Supplementary Information for

Myo-differentiation reporter screen reveals NF-Y as an activator of PAX3-FOXO1 in rhabdomyosarcoma

Martyna W. Sroka[1], Damianos Skopelitis[1], Marit W. Vermunt[2], Jonathan B. Preall[1], Osama El Demerdash[1], Larissa M. N. de Almeida[1], Kenneth Chang[1], Raditya Utama[1], Berkley Gryder[3], Giuseppina Caligiuri[1], Diqiu Ren[4], Benan Nalbant[1], Joseph P. Milazzo[1], David A. Tuveson[1], Alexander Dobin[1], Scott W. Hiebert[5], Kristy R. Stengel[6], Roberto Mantovani[7], Javed Khan[8], Rahul M. Kohli[9], Junwei Shi[4], Gerd A. Blobel[2], Christopher R. Vakoc[1,*]

[1] Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 11724, USA

[2] Division of Hematology, The Children's Hospital of Philadelphia, Philadelphia, PA, 19104, USA

[3] Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, 44106, USA

[4] Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

[5] Department of Biochemistry, Vanderbilt University School of Medicine, Nashville, TN 37232, USA

[6] Department of Cell Biology, Albert Einstein College of Medicine, New York, NY, USA

[7] Dipartimento di Bioscienze, Università degli Studi di Milano, Via Celoria 26, 20133 Milano, Italy

[8] Genetics Branch, National Cancer Institute, National Institute of Health, Bethesda, MD, 20892, USA

[9] Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, Pennsylvania, PA 19104, USA

* correspondence: vakoc@cshl.edu

This PDF file includes:

Supplementary Materials and Methods
Figures S1 to S10
Legends for Datasets S1 to S8 (Datasets provided as separate Excel files)
Supplementary References

Supplementary Materials and Methods

Cell culture and generation of stable cell lines

The identity of all of the cell lines used in this study has been validated by short tandem repeat analysis. All cell lines were cultured at 37°C with 5% CO2 and regularly tested negative for mycoplasma contamination. All experiments were performed within 1 month of thawing of cryopreserved cells.

Cell line	Source	Growth media
HEK293T	CSHL stocks	DMEM + 10% FBS + PS
RH4	Dr. Peter Houghton (Greehey Children's Cancer Research Institute)	RPMI + 10% FBS + PS
RH41	Dr. Peter Houghton (Greehey Children's Cancer Research Institute)	RPMI + 10% FBS + PS
RH30	Dr. Peter Houghton (Greehey Children's Cancer Research Institute)	RPMI + 10% FBS + PS
Dbt-MYCN-indP3F1 (referred to as Dbt-P3F1 in this work)	Dr. Frederic Barr (National Cancer Institute)	F10 + 15% FBS + 1x Pyruvate (11360-070), 1 mM Creatinine Monohydrate (Sigma, C3630) + 50 ug/ml uridine (Sigma, U3003) + PS + 500 ng/ml doxycycline (Sigma, D5207)
RD	American Type Culture Collection (ATCC) CCL-136	RPMI + 10% FBS + PS
CTR	Dr. Javed Khan (National Institutes of Health)	RPMI + 10% FBS + PS
RH18	Dr. Peter Houghton (Greehey Children's Cancer Research Institute)	RPMI + 10% FBS + PS
RH4-FKBP12^{F36V}-2xHA	Dr. Kristy Stengel (Albert Einstein School of Medicine) and Dr. Scott Hiebert (Vanderbilt University School of Medicine)	RPMI + 10% FBS + PS

DMEM - Dulbecco's Modified Eagle Medium (Corning, 10013CV); RPMI - Gibco Roswell Park Memorial Institute 1640 Medium (Gibco, 11875176); F10 - Ham's F10 Nutrient Mix (ThermoFisher, 11550-043); PS - Penicillin-Streptomycin (Gibco, 15070063); FBS - fetal bovine serum (Corning, 35073CV).

Cas9 expressing cell lines were generated by lentiviral transduction with a LentiV_Cas9_blast plasmid (based on Addgene #108100; puromycin was replaced with blasticidin resistance gene using in-fusion (Takara Bioscience)) and drug selected. Cutting efficiency was confirmed in competition assays. Stable cell lines used for inducible base editing (1) was established by transducing RH30 cells with EFS-A3An-FRB-IRES-mCherry. Cells expressing high mCherry were then sorted using FACS, transduced with EFS-FLAG-FKBP12-A3Ac-nCAS9-2XUGI-NLS-P2A-PURO and drug selected. Destabilized GFP reporter lines were established as above, in addition with transducing the cells with a destabilized GFP reporter (destabilized GFP from Addgene #14760 was cloned into LentiV backbone using in-fusion (Takara Bioscience)) and cell sorting of cell populations expressing low levels of GFP to enrich for single copy reporter integration.

sgRNA design and cloning

Single guide RNAs were designed using a custom CSHL CRISPR sgRNA design algorithm (https://crispr.cshl.edu/); all sgRNA sequences used in this study are available in Dataset S1. Whenever possible, guides with high CRISPR score (>0.7; high predicted on-target cutting and low off-target cutting) that target annotated functional protein domains were selected. For competition assays, the DNA oligos of the sgRNAs were cloned into LRG2.1T vector (Addgene #108098) using a BsmBI restriction site. For all other assays where individual gene knockouts were performed, the sgRNAs were cloned into LR2.1Tpuro vector (modified from Addgene #108098; GFP replaced by puromycin resistance through in-fusion). For base editing, sgRNAs were cloned into LR2.1Tblast (modified from Addgene #108098; GFP replaced by blasticidin resistance through in-fusion). For the test library, sgRNAs targeting N-PAX3, C-PAX3, N-FOXO1 and C-FOXO1 were mixed with control sgRNAs. For the exon scan library, all possible sgRNAs that target PAX3, FOXO1 and MYOD1 were designed and mixed with control sgRNAs. For the phenotypic screens, genes were pre-selected based on the magnitude of negative selection (from DepMap (2, 3)), and literature search for genes relevant to RMS biology, prioritizing genes that belong to TFs, cofactors, epigenetic modifiers, kinases and phosphatases protein families. For these candidate genes, whenever possible, 10 sgRNAs targeting functional domains were chosen for each gene. Non-targeting sgRNAs and sgRNAs targeting essential genes were added into the pooled sgRNA libraries. All pooled libraries were synthesized on an array platform (Twist Bioscience) and then cloned into the BsmBI restriction site of LR2.1puro vector using Gibson Assembly (NEB). Representation of sgRNAs in the pools was checked by Illumina next generation sequencing.

Lentivirus production and transduction

HEK293T cells were co-transfected with pVSVG and pPAX2 lentiviral helper plasmids and a plasmid of interest using polyethylenimine (PEI). Media was changed 12 hours post-transfection and supernatant containing the lentivirus was harvested at 48, 72, and 96 hours post-transfection. Pooled supernatants were centrifuged at 8,000 g for 5 minutes at 4ºC to pellet debris and passed through an 0.45 µm filter. To transduce, target cells were mixed with complete media containing 4 mg/mL polybrene and the virus. Media was changed 24 hours post-infection. Where appropriate, antibiotic selection was introduced 48 hours post-infection at the following concentrations: puromycin 1µg/ml, blasticidin 10 µg/ml, G418 1 mg/ml. For preparation of virus containing sgRNA libraries, sufficient cell numbers were used to maintain sgRNA representation at least at 1000x.

Competition-based cell proliferation assay

Cas9-expressing cells were infected with a lentivirus carrying sgRNA cloned into a GFP-containing LRG2.1T backbone at low multiplicity of infection (MOI) (~0.3). The percentage of GFP-positive cells was first measured at day 3 using the Guava EasyCyte flow cytometer (Millipore) and then at regular intervals during cell passage. The relative fluorescence at each timepoint was calculated using the following formula: Relative fluorescence = [% GFP+ day X] / [%GFP+ day 3] × 100%. In Figure 1B and Figure S1, the following sgRNAs were averaged in each category:
PAX3-N: sgPAX3_e1.1, sgPAX3_e1.2, sgPAX3_e2.1 (sgPAX3-FOXO1), sgPAX3_e2.2, sgPAX3_e2.3, sgPAX3_e3.1, sgPAX3_e3.2, sgPAX3_e5.1, sgPAX3_e5.2
PAX3-C: sgPAX3_e8.1, sgPAX3_e8.2, sgPAX3_e8.3
FOXO1-N: sgFOXO1_e1.1, sgFOXO1_e1.2, sgFOXO1_e1.3
FOXO1-C: sgFOXO1_e2.1, sgFOXO1_e2.2, sgFOXO1_e3.1
Pos: sgRosa
Neg: sgRPA3, sgPCNA

Protein extraction, quantification and western blotting

Cells were washed with PBS and lysed directly on tissue culture plates through incubation in RIPA buffer (Thermo Fisher Scientific, 89901) supplemented with 50x protease inhibitor cocktail (Sigma, P8340) for 5 minutes at 4ºC. Lysates were then scraped and sonicated on a Diagenode Bioruptor (medium power, three times for 5 minutes in 30 seconds ON, 30 seconds OFF intervals) to solubilize proteins. Debris was removed by centrifugation at 8,000 g for 10 minutes at 4ºC. Protein concentration was quantified using Pierce BCA Protein Assay Kit (ThermoFisher, 23227) following manufacturer's protocol. Protein samples were then boiled in 6x Laemmli sample buffer (Thermo Fisher Scientific) containing β-mercaptoethanol and boiled at 98 °C for 5 minutes. Western blotting was performed by separation of proteins on NuPAGE 4 to 12%, Bis-Tris Gels (NP0321BOX) in MOPS buffer and electrophoretic transfer onto a nitrocellulose membrane. Proteins were visualized using Ponceau S stain, following membrane blocking in 5% milk in TBST and incubation with primary and secondary antibody staining solution. Horseradish peroxidase signal was visualized by incubation with Pierce ECL Western Blotting Substrate (ThermoFisher, 32106) and detection on radiography films. To strip and re-probe, membranes were incubated in stripping buffer (100 mM 2-mercaptoethanol (M3148), 2% (w/v) SDS (L3771), 62.5 mM TrisHCl, pH 6.8) for 1 hr at 60ºC, blocked in 5% milk solution and stained.

Antibody name	**Catalog #, manufacturer**	**Dilution**	**Staining condition**
anti-MYH	MF20, DSHB	1:1000	4ºC, O/N
anti-FOXO1	sc-374427, Santa Cruz	1:1000	4ºC, O/N
anti-MYOM3	17692-1-AP, Proteintech	1:1000	4ºC, O/N
Anti-TAGLN	ab14106, abcam	1:1000	4ºC, O/N
anti-TNNT3	JLT12, DSHB	1:1000	4ºC, O/N
Anti-TNNT2	MA5-12960, Invitrogen	1:1000	4ºC, O/N
anti-MYL1	ab151749, abcam	1:1000	4ºC, O/N
Anti-MYL1	ab228727, abcam	1:1000	4ºC, O/N
anti-MYOT	ab68915, abcam	1:1000	4ºC, O/N
anti-ACTN2	ab9465, abcam	1:1000	4ºC, O/N
anti-VINC	#4650, Cell Signalling	1:1000	4ºC, O/N
anti-ACTB	#4967, Cel Signalling	1:1000	4ºC, O/N
anti-HA (HRP)	12013819001, Sigma	1:1000	4ºC, O/N
Goat Anti-Mouse IgG H&L (HRP)	ab97023, abcam	1:5000	2hr, RT
Goat Anti-Rabbit IgG H&L (HRP)	ab6721, abcam	1:5000	2hr, RT

Immunofluorescence

Cells were seeded on glass coverslips, infected with lentivirus to generate knockouts and drug selected. Seven days post-transduction, cells were washed with PBS and fixed in 4% paraformaldehyde for 10 minutes at RT, permeabilized in 0.3% Triton X-100 in PBS for 5 minutes at RT, blocked in 5% BSA, 0.3% Triton X-100 in PBS for 1 hour at RT. To stain, samples were incubated in primary antibody solution (antibodies diluted in 1% BSA, 0.3% Triton X-100 in PBS) overnight at 4ºC, washed in PBS and incubated in secondary antibody solution (antibodies diluted in 1% BSA, 0.3% Triton X-100 in PBS) for 2 hours at RT kept in the dark. Anti-beta Tubulin primary antibody was used for all single-color control stains. Following PBS washes, coverslips were mounted on slides on a drop of ProLong Diamond Antifade Mountant (with (fully stained samples) or without (single color controls) DAPI; ThermoFisher P36962 or P36961, respectively), cured at RT in the dark for 1 hour, sealed with nail polish and dried O/N at RT. Samples were imaged on Zeiss LSM 710 Confocal Microscope at 40x magnification. For visualization, all images were processed in the same way in FIJI/ImageJ (4) to enhance visibility. For quantification, multiplexed fluorescence images were analyzed with QuPath version 0.3.2 (5). In brief, following detection of DAPI+ nuclei, single measurement classifiers for DAPI+, TNNT3+ and MYH+

cells were optimized on training images to ensure robust and reliable detection of true positives. The three classifiers were then applied simultaneously to perform single-, double- and triple-positive measurements.

Antibody name	Catalog #, manufacturer	Dilution	Staining condition
anti-ACTB	ab6046, abcam	1:200	4°C, O/N
anti-MYH	MF 20, DSHB	1:25	4°C, O/N
anti-TNNT3	ab175058, abcam	1:25	4°C, O/N
anti- Ki-67	MA5-14520, ThermoFisher	1:200	4°C, O/N
Alexa Fluor™ 568 Phalloidin	A12380, ThermoFisher	1:50	4°C, O/N
anti-Rabbit IgG, Alexa Fluor 647	A-21244, Invitrogen	1:500	2 hrs, RT
anti-Rabbit IgG, Alexa Fluor 488	A-11034, Invitrogen	1:500	2 hrs, RT
anti-Rabbit IgG, Alexa Fluor 568	A-11036, Invitrogen	1:500	2 hrs, RT
anti-Rabbit IgG, Cyanine3	A-10520, Invitrogen	1:500	2 hrs, RT
anti-Mouse IgG, Alexa Fluor 647	A-21236, Invitrogen	1:500	2 hrs, RT
anti-Goat IgG, Alexa Fluor Plus 488	A-32814, Invitrogen	1:500	2 hrs, RT

Inducible base in GFP reporter cell lines
Sorted RH30 cells expressing EFS-A3An-FRB-IRES-mCherry, EFS-FLAG-FKBP12-A3Ac-nCAS9-2XUGI-NLS-P2A-PURO (1) and destabilized GFP reporter were transduced with sgRNAs. At day 2 post-transduction, blasticidin drug selection was implemented to enrich for transduced cells. Two days post blasticidin addition, the cells were split and incubated either 200 nM rapamycin (Research Product International; R64500-0.001) to induce base editing or in DMSO (negative control). GFP% was measured every two days using Guava EasyCyte flow cytometer (Millipore) and normalized, setting the starting GFP percentage as 100%.

Inducible base editing
RH30 stable cell line expressing the two components of the split base editing system (without the GFP reporter) was transduced with sgRNAs in LR2.1Tblast backbone and drug selected with blasticidin starting at day 2 post-transduction for 48 hours. At day 4, rapamycin (Research Products International, R64500-0.001) was added at 200 nM concentration to induce base editing. Cells were harvested and analyzed at day 8 post rapamycin addition. To estimate base editing efficiency, genomic DNA was harvested from edited cells using the phenol-chlorophorm method and a segment of DNA spanning the edited site was PCR amplified with gDNA_CCAAT2_fwd2 primer: CTCGGCACCACCAGAGATG
And gDNA_CCAAT2_rev2 primer: GAGCCTGGTGAGGCTGGA using Phusion Flash High-Fidelity PCR Master Mix (ThermoFisher, F548S). PCR products were then separated on agarose gel, purified using the QIAquick Gel Extraction Kit (Qiagen, 28704) and Sanger sequenced. Deconvolution of sequencing traces was performed using EditR (6).

Dual luciferase assays
The human PAX3 promoter sequence annotation was retrieved from the Eukaryotic Promoter Database (promoter ID: PAX3_1 and PAX3_2) (7) and amplified from gDNA of RH4 RMS cells. Lentiviral plasmid encoding Firefly and Renilla luciferase under the control of a minimal promoter (Addgene #138368 (8)) was modified to remove an enhancer element and replace the minimal promoter with the PAX3 promoter using in-fusion (Takara Biosciences). Mutation of each CCAAT site alone or in

combination was done through in-fusion (CCAAT motif was replaced by TACTA randomly generated sequence through a primer overhang). All constructs were validated by Sanger sequencing. The reporter plasmids were delivered to RH4 and RD RMS cells grown in 24-well plates through lentiviral transduction. Media was changed 24 hours post-transduction. Cells were analyzed 72 hrs post-transduction by performing Dual-Glo Luciferase Assay (Promega, E2920) following manufacturer's protocol (https://www.promega.com/-/media/files/resources/protocols/technical-manuals/0/dual-glo-luciferase-assay-system-protocol.pdf?rev=3f5ca2862b6047a8b4c3a31126ea310e&sc_lang=en). Fluorescence was measured on a Molecular Devices SpectraMax i3x Multi-Mode Microplate Reader. In each sample, Firefly fluorescence signal was normalized to Renilla fluorescence to account for transduction efficiency and the normalized signal was expressed as a fold change of the wild-type promoter activity in a given cell line. Experiments were performed in 3 biological replicates, with 3 technical replicates each.

PAX3 promoter sequence used in this study (CCAAT NF-Y binding sites indicated in **bold**; changed to TACTA in mutant constructs):
aaatgagacatagagagagacacaggaaatcacaagaggaatagaggctgagcgagacacacacacagaggcacagaaagagacagagagg
gaaatagaaagtcaaggaaagagtgatcagagaaagacacacacacacacacacacacacacacacacacacacacacacacacacacacagagtga
cacagacagagagacagagacagagagacaggaacttctccgccctcagcaactgccatctccctggggctgtctctctcagtttccaccgggcca
accttctctcctgggcaaggggcgcagcgcgggtccccctcggggccagcagaggcctcggcaccaccagagatgggaagagaaagtggtcgc
tgttgc**CCAAT**cagcgcgtgtctccgccacccgggacggtctacccgtcgg**CCAAT**cgcagctcagggctcctgaccaagctttgggtaa
aagaactaataaatgctcccgagcccggatccccgcactcggtgtcaccacaggaggagactcaggcaggccgcgctccagcctcaccaggctc
cccggctcgccgtggctctctgagccccctttcagggaccccagtcgctggaacatttgcccagactcgtaccaaacttttccgccctgggctcggg
atcctggactccggggcctccccgtcctccccctttcccgggttccagctccggcctctggactaggaaccgacagccccctccccgcgtccctccc
tctctctccagccgttttggggaggggctctccacgctccggatagttcccgagggtcatccgcgccgcactcgcctttccgtttcgccttcacctggat
ataatttccgagcgaagctgcccccagg

RNA isolation, cDNA synthesis and RT-qPCR
Cas9-expressing RMS cells were transduced with sgRNAs and drug selected. Knockout cells were harvested at a specified timepoint (day 4 or day 7) and RNA was extracted using TRIzol (Thermo Scientific; Cat. No. 15596018) following manufacturer's protocol (https://tools.thermofisher.com/content/sfs/manuals/trizol_reagent.pdf). The extracted RNA was treated with DNase I (NEB, M0303S) and reverse transcription was performed using the qScript cDNA Synthesis Kit (Quantabio, 95047-025) following manufacturer's protocol. RT-qPCR reactions with the SYBR Green PCR Master Mix (ThermoFisher, 4309155) were set up following manufacturer's protocol, using 25 ng template cDNA. Each reaction was set-up in triplicates. The assay was run on QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) and analyzed in the QuantStudio 6 and 7 Pro Real-Time PCR Systems Software (Thermo Fisher Scientific). Data was analyzed using the ΔΔCt method, using three housekeeping gene controls (ACTB, GAPDH, ATP5F1), and gene expression of each gene knockout was normalized to sgRosa control condition. The number of biological and technical replicates is indicated in figure legends.

Target gene	**Forward primer sequence**	**Reverse primer sequence**
ACTB	TCTTCCAGCCTTCCTTCCTG	CAATGCCAGGGTACATGGTG
GAPDH	GCCATCAATGACCCCTTCAT	TGACAAGCTTCCCGTTCTCA
ATP5F1	TTTCGTTGACCATGCTGTCC	GTGTGGCTGCCCTGTATGAA
PAX3-FOXO1	AATGGCCTCTCACCTCAGAATTC	CTTGCCACCCTCTGGATTGA
MYH3	AGCTCGAGGCCAAGATCAAG	GCCAGGGTCAACTCAAGGTC

RNA sequencing

Cells were transduced with sgRNAs and drug selected. Knockout cells were harvested at day 4 or day 7 post-transduction and RNA was extracted using TRIzol (Thermo Scientific; Cat. No. 15596018) as above. Sequencing libraries were prepared using the TruSeq RNA Library Prep Kit v2 (Illumina) following the manufacturer's Low Sample (LS) protocol (https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_truseq/truseqrna/truseq-rna-sample-prep-v2-guide-15026495-f.pdf), starting with 1ug of total RNA as input. The quality of libraries was assessed on Agilent 2100 Bioanalyzer using the Agilent DNA 1000 Assay (Agilent, 5067-1504), quantified using Qubit dsDNA BR Assay (Thermo Fisher Scientific, Q32850) and KAPA Library Quantification Kit (KK4824), diluted, pooled at equimolar concentrations and sequenced on the Illumina NextSeq500 platform using the single end SE75 bp mode of sequencing. All RNA sequencing experiments were performed in two biological replicates, with the exception of RH30 and Dbt-P3F1, for which one replicate was sequenced. All knockout conditions that were directly compared to one another (e.g. RH4 sgRosa, sgPAX3-FOXO1, sgNFYC) were sequenced on the same flow cell.

RNA sequencing data analysis

Sequencing reads were debarcoded and quality was checked using FastQC (9). Reads were then mapped onto the reference human genome (hg38) using HISAT2 (10). Reads within genes were then counted using HTSeq-count (11). Differential gene expression analysis was performed using DESeq2 (12). For quantification of PAX3-FOXO1 fusion and WT PAX3 and FOXO1 transcripts abundance, sequence alignment was performed using STAR v2.7.8a (13) on the whole genome with default parameters. Human genome reference (FASTA assembly and GTF annotation) was obtained from Gencode GRCh38.p13 version 43 release. Fusion transcript from partial PAX3 wild type isoform (exon 1 to 7) and FOXO1 wild type isoform (exon 2 to 3) was concatenated into the reference genome. Genome index was built with additional parameter '--sjdbOverhang 71'. Gene quantification was done by using two comparative approaches as follows: 1. Exon junction counts in fusion and wild type transcripts were directly calculated using featureCounts from Subread v2.0.2 (14); 2. Estimated transcript counts/abundance in fusion and wild type transcripts were calculated using RSEM v1.3.3 (15) based on expectation minimization method. Differential genes, transcripts, and fusion were done using DESeq2 v1.38.3 to generate p-values and log fold change. Comparisons were generated for: 1. NFYC samples (2 replicates) vs Rosa samples (2 replicates); 2. PAX3-FOXO1 samples (2 replicates) vs Rosa samples (2 replicates). Data visualization and statistical analysis were performed in custom python scripts. Gene set enrichment analysis (16) was performed using gene lists ranked by $\log_2$ fold change as input.

Single-cell RNA sequencing

RH4 Cas9 expressing cells were transduced with sgRNAs and drug selected. Knockout cells were harvested at day 7 post transduction and dissociated into a single cell suspension with TrypLE express (Thermo Fisher Scientific, #12604013), then washed and resuspended in PBS containing 0.04% BSA. To assess viability, cells were stained ViaStain AOPI (Nexcelom, #CS2-0106-5mL) and counted using a Countess FL II automated cell counter. The suspension was diluted according to the manufacturer's instructions to target a yield of 5,000 cells per sample, and then loaded into a 10X Chromium microfluidic chip. Single cell capture, barcoding and library preparation were performed using the 10X Chromium chemistry, using the NextGEM Single-Cell 3' Library Kit v3.1 (10X Genomics; 1000121). cDNA and libraries were checked for quality on Agilent 2100 Bioanalyzer, quantified by KAPA Library Quantification Kit (KK4824), and sequenced on a Illumina NextSeq500 to an average depth of ~33,000 reads per cell.

Single-cell RNA sequencing data analysis

The CellRanger count pipeline (v6.0.0, 10X Genomics) was used to align FASTQs to the human reference genome (gex-GRCh38-2020-A, 10X Genomics) and produce digital gene-cell counts matrices. For integrated analysis, the samples were combined into a single matrix using the CellRanger aggr pipeline using the parameter "normalize=mapped" to match cross-sample read depth by downsampling. The resulting gene-barcode matrix was fed into Scanpy version 1.6 for secondary analysis (17).

Cells with fewer than 10,000 UMIs or greater than 10% mitochondrial gene content were discarded. This stringent cutoff served to filter droplets containing significant UMI counts but lacking abundant nuclear noncoding RNAs MALAT1 and NEAT1, suggesting they were non-nucleated cell fragments generated during the cell harvesting step. During our analysis, we also noted a small cluster of cells present in all conditions that was marked by the expression of the dosage compensation noncoding RNA XIST, whereas the majority of cells in the matrix lacked its expression. The RH4 RMS cell line (female) has not previously been reported to express XIST, and we reasoned that these few cells might represent either a spontaneous mutation, minor contamination, or other artefact. This cluster did not notably change in abundance or expression pattern in response to our CRISPR treatments, so we chose to note its existence but mask it from further analysis under the scope of this report.

Gene expression was normalized after masking highly expressed genes using the function scanpy.pp.normalize_total() with the flag exclude_highly_expressed=True, followed by log-transformation. Principal component analysis was restricted to the top 4,000 most highly expressed genes, excluding mitochondrial and ribosomal genes. The top 50 principal components were used to compute the KNN graph and perform UMAP dimensionality reduction and graph-based clustering via the Leiden modularity optimization algorithm using a default resolution of 0.5. Differential expression was calculated between clusters using a Student's t-test via scanpy.pp.rank_genes_groups() function. Upon initial clustering, we noted that one large cluster, mainly comprised of non-cycling cells (see below for methods), was jointly populated by cells from each sample in roughly equal proportions. Minimal marker genes could be observed between CRISPR sgRNA treatments within this cluster, and no significant differentiation programs could be detected. We reasoned that these cells had undergone some nonspecific form of cell-cycle arrest, and were excluded from our downstream analysis comparing the different treatments.

The gseapy Python package was used to carry out Enrichr (18, 19) analysis as follows: for each cluster, the top differentially expressed genes were computed against the remaining cells in the dataset and filtered using a cutoff of log(fold change) > 1 and Padj (BH corrected) < 1e-12. Common highly-expressed genes that have low predictive value, including ribosomal protein genes and mitochondrial genes, were dropped from the analysis. These gene lists were then used query Enrichr across relevant gene sets, including Descartes_Cell_Types_and_Tissue_2021 (20) and Human_Gene_Atlas (21), and Tabula Sapiens (22). Hits with FDR < 0.01 were considered for annotating cluster phenotypes. For GSEA analysis (16), each cluster's differential expression statistics were used as preranked input to compute NES. To visualize expression of these signatures, scanpy.tl.score_genes() was performed using the significantly enriched genes intersecting with the indicated gene sets.

For pseudotime analysis, Palantir (23) was called via the scanpy.external API. Briefly, gene expression data was first denoised by conducting principal component analysis on the top 4,000 highly variable genes. Scanpy.external.palantir() was used to compute a nearest-neighbor graph (k=30 neighbors) and diffusion maps based on the top 50 principal components. To determine the starting cell for pseudotime calculation, we arbitrarily chose the cell with the most negative UMAP X-coordinate, farthest from the two differentiated muscle-type cell clusters in the 2D projection. This was fed into the Palantir "early_cell" parameter, which is used to determine an optimal nearby root cell for pseudotime calculation. For visualization, cells were embedded in a 2D plane using t-SNE computed from the first 5 Palantir diffusion components.

Gene expression data was smoothed for the purposes of visualization using the MAGIC imputation algorithm built into the scanpy.external API (24). Log-normalized gene expression data was fed into

scanpy.external.pp.magic() and values were imputed with the default KNN kernel size of 5 nearest-neighbors. Imputed values were only used to visualize expression on tSNE, UMAP, heatmap, etc. plots, rather than for downstream dimensionality reduction or differential expression
For cell-cycle analysis, a list of genes reported to be upregulated during S and G2/M phases were compiled from (25) and used to compute a generic cell-cycle score using the function scanpy.tl.score_genes(). This score is meant to differentiate between G0 and G1 cells from these likely to be currently in S, G2, or M phase.

CUT&RUN
CUT&RUN assay (26) was performed following the modified Epicypher CUT&RUN Protocol v2.0 (https://www.epicypher.com/content/documents/protocols/cutana-cut&run-protocol.pdf. In short, RH4 cells were transduced with sgRNAs and drug selected. Knockout cells were analyzed at day 7 post-transduction. One plate of cells for each condition was trypsinized and the cells were counted using the Invitrogen Countess Automated Cell Counter to estimate cell numbers. The assay cells were then harvested by scraping and appropriate cell numbers were aliquoted (500,000 cells per condition; batch processed). The cells were washed and bound to activated concanavalin A beads (Epicypher, 21-1401) (10ul per reaction) for 10 min at RT. No cross-linking was performed. The bead-bound cells were retrieved through magnetic separation, resuspended in antibody binding buffer containing digitonin and aliquoted into individual staining conditions. The appropriate amount of each antibody was added and the samples were incubated at 4ºC overnight on a nutator to ensure constant agitation. Following staining, the bead-bound cells were washed and resuspended in digitonin buffer. 2.5 µl of protein A/G micrococcal nuclease fusion enzyme (Epicypher, 15-1016) was added to each sample and incubated for 10 min at RT to allow binding to the fc region of the antibodies used for staining. The samples were then washed to remove the unbound enzyme and chromatin cleavage was activated through addition of 1µl of 100 mM CaCl2. The digestion was performed at 4ºC for 2 hrs under constant agitation. The reaction was stopped by addition of the stop buffer and incubation for 10 min at 37ºC. The samples were then placed on a magnet and the cleaved chromatin present in the supernatant was transferred to new tubes and purified using the phenol-chlorophorm-isoamyl alcohol method. Sequencing libraries were constructed using the entire volume of the sample as a starting material using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S) and the NEBNext Multiplex Oligos for Illumina (NEB E7335S) following the manufacturer's protocol (https://www.neb.com/-/media/nebus/files/manuals/manuale7103-e7645.pdf?rev=09ba7d304b454b28b44dfdcec1eb33e3&hash=BD108A19BADA62177E5DBDA331FEF506) with the following modifications: adaptors were diluted at 1:10; Ampure bead clean-up was performed without size selection, using 1.1x sample volume ratio. PCR was performed using the following cycling conditions: 45 seconds at 98ºC; 14 cycles of 15 seconds at 98ºC, 10 seconds at 60ºC; 1 minute at 72 ºC. The libraries were quantified using the Qubit dsDNA BR assay (Thermo Fisher Scientific) as per the manufacturer's instructions and the quality was assessed using Bioanalyzer dsDNA High Sensitivity assay (Agilent, 5067-4626). Sequencing libraries were pooled at the equimolar ratio and sequenced on a Illumina NextSeq500 using the paired end PE26 bp mode of sequencing.

Antibody	**Catalogue #, manufacturer**	**Amount**
anti-IgG	13-0042, Epicypher	1ul
anti-H3K27ac	ab4729, abcam	1ul
anti-H3K27me3	#9733, Cell Signaling Technology	1ul
anti-NF-YB	custom-made	5ul
anti-NF-YC	custom-made	5ul

* NF-YA subunit of the complex was not assessed due to poor performance of anti-NFYA antibodies in CUT&RUN assay

CUT&RUN data processing and analysis

Quality control was performed using FastQC (9). Adapters were trimmed using Trimmomatic (27) and the reads were mapped to the reference genome (hg38) using bowtie2 (28). Peaks were called using MACS2 algorithm (29) at a threshold of $p < 0.0000001$. NF-Y consensus peak set was derived from NF-YB and NF-YC peak sets using bedtools intersect (30). Peaks were annotated using the annotatePeaks command from the Homer suite (31). Motif analysis was performed using MEME AME (32). Gene Ontology analysis was performed using Metascape (1). Distances between peaks were calculated using bedtools closest (30). For assessment of differential NF-Y binding following base editing, data was trimmed and aligned as above. Peaks were called using Homer findPeaks. Differentially bound sites were identified using Homer getDifferentialPeaks (pval ≤ 0.001, fold change 2.5 in sgCtrl1). For visualization, bam files were CPM-normalized using bamCoverage.

Looping between PAX3-FOXO1 at distal enhancers and NF-Y at gene promoters

First, we identified high-confidence P3F1 binding sites using a previously published dataset comprised of 4 biological replicate ChIP-seq experiments in RH4 cells (GSE116344) (1) by calling peaks with MACS2 at a threshold of $p < 0.0000001$. For NF-Y, the consensus peak set described above was used. To find PAX3-FOXO1 associations with NF-Y bound and unbound gene promoters, we identified all gene transcriptional start sites within a topologically associated domains (TADs) containing PAX3-FOXO1 and made all possible intra-TAD coordinate pairs, annotating these gene promoters that were NF-Y bound. Next, using Juicer tools -dump command (wrapped in the script annotate_loops.sh, available here: https://github.com/orgs/axiotl/repositories), we extracted the H3K27ac HiChIP contact frequency between each of these coordinate pairs. For each candidate gene, we assigned PAX3-FOXO1 target genes as these that were in the top quantile by contacts per million, and that were also expressed in the top 40% of genes by transcripts per million (TPM, from RNA-seq in RH4 cells). Thus, four categories were compared: PAX3-FOXO1 connected genes with/without NFY in the promoter, and genes not connected to PAX3-FOXO1 with/without NF-Y in the promoter. These categories were then compared at the gene expression and HiChIP contact frequency levels using custom R scripts and ggplot2.

4C-seq

4C was performed as previously described (35). First digestion was performed with DpnII and the second digestion with Csp6I. Four separate PCR reactions for each viewpoint were carried out using 150-200 ng 4C template input material. The following primers were used (viewpoint specific primers in **bold**, adapters underlined):

ERRFI1_R TACACGACGCTCTTCCGATCT**GCCTGATAGTTAGACACTGATC**
ERRFI1_NR ACTGGAGTTCAGACGTGTGCTCTTCCGATC**AGGACCTTTCCTCCATATTC**
Jun_R TACACGACGCTCTTCCGATCT**CAAGCGTGTAGGCGATC**
Jun_NR ACTGGAGTTCAGACGTGTGCTCTTCCGATC**AAACTTAAGTCCCCTTAGGC**
PEG3_R TACACGACGCTCTTCCGATCT**GGTAGCATTAGATGTGACGATC**
PEG3_NR ACTGGAGTTCAGACGTGTGCTCTTCCGATC**GGCTGTCTGCCCTAATGC**
Six1_R TACACGACGCTCTTCCGATCT**GGGTTGTTGAGACCAGATC**
Six1_NR ACTGGAGTTCAGACGTGTGCTCTTCCGATC**TTTTCCAATAAGAAGTGACCA**

*R = reading primer, NR = non-reading primer

Reactions were pooled and purified using Ampure XP beads (Beckman Coulter, #A63880), eluting in 35 µl milliQ water. Ten µl purified PCR1 product was used as input for a second PCR to obtain full Illumina adapter sequences using the universal forward primer (AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT) and a barcoded reverse primer (CAAGCAGAAGACGGCATACGAGATXXXXXXGTGACTGGAGTTCAGACGTGTGCT) with a 6 nucleotide barcode (XXXXXX). PCR reactions were then purified using a Roche high pure PCR purification kit (#11732676001). Pooled 4C reactions were purified once more with Ampure XP beads and sequenced on the Illumina NextSeq 2000 platform using the single end SE75 bp mode of sequencing.

4C-seq data processing and analysis

4C mapping was done as described in https://github.com/deWitLab/4C_mapping. PeakC (36) was used to call significant 4C peaks https://github.com/deWitLab/peakC. Signal for two biological replicates was averaged using the wiggletools mean command (37).

PAX3-FOXO1 promoter motif analysis

Identification of transcription factor binding motifs present within the PAX3-FOXO1 promoter was performed using Find Individual Motif Occurrences (FIMO) Version 5.5.2 from the MEME suite. HOCOMOCOv11_core_HUMAN collection of position weight matrices was queried, using a significance threshold of 1.0E-4. PAX3-FOXO1 promoter sequence queried is identical to the sequence used in dual luciferase assays and listed above.

Comparison of NF-Y subunit expression across RMS cell lines and tumors

Expression data for RMS cell lines was retrieved from DepMap (Expression 21Q1). Expression data for RMS tumors was retrieved from The Database of Genotypes and Phenotypes (dbGaP) (38).

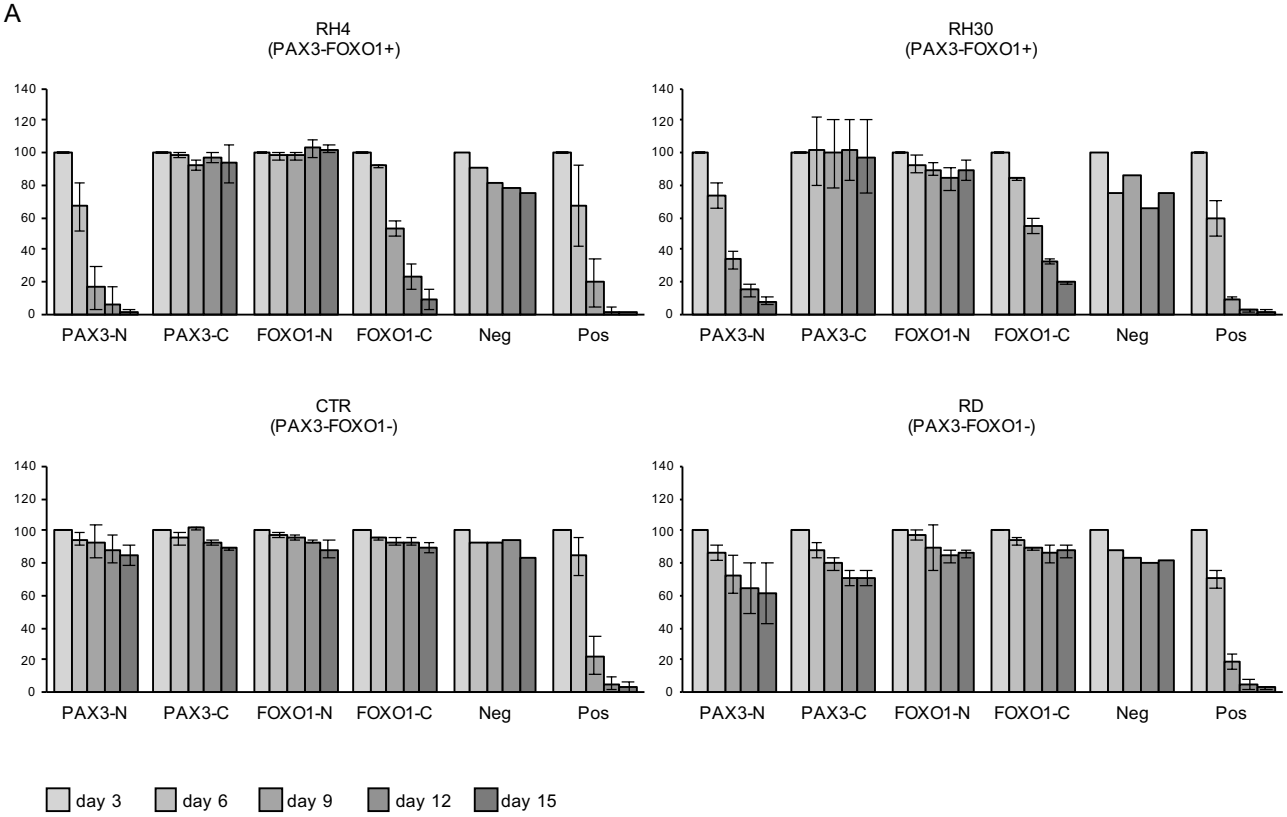
Quantification and statistical analysis

All statistical analyses were performed in python, with details provided in the corresponding figure legends, such as the number of replicates and statistical tests employed. Error bars represent standard deviation, unless otherwise stated.

Data visualization

Cartoon representation of the screening procedure in Fig. 2A was created in BioRender. Volcano plots, heatmaps, dot plots, bubble plots and box-and-whiskers plots were generated in python matplotlib and seaborn. Bar graphs and pie charts were created in Microsoft Excel Spreadsheet Software. Flow cytometry plots were created in FlowJo (BD Life Sciences). Figures 5D, 6F and Figure S9C were generated in Integrated Genome Viewer (39).

A



B

Gene	Rosa_rep1	Rosa_rep2	NFYC_rep1	NFYC_rep2	PAX3-FOXO1_rep1	PAX3-FOXO1_rep2
RSEM (counts)						
PAX3-FOXO1	3181.09	3412.78	2161.52	1824.01	2647.25	1806.32
			p-val = 2.42e-9 ; logFC = - 0.52		p-val = 0.0029 ; logFC = - 0.39	
PAX3	717.26	714.12	407.1	292.33	338.26	218.5
PAX3 WT only	0	0	3.91	0	6.18	3.44
FOXO1	509.27	407.11	794.38	532.66	1430.66	902.18
FOXO1 WT only	509.27	407.11	790.91	532.66	1429.25	900.85
			p-val = 0.00035 ; logFC = 0.58		p-val = 1.32e-8 ; logFC = 1.39	

Figure S1. The effects of inactivating PAX3-FOXO1 in fusion-positive RMS. A. GFP competition assays in PAX3-FOXO1 positive and negative RMS cell lines. The cells were lentivirally transduced with GFP-linked sgRNAs targeting N- or C-terminus of PAX3 and FOXO1. The effect on cell fitness were tracked by assessing the fraction of GFP+ cells over time using flow cytometry. SgRNAs targeting PCNA and RPA3 were used as positive controls, sgRosa was used as a negative control. **B.** Quantification of the PAX3-FOXO1 and WT PAX and FOXO1 transcripts in the negative control, PAX3-FOXO1 knockout and NFYC knockout based on bulk RNA-seq performed at day 4 poi with sgRNAs.

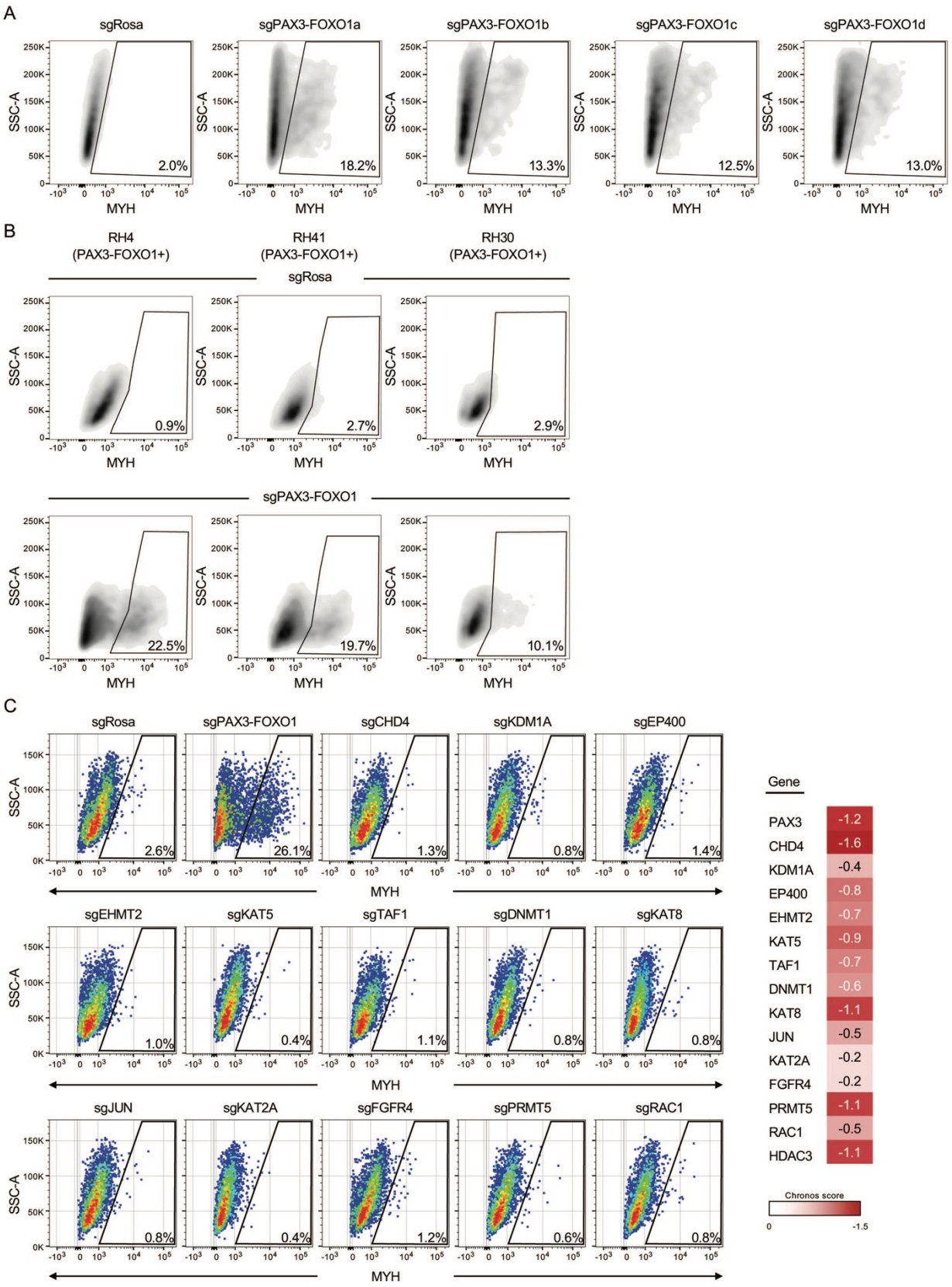


Figure S2. CRISPR-mediated targeting of PAX3-FOXO1 results in upregulation of MYH, which can be tracked by flow cytometry. A. RH4 PAX3-FOXO1+ RMS cells stably expressing Cas9 were

transduced with a control sgRNA (sgRosa) or multiple sgRNAs targeting the PAX3-FOXO1 fusion (sgPAX3-FOXO1a-d). MYH expression was assessed by flow cytometry 7 days poi. **B.** RH4, RH41 and RH30 PAX3-FOXO1+ RMS cells were transduced with control or PAX3-FOXO1 targeting sgRNAs. Expression of MYH was assessed 7 days poi by flow cytometry. **C.** Flow cytometry analysis and quantification of MYH expression following knockout of multiple essential genes in RMS. sgRosa is used as a negative control, sgPAX3-FOXO1 as positive control. Heatmap represents gene essentiality score (Chronos) in RH4 RMS cell line, based on DepMap CRISPR Avana Public 21Q1 data.

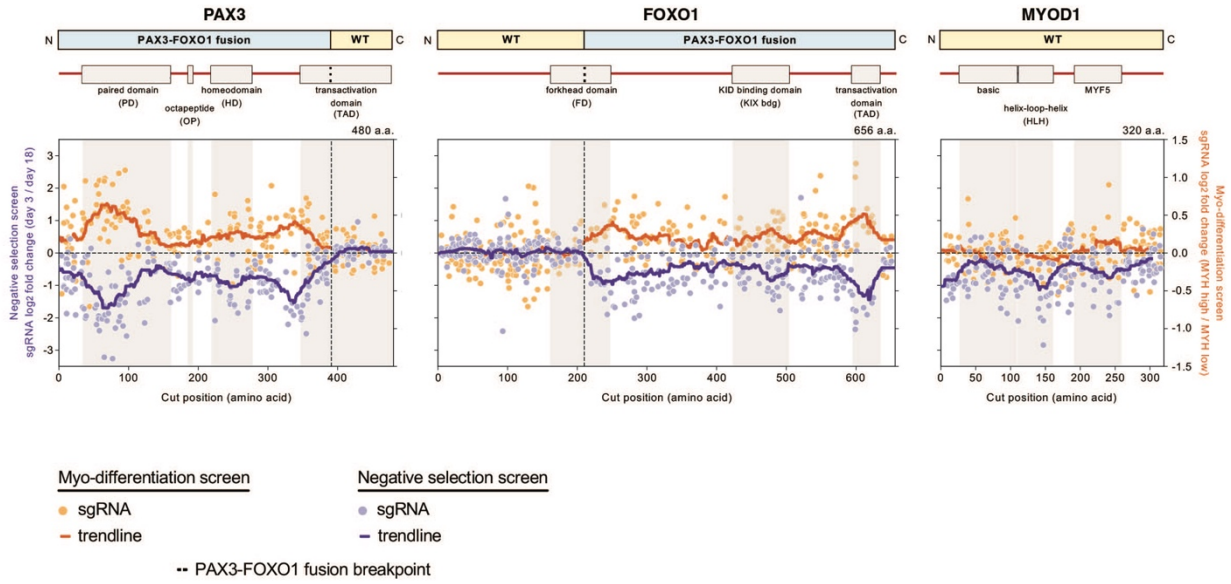


Figure S3. Negative selection and myo-differentiation CRISPR tiling scans. Expanded version of Fig. 2D. The results of the exon scan are plotted for PAX3 and FOXO1 polypeptides, including the segments absent from the PAX3-FOXO1 fusion. MYOD1 was used as a positive control for cell fitness and negative control for myo-differentiation. For the negative selection screen, the cells were harvested at day 3 poi (starting pool of sgRNAs) and at multiple timepoints thereafter to track negative selection of sgRNAs. The data depict $\log_2$-transformed fold change of sgRNA abundance between day 3 and day 18 samples plotted along the length of the PAX3-FOXO1 fusion polypeptide chain. Pale purple dots – individual sgRNAs; dark purple line – a trendline depicting rolling average over 35 amino acid window. For the myo-differentiation screen, the abundance of sgRNAs was compared between the MYH high and MYH low pools following cell sorting of the cells harvested 7 days poi. Pale orange dots - individual sgRNAs; dark orange line – a trendline depicting rolling average over 35 amino acid window.

Library name	Total # sgRNAs	# control sgRNAs	# targeted genes	Main class of targeted genes	# cells in MYH high gate	sgRNA representation
TF1	1096	67	104	transcription factors	1,180,820	1,077x
TF2	1139	67	108	transcription factors	1,555,884	1,366x
TF3	1141	67	108	transcription factors	1,149,607	1,007x
DR1	818	67	75	genes linked to myodifferentiation, PAX3-FOXO1 and/or RMS biology	870,306	1,063x
DR2	817	67	74	genes linked to myodifferentiation, PAX3-FOXO1 and/or RMS biology	1,149,319	1,406x
CF1	1188	67	107	cofactors, epigenetic modifiers	1,323,420	1,113x
CF2	1182	67	105	cofactors, epigenetic modifiers	1,145,196	968x
CF3	1148	67	106	cofactors, epigenetic modifiers	1,148,723	1,000x
KP1	1062	67	99	kinases, phosphatases, signaling	1,081,679	1,018x
KP2	1052	67	98	kinases, phosphatases, signaling	756,921	719x

Figure S4. Libraries used in myo-differentiation screens. sgRNAs targeting 967 genes essential in RMS were divided into 10 sub-libraries and screened for the expression of MYH reporter. Each sub-library contains universal positive and negative controls. Numbers of cells recovered in the MYH positive gate are indicated. Extended library information in Dataset S1.

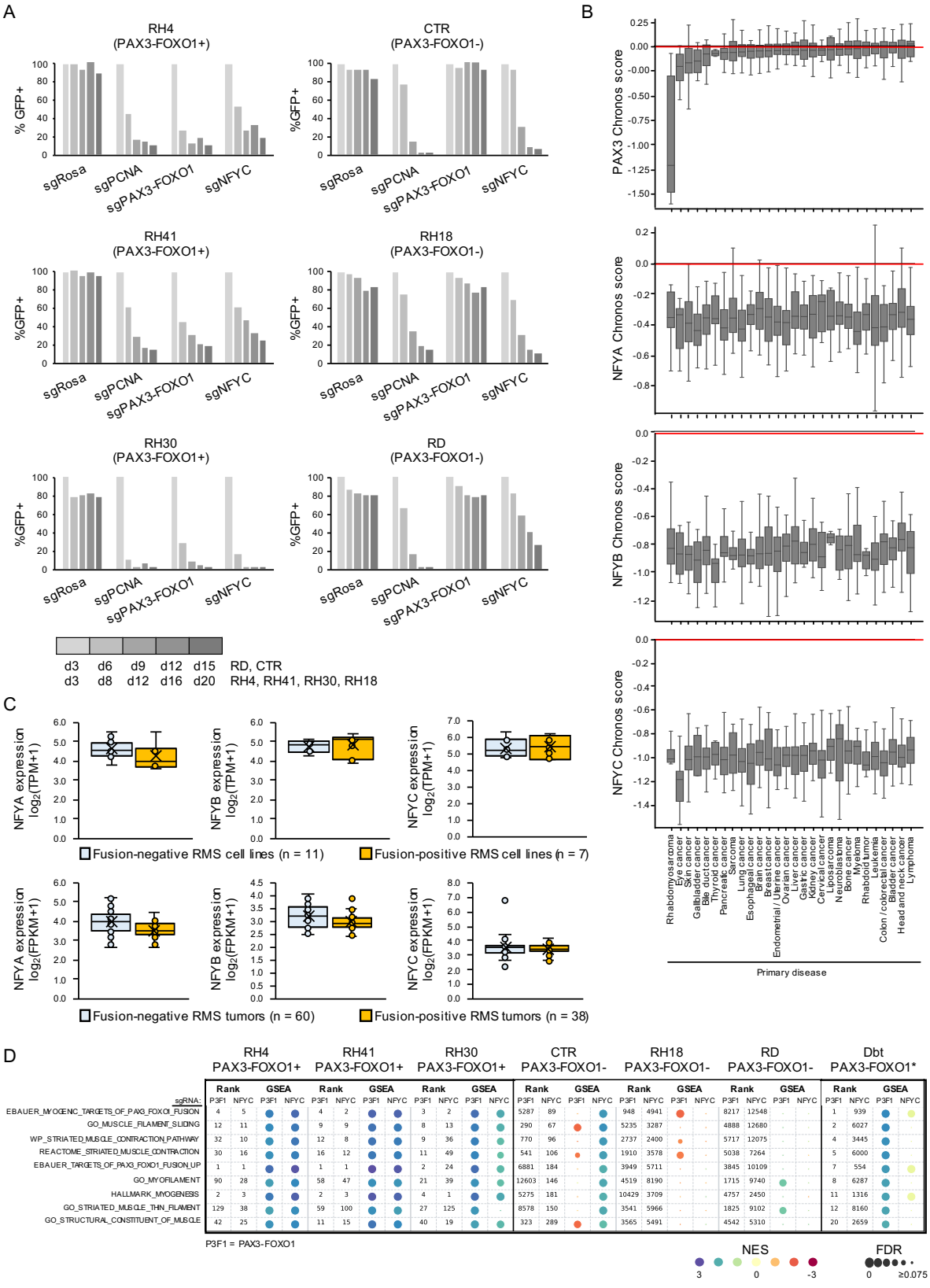


Figure S5. The effects of NF-Y loss in PAX3-FOXO1+ and – RMS cell lines. A. The results of GFP competition assays (n = 1) in 6 RMS cell lines following targeting of NFYC. Guides targeting PCNA

and Rosa were used as positive and negative controls respectively. sgRNA against PAX3-FOXO1 (targets wild-type PAX3 in fusion negative RMS) was used to distinguish between fusion positive and negative cell lines. **B.** Box-and-whiskers plots depicting essentiality scores of PAX3, NFYA, NFYB and NFYC in different types of cancers. CRISPR essentiality data from DepMap Public 22Q4+Score, Chronos. **C.** Quantification of NFYA, NFYB and NFYC transcript abundance in a panel of fusion-positive and fusion-negative RMS cell lines and tumors. **D.** A summary of GSEA analysis following NF-Y and PAX3-FOXO1 knockout in a panel of PAX3-FOXO1+ and – cell lines. NES, FDR and rank are shown for selected muscle-related gene signatures are shown for each condition.

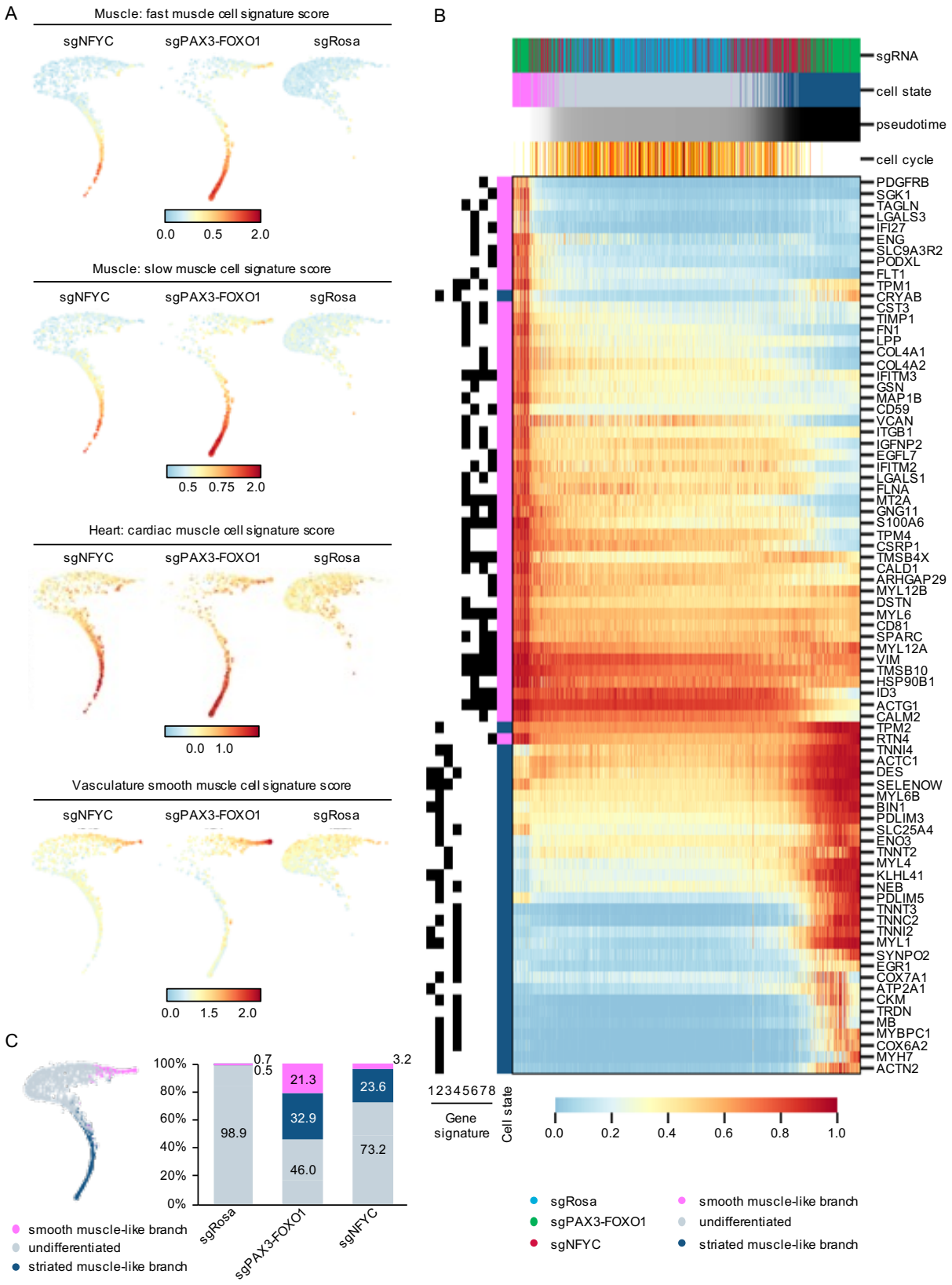


Figure S6. Gene expression changes in pseudotime following PAX3-FOXO1 and NF-Y knockout. **A.** Expanded version of Fig. 4H. Scores for selected muscle-relevant gene signatures are shown on t-

SNE plots to characterize the identity of each branch. **B.** Heatmap depicting gene expression changes in pseudotime following targeting with sgRosa, sgPAX3-FOXO1 and sgNFYC. Selected muscle-relevant genes, and signatures they are included in, are shown. Gene signatures: 1. Muscle-Fast Muscle Cell; 2. Muscle-Slow Muscle Cell; 3. Heart-Cardiac Muscle Cell; 4. Thymus-Fast Muscle Cell; 5. Vasculature-Smooth Muscle Cell; 6. Heart-Cardiac Endothelia Cell; 7. Lung-pericyte cell; 8. Lung-Capillary Aerocyte. **C.** Quantification of the smooth muscle-like, striated muscle-like and undifferentiated cells based on scRNA-seq signatures.

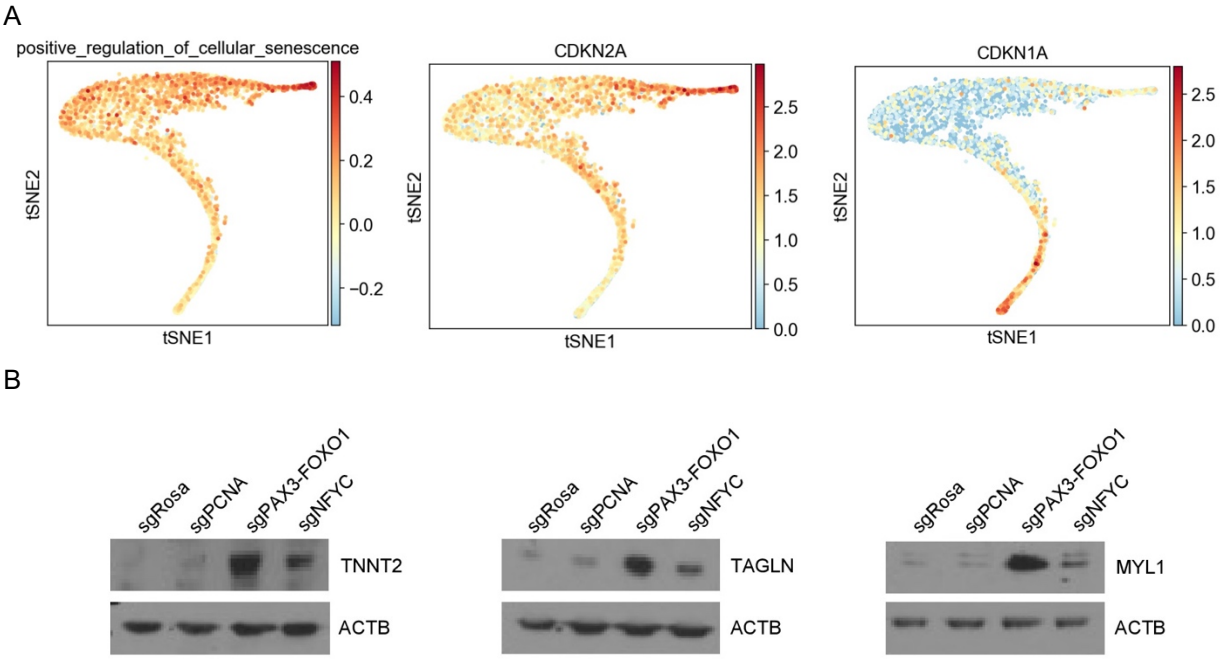


Figure S7. Additional data evaluating cell fate of RH4 cells following PAX3-FOXO1 and NF-Y inactivation. A. scRNA-seq evaluation of cellular senescence in perturbed RH4 RMS cells. The positive_regulation_of_cellular_senescence signature was obtained from the 2021 Gene Ontology database (GO:2000774). **B.** Western blot analysis of RH4 PAX3-FOXO1+ RMS cells 7 days post lentiviral transduction with sgRNAs targeting PAX3-FOXO1, NFYC, PCNA, or negative control sgRNA targeting the ROSA26 locus. Beta-actin (ACTB) antibody was used as a loading control. TNNT2 is a cardiac muscle marker. TAGLN is a smooth muscle marker. MYL1 is a skeletal muscle marker.

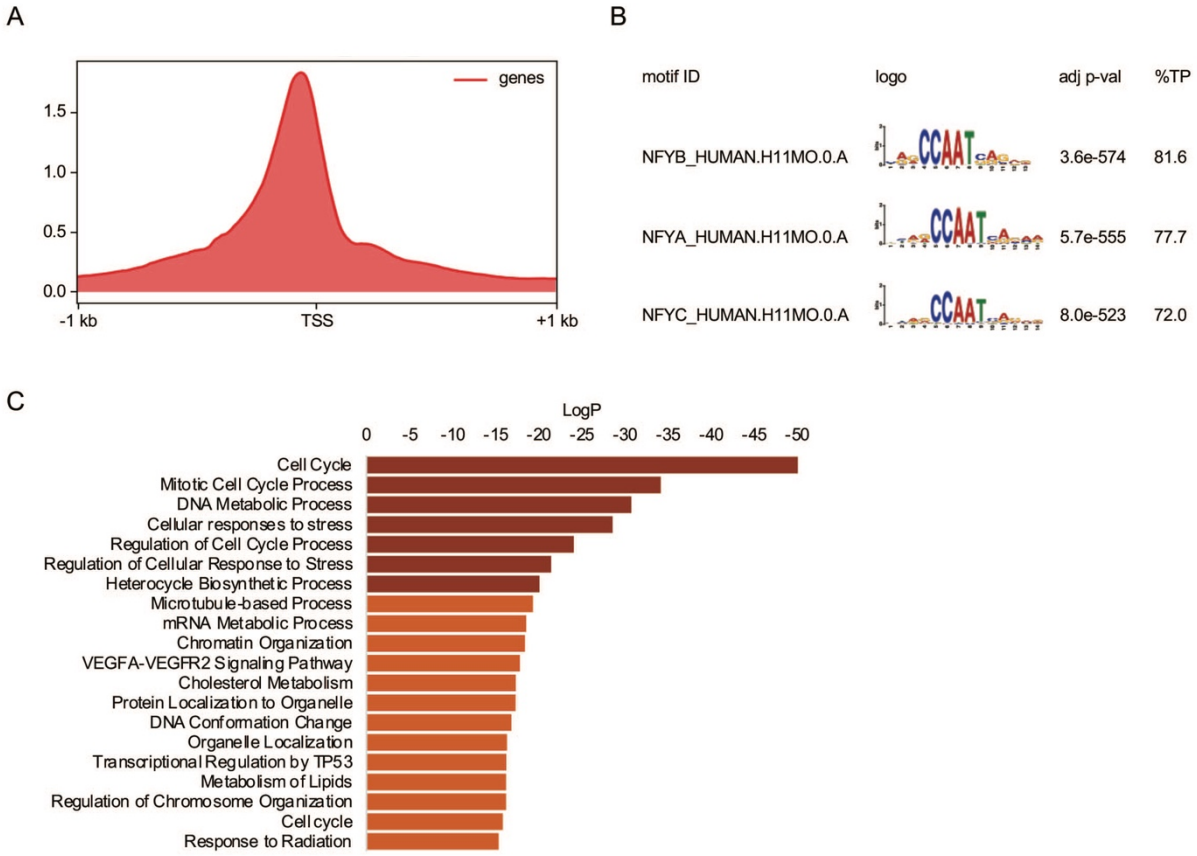


Figure S8. NF-Y does not function in the same transcriptional complex as PAX3-FOXO1. A. Profile of read density for NF-YB CUT&RUN, displaying a peak ~100 bp upstream of TSS. **B.** Motif analysis of the NFYB CUT&RUN data, displaying enrichment of the known NF-Y CCAAT binding motif within called peaks. **C.** Metascape analysis of genes associated with NF-Y binding.

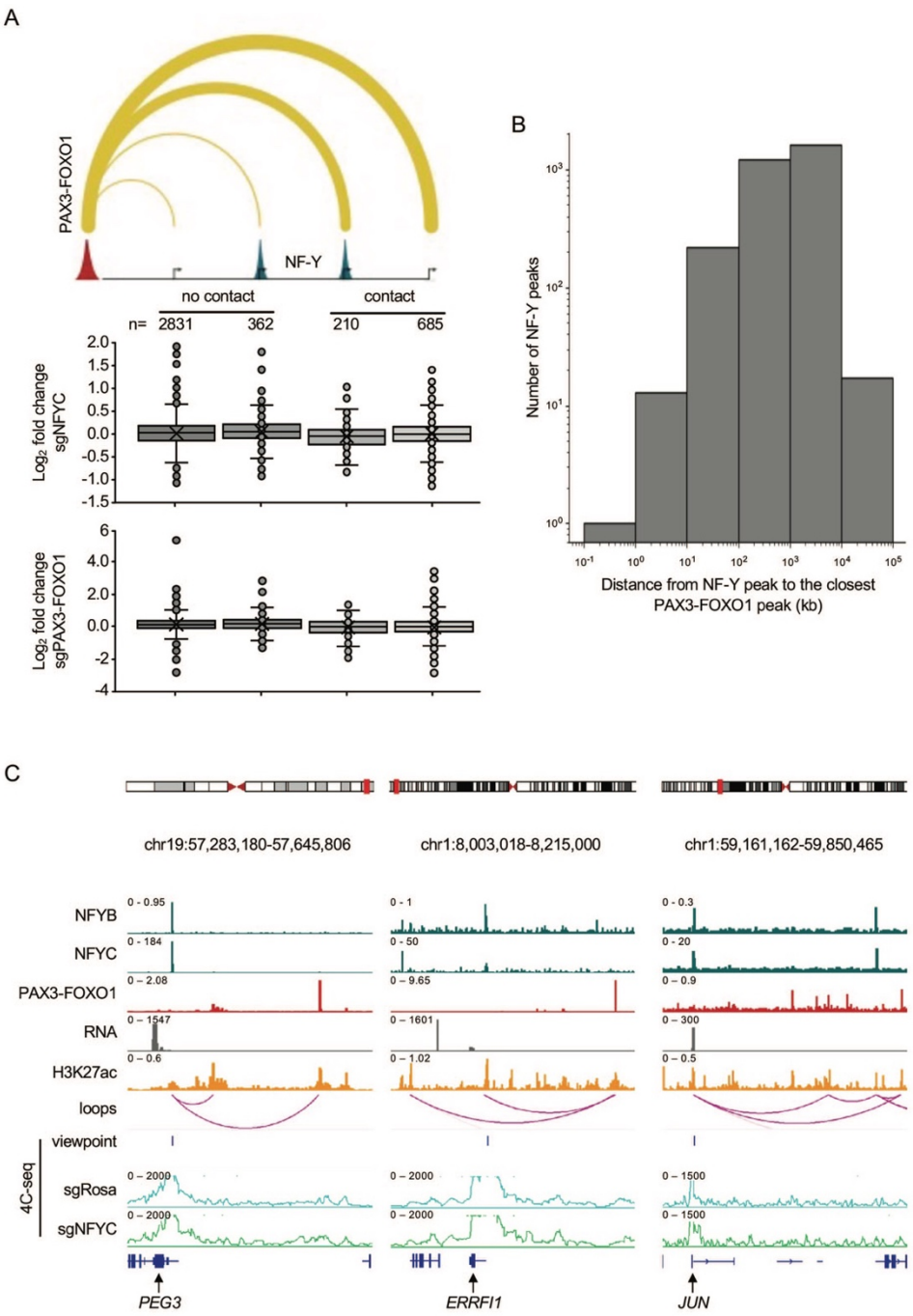


Figure S9. PAX3-FOXO1 and NF-Y interact on the genome in 3D via long range enhancer-promoter contacts. A. Cartoon quantifying the presence or absence of H3K27ac-mediated interactions genome-wide between sites bound by or lacking PAX3-FOXO1 and NF-Y. Box-and-whiskers blots depict gene expression changes following knockout of either PAX3-FOXO1 or NF-Y (day 7 poi) for each category of genes. **B.** Histogram representing $\log_{10}$-transformed distance in kilobases from each NF-Y peak to the closest PAX3-FOXO1 peak. **C.** Extension of Figure 5D, depicting NF-Y and PAX3-FOXO1 binding at 3 additional selected loci. H3K27ac marks and expression levels of neighboring genes are shown. 3D chromatin loops were identified based on AQuA-HiChIP data and validated using 4C-seq. The position of the viewpoint is indicated, as well as normalized read distribution averaged across two biological replicates for control cells (sgRosa) and NF-Y knockout cells (sgNFYC) analyzed at day 7 poi.

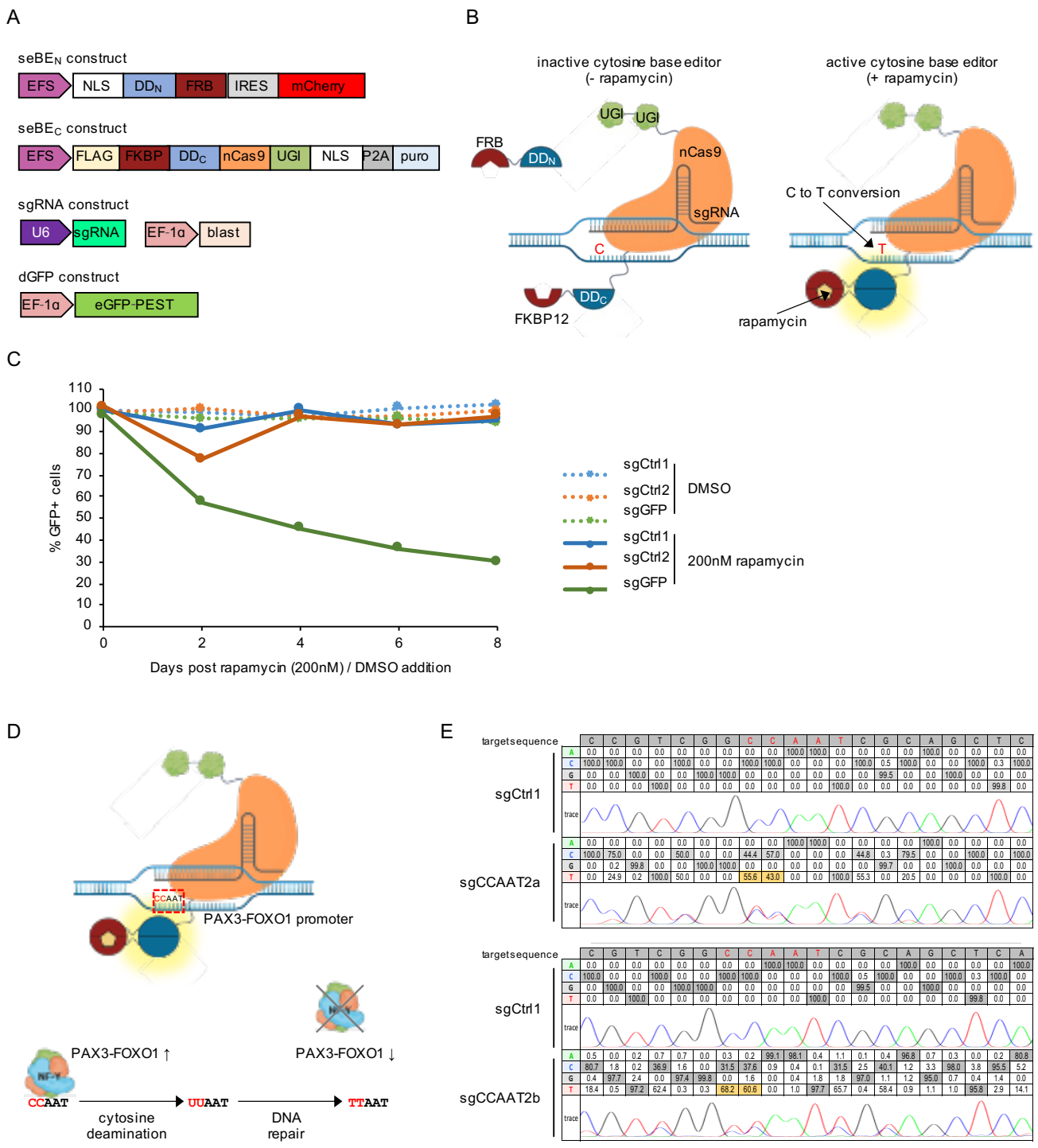


Figure S10. Inducible base editing in RMS cells. A. Schematic representation of the lentiviral constructs used in base editing experiments. **B.** Schematic representation of the rapamycin-controlled split base editing system. **C.** Validation of the inducible base editing system in RMS cells. RH30 PAX3-FOXO1+ RMS cells were transduced with a destabilized GFP reporter and sorted based on GFP signal to enrich for single copy reporter integration. The cells were then transduced with the two components of the split base editing system. A stable cell line was established by puromycin selection and sorting of mCherry+ cells. The cells were then transduced with a blasticidin resistance coupled sgRNA that results in a C to T conversion that introduces a premature stop codon within the destabilized GFP reporter or

control sgRNAs that result in point mutations within IDH2 or ROSA26 neutral loci that do not alter reporter expression. Cells transduced with sgRNAs were drug selected and subsequently cultured in the presence of DMSO or rapamycin. Base editing was monitored through tracking of GFP. The data was normalized to GFP% at the day of DMSO or rapamycin addition. **D.** Schematic representation of targeted mutagenesis of the CCAAT NF-Y binding motif within the promoter of PAX3-FOXO1 using inducible base editing. Editing of the NF-Y binding motif within the PAX3-FOXO1 promoter. RH30 cells carrying the two components of the split base editing system were transduced with a control sgRNA or a sgRNA that targets the second CCAAT motif within the PAX3-FOXO1 promoter and introduces targeted C to T conversions to ablate this NF-Y binding site. Transduced cells were then drug selected, and cultured in the presence of rapamycin to induce base editing. **E.** At day 8 post rapamycin addition, the cells were harvested, genomic DNA was extracted, and the region spanning the targeted CCAAT motif of interest was PCR amplified and Sanger sequenced. Trace deconvolution was performed in EditR to quantify the frequency of point mutations.

Supplementary Datasets:

Dataset S1. sgRNA sequences.

Dataset S2. Results of myo-differentiation screens.

Dataset S3. Results of RNA-seq analysis.

Dataset S4. NF-Y and PAX3-FOXO1 peaks.

Dataset S5. Chromatin looping analysis.

Dataset S6. Gene lists used to generate Fig. 6A.

Dataset S7. Signatures used in scRNA-seq analysis.

Dataset S8. Transcription factor binding motif analysis of the PAX3-FOXO1 promoter.

References

1. K. N. Berríos, *et al.*, Controllable genome editing with split-engineered base editors. *Nat. Chem. Biol.* **17**, 1262–1270 (2021).

2. F. M. Behan, *et al.*, Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).

3. N. V. Dharia, *et al.*, A first-generation pediatric cancer dependency map. *Nat. Genet.* **53**, 529–538 (2021).

4. J. Schindelin, *et al.*, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

5. P. Bankhead, *et al.*, QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

6. M. G. Kluesner, *et al.*, EditR: A Method to Quantify Base Editing from Sanger Sequencing. *CRISPR J.* **1**, 239–250 (2018).

7. R. C. Périer, T. Junier, C. Bonnard, P. Bucher, The eukaryotic promoter database (EPD): Recent developments. *Nucleic Acids Res.* **27**, 307–309 (1999).

8. S. Sissaoui, *et al.*, Genomic Characterization of Endothelial Enhancers Reveals a Multifunctional Role for NR2F2 in Regulation of Arteriovenous Gene Expression. *Circ. Res.* **126**, 875–888 (2020).

9. S. Andrews, FastQC: A Quality Control Tool for High Throughput Sequence Data.

10. D. Kim, J. M. Paggi, C. Park, C. Bennett, S. L. Salzberg, Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

11. S. Anders, P. T. Pyl, W. Huber, HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

12. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 1–21 (2014).

13. A. Dobin, *et al.*, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

14. Y. Liao, G. K. Smyth, W. Shi, The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res.* **47**, e47 (2019).

15. B. Li, C. N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011).

16. A. Subramanian, *et al.*, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

17. G. Biffi, *et al.*, Il1-induced Jak/STAT signaling is antagonized by TGFβ to shape CAF heterogeneity in pancreatic ductal adenocarcinoma. *Cancer Discov.* **9**, 282–301 (2019).

18. Z. Xie, *et al.*, Gene Set Knowledge Discovery with Enrichr. *Curr. Protoc.* **1**, e90 (2021).

19. E. Y. Chen, *et al.*, Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).

20. J. Cao, *et al.*, A human cell atlas of fetal gene expression. *Science* **370** (2020).

21. A. I. Su, *et al.*, A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **101**, 6062–6067 (2004).

22. The Tabula Sapiens Consortium, The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. *Science* (2022) (March 8, 2023).

23. M. Setty, *et al.*, Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).

24. D. van Dijk, *et al.*, Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell* **174**, 716-729.e27 (2018).

25. I. Tirosh, *et al.*, Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

26. P. J. Skene, S. Henikoff, An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856 (2017).

27. A. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence (2014).

28. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

29. Y. Zhang, *et al.*, Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

30. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

31. 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).

32. R. C. McLeay, T. L. Bailey, Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).

33. Y. Zhou, *et al.*, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10** (2019).

34. B. E. Gryder, *et al.*, Histone hyperacetylation disrupts core gene regulatory architecture in rhabdomyosarcoma. *Nat. Genet.* **51**, 1714–1722 (2019).

35. G. Geeven, H. Teunissen, W. de Laat, E. de Wit, peakC: a flexible, non-parametric peak calling package for 4C and Capture-C data. *Nucleic Acids Res.* **46**, e91 (2018).

36. P. H. L. Krijger, G. Geeven, V. Bianchi, C. R. E. Hilvering, W. de Laat, 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. *Methods* **170**, 17–32 (2020).

37. D. R. Zerbino, N. Johnson, T. Juettemann, S. P. Wilder, P. Flicek, WiggleTools: parallel processing of large collections of genome-wide datasets for visualization and statistical analysis. *Bioinformatics* **30**, 1008–1009 (2014).

38. A. S. Brohl, *et al.*, Immuno-transcriptomic profiling of extracranial pediatric solid malignancies. *Cell Rep.* **37**, 110047 (2021).

39. J. T. Robinson, *et al.*, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).