Figure S1

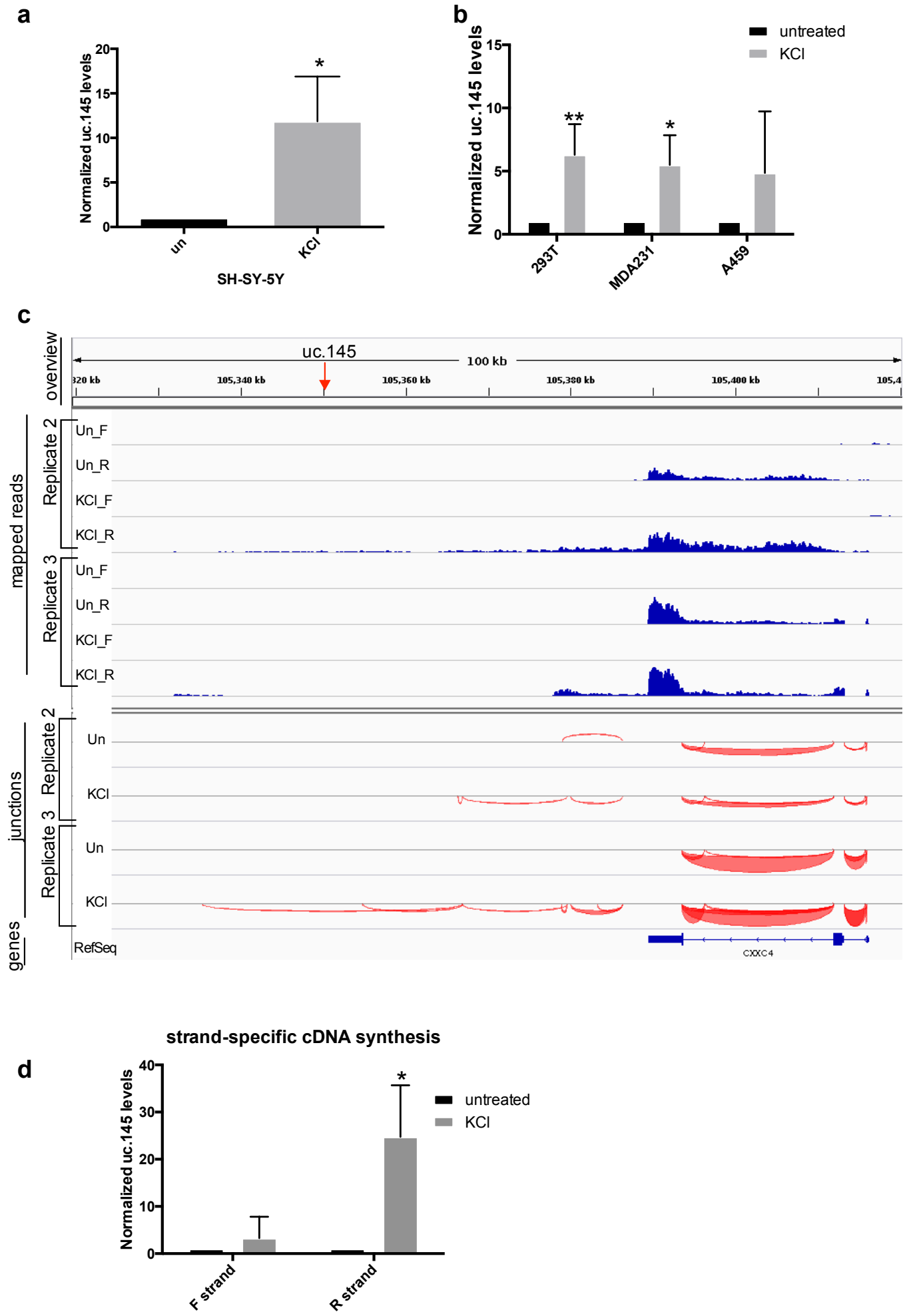
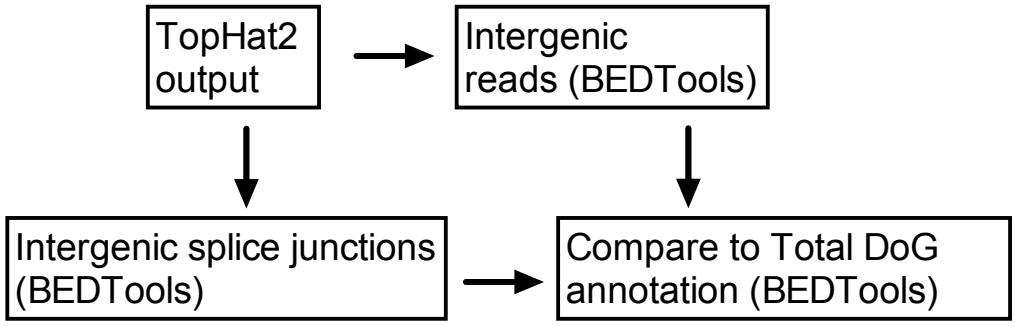
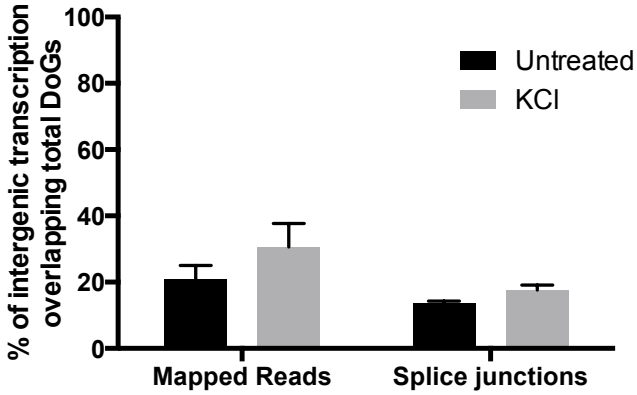

a
Normalized uc.145 levels
20
15
10
5
0
*
un
KCl
SH-SY-5Y
b
Normalized uc.145 levels
15
10
5
0
untreated
KCl
**
*
293T
MDA231
A459
c
overview
uc.145
100 kb
320 kb
105,340 kb
105,360 kb
105,380 kb
105,400 kb
mapped reads
Replicate 2
Un_F
Un_R
KCl_F
KCl_R
Replicate 3
Un_F
Un_R
KCl_F
KCl_R
junctions
Replicate 2
Un
KCl
Replicate 3
Un
KCl
genes
RefSeq
CXXC4
strand-specific cDNA synthesis
d
Normalized uc.145 levels
40
30
20
10
0
*
untreated
KCl
F strand
R strand

Figure S2

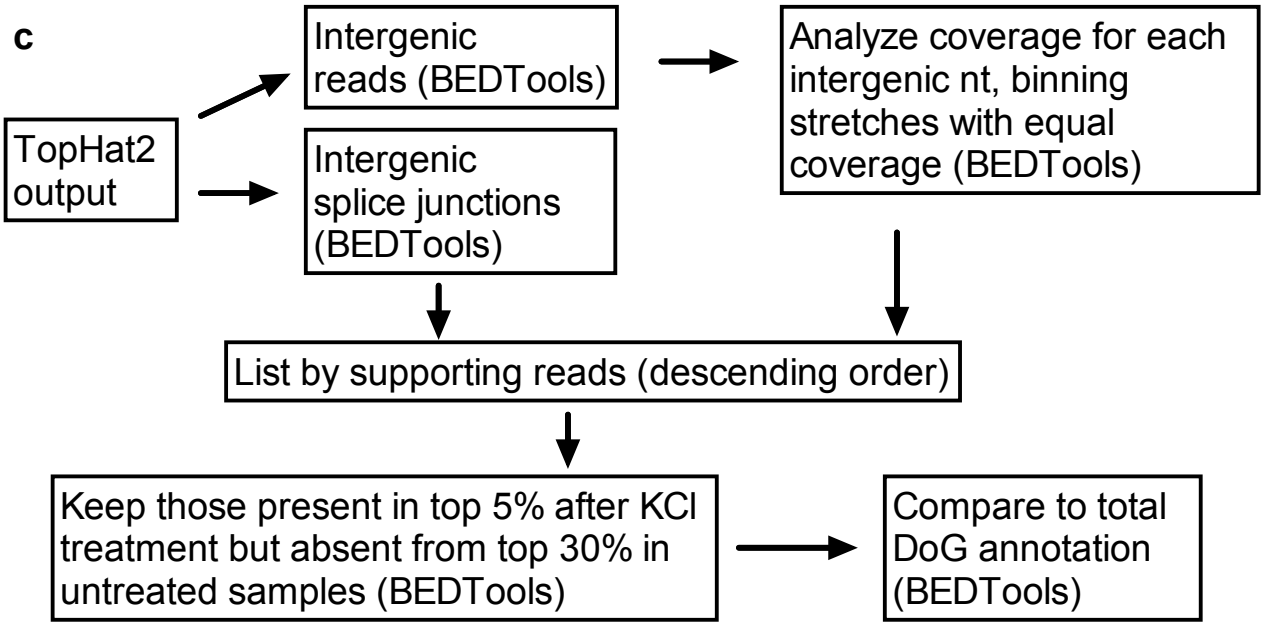
a



b



c



d

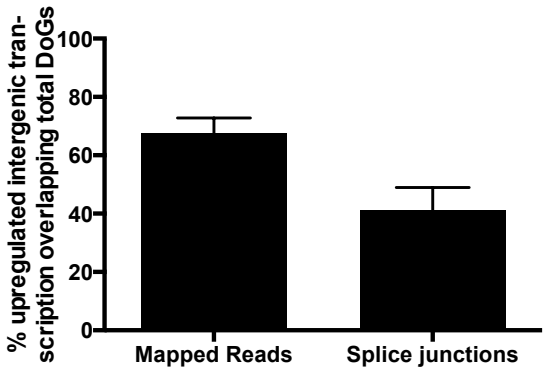
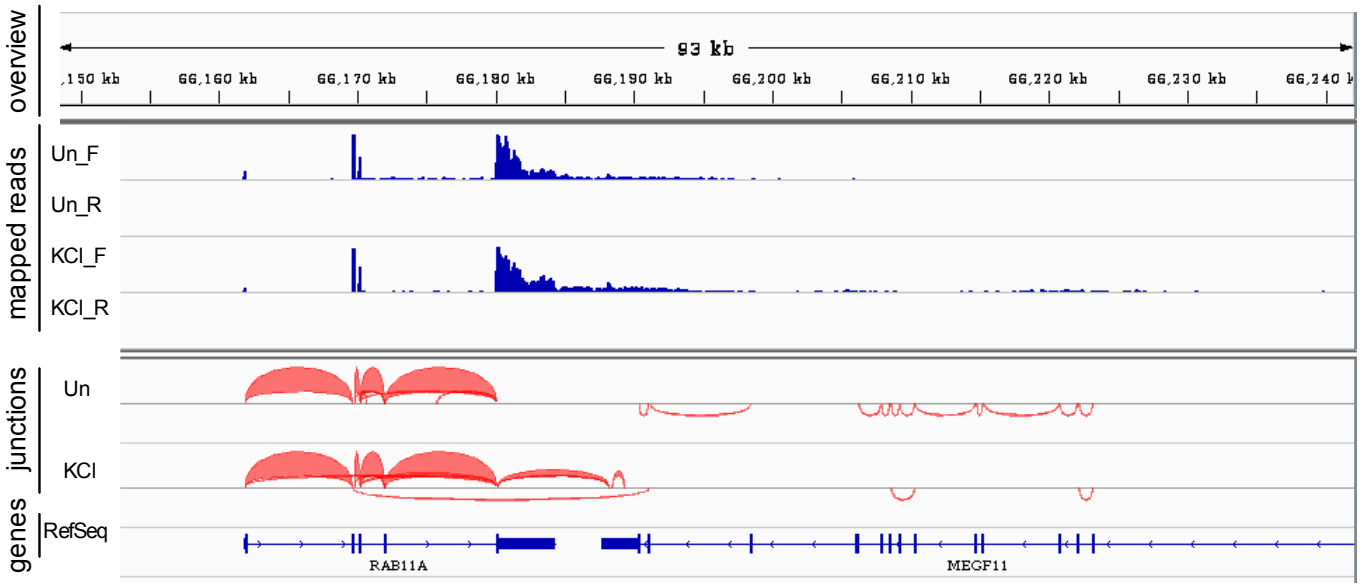


Figure S3

a Mapped reads and splice junctions for doRAB11A



b Mapped reads and splice junctions for doSERBP1

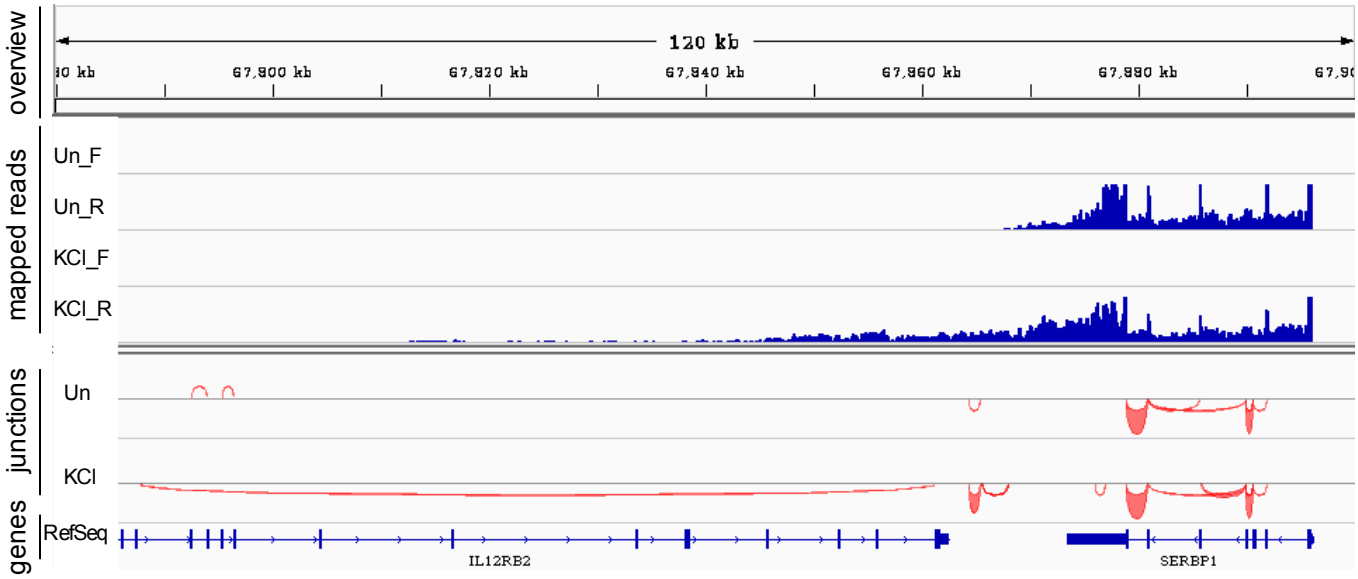


Figure S4

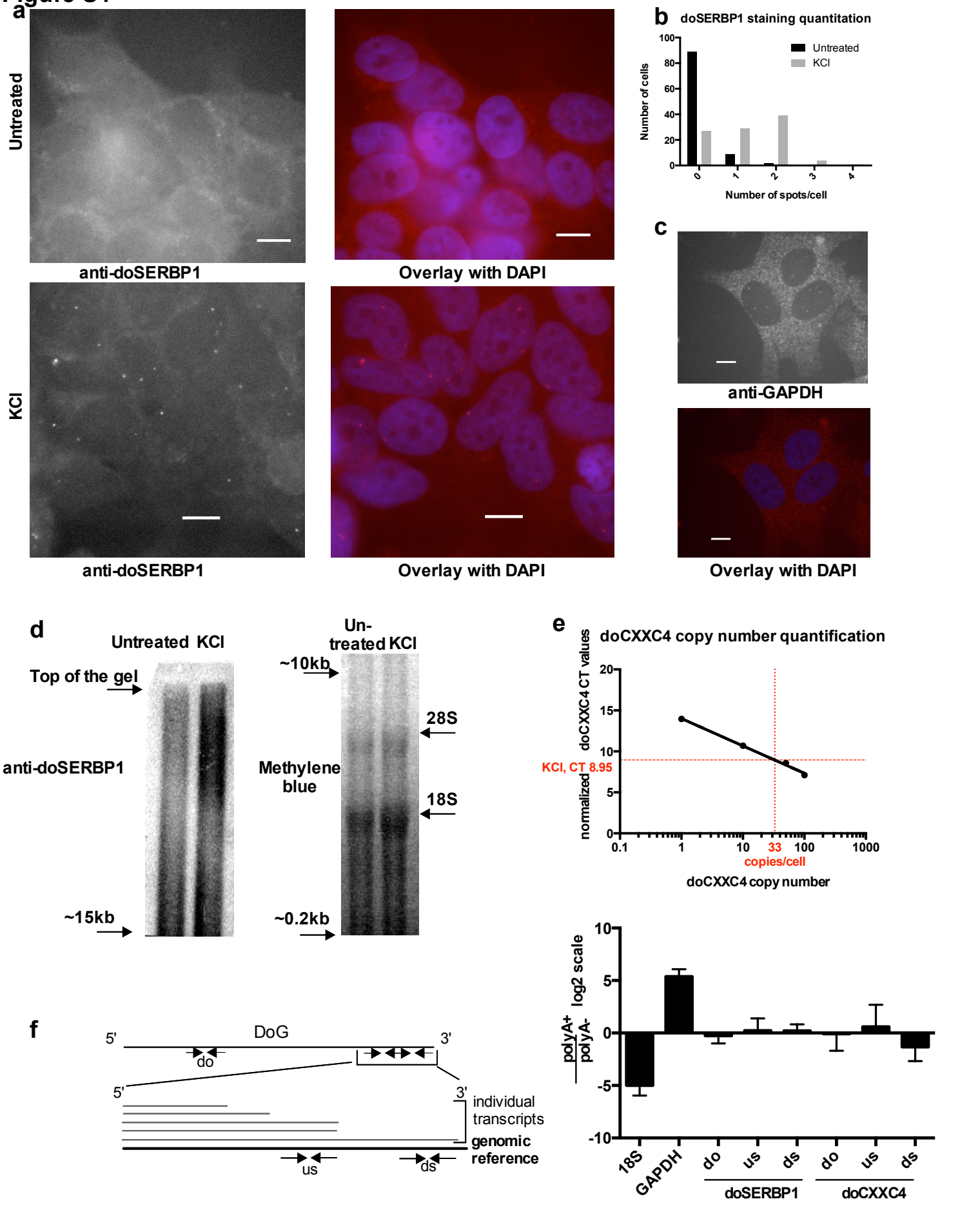
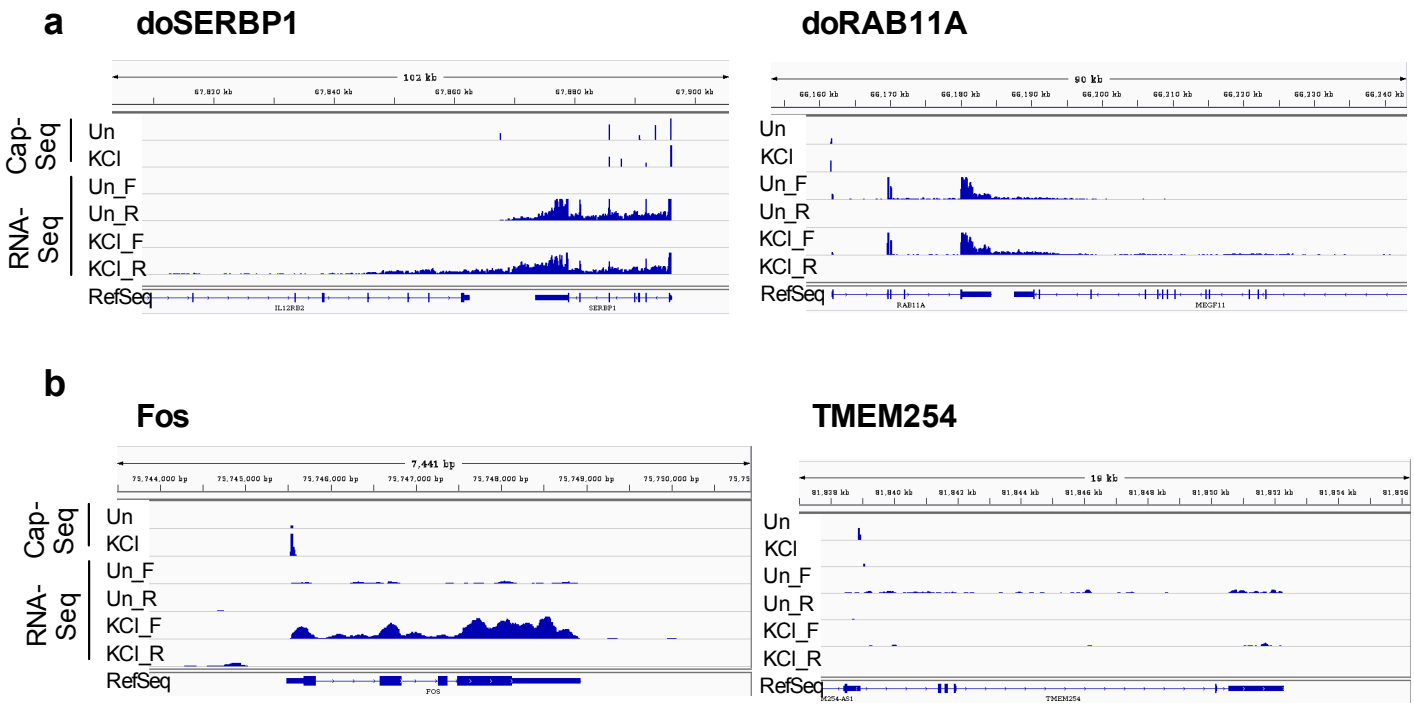
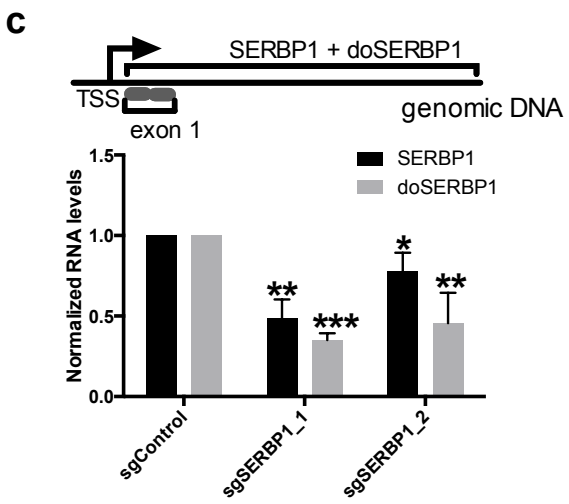

a
Untreated
KCl
anti-doSERBP1
Overlay with DAPI
anti-doSERBP1
Overlay with DAPI
b
doSERBP1 staining quantitation
Untreated
KCl
Number of cells
Number of spots/cell
c
anti-GAPDH
Overlay with DAPI
d
Untreated KCl
Top of the gel
anti-doSERBP1
~15kb
Un-
treated KCl
~10kb
28S
Methylene
blue
18S
~0.2kb
e
doCXXC4 copy number quantification
doCXXC4 CT values
normalized
KCl, CT 8.95
copies/cell
doCXXC4 copy number
f
DoG
5'
3'
do
individual
transcripts
genomic
reference
us
ds
polyA+ polyA- log2 scale
18S
GAPDH
do
us
ds
doSERBP1
doCXXC4

Figure S5


a
doSERBP1
doRAB11A
Cap-Seq
RNA-Seq
Un
KCl
Un_F
Un_R
KCl_F
KCl_R
RefSeq
b
Fos
TMEM254


c
SERBP1 + doSERBP1
TSS
exon 1
genomic DNA
SERBP1
doSERBP1
Normalized RNA levels
sgControl
sgSERBP1_1
sgSERBP1_2

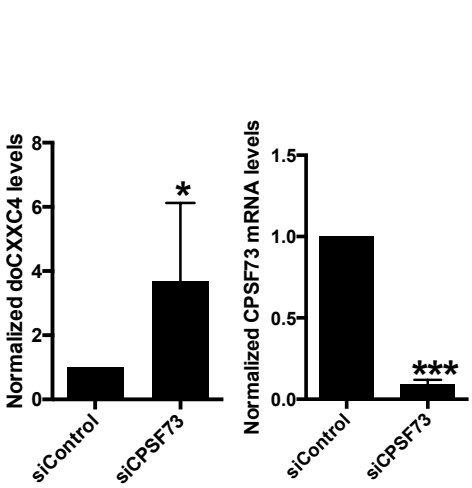

d
Normalized doCXXC4 levels
siControl
siCPSF73
Normalized CPSF73 mRNA levels

Figure S6

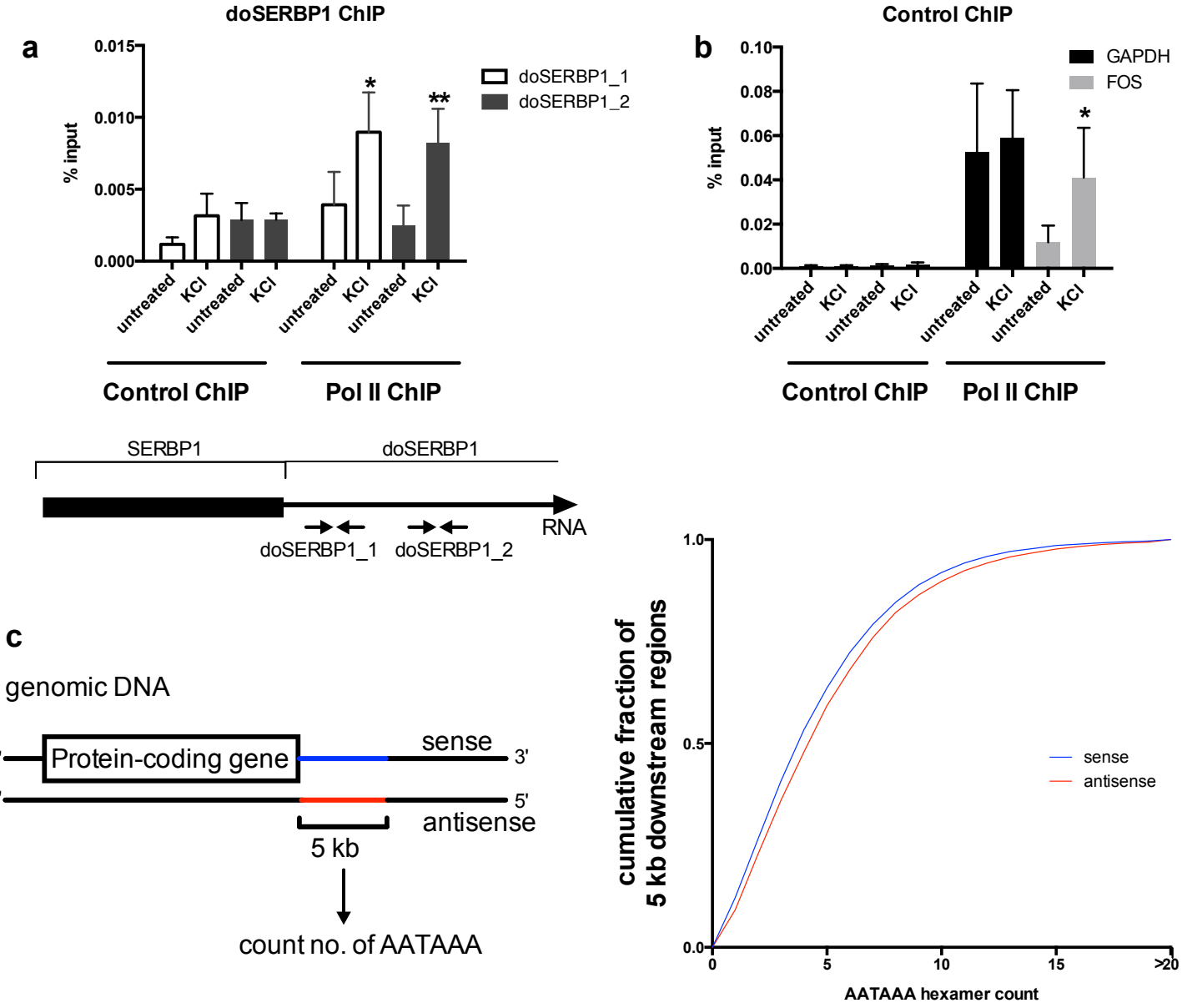
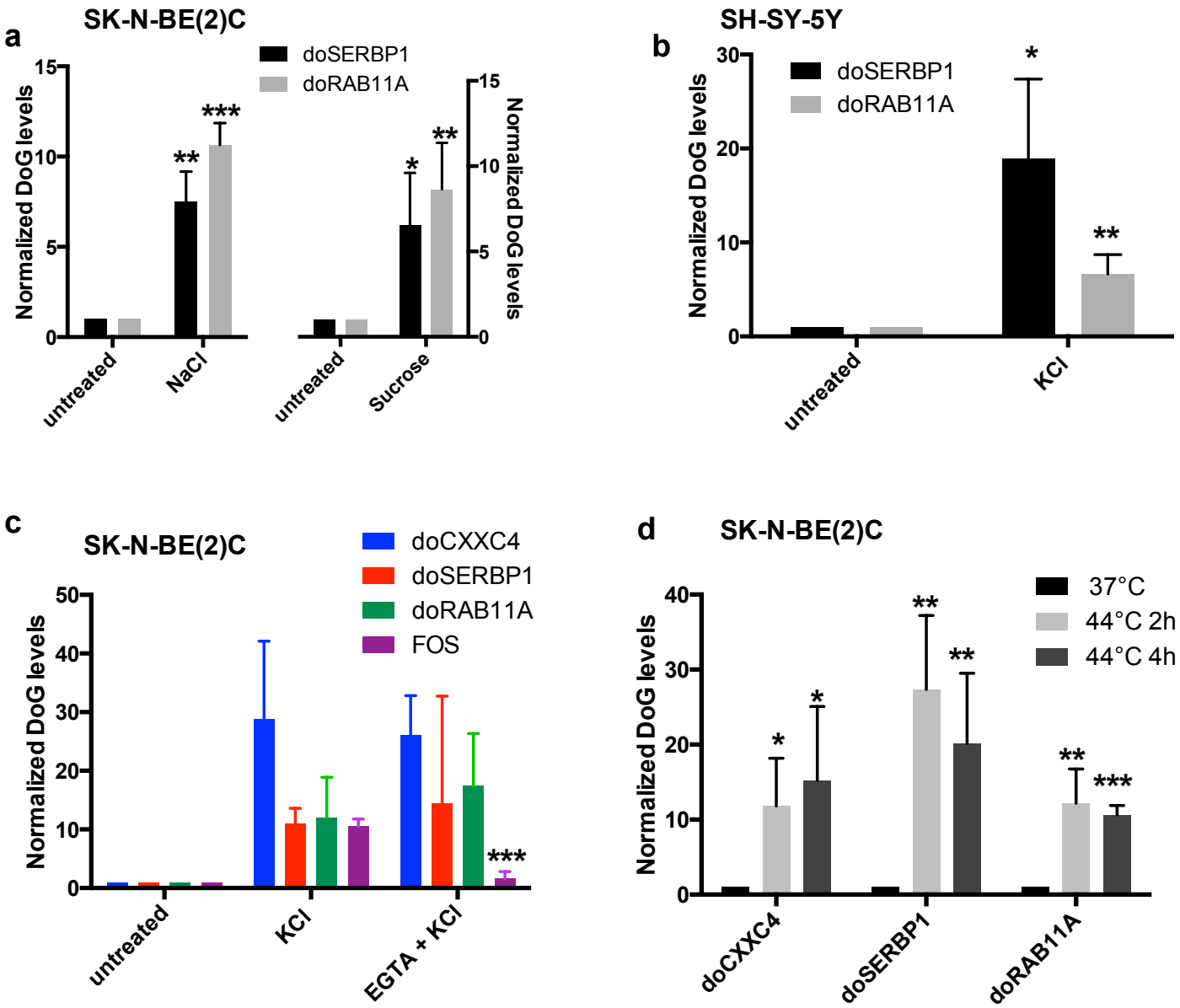

doSERBP1 ChIP
a
% input
0.015
0.010
0.005
0.000
doSERBP1_1
doSERBP1_2
*
**
untreated
KCl
Control ChIP
Pol II ChIP
Control ChIP
b
% input
0.10
0.08
0.06
0.04
0.02
0.00
GAPDH
FOS
*
Control ChIP
Pol II ChIP
SERBP1
doSERBP1
RNA
doSERBP1_1
doSERBP1_2
c
genomic DNA
5'
Protein-coding gene
3'
sense
antisense
5 kb
count no. of AATAAA
cumulative fraction of 5 kb downstream regions
1.0
0.5
0.0
sense
antisense
0
5
10
15
>20
AATAAA hexamer count

Figure S7


a
SK-N-BE(2)C
doSERBP1
doRAB11A
Normalized DoG levels
15
10
5
0
**
***
*
**
untreated
NaCl
Sucrose
b
SH-SY-5Y
30
20
10
0
*
**
KCl
c
SK-N-BE(2)C
doCXXC4
doSERBP1
doRAB11A
FOS
50
40
30
20
10
0
***
EGTA + KCl
d
SK-N-BE(2)C
37°C
44°C 2h
44°C 4h
doCXXC4
doSERBP1
doRAB11A

Supplemental Figure legends

Figure S1: lncRNA uc.145 is induced by osmotic stress and is part of a DoG transcript. Related to Figure 1.

a) uc.145 is upregulated in the neuroblastoma cell line SH-SY-5Y cells incubated with 120 mM KCl for 1 h, analyzed by qRT-PCR, n=3. All figures show mean and SD, *p*-values of 0.05-0.01 are denoted by *, 0.01-0.001 by **, and ≤ 0.001 by ***. We note that a slightly higher KCl dose is required for maximal induction in SH-SY-5Y (120 mM versus 80 mM in SK-N-BE(2)C) cells and that the average induction obtained in SH-SY-5Y cells is 10-fold, rather than 20- to 30-fold as in SK-N-BE(2)C. b) qRT-PCR quantitation showing significant KCl induction, albeit at lower levels than in neuroblastoma cell lines, of uc.145 in 293T (human embryonic kidney-derived) and MDA231 (breast cancer) but no significant induction in A459 (lung cancer) cells. c) Mapped reads and splice junctions from TopHat2 surrounding uc.145 (indicated by a red arrow) in two additional biological replicates (not shown in Figure 1) of the RNA-Seq experiment. Top panel ("overview") shows genomic coordinates and the red arrow indicates the location of uc.145. Second panel shows mapped reads, with reads mapping to the forward (_F) and reverse (_R) strands indicated. Third panel shows splice junctions ("junctions"), with splice junctions from the reverse strand depicted below and splice junctions from the forward strand above the indicator line. The bottom panel shows RefSeq genes. Un stands for untreated. d) qPCR quantitation following strand-specific cDNA synthesis, using primers specific for the uc.145 region transcribed either from the strand opposite to *CXXC4* (the forward DNA strand, F) or the same strand as *CXXC4* (the reverse DNA strand, R) normalized to *GAPDH* mRNA primed by random primers. Transcripts are detectible

from both strands, in line with previous reports of uc.145 expression; however, only the transcript from the same strand as *CXXC4* (the reverse strand) is induced by KCl (80 mM, 1 h), n=3.

Figure S2: A significant percentage of intergenic transcription is attributable to DoGs. Related to Figure 2.

a) Flowchart describing how intergenic transcription was compared to the total DoG annotation (see Fig. 2c). b) Percentage of intergenic mapped reads and predicted splice junctions overlapping DoG regions, n=3. c) Flowchart describing how KCl-induced intergenic transcription was identified and compared to the total DoG annotation. d) Fraction of upregulated intergenic transcription defined by mapped reads and predicted splice junctions that overlaps DoG regions, n=3.

Figure S3: Mapped reads and splice junctions for *doRAB11A* and *doSERBP1.* Related to Figure 3.

a) Mapped reads and splice junctions for *doRAB11A*. Top panel ("overview") shows genomic coordinates. Second panel shows mapped reads, with those from the forward (_F) and reverse (_R) strands indicated. Third panel shows splice junctions ("junctions"), with those from the reverse strand depicted below and those from the forward strand above the indicator line. Bottom panel shows RefSeq genes. *RAB11A* is encoded on the forward strand, whereas *MEGF11* is encoded on the reverse strand relative to *doRAB11A*. Un stands for untreated. b) Mapped reads and splice junctions for *doSERBP1*, shown as for *doRAB11A* in a.

Figure S4: Further characterization of selected DoGs. Related to Figure 3.

a) FISH using Stellaris probes for *doSERBP1* in untreated and KCl-treated SK-N-BE(2)C cells confirms KCl-mediated induction and nuclear localization of *doSERBP1*. Images are at 40x magnification and scale bars indicate 10 μM. See Extended Experimental Procedures for more information on FISH probes. b) Quantitation of number of nuclear spots observed per cell for 100 counted cells after *doSERBP1* staining in untreated or KCl-treated cells. c) Stainings for *GAPDH* in untreated cells as an additional control for staining protocol efficiency. Images are at 40x magnification and scale bars indicate 10 μM. See Extended Experimental Procedures for more information on FISH probes. d) Northern blot for *doSERBP1* on nuclear RNA from untreated or KCl-treated SK-N-BE(2)C cells detects a smear at the top of the gel after KCl-treatment, in line with the large size (~80 kb) and heterogeneous end-points of *doSERBP1*. Note that because of the large size of *doSERBP1,* it is unclear whether full-length transcripts enter the gel or whether the transcripts detected represent shorter variants and/or degradation products. However, the clearly detectable signal after KCl-treatment validates a large, heterogenous, KCl-inducible transcript containing the *doSERBP1* sequence. e) Copy number estimation for *doCXXC4.* A portion of *doCXXC4* RNA was *in vitro* transcribed in the presence of trace amounts of ^{32}P-labeled uridine to enable calculation of the number of *in vitro* products and then spiked into RNA from untreated SK-N-BE(2)C cells at 1, 10, 50 or 100 copies/cell. The RNA was reverse transcribed to cDNA and analyzed by qPCR using primers for *doCXXC4, GAPDH* and *GUSB*. CT values for *doCXXC4* were normalized to the mean of the CT values for *GAPDH* and *GUSB*. The resulting CT values were plotted to generate a standard curve that was used to correlate the normalized CT value of 8.95 obtained from RT-qPCR of *doCXXC4* from a KCl-treated sample with a copy number of 33 molecules of *doCXXC4*/cell. The ~40-fold

induction of *doCXXC4* in KCl-treated compared to untreated cells in this experiment (data not shown) indicates that the copy number of *doCXXC4* in untreated cells is close to or less than 1. N=1. f) DoGs are partially polyadenylated. polyA$^+$ and polyA$^-$ RNA was purified and analyzed by qRT-PCR. Designing primers to detect the ends of DoGs was difficult because of their heterogeneous endpoints; their length (~50-80 kb) also increased the likelihood of RNA breakage during purification. Therefore, we used three primer pairs. One primer pair (downstream, "ds") recognized the extreme 3′ end of each DoG, after which there were no mapped reads in any sample. A second primer pair was located just upstream ("us") of a significant drop in read support, and a third primer pair ("do") spanned a DoG splice junction close to the end of the associated mRNA. Left panel is a schematic of the 3′ portions of *doCXXC4* and *doSERBP1* with all primer pairs indicated by arrows in the top panel and a zoomed-in view below highlighting their heterogeneous 3′-termini; with the "us" and "ds" primer pairs indicated. The diagram is not to scale. While *18S* rRNA was clearly enriched in polyA$^-$ and *GAPDH* mRNA in polyA$^+$ RNA, all primer pairs targeting DoGs detected approximately equal levels in both fractions. N=3.

Figure S5: DoGs are generated by readthrough transcription. Related to Figure 4.

a) CapSeq data for *SERBP1-doSERBP1* (left panel) and *RAB11A-doRAB11A* identify the annotated TSS for *SERBP1* and *RAB11A* both with and without KCl treatment. For *RAB11A*, no downstream TSSs are detected. For *SERBP1*, 4 downstream TSSs are found in the untreated and 3 in the KCl-treated samples; of these, 1 is found in both samples while the others are unique. Because additional TSSs are found in the untreated sample, and the TSS found in both samples is not upregulated after KCl

treatment, it is unlikely that *doSERBP1* is generated by KCl-induced transcriptional initiation at any of these TSSs. Genomic coordinates are shown on top, CapSeq data in the upper panel, mapped reads from the RNA-Seq experiment in the following panel, and RefSeq genes in the bottom panel, annotated as in Figure S1c. Un stands for untreated. b) CapSeq data displayed as in a for two control genes: peak reads mapping to the *FOS* TSS (left panel) demonstrate detection of the expected KCl-mediated increase in transcriptional initiation at the *FOS* TSS. Peak reads mapping to the TMEM254 TSS (right panel) demonstrate that transcripts expressed at less than 0.5 RPKM can be detected by the CapSeq method (the RPKM for the first 100 nt of TMEM254 is 0.5 while the RPKM over the whole TMEM254 transcript, including introns, is 0.03). c) Inhibiting transcription through the *SERBP1* gene by targeting mutant Cas9 fused to KRAB to the first exon of *SERBP1* in 293T cells using two different sgRNAs reduces the levels of both *SERBP1* and *doSERBP1*, assessed by qRT-PCR quantitation. n=3. All figures show mean and SD, *p*-values are denoted as in Figure S1. RNA levels are normalized to the average of *GAPDH* and *GUSB* mRNAs in a, and to *GAPDH* mRNA in b. Top panel shows the locations of the sgRNAs (short grey bars) relative to the *SERBP1* TSS. Diagram is not to scale. d) *doCXXC4* is induced about 4-fold upon siRNA knockdown of CPSF73, as analyzed by qRT-PCR, n=5. Left panel shows *doCXXC4* levels, while right panel shows efficient knockdown of *CPSF73* mRNA levels.

Figure S6: DoGs are induced at the level of transcription, not RNA stability. Related to Figure 5.

a) Pol II ChIP followed by qPCR for *doSERBP1* using 2 separate primer pairs (arrows in the bottom panel) detect increased Pol II occupancy of the *doSERBP1* region after

KCl treatment (80 mM, 1 h). N=4, p-values denoted as in Figure S1, all panels show mean and SD. The figure shows % input. The diagram in the bottom panel is not to scale. b) as in a, but using primers detecting promoters of the control genes *FOS* and *GAPDH*: Pol II occupancy increases at the *FOS,* but not the *GAPDH*, promoter after KCl treatment, as expected. c) Cumulative plot (right panel) of the AATAAA count in 5 kb regions downstream of all human protein-coding genes on the sense or antisense strand (left panel). The curve for sense regions is shifted to the left, indicating fewer AATAAA signals per region on the sense versus the antisense strand.

Figure S7: DoGs are induced by osmotic stress in SH-SY-5Y cells, independent of extracellular Ca^{2+}, and by heat shock. Related to Figure 6 and Discussion.
a) *doSERBP1* and *doRAB11A* are induced after treatment in SK-N-BE(2)C cells with 80 mM NaCl and 200 mM sucrose for 1 h, analyzed by qRT-PCR, n=3. All panels show mean and SD, *p*-values are denoted as in Figure S1. RNA levels were normalized to *GAPDH* mRNA in a and c and to the average of *GAPDH* and *GUSB* mRNAs in b and d. b) *doSERBP1* and *doRAB11A* are upregulated in SH-SY-5Y cells incubated with 120 mM KCl for 1 h, analyzed by qRT-PCR, n=3. c) Pretreatment of SK-N-BE(2)C cells with EGTA (6 mM, 5 min) followed by KCl treatment (80 mM, 1 h) does not affect induction of *doCXXC4, doSERBP1* or *doRAB11A*, while induction of *FOS*, a known KCl-responsive gene in neuronal cells, is reduced, as determined by qRT-PCR quantitation; n=3, statistical significance indicated compares KCl and EGTA+KCl. d) DoG levels are induced by heat shock, as demonstrated by qRT-PCR quantitation of SK-N-BE(2)C cells exposed to 44°C for 2 or 4 h, n=3.

Table S1: Coordinates and information for all annotated DoGs. Related to Figure 2.

Table S2: Statistics of TSSs. Related to Figure 4.

	TSS anywhere in gene/region (+ 2 kb downstream*)	TSS in 3′-UTR + 2 kb downstream
% of TSS in the untreated sample that was found only in or was upregulated >3-fold in the untreated sample **Total DoG associated genes**	49%	62%
% of TSS in the KCl-treated sample that was found only in or was upregulated >3-fold in the KCl-treated sample **Total DoG associated genes**	31%	41%
% of TSS in the untreated sample that was found only in or was upregulated >3-fold in the untreated sample **Stringent DoG associated genes**	48%	59%
% of TSS in the KCl-treated sample that was found only in or was upregulated >3-fold in the KCl-treated sample **Stringent DoG associated genes**	36%	44%
% of TSS in the untreated sample that was only in or was upregulated >3-fold in the untreated sample **Total DoG regions**	48%	N/A
% of TSS in the KCl-treated sample that was only in or was upregulated >3-fold in the KCl-treated sample **Total DoG regions**	30%	N/A
% of TSS in the untreated sample that was only in or was upregulated >3-fold in the untreated sample	63%	N/A

Stringent DoG regions		
% of TSS in the KCl-treated sample that was only in or was upregulated >3-fold in the KCl-treated sample **Stringent DoG regions**	47%	N/A

*****for DoG-associated genes, TSSs are mapped anywhere in the gene + 2 kb downstream, or anywhere in the 3′-UTR and 2 kb downstream. For DoG regions, TSSs are mapped anywhere in the DoG region from the endpoint of the associated gene to the annotated endpoint of the DoG.

Table S3: List of control genes. Related to Figure 5.

Table S4: PAS counts and scores. Related to Figure 5.

	Non-DoGs	Stringent DoGs	All protein-coding genes
% with at least one AATAAA[1] in the 200 nts surrounding annotated end-of-gene	79%	63%	61%
% with at least one AATAAA[1] or ATTAAA[1] in the 200 nts surrounding annotated end-of-gene	91%	79%	75%
pA site score AATAAA/ATTAAA[1] based*	4.62	3.71	3.68
pA site score including alt. signals[2]	7.19	6.29	6.11
ratio AATAAA[1] in 5 kb downstream of main PAS, sense/antisense	1.1	0.8	0.9

[1]AATAAA represents ~50% and ATTAAA ~15% of PASs in mammalian cells (Almada et al., 2013; Beaudoing et al., 2000)

[2]See Extended Experimental Procedures

Table S5: Primer and probe sequences. Related to Extended Experimental Procedures.

Extended Experimental Procedures

Cell culture, transfections and treatments

Cell lines included in this study are SK-N-BE(2)C, SH-SY-5Y (both human neuroblastoma), 293T (human embryonic kidney-derived), MD231 (human breast cancer) and A459 (human lung cancer). All cell lines were cultured in a 1:1 mix of Dulbecco's Modified Eagle Medium and F-12 Nutrient's mixture (both Gibco) substituted by 10% Fetal Bovine Serum (Atlanta Biologicals), 1% L-glutamine (Gibco) and 1% Penicillin/Streptomycin (Sigma-Aldrich). For transfections, media without antibiotics was used. For KCl treatments, cells were cultured in the presence of 80 mM KCl for 1 h if not otherwise stated.

RNA preparations, RT-PCR, qRT-PCR

All primers are listed in Table S5.

If not otherwise stated, total RNA was used, and was prepared by harvesting cells in Trizol (Ambion) following the manufacturer's instructions. For preparations of nuclear-enriched RNA, cells were grown to confluency on 15 cm plates (Falcon), washed in PBS, scraped in PBS, collected by centrifugation (500 x g 3 min 4°C) and lysed with 400 µl cytoplasmic lysis buffer (CLB): 0.32 M sucrose, 3 mM $CaCl_2$, 2 mM Mg Acetate, 0.1 mM EDTA, 10 mM Tris-HCl, pH 8.0, 1 mM DTT, 0.5% NP-40 and 100U/ml RNaseIN (Roche). Samples were incubated for 5 min on ice and nuclei were recovered by centrifugation (500 x g 5 min 4°C). Supernatant was removed and discarded, and nuclei were resuspended in 1 ml CLB without NP40 and centrifuged again. The supernatant was discarded; the nuclear pellet was resuspended in 100 µl CLB without NP40; and nuclear RNA was recovered by addition of 1 ml Trizol

followed by RNA preparation as above. RNA (total or nuclear) was then treated with DNase RQ (Promega) using 0.5-1 units DNase/μg RNA following the manufacturer's instructions. RNA was recovered by phenol/chloroform/isoamyl alcohol (PCA) extraction and ethanol precipitation, following standard protocols. RNA concentrations were analyzed on a NanoDrop 8000 (Thermo Scientific) and 1-4 μg was used for cDNA synthesis using SuperScript III and random primers (Invitrogen) according to the manufacturer's protocol. Minus-RT controls were included and analyzed alongside other samples to ensure efficient DNase treatment. For gene-specific cDNA synthesis, gene-specific primers (see Table S5) were used instead of random primers at 1 μM. PCR was performed using Phusion (NEB), according to the manufacturer's recommendations. qPCR was performed on a StepOnePlus instrument (Applied Biosystems) with FastStart Universal SYBR Green Master (ROX) mix (Roche) according to the manufacturer's recommendation. Relative RNA levels were calculated by the ΔΔCT method and target RNAs were normalized either to the mean value of two control genes (*GAPDH* and *GUSB*) or to *GAPDH* alone. All primer pairs used for qRT-PCR were validated by agarose gel electrophoresis of the qRT-PCR products to ensure expected product size and absence of other amplification products and sequencing of the qRT-PCR product to ensure amplification of the expected target.

RNA-Seq and data analysis

Nuclear-enriched RNA was prepared as above from untreated cells or cells treated with 80 mM KCl for 1 h in 3 biological replicates, with the exception that after DNase treatment, RNA was recovered by RNAEasy mini-kit column purification (Qiagen). RNA quality was assessed by NanoDrop 8000 (Thermo Scientific) and agarose gel

electrophoresis. RNA quality was further established at the Yale Center for Genome Analysis (YCGA) by running an Agilent Bioanalyzer gel. Ribosomal RNA was then depleted using the Ribo-Zero Gold Kit (MRZG12324, Epicentre) and remaining RNA fragmented by incubation at 94°C. Libraries were prepared using Illumina reagents. Following first-strand synthesis with random primers, second-strand synthesis was performed with dUTP for generating strand-specific sequencing libraries. The cDNA library was end-repaired and A-tailed, adapters were ligated and second-strand digestion was performed by Uricil-DNA-Glycosylase. Quality of indexed libraries was validated by quantification by qRT-PCR (KAPA Biosystems) and by insert size distribution determination (LabChip GX). Two samples were multiplexed per sequencing lane, and samples were sequenced using 75 bp paired-end sequencing on an Illumina HiSeq 2000 according to Illumina protocols. Signal intensities were converted to individual base calls during the run using the system's Real-Time Analysis (RTA) software. Sample de-multiplexing and splitting of the paired-end reads were done using Illumina's CASAVA 1.8.2 software suite. At this point, we obtained the data from YCGA and plotted it to hg19 using TopHat2 (Kim et al., 2013) following protocol for mapping paired-end, strand-specific data (see TopHat manual, http://ccb.jhu.edu/software/tophat/manual.shtml) allowing only the best match for each read (-g 1). We used RefSeq genes downloaded from UCSC in May 2012 as a reference transcriptome (-G). TopHat2 output "accepted_reads" bam files were edited with biotoolbox (https://code.google.com/p/biotoolbox/) split_bam_by_isize.pl with default settings (removing reads for which the insert size was <100 or >200) and then converted to strand-specific BigWig files using biotoolbox bam2wig.pl with options --pe --pos span --strand --bw for inspection in the IGV browser (Robinson et al., 2011; Thorvaldsdottir et al., 2013).

Analysis of reads and splice junctions downstream of genes

To analyze the percentage of intergenic reads mapping downstream of genes, we generated bed files of intergenic reads within 45 kb downstream of RefSeq genes as follows: we converted our "accepted_hits" bam files outputted from TopHat2 to bed files by first editing the bam files using biotoolbox split_by_bam_isize.pl as described above and then converting them to bed files using biotoolbox bam2gff_bed.pl with the option --pe. To keep only intergenic reads, we removed all reads overlapping exonic or intronic sequences using BEDTools (v2.17.0) intersect (Quinlan and Hall, 2010) with a Knowngenes gtf file with annotated exons and an intron bed file with annotated introns, both downloaded from UCSC in May 2014. Intergenic reads 45 kb downstream of genes were then obtained using BEDTools window, which with the settings used will search for genes overlapping with or present within a window of 45 kb upstream of each read, and will report the read if an overlap is found. Because genic reads had already been removed, this analysis returned only intergenic reads within 45 kb downstream of RefSeq genes. The resulting number of reads was then normalized to the total number of intergenic reads for that sample. A similar analysis for splice junctions was performed starting from bedfiles with predicted splice junctions outputted from TopHat2. These files contain predictions of splice junctions based on reads spanning the junction. To generate bed files containing only intergenic splice junctions, we removed splice junctions mapping to exonic and intronic sequences by BEDTools intersect as described above using RefSeq genes and UCSC introns as references, respectively. We then generated files of intergenic splice junctions found only in untreated samples, only in KCl-treated samples or found in both samples using BEDTools intersect. Next, we used BEDTools window as above

to return splice junctions within 15 kb downstream of RefSeq genes. We normalized the number of intergenic splice junctions downstream of genes with the corresponding number of total intergenic splice junctions in the corresponding input file (e.g., the number of splice junctions within 15 kb downstream of genes present in "KCl only" in replicate 1 was normalized to the total number of intergenic splice junctions present in "KCl only" in replicate 1).

DoG-finding pipeline

To identify DoGs genome wide, we defined DoGs as transcripts continuous with the associated genes for 5 or 10 kb (at low or high stringency, respectively) and that were longer after KCl treatment. We created a pipeline that employed BEDTools and custom-written Perl scripts to identify DoGs starting from a bed file with all protein-coding human genes (downloaded from ENSEMBL BioMart GRCh37). The pipeline first used BEDTools coverage with strand specific-settings (-s) to determine the coverage over the first 5 or 10 kb downstream of protein-coding genes using the mapped reads from KCl-treated samples. Transcripts with >95% coverage were selected as candidate DoGs. The pipeline next calculated the RPKM over the last 1 kb of the associated gene in both the KCl-treated sample and the corresponding untreated sample, and then iteratively measured the RPKM over each kb in the DoG. The script defined as the DoG endpoint the region at which the RPKM of the DoG dropped to <1% of the last 1 kb of the associated gene. It then compared DoG endpoints between untreated and treated samples and defined as DoGs those transcripts that were longer after KCl treatment. We next used BEDTools intersect to create a list of stringent DoGs found in all 3 biological replicates based on continuous transcription at least 10 kb downstream of associated genes, and a list of total DoGs found in at least 1

biological replicate based on continuous transcription at least 5 kb downstream of the associated genes. Total and stringent DoGs are listed in Table S1. To calculate the RPKM of the first 5kb of total DoG regions (included in Table S1), we used custom-written Perl scripts and BEDTools.

To determine the percentage of intergenic transcription attributable to DoGs, we used BEDTools intersect (in strand-specific setting, -s) to compare intergenic reads and splice junctions from each sample (see above) with the annotation of total DoGs.

Analysis of transcripts upregulated by KCl

To study the nature of KCl-upregulated intergenic transcripts, we first identified intergenic transcription upregulated by KCl, defined either as intergenic regions or predicted splice junctions with more reads mapping to them in KCl-treated than in untreated samples. To identify intergenic regions with more mapped reads after treatment, we started with bed files with intergenic reads generated as described above. We used BEDTools genomecov to create lists of the read coverage for each intergenic base, binning regions with equal coverage together. We performed the BEDTools genomecov analysis twice for each file with either the option -strand - or -strand + to obtain separate results for either strand. We then sorted these files in descending order of the number of mapped reads by sorting with regard to the column containing this information using the UNIX sort command. We made separate lists containing either the top 5% genomic regions in the KCl-treated samples, or the top 30% in the untreated samples. These lists were compared to each other using BEDTools intersect for each biological replicate, resulting in genomic regions with mapped reads present in the top 5% of KCl-treated samples but absent from the top

30% of untreated samples. We used a similar approach to find splice junctions with increased read support in KCl-treated samples. We started from the junction bed files outputted from TopHat2. Exonic and intronic splice junctions were removed as described above. We sorted these junction files and performed the same analysis as described above for genomic regions. Once bed files listing upregulated regions of reads and splice junctions had been generated, we compared these bed files to our annotation of total DoGs (see above) using BEDTools intersect (using the -s option for strand-specific setting, strand information had been added to the files listing intergenic regions with mapped reads using custom-written Perl scripts). In this way, we established the percentage of upregulated intergenic regions of reads and splice junctions attributable to DoGs.

Selection of DoG candidates for validation

Candidates were selected based on the presence of a splice junction within 15 kb downstream of a RefSeq gene detectable only after KCl treatment (identified in the analysis of splice junctions downstream of genes shown in Figure 2b). We further stipulated that the splice junction should appear in all biological replicates and be among the top 20 most-supported splice junctions in at least one biological replicate.

CapSeq

We used a modified version of the CapSeq protocol described in (Xie et al., 2013). In short, we isolated nuclear RNA from untreated and KCl-treated SK-N-BE(2)C cells as described above and fragmented this RNA by incubating 10-20 µg of RNA in a buffer containing 40 mM Tris pH 8.0, 100 mM KOAc and 30 mM $MgCl_2$ for 5 min at 94°C. We performed these reactions in aliquots of 20 µl in thin-wall 200 µl PCR tubes

to allow for rapid temperature adjustment. Reactions were stopped by the addition of 5 µl 0.2 M EDTA, and RNA was purified by RNAEasy mini-kit column purification (Qiagen) to remove excess salt. RNA was then separated on an 8% PAGE gel run in 1 x TBE according to standard protocols. The RNA was visualized by Ethidium bromide staining, RNA fragments of 90-250 nt were excised and extracted from the gel by adding equal volumes of RNA extraction buffer (25 mM Tris pH 8.0, 0.3 mM NaOAc) and PCA followed by incubation at RT over night under rotation. The next day, PCA extraction was completed, RNA was recovered by ethanol precipitation, and RNA from several preparations was pooled. ~18 ug of RNA was then subjected to pulldown with eIF-4E cap binding protein as described (Xie et al., 2013). After washing (Xie et al., 2013), pulled-down RNA was eluted from the beads by PCA extraction and subsequently subjected to treatment with Calf Intestine Phosphatase (CIP, Roche) according to the manufacturer's recommendation, to remove the 3′-phosphates generated by the fragmentation and any 5′-phosphates present. RNA was recovered by 2 rounds of PCA extraction to thoroughly remove any traces of CIP, and was then subjected to cap removal by Tobacco Acid Phosphatase (Epicentre) according to the manufacturer's recommendation. The RNA was recovered by PCA extraction. At this point, the RNA was submitted to YCGA for further processing. At YCGA, the RNA was cloned using library preps for small RNA (NEB) adjusting size extraction steps for the expected size of the cloned RNAs, and the two samples were multiplexed in one sequencing lane and sequenced as above using 75 bp single-end sequencing on an Illumina HiSeq 2000 according to Illumina protocols.

Analysis of CapSeq data

De-multiplexed raw sequencing reads were obtained from YCGA and further processed by removing Illumina adapter sequences using fastx_clipper followed by trimming reads to 25 nt/read (containing the 3rd to the 27th nt of each read) using fastx_trimmer (hannonlab.cshl.edu/ fastx_toolkit). The resulting reads were mapped to hg19 using TopHat2. Mapped reads were then filtered to reduce background and identify TSS peaks using a custom-written pipeline that utilized BEDTools and custom-written Perl scripts to identify peaks within any region of 200 nt or less that is increased at least 5-fold over background. The input to the script was a bedgraph file generated by converting the TopHat2 output accepted_hits.bam to bed and then to bedgraph using BEDTools. Next, BEDTools intersect was used to select reads from the bedgraph file mapping to the annotation of total DoG associated genes + 2 kb downstream, and annotations of control genes (including *FOS* and *TMEM254*). The peak-finding pipeline then calculated the RPKM over the 25 nt downstream of each starting nucleotide in the bedgraph file and compared that to the RPKM of one upstream and one downstream region. The upstream region was defined as the 25 nt starting 125 nt upstream of, and the downstream region was defined as the 25 nt starting 100 nt downstream of, the coordinate for each starting nucleotide in the bedgraph file. 25 nt regions with an RPKM at least 5-fold higher than both upstream and downstream regions were kept as peaks (background of 0 was set to 1 to enable fold-change calculations). The resulting peak-containing bed files were converted to BigWig format for viewing in the IGV browser. To identify KCl-upregulated peaks, we first used BEDTools intersect to identify any peak reads that were uniquely found in the untreated or in the KCl-treated sample, and calculated the % of unique reads for each sample. We next used custom-written Perl scripts to compare the RPKMs for the peaks that were common for both samples, and calculated the percentage of peaks that

was upregulated >3-fold either in the untreated sample compared to the KCl-treated sample, or vice versa. We then calculated the sum of the percentages for unique and upregulated peaks in each sample to get a final percentage of unique or upregulated peaks in that sample. Next, we used BedTools intersect to get only peak reads mapping to our stringent DoGs + 2 kb downstream or to the 3′-UTRs + 2 kb downstream of total or stringent DoGs and repeated the analysis to calculate percentage of unique and upregulated peaks. 3′-UTR annotations were downloaded from ENSEMBL BioMart GRCh37. For genes with multiple 3′-UTR isoforms, the 3′-UTRs with most distal endpoints were selected for analysis using custom-written Perl scripts. Finally, we re-applied our peak-identifying script to instead select reads from the original bedgraph file of mapped CapSeq reads that overlapped with the annotation of total or stringent DoG regions (starting from the end of the associated gene and continuing through the length of the annotated DoG).

polyA site bioinformatics analysis

Sequences of the 5 kb 3′ of, or of the 200 nt surrounding, the annotated endpoints of either genes upstream of stringent DoGs or genes that were expressed in our cell line but showed no evidence of readthrough transcription (non-DoG genes, listed in Table S3), or all protein-coding genes, were downloaded from Ensembl (GRCh38). First, we used custom-written Perl scripts to calculate the number of genes that had at least 1 AATAAA or ATTAAA in the 200 nt surrounding their annotated endpoint. Next, we expanded our analysis to also include adjacent regions reported to enhance cleavage and polyadenylation (Tian and Graber, 2012), giving a high score for presence of AATAAA, slightly lower score for ATTAAA and lower score for other

known PAS sequences (Almada et al., 2013) and for the presence of other upstream and downstream enhancing signals (Tian and Graber, 2012).

We used custom-written Perl scripts to count the occurrence of each possible hexamer in the 5 kb regions of sequence 3′ of the genes associated with the stringent DoGs and the non-DoG genes. We calculated the frequency of sense over antisense for each hexamer using Excel. To analyze AATAAA hexamers in 5 kb regions downstream of all human protein-coding genes, we used custom-written Perl scripts to count the number of AATAAAs found downstream of each gene and in the corresponding antisense regions. See Table S4 for scores and percentages.

Cellular fractionation

SK-N-BE(2)C cells were grown to confluency on 10 cm plates (Falcon) and treated with 80 mM KCl for 1 h before harvest to boost DoG levels, thereby facilitating detection by qRT-PCR. Plates were washed 2 times in PBS, cells were scraped in 1ml PBS, collected by centrifugation (500 x g 3 min 4°C) and lysed in 450 µl of buffer A: 10 mM HEPES pH 7.9, 10 mM KCl, 1.5 mM $MgCl_2$, 0.34 M sucrose, 10% glycerol, 1 mM DTT, 100 U/ml RNaseIN (Roche). After resuspension, Triton X-100 was added to a final concentration of 0.1% followed by 5 min incubation on ice. Nuclei were collected by centrifugation (1300 x g 4 min 4°C). Supernatant (cytoplasmic fraction) was moved to a fresh tube and cleared by centrifugation at 20000 x g for 15 min 4°C. Nuclei were washed with 500 µl buffer A without Triton X-100 and then resuspended in 400 µl buffer B: 3 mM EDTA, 0.1 mM EGTA, 1 mM DTT, 100 U RNAseIN/ml. The chromatin fraction was collected by centrifugation (1700 x g 4 min 4°C). The supernatant (soluble nuclear extract) was cleared by centrifugation at 20000

x g for 15 min 4°C. Chromatin was washed once in 500 µl buffer B. Then, 250 µl chromatin extract buffer (50 mM NaAcetate pH 5, 50 mM NaCl, 1% SDS), and 250 µl PCA was added followed by incubation at 37°C for 1 h. RNA was recovered from the chromatin fraction by completing the PCA extraction followed by ethanol precipitation in the presence of 1 µl glycogen (Roche). RNA from other fractions was similarly recovered by PCA extraction and ethanol precipitation. RNA was resuspended in water and DNase treated. For cytoplasmic and soluble nuclear fractions, 1 U DNase/µg RNA was used. For chromatin fractions, 40 U DNase was used although the RNA amount was usually estimated to about 10 µg, because genomic DNA contamination was expected to be greatest in the chromatin fraction. DNase treatment was carried out as above, and RNA was recovered by PCA extraction and ethanol precipitation. RNA was resuspended in equal volumes of water and equal volumes from all fractions were used for cDNA synthesis as described above (maximum RNA amount for any fraction – the cytoplasmic fraction typically had the most RNA – was 5 µg) followed by qPCR analysis as described above. Minus-RT controls for chromatin fractions were included to ensure effective DNase treatment.

Stellaris RNA FISH and DAPI nuclear staining/analysis of nuclear phenotype

We used a Quasar 570 labeled *GAPDH* probe (Biosearch Technologies, SMF-2026-1), custom-designed Stellaris probes for *doSERBP1* (designed by the Biosearch design tools using the 7 kb downstream of the end of *SERBP1* as input) labeled with Quasar 570, and custom-designed Stellaris probes for intron 1 and intron 5 of *SERBP1* (designed by the Biosearch design tools using the entire sequences of *SERBP1* intron 1 and 5 as input, respectively) labeled with FAM (Biosearch

Technologies) to perform FISH in SK-N-BE(2)C cells either untreated or treated with 80 mM KCl for 1 h prior to fixation using standard Stellaris protocols (Biosearch Technologies). FISH staining was visualized on a Diaphot 300 widefield microscope (Nikon), captured using MetaMorph v 7.6.1.0 software and further processed in ImageJ. For DAPI staining of SK-N-BE(2)C nuclei, cells were treated with 100 μM 2-APB (Sigma-Aldrich) for 30 min followed by 80 mM KCl for 1 h, or treated with either 2-APB or KCl alone, or left untreated. Cells were then fixed and processed for imaging using the same protocol as for Stellaris FISH above omitting the probe hybridization steps. Nuclei were visualized on a Zeiss Axioplan microscope using AxioVision 4.8 software, and further processed in ImageJ. Nuclear areas were measured using ImageJ on 20 cells each from 5 individual images (100 cells in total) for each treatment. For phenotype quantification, images were assigned random labels and numbers of cells with each phenotype were counted by a person different from the person capturing the microscope images. Phenotypes were classified as no phenotype (uniform DAPI staining), mild phenotype (one small hole in a nucleus with otherwise uniform staining), or severe phenotype (one or several big holes accompanied by uneven DAPI staining in the remainder of the nucleus). Quantifications were based on 20 cells each from 5 individual images (100 cells in total) for each treatment.

Northern blotting

Northern blotting was performed according to standard protocols. Nuclear RNA from SK-N-BE(2)C cells treated with 80 mM KCl for 1 h or left untreated was purified and DNase treated as described above, except that RNA was recovered by RNAEasy mini-kit column purification (Qiagen) after DNase treatment. Purified RNA was run

on a 1% agarose gel for ~30 h, after which the gel was soaked in 0.05 M NaOH for 20 min and then transferred using standard protocols to a Zetaprobe GT membrane (Biorad). RNA probes were *in vitro* transcribed in the presence of ^{32}P-αUTP (Perkin-Elmer) using Sp6 polymerase (NEB) from a ~1 kb PCR product generated by PCR of genomic DNA from SK-N-BE(2)C cells with primers listed in Table S5. Membranes were stained with 0.03% Methylene blue in 0.3M NaOAC (pH 5.2) and imaged using a G-Box (Syngene). Membranes were then pre-hybridized in a hybridization buffer containing 25 mM Na-phosphate (pH 6.5), 6 x SSC, 5 x Denhardt's solution, 0.5% SDS and 50% formamide at 65°C for 1 h. Probe was added to the hybridization buffer and incubated at 65°C degrees overnight. Membranes were washed according to standard protocols, exposed and visualized on a Storm 860 Phosphoimager.

doCXXC4 *copy number estimation*

248 nt of *doCXXC4* was PCR-amplified from nuclear SK-N-BE(2)C RNA using primers indicated in Table S5, cloned into pCDNA3 and used as a template for *in vitro* transcription by T7 polymerase (lab stock) in the presence of trace amounts of ^{32}P -αUTP (Perkin-Elmer) (1 µl of 1:10 dilution of stock ^{32}P -αUTP was used for a 20 µl reaction) according to manufacturer's recommendation. RNA product was purified through a G-50 column (GE Healthcare), and sample activity was measured by a scintillation counter (Packard). Scintillation counts were used to calculate the incorporation rate of ^{32}P -αUTP, which in turn was used to calculate the moles and thereby molecules of RNA in the sample. *In vitro* transcribed molecules were spiked into untreated SK-N-BE(2)C Trizol extracts at 1, 10, 50 and 100 molecules per cell equivalent directly after harvesting cells in Trizol (see above). RNA was extracted and quantified by qRT-PCR as described above. CT values for *doCXXC4* normalized

to the mean of *GAPDH* and *GUSB* CT values were plotted against the number of spiked-in molecules to generate a standard curve correlating qRT-PCR CT values to number of molecules. Untreated and KCl-treated SK-N-BE(2)C cells were harvested and processed in parallel without the addition of *in vitro* transcribed RNA for comparison.

Purification of polyA$^+$ and polyA$^-$ RNA

SK-N-BE(2)C cells were grown to confluency on 10 cm plates and treated with 120 mM KCl for 2 h before harvesting to boost DoG levels, thereby facilitating detection by qRT-PCR. RNA was harvested in Trizol, and total RNA was purified and DNase treated as above. Oligo-dT cellulose (Amersham Biosciences) (15 mg/sample) was prepared by swelling for 15 min at RT in elution buffer: 10 mM Tris pH 7.5, 1 mM EDTA, 0.05% SDS, washed 4 times with 1 ml elution buffer (30 sec 20000 x g spin in between), and equilibrated with 3 washing steps using 1 x binding buffer: 10 mM Tris pH 7.5, 1 mM EDTA, 0.05% SDS, 0.5 M NaCl. RNA was brought to a volume of 600 µl with water and denatured at 65°C for 5 min. After mixing with 600 µl 2 x binding buffer, the sample was added to the prepared oligo-dT cellulose and incubated rotating at RT for 30 min. The oligo-dT cellulose was spun down, and supernatant was kept as the polyA$^-$ fraction. The oligo-dT cellulose was washed 6 times with 1 x binding buffer and 7 times with wash buffer (10 mM Tris pH 7.5, 1 mM EDTA, 0.05% SDS, 0.2M NaCl), 3 min rotation for each wash. polyA$^+$ RNA was eluted with 250 µl elution buffer at 37°C for 5 min 3 times. Eluates were combined as the polyA$^+$ fraction. polyA$^+$ and $^-$ fractions were ethanol precipitated, resuspended in 600 µl water and subjected to an additional round of purification, keeping the polyA$^-$ fraction when starting with the previous polyA$^-$ fraction and the

polyA$^+$ fraction when starting with the previous polyA$^+$ fraction. RNA was then ethanol precipitated and resuspended in equal volumes, with equal volumes of each fraction used for cDNA synthesis (using no more than 5 μg of the fraction (polyA$^-$) with the highest RNA concentration). cDNA was then analyzed by qPCR as described above.

CRISPR

We used the mutant CRISPR system to block transcription as described (Chen et al., 2013; Gilbert et al., 2013). Our sgRNA sequences are listed in Table S5. For targeting *CXXC4*, we used the sgRNA design described (Gilbert et al., 2013) with BFP-tagged dCas9. For *SERBP1,* we used the improved sgRNA design described (Chen et al., 2013) together with BFP-tagged dCas9 fused with KRAB (described in (Gilbert et al., 2013)). The experiments targeting *SERBP1* were performed at a later time when further data on optimizing the mutant CRISPR system was available; hence the different choice of constructs. In each case, sgRNAs were designed using www.genome-engineering.org/crispr/ (Hsu et al., 2013) to target the non-template strand of the first exon of the gene of interest. sgRNAs were cloned into Mp177 plasmids that express mCherry (as described (Larson et al., 2013). BFP-dCas9 and BFP-dCas9-KRAB in pHR plasmids (pHR-cas9 and pHR-cas9-KRAB) and Mp177-expressing control sgRNA targeting GFP (Gilbert et al., 2013) were kind gifts from Luke Gilbert and Jonathan Weissman (UCSF). We established 293T cell lines expressing BFP-dCas9 or BFP-dCas9-KRAB through lentiviral transduction (using lentivirus produced in 293T cells by co-transfecting pHP and pHEF-VSV-G with either pHR-cas9 or pHR-cas9-KRAB) followed by sorting of BFP positive cells on a FACS ARIA cell sorter (BD) using FACS DIVA v6.1.3 software (BD). We chose

293T cells for these experiments because they are readily manipulated by infection and electroporation, while SK-N-BE(2)C cells are difficult to transfect with plasmids by any method. Once Cas9-expressing cell lines were established, they were electroporated with sgRNA plasmids according to standard protocols using a Genepulser II instrument (Biorad) and sorted the following day on a FACS ARIA cell sorter (BD) using FACS DIVA software (BD). Cells positive for both BFP and mCherry were cultured for one additional day, treated with 120 mM KCl to induce DoG levels followed by a washout period of 15 min in media without KCl. Cells were harvested in Trizol, and total RNA was DNase treated and subjected to analysis by qRT-PCR as decribed above.

ASO and siRNA transfections

Oligonucleotide sequences or accession numbers are found in Table S5. For all transfections, SK-N-BE(2)C cells were plated in 6-well plates at 55,000 cells/well and transfected the following day. ASOs had been selected after performing RNaseH mapping (Simon, 2013) to identify accessible sites. Control ASO was designed to target GFP. For ASOs, 70 pmol ASO was used per well (final volume in well: 2 ml). For siRNAs from IDT (targeting *CPSF73*), we used 50 nM and for siRNAs from Qiagen (targeting *IP3R1*), we used 20 nM final conc. Transfections were performed using RNAiMAX (Invitrogen) according to the manufacturer's recommendations. For transfection with ASOs targeting *CXXC4*, cells were cultured 48 h post transfection, treated with 120 mM KCl to induce DoG expression followed by a washout period of 15 min in media without KCl, and harvested in Trizol. For CPSF73 knockdown, cells were cultured 48 h post transfection and harvested in Trizol; and for IP3R1 knockdown, cells were cultured for 72 h, treated with 80 mM KCl for 1 h or left

untreated, and harvested in Trizol. Total RNA was DNase treated and subjected to analysis by qRT-PCR as described above.

Selection of metabolically-labeled RNA and of CXXC4 *RNA*

Metabolically-labeled RNA was prepared and isolated according to the Click-iT Nascent RNA Capture Kit manual (Life Technologies) using the maximum recommended amount of EU and RNA according to the manufacturer's protocol. For *CXXC4* RNA pulldown, streptavidin beads bound to biotin oligonuceotides were prepared as follows: MyOne C1 Streptavidin dynabeads (Life Technologies) were washed with binding buffer: 10 mM Tris pH 7.5, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween-20. Then, 250 µl of beads were mixed with 500 µl binding buffer and 2.5 µl 100 µM stock oligonucleotides (either 2.5 µl control oligonucleotide (targeting *EBER2,* a viral RNA not expressed in these cells) or 1.25 µl of each *CXXC4* oligonucleotide, sequences in Table S5). Beads were incubated with oligonucleotides while rotating at RT for 30 min, then washed 3 times in binding buffer and resuspended in 250 µl PBS and stored at 4°C until used. *CXXC4* RNA was selected from total DNase-treated RNA (prepared as described above) using buffers and protocol from the Nascent RNA Capture Kit (Life Technologies) in the same way as for selection of metabolically-labeled RNA. Specifically, denatured RNA in the kit binding buffer was exposed to 20 µl streptavidin beads attached to biotinylated oligonucleotides, which had been washed in the kit wash buffer 2 and resuspended in 20 µl of the kit wash buffer 2. For both the metabolically-labeled RNA pulldown and the biotin pulldown of *CXXC4* RNA, cDNA reactions were performed on beads after washing as described in the kit manual, using Superscript III and random primers (Invitrogen).

Pol II ChIP

SK-N-BE(2)C cells were grown to semi-confluency on 15-cm plates (~10 million cells/plate), treated with 80 mM KCl for 1 h or left untreated, and were crosslinked by addition of 1% final volume formaldehyde to the culturing media for 10 min. Cells were then washed 2 times in PBS and scraped in PBS. Cells were pelleted by centrifugation at 500 x g for 5 min, resuspended in 800 µl SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris, pH 8.0, and 1 x protease inhibitors (Calbiochem)) and incubated for 10 min on ice. Next, chromatin was fractionated by sonication on a Sonifier 250 (Branson) sonicator using constant duty cycle and 1.5 amplitude for 2 x 8 sec with a 30 sec incubation on ice in between. This sonication protocol was optimized to generate ~350-550 nt DNA fragments. Lysates were cleared by centrifugation at 13,000 rpm for 10 min at 4°C. Next, 2 x 150 µl of each sample was moved to two separate tubes per ChIP reaction and diluted with dilution buffer (0.01% SDS, 1.1% Triton-X, 1.2 mM EDTA, 16.7 mM Tris pH 8.0, 167 mM NaCl) to a final volume of 1.5 ml. At this point, 2% input was taken out and stored at -20°C. Next, lysates were precleared by the addition of 30 µl of protein A sepahrose beads. The protein A sepharose beads had first been prepared by mixing ~150 µl of dry protein A sepharose (GE Healthcare) with 1 ml of double distilled water followed by swelling under rotation for 5 min at RT. Water was then removed and beads were washed 2 times with 1 ml PBS and 2 times with 1 ml PBS containing 20 µl 10 µg/µl BSA and 100 µl 20 µg/µl salmon sperm DNA (ssDNA) per ml of PBS. After the last wash, beads were resuspended in 1 ml of PBS with BSA and ssDNA and left to rotate for 1 h at RT, followed by 5 washes with PBS. After the last wash, beads were resuspended in the same amount PBS as bead volume (approx. 250 µl), and were

ready to use. After 1 h of preclearing during rotation at 4 °C, beads in ChIP samples were pelleted by centrifugation at 1000 x g for 1 min, and supernatants were moved to fresh tubes. Antibodies (either 10 µg N-terminal Pol II antibody (H-224, Santa Cruz Biotechnology) or 4 µl GAPDH antibody (Cell Signaling)) were added to each tube and samples were incubated rotating at 4°C overnight. The following day, 30 µl Protein A sepharose beads were added and samples were incubated rotating at 4°C for 1 h. Next, beads were pelleted and washed 2 times each in the following buffers in the order listed: a) Low Salt Wash: 0.1% SDS, 1% Triton-X, 2mM EDTA, 20 mM Tris pH 8.0, 150 mM NaCl. b) High Salt Wash: as Low Salt Wash, but with 500 mM NaCl instead of 150 mM. c) LiCl Wash: 0.25 M LiCl, 1% IGEPAL-CA630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris pH 8.0. d) 1xTE Wash: 10 mM Tris pH 8.0, 1mM EDTA pH, 0.05% Tween-20. Each wash was performed by rotation at RT for 3 min. The duplicate samples for each ChIP reaction were pooled in the first wash. After the last wash, precipitated DNA was eluted by 2 rounds of elution in 200 µl of freshly made 1% SDS and 0.1M $NaHCO_3$. Eluates from the same samples were combined and crosslinks were reversed by addition of NaCl to a final concentration of 200 mM to eluates and input samples and incubation at 65°C for 4 h followed by treatment with first 50 µg/ml RNase A at 37°C for 30 min and then 20 µg/ml Proteinase K and 0.005% SDS at 45°C for 1 h. DNA was recovered by PCA extraction and ethanol precipitation followed by qPCR with the primers listed in Table S5.

Treatments

Actinomycin D (ActD, Sigma-Aldrich) treatment: SK-N-BE(2)C cells were plated in 6-well plates at 300,000 cells per well 1 day prior to the experiment. On the day of the

experiment, they were treated with 80 mM KCl or left untreated for 1 h, then treated with 1 μg/ml ActD for the indicated amounts of time and harvested in Trizol. Total RNA was DNase treated and analyzed by qRT-PCR as described above. EGTA, BAPTA or inhibitor treatments: SK-N-BE(2)C cells were plated in 6-well plates at 300,000 cells/well 1 day prior to the experiment. On the day of the experiment, they were treated with 6 mM EGTA for 5 min, or 50 mM BAPTA (Abcam) or 100 μM 2-ABP (Sigma-Aldrich) for 30 min, or 10 μM Gö6976 (Abcam) or 10 μM KN-93 (Abcam) or 10 μM of both Gö6976 and KN-93 for 1 h, and then treated with 80 mM KCl or left untreated for 1 h. In summary, 3 samples were included in the cases of a single chelator/inhibitor: no chelator/inhibitor followed by no KCl, no chelator/inhibitor followed by KCl, and chelator/inhibitor as above followed by KCl. For the experiments using Gö6976 and KN-93, 5 samples were included: no inhibitor followed by no KCl, no inhibitor followed by KCl, either inhibitor alone followed by KCl, or both inhibitors together followed by KCl. Cells were harvested in Trizol and total RNA was DNase treated and analyzed by qRT-PCR as described above. For NaCl or sucrose treatments, SK-N-BE(2)C cells were plated in 6-well plates at 300,000 cells/well 1 day prior to the experiment. On the day of the experiment, they were treated with 80 mM NaCl or 200 mM sucrose, or left untreated for 1 h: cells were harvested in Trizol and total RNA was DNase treated and analyzed by qRT-PCR as described above. We observed that a higher dose of sucrose than NaCl or KCl was required for similar induction, likely due to the fact that NaCl and KCl gives rise to two solutes in solution, while sucrose give rise to one. For KCl treatments in cell lines other than SK-N-BE(2)C, SH-SY-5Y cells were plated in 6-well plates at 500,000 cells/well and treated the next day with 120 mM KCl for 1 h, or 293T, MDA231, or A459 cells were plated in 6-well plates at 300,000 cells/well and treated the following

day with 80 mM KCl for 1 h. RNA was harvested and analyzed as described above. For heat-shock experiments, SK-N-BE(2)C cells were plated in 6-well plates at 300,000 cells/well. The following day, they were moved to 44°C for 2 or 4 h, or left at 37°C as control, followed by harvesting in Trizol. RNA was prepared and analyzed as described above.

Supplemental References

Almada, A.E., Wu, X., Kriz, A.J., Burge, C.B., and Sharp, P.A. (2013). Promoter directionality is controlled by U1 snRNP and polyadenylation signals. Nature *499*, 360-363.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D. (2000). Patterns of variant polyadenylation signal usage in human genes. Genome Res *10*, 1001-1010.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S.*, et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479-1491.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A.*, et al.* (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol *31*, 827-832.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol *14*, R36.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat Protoc *8*, 2180-2196.

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

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat Biotechnol *29*, 24-26.

Simon, M.D. (2013). Capture hybridization analysis of RNA targets (CHART). Curr Protoc Mol Biol *Chapter 21*, Unit 21 25.

Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.

Tian, B., and Graber, J.H. (2012). Signals for pre-mRNA cleavage and polyadenylation. Wiley Interdiscip Rev RNA *3*, 385-396.

Xie, M., Li, M., Vilborg, A., Lee, N., Shu, M.D., Yartseva, V., Sestan, N., and Steitz, J.A. (2013). Mammalian 5'-capped microRNA precursors that generate a single microRNA. Cell *155*, 1568-1580.