Cell Reports, Volume *23*

Supplemental Information

The RNA Exosome Adaptor ZFC3H1 Functionally Competes with Nuclear Export Activity to Retain Target Transcripts

Toomas Silla, Evdoxia Karadoulama, Dawid Mąkosa, Michal Lubas, and Torben Heick Jensen

Figure S1. Related to Figure 1

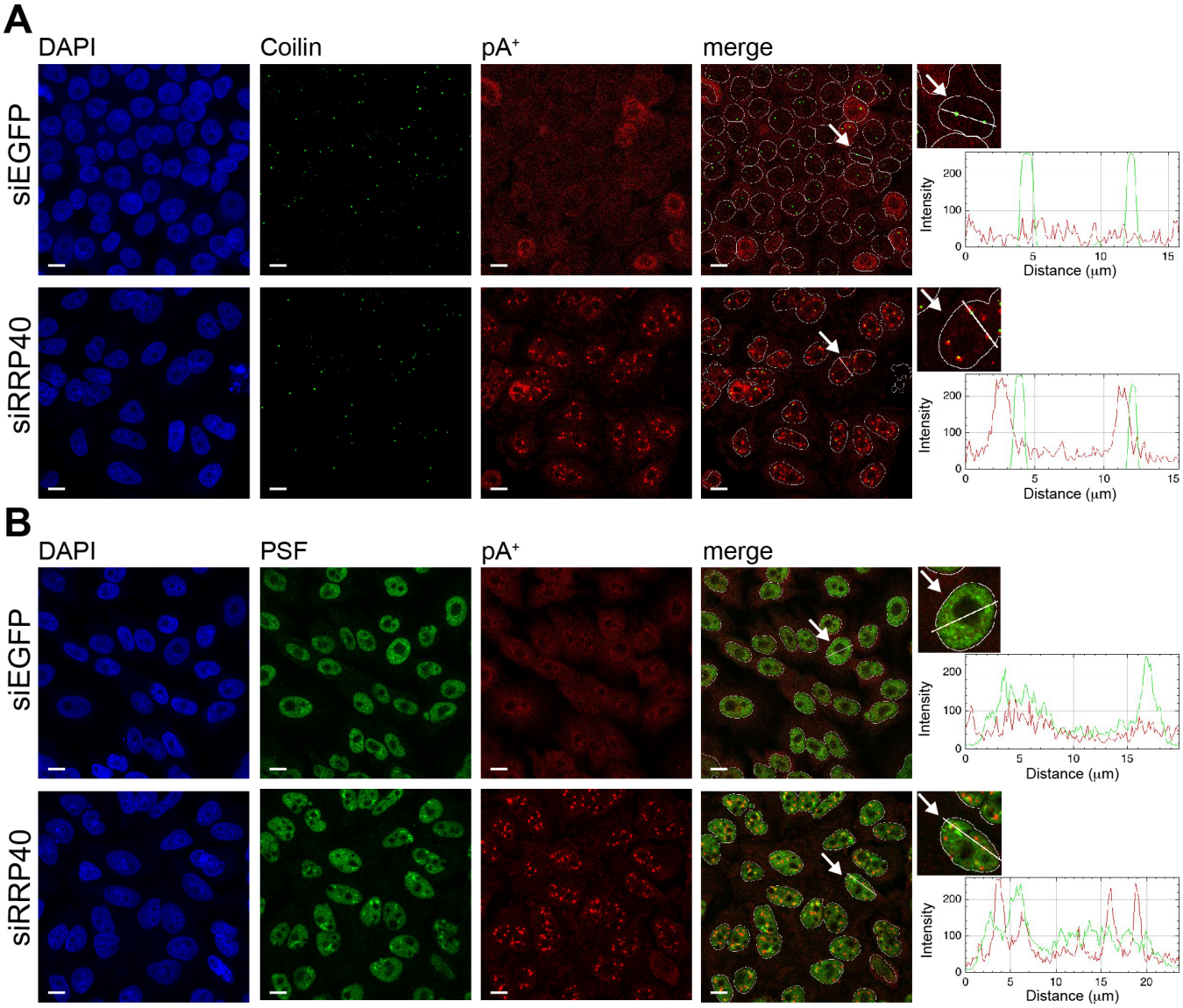

FIGURE S1. Related to Figure 1.
(A-B) Co-localization analysis of pA^+ RNA foci with the Cajal body marker Coilin (A) or the paraspeckle marker PSF (B). Images are displayed as in Figure 1C.

Figure S2. Related to Figure 2

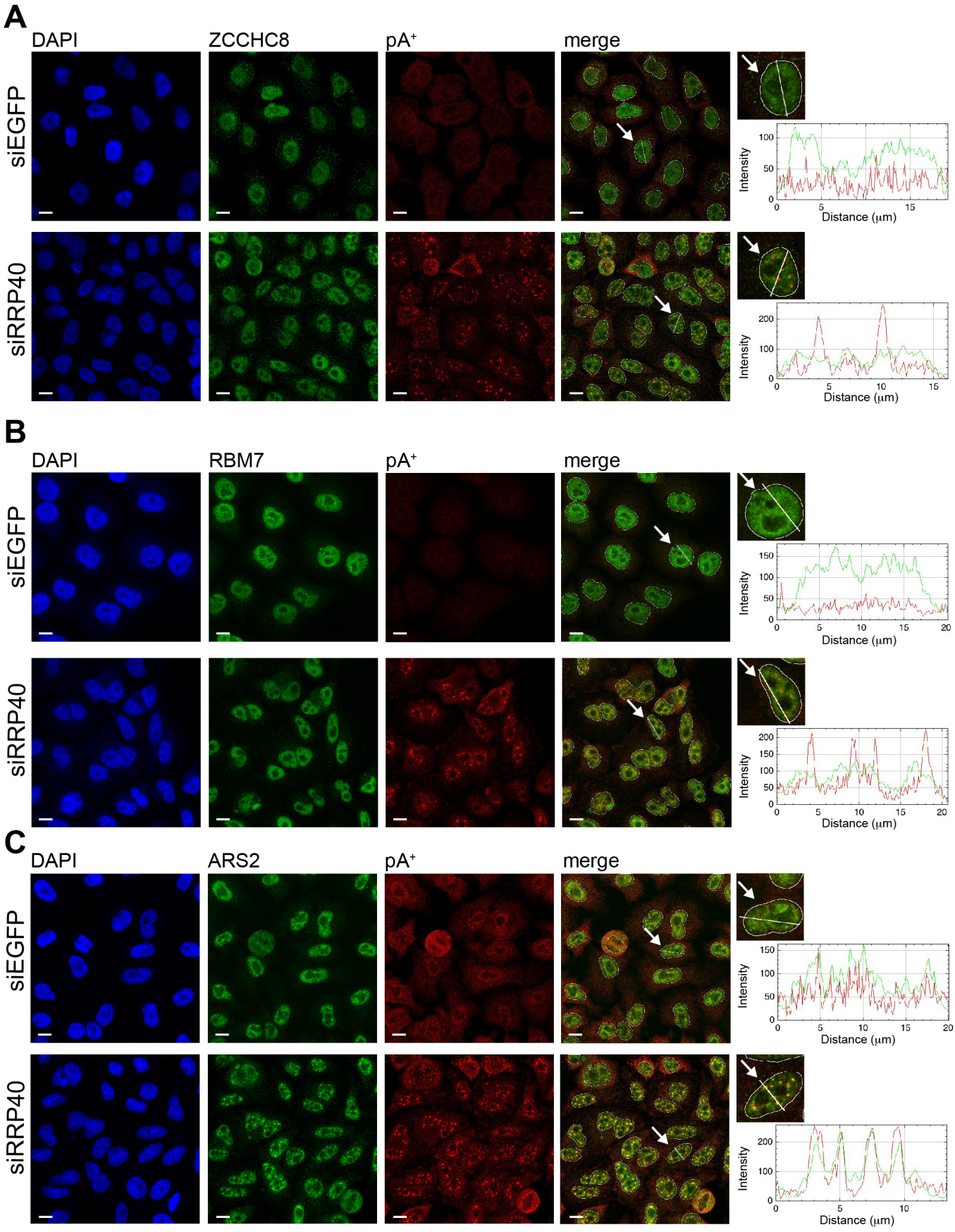

FIGURE S2. Related to Figure 2.
(A-C) Co-localization analysis of NEXT complex components ZCCHC8 (A) and RBM7 (B) as well as the CBCA component ARS2 (C) with pA$^+$ RNA in control (siEGFP) or RRP40 depleted (siRRP40) HeLa cells. Image display as in Figure 2.

Figure S3. Related to Figure 3

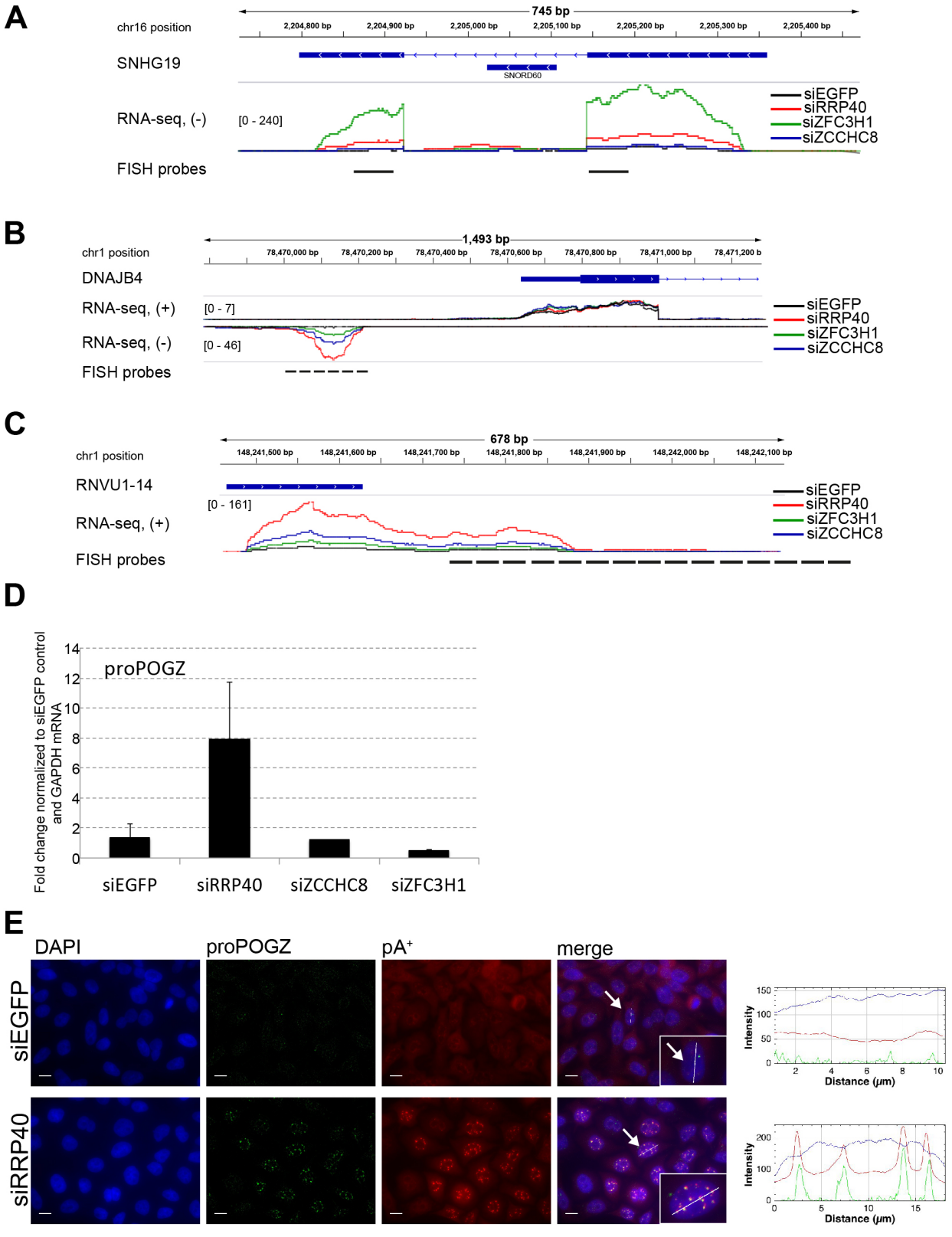

FIGURE S3. Related to Figure 3.
(A-C) Genome browser view of the SNHG19 (A), DNAJB4 (A) and RNVU1-14 (C) loci, showing tracks of normalized total RNA-seq reads from control (siEGFP, black line), RRP40 (red line), ZFC3H1 (green line) and ZCCHC8 (blue line) depleted HeLa cells (Meola et al., 2016). All tracks represent an average of three biological replicates. Solid black lines below the sequencing tracks indicate target sites of the used RNA FISH probes. Blue horizontal bars represent the Refseq annotation of the SNHG19 snoRNA host gene and the SNORD60 snoRNA (A), the 5` end of the DNAJB4 (B) and the RNVU1-14 (C) genes.
(D) RT-qPCR analysis of proDNAJB4, RNVU1-14 3`extension and exogenous proPOGZ transcript levels using total RNA harvested from control (EGFP), RRP40, ZCCHC8 and ZFC3H1, siRNA treated Hela cells as indicated. Data are displayed as mean values with error bars denoting SDs (n = 3 biological replicates).
(E) RNA FISH co-localization analyses of pA^+ RNA with exogenous proPOGZ transcript in control (siEGFP) or RRP40 (siRRP40) depleted HeLa cells. Image display as in Figure 3 A-C.

Figure S4. Related to Figure 4

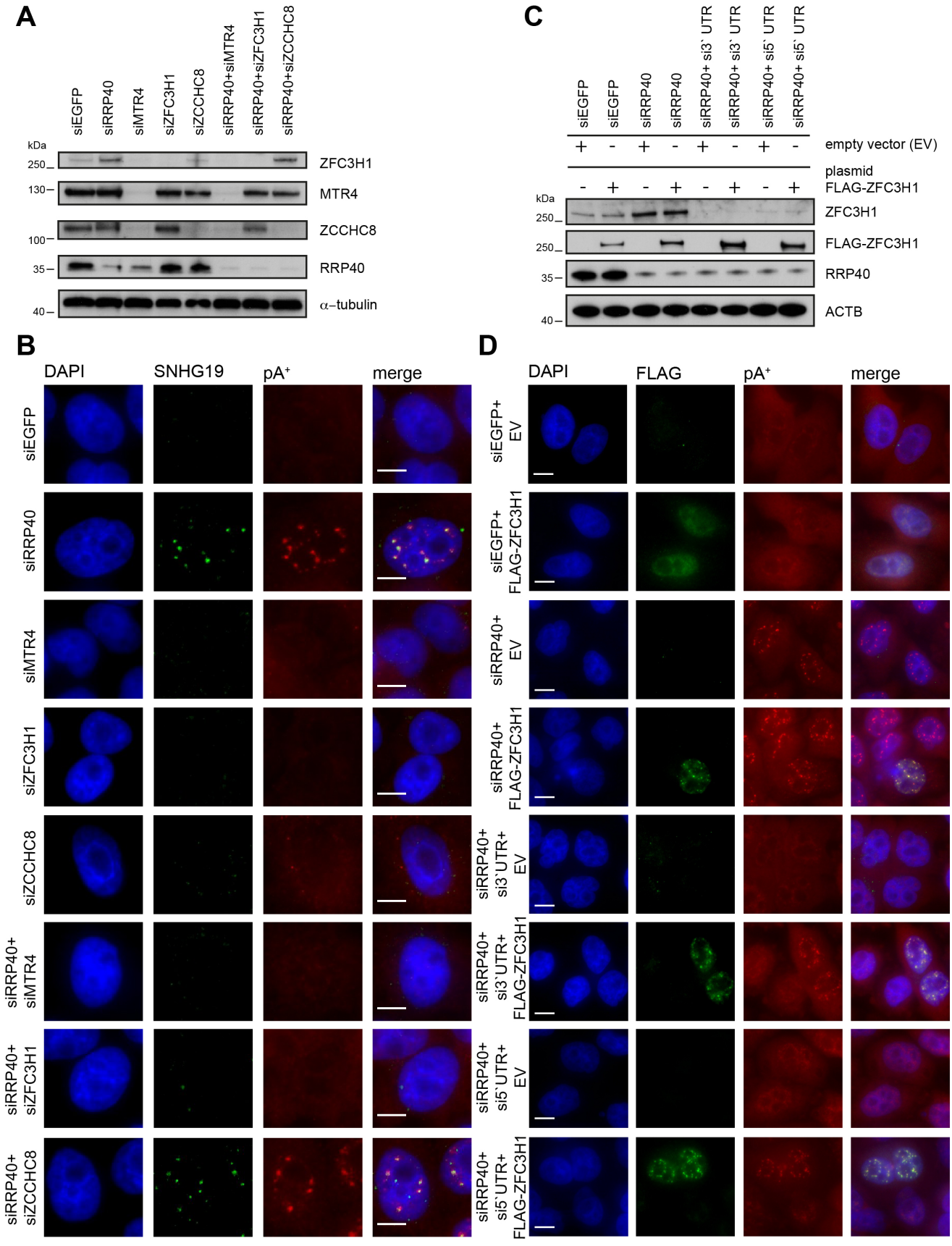

FIGURE S4. Related to Figure 4.
(A) Western blotting analysis using extracts from HeLa cells depleted for the indicated proteins. α-tubulin was used as loading control. Note that MTR4 depletion also affects ZFC3H1, ZCCHC8 and RRP40 protein levels.
(B) Dual SNHG19- and pA^+ RNA-FISH analysis from Figure 4A but with respective signals shown separately. pA^+ RNA was detected with an oligo(dT) 50-mer probe.
(C) Western blotting analysis using extracts from HeLa cells depleted for the indicated proteins and expressing FLAG-tagged exogenous ZFC3H1 or not as indicated. HeLa cells were co-transfected with the indicated siRNAs and the empty vector (EV) or FLAG-ZFC3H1 constructs. Beta-actin (ACTB) was used as loading control. si3'UTR and si5'UTR target the ZFC3H1 RNA 3' and 5' UTRs, respectively. Note that the FLAG-ZFC3H1 expression is not discernable using the ZFC3H1 antibody due to a low plasmid transfection efficiency.
(D) Parallel pA^+ RNA FISH analysis of HeLa cells from (C). Images of DAPI (blue), anti-FLAG antibody (green) and pA^+ RNA (red) are shown together with their merged images. pA^+ RNA was detected with an oligo(dT)-LNA probe. Note that and the pA^+ RNA signal is less exposed in panel (B) compared to (D). Scale bar: 10 μm.

Figure S5. Related to Figure 5

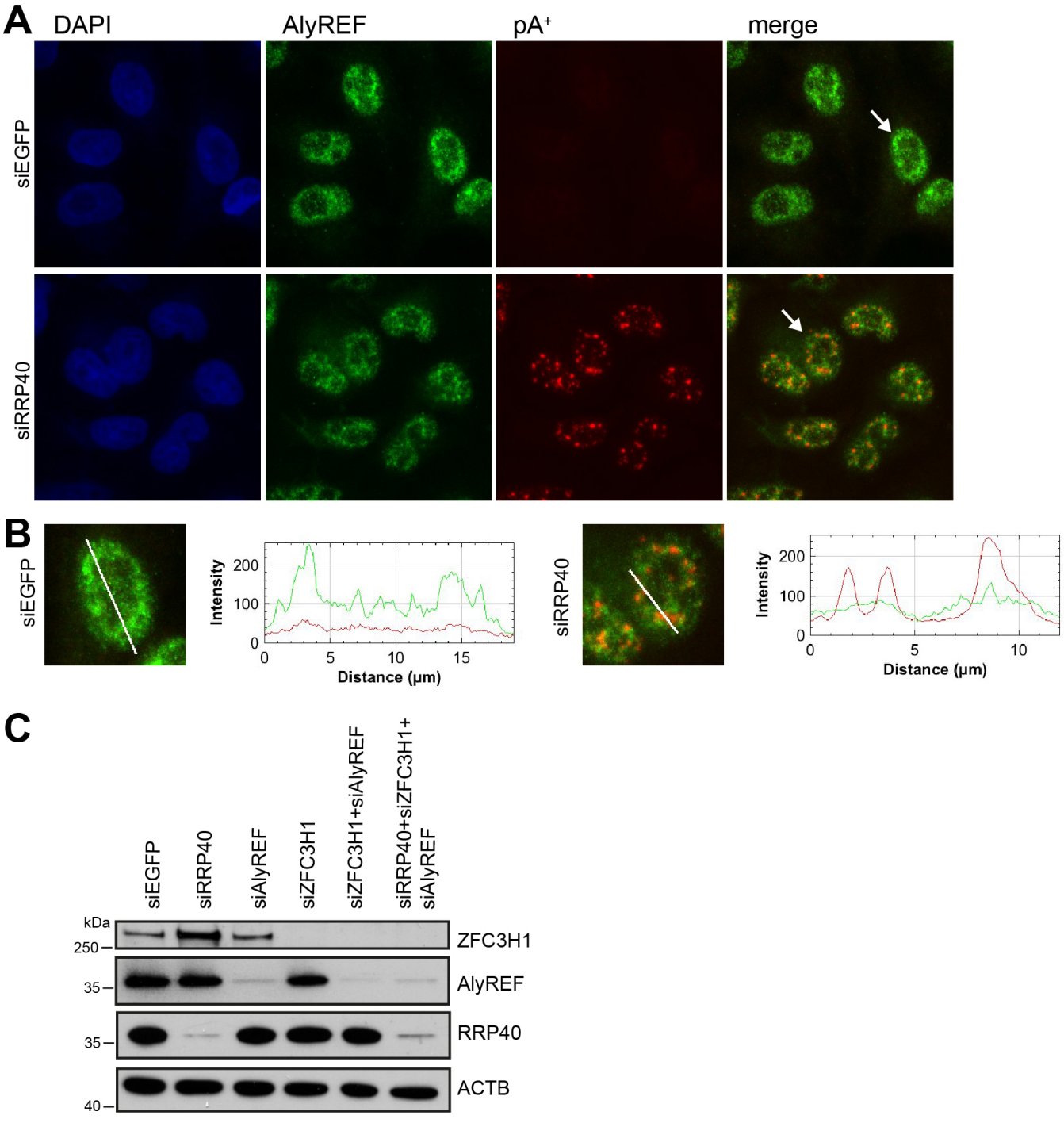

FIGURE S5. Related to Figure 5.
(A) Co-localization analysis of AlyREF and pA$^+$ RNA in control (siEGFP) or RRP40 (siRRP40) depleted HeLa cells. Image display as in Figure 1C.
(B) Zoom-in of cells pointed to by arrows in (A). Plot profiles as in Figure 1C.
(C) Western blotting analysis of levels of the indicated factors using extracts from HeLa cells depleted for the indicated proteins. ACTB (beta-actin) was used as loading control.

Figure S6. Related to Figure 6

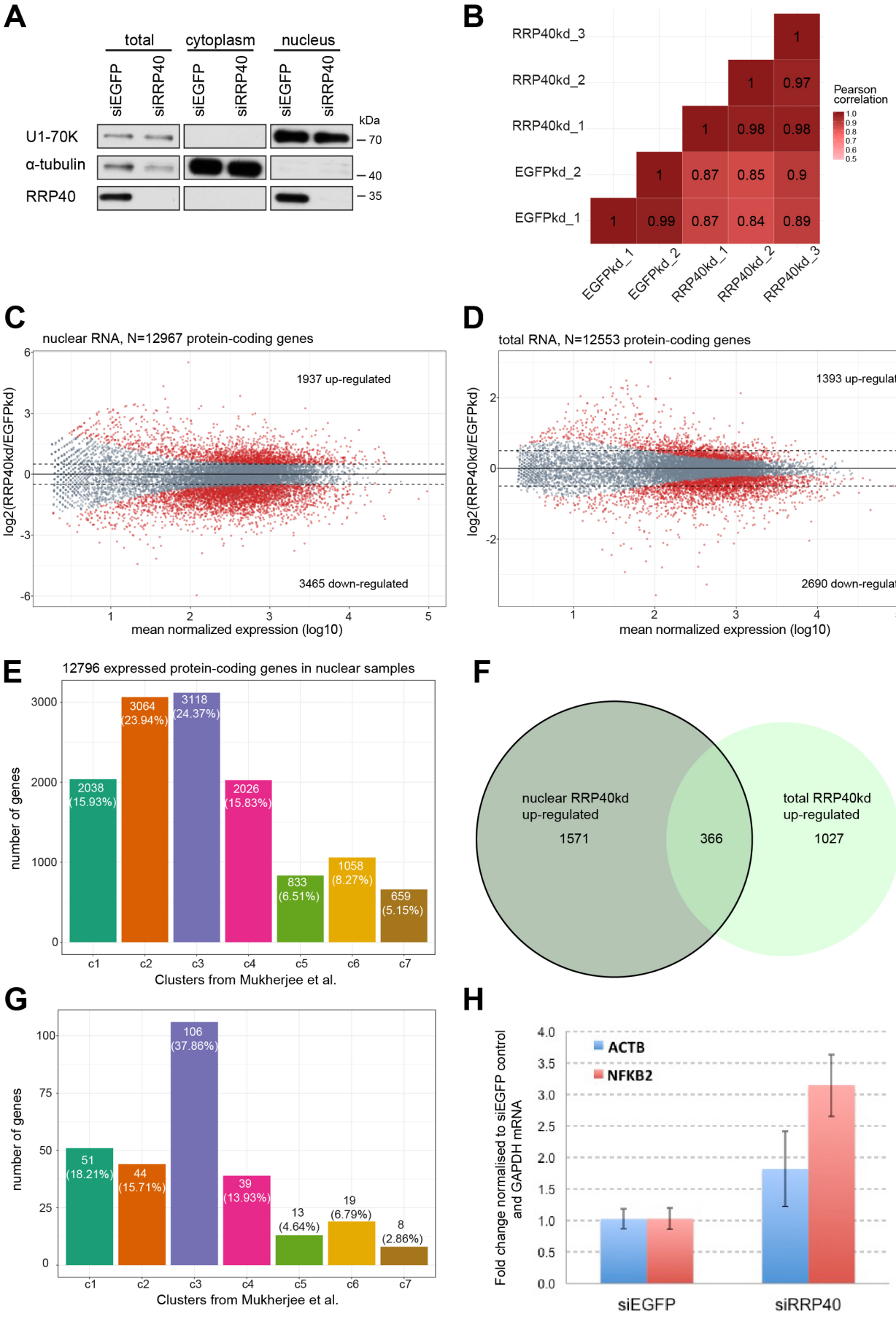

FIGURE S6. Related to Figure 6.

(A) Western blotting analysis of total, cytoplasmic and nuclear fractions extracted from control (siEGFP) and RRP40 (siRRP40) depleted HeLa cells. U1-70K and α-tubulin are used as nuclear and cytoplasmic markers, respectively. (B) Pearson correlation matrix of expression values of all expressed GENCODE annotated genes between the two control (EGFPkd) and three RRP40kd samples. (C and D) MAplot of all expressed protein-coding genes in nuclear (C) and total (D) RNAseq libraries. Results represent differential expression of RRP40kd compared to control samples. Statistically significant hits (padj<0.05) are color-coded red.

(E) Gene classification of all expressed genes in nuclear HeLa samples according to Mukherjee et.al (Mukherjee et al., 2016)

(F) Venn diagram displaying the overlap between significantly up-regulated (log2(RRP40kd/EGFPkd)>0, padj<0.05) protein coding genes in the nuclear and total RNA libraries.

(G) Bar plots showing the distribution of the overlapping genes from (E) according to a previously published coding and non-coding RNA classification (Mukherjee et al., 2016). Note the enrichment of genes belonging to cluster c3 (Pearson's Chi-squared test, p-value = 3.207e-07; against background of all expressed genes in nuclear samples shown in (E)).

(H) RT-qPCR analysis of ACTB and NFKB2 transcripts from nuclear RNA fractions expressed as ratios between RRP40-depleted and control HeLa cells. Data were displayed as mean values with error bars denoting s.e.m (n = 3 biological replicates).

Figure S7. Related to Figure 6

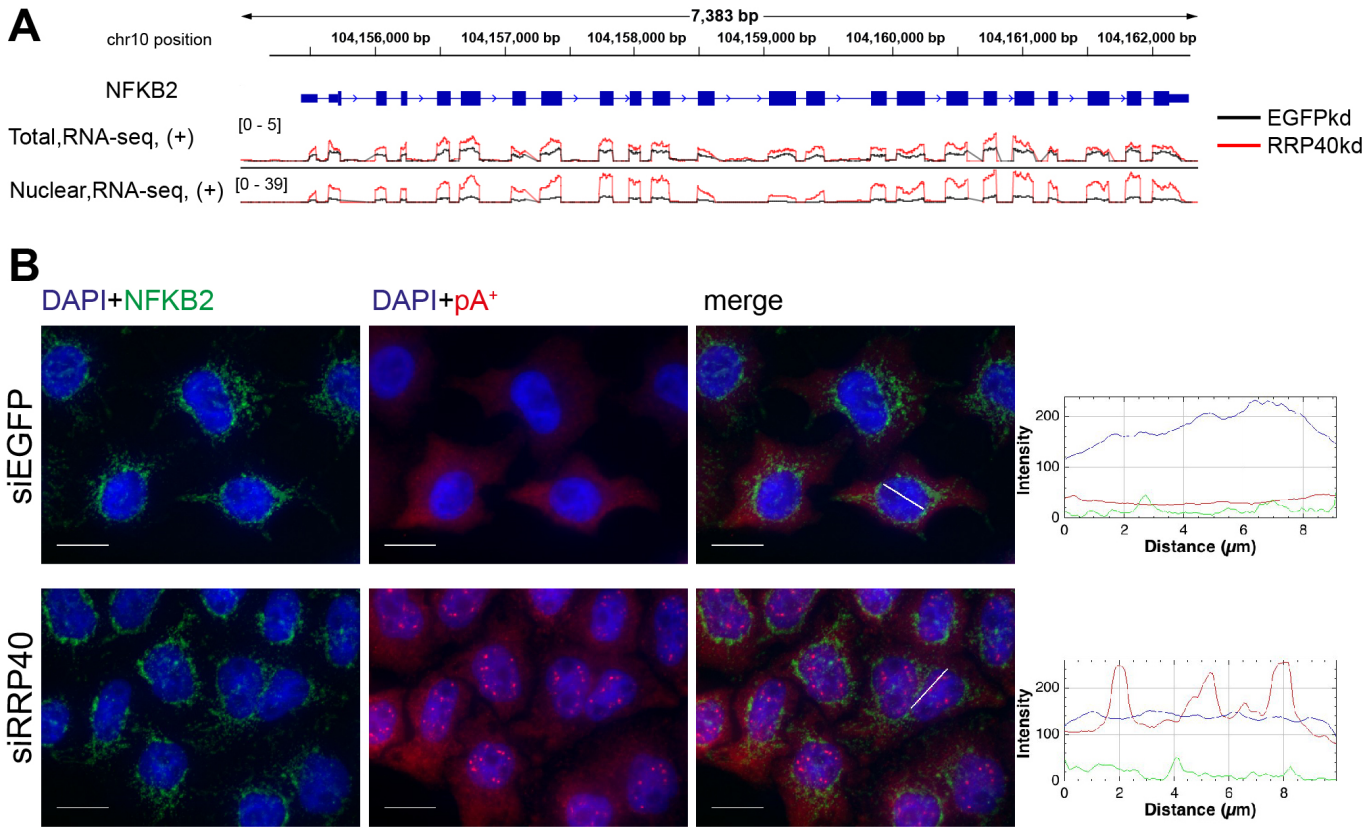

FIGURE S7. Related to Figure 6.
(A) Genome browser view of the NFKB2 locus, showing tracks of normalized total and nuclear RNA-seq data from control (siEGFP, black line) and RRP40 (red line) depleted HeLa cells. Tracks represent average of three biological replicates.
(B) Dual NFKB2- and pA^+ RNA-FISH analysis in the indicated factor-depleted HeLa cells. Image display as in Figure 1C.

Table S1. siRNA oligonucleotides. Related to Experimental Procedures

Target	Sense (5'-3')	Antisense (5'-3')
EGFP	GACGUAAACGGCCACAAGUdTdT	ACUUGUGGCCGUUUACGUCdTdT
ZCCHC8	GGAAUGUACCUCAGGAUAAdTdT	UUAUCCUGAGGUACAUUCCdTdT
ZFC3H1	GAUUAGAGUCCAUGAUUAAdTdT	UUAAUCAUGGACUCUAAUCdTdT
RRP40	CACGCACAGUACUAGGUCAdTdT	UGACCUAGUACUGUGCGUGdTdT
AlyREF	CGUGCACUUUGAGCGGAAGdTdT	CUUCCGCUCAAAGUGCACGdTdT
MTR4	CACGCACAGUACUAGGUCAdTdT	UGACCUAGUACUGUGCGUGdTdT
PABPN1	GUAGAGAAGCAGAUGAAUAdTdT	UAUUCAUCUGCUUCUCUACdTdT
ZFC3H1 5`UTR	GGGCUAAGGUUGUGUGGAAdTdT	UUCCACACAACCUUAGCCCdTdT
ZFC3H1 3`UTR	GGUCAAAUAUUAUGUGCAAdTdT	UUGCACAUAAUAUUUGACCdTdT

Table S2. FISH probes. Related to Experimental Procedures (Excel sheet)

Table S3. Antibodies used in sequential RNA-FISH and Immunofluorescence. Related to Experimental Procedures

Target	**Source**	**Cat # reference**	**Working dilution**
MTR4	Abcam	ab70551	1:1000
ZFC3H1	Novus Biologicals	NB100-68267	1:500
PABPN1	Abcam	ab75855	1:1000
ARS2	Abcam	ab88392	1:1000
ZCCHC8	Abcam	ab68739	1:500
RBM7	Sigma-Aldrich	HPA013993	1:1000
AlyREF	Sigma-Aldrich	HPA019799	1:1000
PSF	Sigma-Aldrich	P2860	1:1000
Coilin	Abcam	ab87913	1:1000
SC35	Abcam	ab11826	1:1000
FLAG	Sigma-Aldrich	F3165-2MG	1:2000

Table S4. Antibodies used in western experiments. Related to Experimental Procedures

Target	**Source**	**Cat # reference**	**Working dilution**
MTR4	Abcam	Ab70551	1:2500
ZFC3H1	Novus Biologicals	NB100-68267	1:1000
ZCCHC8	Abcam	Ab68739	1:2500
β-ACTIN	Sigma-Aldrich	A2228	1:100000
RRP40	Proteintech	15062–1-AP	1:4000
α-TUBULIN	Rockland	600-401-880	1:2500
U1-70K	custom made		1:10000
AlyREF	Abcam	Ab202894	1:10000
FLAG	Sigma-Aldrich	F3165-2MG	1:10000

Table S5. qPCR oligonucleotides. Related to Experimental Procedures

Target	Forward (5'-3')	Reverse (5'-3')
GAPDH	GTCAGCCGCATCTTCTTTTG	GCGCCCAATACGACCAAATC
SNHG19	CGTCCAGGCCTGGCCTAC	GCTCGCGACGAAACCTGC
RNVU1-14 3` ext	CTGTCCGGTCAGTTGTTTCC	GGAGAACAACCCAACCAACA
proPOGZ	GCTGGGCCTGCAAATAAATA	TTGGCAAGTAAATTGGGGAATA
proDNAJB4	TTTCTGGCGTTTCTGATTGA	ACCAAAACGCAGGTTGTTTA
ACTB mRNA	GAGACCGCGTCCGCC	ATCATCCATGGTGAGCTGGC
NFKB2 mRNA	TCTACTGGAGGCCCTGTCTG	TTCCTTCACCTCTGTGCTGG

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Western blotting analysis
Cell pellets were re-suspended in RSB100 (100 mM Tris, pH 7.5, 100 mM NaCl, 2.5 mM MgCl2, 0.5% Triton X-100) and incubated on ice for 10 min. Cell debris was removed by centrifugation at 4000g, 4°C for 15 min. Samples were separated by 10% denaturing PAGE and transferred to PVDF membranes (Millipore), which were blocked in 5% skimmed-milk powder (SMP) in PBS for one hour at room temperature. Primary antibody solutions in 5% SMP in PBS were added to the membranes and incubated one hour at room temperature. Membranes were then washed three times for 5 min in PBS/0.05% Tween 20 and incubated in horseradish peroxidase (HRP)-conjugated goat-anti-rabbit or -mouse secondary antibody (Dako) in 5% SMP in PBS, washed again three times for 5 min in PBS/0.05% Tween 20 and exposed using Supersignal West Femto Substrate (Thermo Fischer Scientific). All used antibodies and applied concentrations are listed in Table S4.

Cellular fractionation
Cells were collected in PBS/3mM EDTA, pelleted by centrifugation and lysed on ice in 1 ml (per 150 mm dish) fractionation buffer (10 mM Tris/HCl [pH 7.4], 10 mM NaCl, 2.5 mM MgCl2, 1mM DTT, 200 µg/ml Digitonin, protease inhibitors [Roche], 40 U RiboLock RNase inhibitors [Thermo Fisher Scientific]) by gentle pipetting. Lysates were left on a tube rotator for 10 min at 4°C. Subsequently, aliquots were taken for total RNA and protein analysis. Nuclei were pelleted by centrifugation at ~1000g for 5 min at 4°C. After centrifugation the supernatant (cytoplasmic fraction) was transferred to a new tube and centrifuged at maximum speed in a tabletop centrifuge at 4°C. The nuclear pellet was washed in 1 ml of homogenization buffer (10 mM Tris/HCl [pH 7.4], 15 mM NaCl, 60 mM KCl, 150mM Sucrose, 0.3% NP40) and centrifuged at ~1000x g for 5 min at 4°C. The nuclear pellet was re-suspended in 1 ml of S1 solution (0.25 M sucrose, 10 mM MgCl2) and added on top of an equal volume of S3 solution (0.88 M sucrose, 0.5 mM MgCl2), and centrifuged at 2800g for 10 min at 4°C. Supernatant was discarded and the nuclear pellet was re-suspend in 1 ml of wash buffer (10 mM Tris/HCl [pH 7.4], 15 mM NaCl, 60 mM KCl) and centrifuged at 1000x g for 5 min. After centrifugation the supernatant was discarded and the nuclear pellet was washed with 1 ml PBS and centrifuged again at 1000x g for 5 min. The resulting supernatant was discarded and the nuclear pellet was re-suspended in PBS and used for RNA extraction and protein analysis.

RNA isolation and RT-qPCR analysis
RNA was purified using TRIzol reagent (Thermo Fisher Scientific) according to the manufacturer's instructions. Purified RNA was DNase I treated (TURBO DNA-free Kit, Thermo Fisher Scientific). For RT-qPCR analysis cDNA was synthesized using random hexamer primers according to standard procedures. qPCR was performed using primers from Table S5 on a AriaMx Real-time qPCR System (Agilent). Data were processed using the ΔΔCt method, with normalization to both GAPDH mRNA levels and EGFP-siRNA control samples.

RNAseq library preparation and data processing
50 µg of nuclear RNA was used for pA^+ RNA purification using Oligo (dT)25 Dynabeads (Thermo Fisher Scientific) according to the manufacturer's instructions. ~750 ng of nuclear pA^+ RNA was treated with Ribo-ZeroTM Magnetic Gold Kit (Human/Mouse/Rat; Illumina) according to the manual. cDNAs were produced using random hexamer-primers. Strand specific libraries and their sequencing on Illumina HiSeq 2000 were performed at BGI Tech Solutions (Hong Kong, China). Sequence reads were quality controlled using FastQC v0.11.2 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmomatic v0.32 was used to remove remainders of illumine adaptors, low quality bases, the first 12 bases of the reads and reads below 25 nucleotides. The settings used were "ILLUMINACLIP:<TrueSeq3_PE_2>:2:30:10 HEADCROP:12 LEADING:22 SLIDINGWINDOW:4:22 MINLEN:25" (Bolger et al., 2014). Trimmed reads were mapped against the human genome 19 (hg19), using HISAT v0.1.6.beta (Kim et al., 2015) by providing a list of GENCODE-annotated splice sites, setting the maximum fragment length to 1000 and using parameter --rf for the upstream/downstream mate orientation, and otherwise

using default settings. Samtools v0.1.17 (Li et al., 2009) were utilized to select uniquely mapped and properly paired reads, which were then used for downstream analysis.

Gene annotation
For annotation all genes from the human GENCODE annotation v19 (Harrow et al., 2012) were used together with PROMPT and eRNA sets from (Meola et al., 2016). Exons and introns were defined from the GENCODE annotation, that is, all nucleotides within a gene that could be annotated as exons in any of the isoforms of the gene, were annotated as exonic and all the remaining nucleotides within the gene region were annotated as intronic. R and python scripts were used and are available upon request.

RNAseq quantification
The Rsubread package (Liao et al., 2013) from the Bioconductor R package (http://www.bioconductor.org) was used to quantify expression in control (EGFPkd) and exosome depleted (RRP40kd) nuclear pA^+ and total libraries (GEO database ID: GSE84172, (Meola et al., 2016)) by counting the RNAseq fragments that overlapped exons on the relevant strand and summarized on the gene level. Fragment counts were converted to tags per million (where tags were defined as uniquely mapped and properly paired fragments). The quantification was done for all genes in the human GENCODE annotation v19, for PROMPTs and eRNAs used previously (Meola et al., 2016) and for all collapsed exons and introns defined using the GENCODE annotation v19. Then, for Figures 6A and S6C-D, the DESeq2 package (Love et al., 2014) was used to perform differential expression analysis between the control and the knock down samples for both total and nuclear $polyA^+$ RNAseq libraries. In Figure 6B, exon and intron exosome sensitivity was calculated based on the strand specific expression using following formula:

$$(1)\ \text{sensitivity}_{\text{KD}} = \frac{\text{expression}_{\text{KD library}} + \varepsilon}{\text{expression}_{\text{CTRL library}} + \varepsilon},$$

where the expression values are in TPMs and ε is a pseudocount defined as the minimum expression value of all RNAs in all libraries min(expression_ALL>0)

Statistics and RNAseq data visualisation
Visualisations were performed using the ggplot2 R package (Wickham et al, 2009). Statistical tests were done in the environment of the R Project for Statistical Computing (https://www.r-project.org).

SUPPLEMENTAL REFERENCES

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data . Bioinformatics 30, 2114–2120.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res 22, 1760–1774.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat Meth 12, 357–360.

Liao, Y., Smyth, G.K., and Shi, W. (2013). The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. Nucleic Acids Research 41, e108–e108.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550–21.

Meola, N., Domanski, M., Karadoulama, E., Chen, Y., Gentil, C., Pultz, D., Vitting-Seerup, K., Lykke-Andersen, S., Andersen, J.S., Sandelin, A., et al. (2016). Identification of a Nuclear Exosome Decay Pathway for Processed Transcripts. Molecular Cell 64, 520–533.

Mukherjee, N., Calviello, L., Hirsekorn, A., de Pretis, S., Pelizzola, M., and Ohler, U. (2016). Integrative classification of human coding and noncoding genes through RNA metabolism profiles. Nat. Struct. Mol. Biol. 24, 86–96.

Wickham H (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York