharmacology & therapeutics* **137**, 64-77 (2013).
22. F. T. Gallo, C. Katche, J. F. Morici, J. H. Medina, N. V. Weisstaub, Immediate early genes, memory and psychiatric disorders: focus on c-Fos, Egr1 and Arc. *Front Behav Neurosci* **12**, 79 (2018).
23. G. Thiel, G. Cibelli, Regulation of life and death by the zinc finger transcription factor Egr-1. *Journal of cellular physiology* **193**, 287-292 (2002).
24. F. Giudicelli, E. Taillebourg, P. Charnay, P. Gilardi-Hebenstreit, Krox-20 patterns the hindbrain through both cell-autonomous and non cell-autonomous mechanisms. *Genes Dev* **15**, 567-580 (2001).
25. R. Poirier *et al.*, Distinct functions of egr gene family members in cognitive processes. *Front Neurosci* **2**, 47-55 (2008).
26. E. Thierion *et al.*, Krox20 hindbrain regulation incorporates multiple modes of cooperation between cis-acting elements. *PLoS genetics* **13**, e1006903 (2017).
27. Prognostic role of EGR1 in breast cancer: a systematic review - PubMed.
28. B. Wang *et al.*, The Role of the Transcription Factor EGR1 in Cancer. *Frontiers in Oncology* **11**, (2021).
29. S.-K. Bae *et al.*, Egr-1 mediates transcriptional activation of IGF-II gene in response to hypoxia. *Cancer research* **59**, 5989-5994 (1999).
30. I. de Belle *et al.*, p53 and Egr-1 additively suppress transformed growth in HT1080 cells but Egr-1 counteracts p53-dependent apoptosis. *Oncogene* **18**, 3633-3642 (1999).
31. J. Svaren *et al.*, Novel mutants of NAB corepressors enhance activation by Egr transactivators. *The EMBO Journal* **17**, 6010-6019 (1998).
32. E. Barbieri *et al.*, Targeted Enhancer Activation by a Subunit of the Integrator Complex. *Mol Cell* **71**, 103-116 e107 (2018).

33. M. Bieg *et al.*, Gene expression in Solitary Fibrous Tumors (SFTs) correlates with anatomical localization and NAB2-STAT6 gene fusion variants. *The American Journal of Pathology*, S0002944021000365 (2021).
34. H. K. Park *et al.*, Molecular changes in solitary fibrous tumor progression. *J Mol Med (Berl)* **97**, 1413-1425 (2019).
35. H.-C. Tai *et al.*, NAB2–STAT6 fusion types account for clinicopathological variations in solitary fibrous tumors. *Modern Pathology* **28**, 1324-1335 (2015).
36. S. Yuzawa *et al.*, Analysis of NAB2-STAT6 Gene Fusion in 17 Cases of Meningeal Solitary Fibrous Tumor/Hemangiopericytoma: Review of the Literature. *Am J Surg Pathol* **40**, 1031-1040 (2016).
37. I. C. Chuang *et al.*, NAB2-STAT6 gene fusion and STAT6 immunoexpression in extrathoracic solitary fibrous tumors: the association between fusion variants and locations. *Pathology International* **66**, 288-296 (2016).
38. E. G. Demicco *et al.*, Solitary fibrous tumor: a clinicopathological study of 110 cases and proposed risk assessment model. *Modern Pathology* **25**, 1298-1306 (2012).
39. Y.-S. Park *et al.*, NAB2-STAT6 fusion protein mediates cell proliferation and oncogenic progression via EGR-1 regulation. *Biochemical and Biophysical Research Communications* **526**, 287-292 (2020).
40. Y. Li *et al.*, Reduction of Tumor Growth with RNA-Targeting Treatment of the NAB2–STAT6 Fusion Transcript in Solitary Fibrous Tumor Models. *Cancers* **15**, 3127 (2023).
41. T. Georgiesh *et al.*, Clinical and molecular implications of NAB2-STAT6 fusion variants in solitary fibrous tumour. *Pathology* **53**, 713-719 (2021).
42. V. Y. Jo, E. G. Demicco, Update from the 5th Edition of the World Health Organization Classification of Head and Neck Tumors: Soft Tissue Tumors. *Head and Neck Pathol*, (2022).
43. K. Kazazian, E. G. Demicco, M. de Perrot, D. Strauss, C. J. Swallow, Toward Better Understanding and Management of Solitary Fibrous Tumor. *Surgical Oncology Clinics of North America* **31**, 459-483 (2022).
44. S. Vejvodova *et al.*, Solitary fibrous tumor–Less common neoplasms of the pleural cavity. *Annals of Thoracic and Cardiovascular Surgery* **23**, 12-18 (2017).
45. Q. Gao *et al.*, Driver fusions and their implications in the development and treatment of human cancers. *Cell reports* **23**, 227-238. e223 (2018).
46. F. Mertens, B. Johansson, T. Fioretos, F. Mitelman, The emerging complexity of gene fusions in cancer. *Nature Reviews Cancer* **15**, 371-381 (2015).
47. C. Shen, C. R. Vakoc, Gain-of-function mutation of chromatin regulators as a tumorigenic mechanism and an opportunity for therapeutic intervention. *Curr Opin Oncol* **27**, 57-63 (2015).
48. J. H. Bushweller, Targeting transcription factors in cancer—from undruggable to reality. *Nature Reviews Cancer* **19**, 611-624 (2019).
49. N. S. Latysheva, M. M. Babu, Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic acids research* **44**, 4487-4503 (2016).
50. J. Chmielecki *et al.*, Whole-exome sequencing identifies a recurrent NAB2-STAT6 fusion in solitary fibrous tumors. *Nat Genet* **45**, 131-132 (2013).
51. J. Kumbrink, M. Gerlinger, J. P. Johnson, Egr-1 Induces the Expression of Its Corepressor Nab2 by Activation of the Nab2 Promoter Thereby Establishing a Negative Feedback Loop. *Journal of Biological Chemistry* **280**, 42785-42793 (2005).
52. C.-T. Chen, J.-C. Lee, S.-C. Huang, T.-W. Chen, Dacarbazine and bevacizumab improved paraneoplastic doege–potter syndrome of malignant solitary fibrous tumor. *Journal of Cancer Research and Practice* **8**, 68 (2021).
53. A. Belfiore *et al.*, IGF2: a role in metastasis and tumor evasion from immune surveillance? *Biomedicines* **11**, 229 (2023).
54. D. B. Brotto *et al.*, Contributions of HOX genes to cancer hallmarks: Enrichment pathway analysis and review. *Tumor Biology* **42**, 1010428320918050 (2020).
55. C. Chen, H. Okayama, High-efficiency transformation of mammalian cells by plasmid DNA. *Mol Cell Biol* **7**, 2745-2752 (1987).
56. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
57. Y. Zhang *et al.*, Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137 (2008).

58. A. Dobin *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
59. P. Danecek *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, (2021).
60. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
61. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842 (2010).
62. N. E. Hamilton, M. Ferry, ggtern: Ternary Diagrams Using ggplot2. *J Stat Softw* **87**, 1-17 (2018).
63. F. Ramirez *et al.*, deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160-165 (2016).
64. R. Stark, G. Brown, DiffBind: differential binding analysis of ChIP-Seq peak data. *R package version* **100**, (2011).
65. L. Kolberg, U. Raudvere, I. Kuzmin, J. Vilo, H. Peterson, gprofiler2 -- an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, (2020).
66. T. M. Therneau, T. Lumley, A package for survival analysis in R. *New York, NY, USA*, (2019).
67. S. Hänzelmann, R. Castelo, J. Guinney, GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC bioinformatics* **14**, 1-15 (2013).
68. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589 (2010).

Supplementary Fig. 1



Supplementary Figure 1. Solitary Fibrous Tumors express a neuronal gene signature.

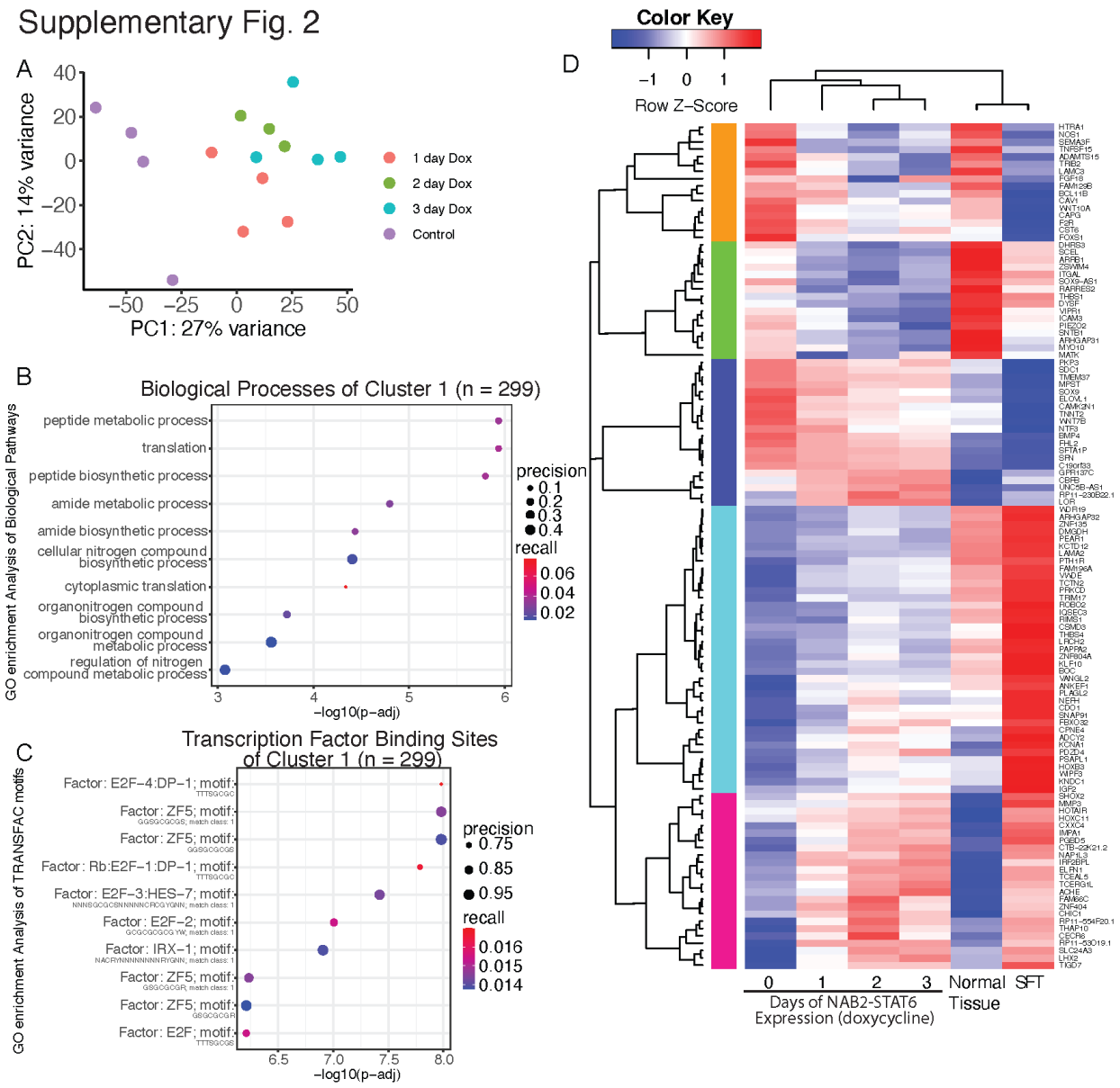
(a) Chromosomal Rearrangements present in SFT 004 detected by Arrriba. The inversion producing NAB2-STAT6 was the most abundant.

(b) Spearman's correlation analysis of gene expression between 26 previously published RNA-seqs of SFTs (Robinson et al. 2013) and our 8 SFTs and Normal Matching Tissues from the FFPE RNA.

(c) Principal Component Analysis (PCA) of FFPE RNA-seq datasets of SFTs and Normal Matching Tissue. All SFTs cluster away from Normal matching tissue along the PC1 axis (27% of variance), suggesting that broad and consistent transcriptome changes in SFTs.

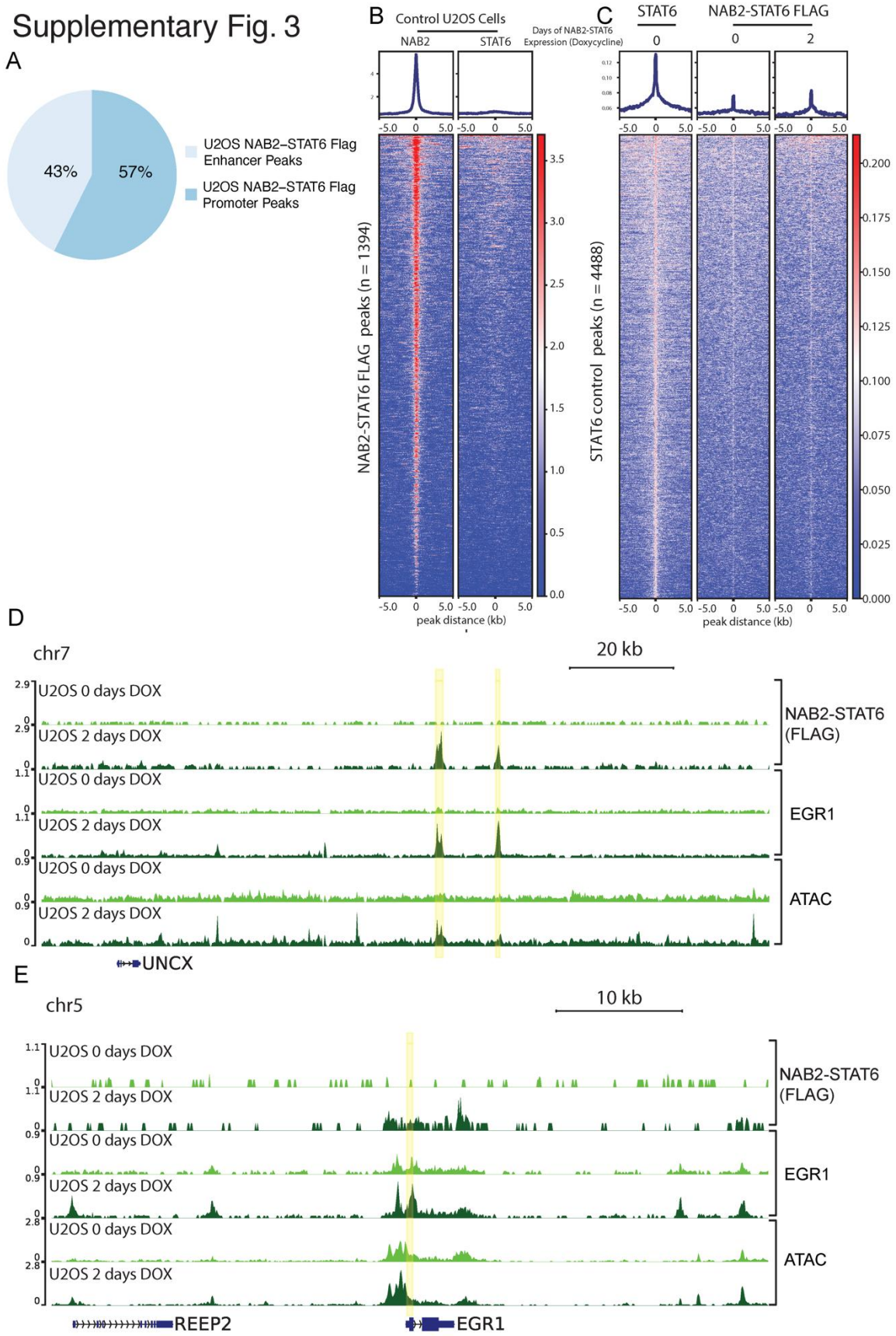
(d) Biological pathway GO analysis of 3,769 upregulated genes in INTS10 KO cells revealed enrichment for immune and cell signaling pathways.

(e) Dot plot of GSEA analysis shows the upregulation of Neuronal signatures in SFT and downregulation of immune signatures.
(f) TRANSFAC motif (transcription factor motifs at +/-1kb from transcription start site) GO analysis of 3,769 upregulated genes in INTS10 KO cells revealed enrichment for HTF4 and MRF4, transcription factors active in mesoderm derived tissues.



Supplementary Figure 2. Generation of an inducible NAB2-STAT6 system to investigate early transcriptional changes.
(a) Principal Component Analysis (PCA) of 3' mRNA Quant-seq (n = 4) datasets of Dox treated (1, 2, or 3 days) and control cells. The longer NAB2-STAT6 is expressed the more cells (Dox) move further along the PC1 axis (27% of variance), suggesting NAB2-STAT6 induces broad and consistent transcriptome changes in U2OS cells.

(b) Biological pathway GO analysis of 299 upregulated genes in cluster 1 of heatmap from Fig 2b. revealed enrichment for translation and biosynthetic pathways.
(c) TRANSFAC motif (transcription factor motifs at +/-1kb from transcription start site) GO analysis 299 upregulated genes in cluster 1 of heatmap from Fig 2b. revealed enrichment for DP-1 and ZF5 motifs.
(d) Heatmap clustering analysis of 166 genes that are differentially expressed (fold change >1, FDR <0.1) across 1, 2, and 3 days of NAB2-STAT6 expression (Dox) and SFTs as determined by 3' mRNA Quant-seq (n = 4). Includes EGR1 targets like IGF2, LHX2, and ROBO2



Supplementary Figure 3. EGR1 targeted promoters and enhancers are activated by NAB2-STAT6.
(a) Pie chart showing the percentage of enhancer (1kb > from nearest TSS) and promoter (1 kb < from nearest TSS) sites from NAB2-STAT6 Flag Peaks (n = 1394).

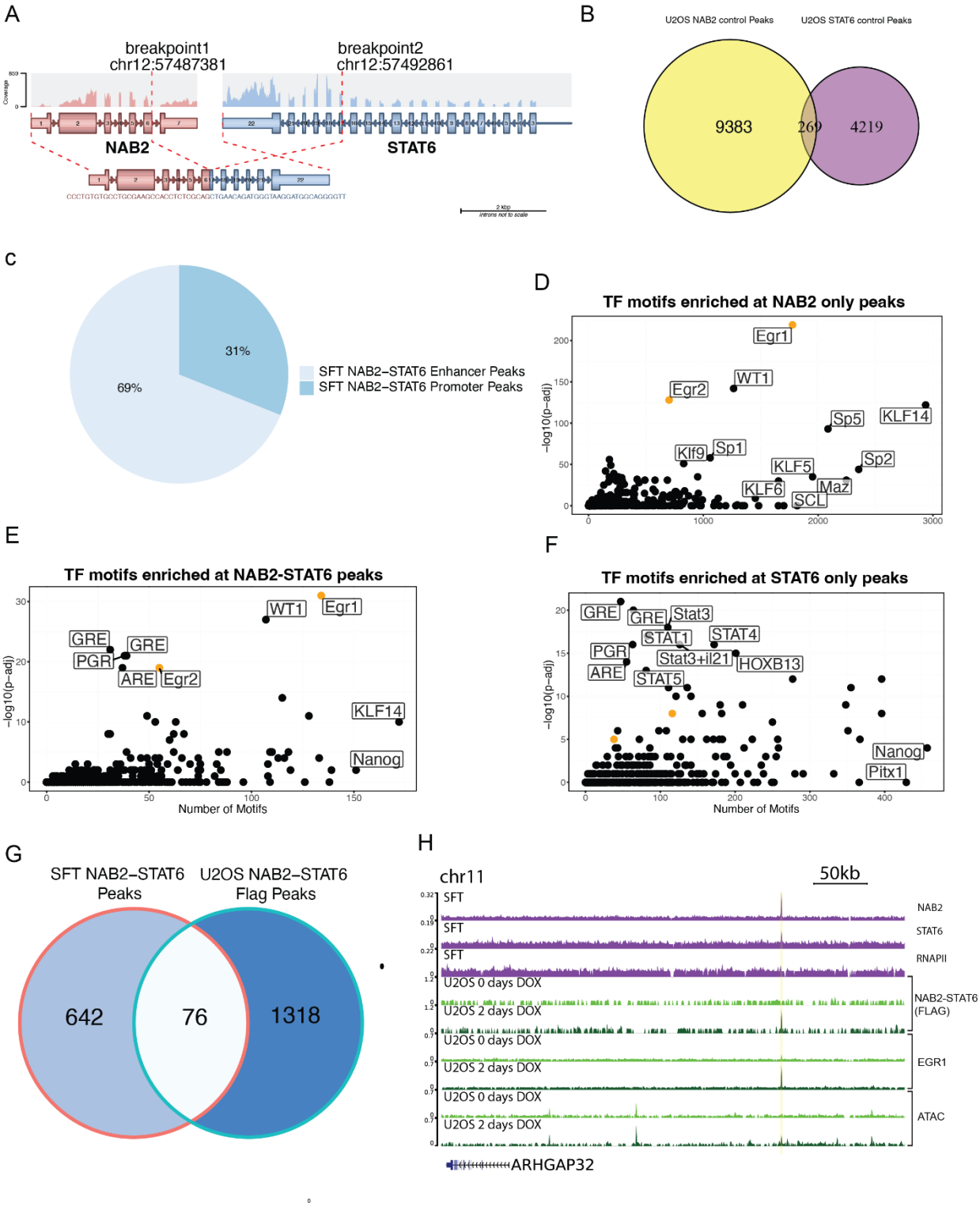
(b) Average profiles and heatmaps of NAB2 and STAT6 in both control U2OS cells at 1394 NAB2-STAT6 FLAG peaks. There is significant NAB2 but no STAT6 localization.

(c) Average profiles and heatmaps of STAT6 in control U2OS cells and NAB2-STAT6 FLAG in control and 2 days NAB2-STAT6 (Dox) expressing U2OS cells at 4488 STAT6 control peaks. There is significant STAT6 but limited NAB2-STAT6 FLAG localization

(d) Screenshot displays two enhancers (highlighted in yellow) of *UNCX* that gains NAB2-STAT6 FLAG localization and have increases in EGR1 localization and accessibility by ATAC-seq.

(e) Screenshot displays the promoter (highlighted in yellow) of *EGR1* that gains NAB2-STAT6 FLAG localization and has increased in EGR1 localization and accessibility by ATAC-seq.

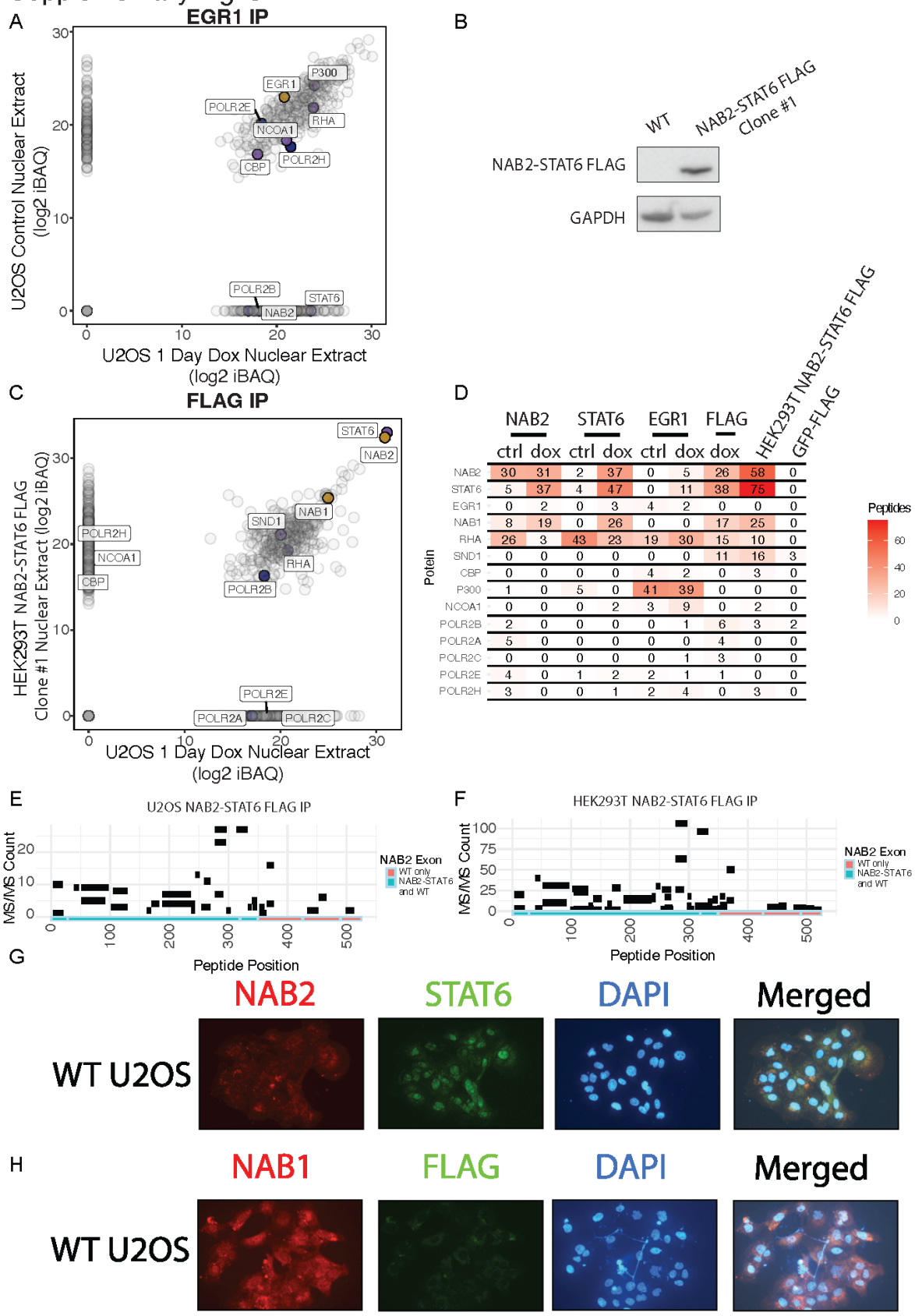
Supplementary Fig. 4



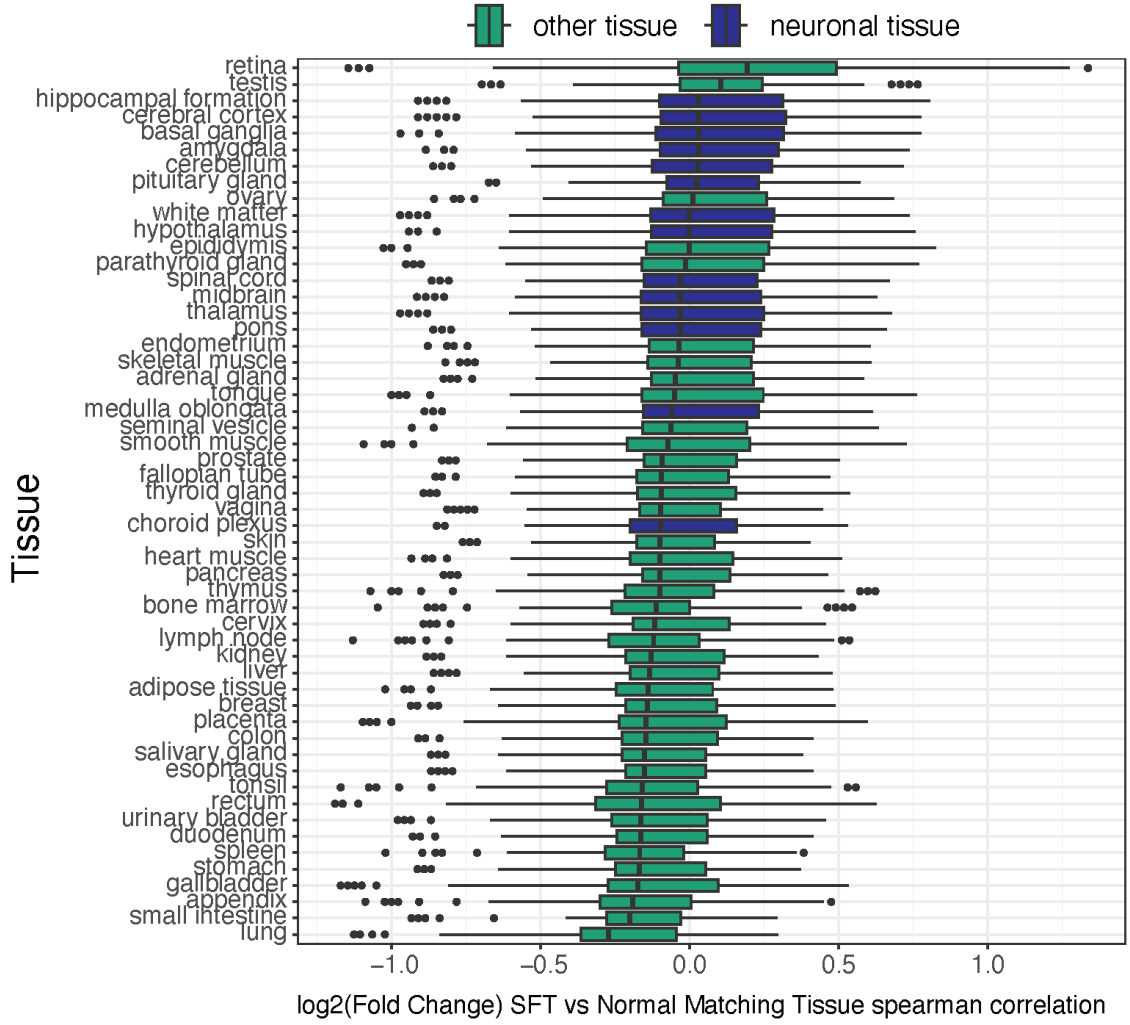
Supplementary Figure 4. NAB2-STAT6 localizes to EGR1 targets in primary tumors.
(a) The most abundant gene fusion present in the primary SFT used in Fig 4 by RNA-seq was NAB2-STAT6 with exons 1-6 of NAB2 and exons 17-22 of STAT6. Graphic generated with Arriba, a fusion detection algorithm.

(b) Venn diagram of overlapping NAB2 and STAT6 peaks from ChIP-seq in control U2OS cells showing minimal overlap in physiological conditions confirming validity of dual antibody from Fig 4a.
(c) Pie chart showing the percentage of enhancer (1kb > from nearest TSS) and promoter (1 kb < from nearest TSS) sites from SFT NAB2-STAT6 Peaks (n = 718).
(d) Motif analysis of 5921 NAB2 only peaks using HOMER shows EGR1, EGR2, and WT1 as the most significantly enriched TF matrices.
(e) Motif analysis of 718 NAB2-STAT6 peaks using HOMER shows EGR1 and WT1 as the most significantly enriched TF matrices.
(f) Motif analysis of 1285 STAT6 only peaks using HOMER shows GRE and STAT3 as the most significantly enriched TF matrices.
(g) Venn diagram of overlapping U2OS NAB2-STAT6 FLAG and SFT NAB2-STAT6 peaks.
(h) Screenshot displays an enhancer (highlighted in yellow) of *ARHGAP32* that has significant NAB2, STAT6, and RNAPII localization in SFTs and in U2OS gains NAB2-STAT6 FLAG localization and has increased EGR1 localization and accessibility by ATAC-seq.

Supplementary Fig. 5



Supplementary Figure 5. NAB2-STAT6 interacts with EGR1 and NAB1 directs them to the nucleus.
(a) Eluates from EGR1 IPs from control U2OS and U2OS nuclear extracts expressing NAB2-STAT6 for 1 day were subjected to MudPIT LC-MS/MS analysis for unbiased identification of the top interactors. Log2 iBAQ

protein scores of control IP interactors are plotted against scores of NAB2-STAT6 expressing IP. NAB2 and STAT6 were only pulled down in NAB2-STAT6 expressing cells.
(b) We generated a clone that constitutively expressed NAB2-STAT6 (NAB2 exons 1-4, STAT6 exons 2-22) with a C-terminal FLAG tag in HEK293T. Immunoblot analysis of whole cell extracts shows strong expression of NAB2-STAT6. GAPDH was used as control.
(C) Eluates from FLAG IPs from U2OS nuclear extracts expressing NAB2-STAT6 for 1 day and HEK293T clone #1 constitutively expressing NAB2-STAT6 were subjected to MudPIT LC-MS/MS analysis for unbiased identification of the top interactors. Log2 iBAQ protein scores of U2OS IP interactors are plotted against scores of HEK293T IP. NAB1 was the top interactor.
(d) Peptide counts pulled down from IP-MS.
(e) Map showing the position of peptides pulled down by U2OS NAB2-STAT6 FLAG IP on NAB2. Peptides were pulled down that are only present in the WT NAB2, not NAB2-STAT6.
(f) Map showing the position of peptides pulled down by HEK293T NAB2-STAT6 FLAG IP on NAB2. Peptides were pulled down that are only present in the WT NAB2, not NAB2-STAT6.
(g) Immunocytochemistry (ICC) of NAB2 (red), STAT6 (green), and DAPI (blue) in WT U2OS cells. NAB2 was primarily cytoplasmic and STAT6 primarily nuclear.
(h) Immunocytochemistry (ICC) of NAB1 (red), FLAG (green), and DAPI (blue) in WT U2OS cells. NAB2 was cytoplasmic and nuclear while FLAG staining was absent.

Supplementary Fig. 6