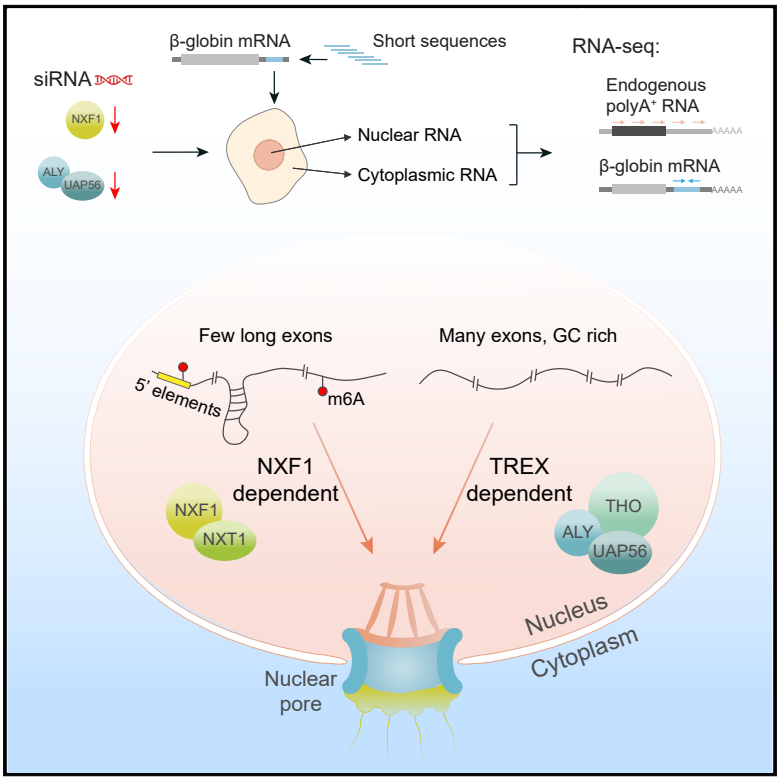
Molecular Cell

Article

Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex

Graphical Abstract



Highlights

- Depletion of NXF1 and TREX retains in the nucleus different transcript groups
- Transcripts with few or long exons are preferentially dependent on NXF1
- G/C-rich, 5′-biased, and m^{6}A-modified regions drive single-exon transcript export
- Splicing efficiency affects export in a largely NXF1-independent manner

Authors

Binyamin Zuckerman, Maya Ron, Martin Mikl, Eran Segal, Igor Ulitsky

Correspondence

igor.ulitsky@weizmann.ac.il

In Brief

Zuckerman et al. study the consequences of depletion of core components of the nuclear RNA export pathway in human cells. Different components are required for nuclear export of distinct transcript sets. Gene architecture, sequence composition, RNA secondary structure, RNA modifications, and certain sequence motifs are associated with this selective dependency.

Zuckerman et al., 2020, Molecular Cell 79, 251–267
July 16, 2020
https://doi.org/10.1016/j.molcel.2020.05.013

CellPress

Molecular Cell

CellPress

Article

Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex

Binyamin Zuckerman,[1] Maya Ron,[1] Martin Mikl,[2,3] Eran Segal,[2,3] and Igor Ulitsky[1,4,*]
[1]Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel
[2]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel
[3]Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot 76100, Israel
[4]Lead Contact
*Correspondence: igor.ulitsky@weizmann.ac.il
https://doi.org/10.1016/j.molcel.2020.05.013

SUMMARY

The core components of the nuclear RNA export pathway are thought to be required for export of virtually all polyadenylated RNAs. Here, we depleted different proteins that act in nuclear export in human cells and quantified the transcriptome-wide consequences on RNA localization. Different genes exhibited substantially variable sensitivities, with depletion of NXF1 and TREX components causing some transcripts to become strongly retained in the nucleus while others were not affected. Specifically, NXF1 is preferentially required for export of single- or few-exon transcripts with long exons or high A/U content, whereas depletion of TREX complex components preferentially affects spliced and G/C-rich transcripts. Using massively parallel reporter assays, we identified short sequence elements that render transcripts dependent on NXF1 for their export and identified synergistic effects of splicing and NXF1. These results revise the current model of how nuclear export shapes the distribution of RNA within human cells.

INTRODUCTION

Transcription occurs primarily in the nucleus, and most well-expressed long RNAs, including presumably all mRNAs, need to move to the cytosol to carry out their functions. Nuclear RNA export is tightly controlled and coupled to transcription and RNA processing in the nucleus (Hocine et al., 2010). mRNAs encoding highly translated and uniformly expressed proteins mostly need to be exported from the nucleus as quickly and as robustly as possible. In contrast, there are other RNA subsets for which export control is desired. For example, some mRNAs are needed only upon specific stimuli and yet may take a long time to transcribe, so regulated export can help uncouple the time needed for transcription and the time to first protein production (Mauger et al., 2016; Ni et al., 2016; Zhou et al., 2017). In various conditions, regulated export can also influence the amount of noise in protein levels (Bahar Halpern et al., 2015; Battich et al., 2015; Hansen et al., 2018)

Splicing density and efficiency, length, sequence composition, and chromatin environment are all implicated in nuclear export efficacy (Palazzo and Lee, 2018; Valencia et al., 2008), and studies using massively parallel assays have identified specific sequences regulating nuclear enrichment (Carlevaro-Fita et al., 2019; Lubelsky and Ulitsky, 2018; Shukla et al., 2018; Yin et al., 2020). We recently combined many of these features, built statistical models that can predict cytoplasmic/nuclear (Cyto/Nuc) ratios in human cells, and found that splicing efficiency alone can explain up to a third of the variance in export ratios (Zuckerman and Ulitsky, 2019). The cellular mechanisms that integrate the different features into export efficiency are largely unknown. Splicing has been shown to contribute to regulation of export both through recruitment of export pathway components by splicing factors and through inhibition of export of incompletely spliced transcripts (Elbarbary and Maquat, 2016; Reed and Hurt, 2002; Yap et al., 2012). Still, most intronless transcripts are efficiently exported, which was proposed to rely on specific sequence elements, and such elements have been studied in several genes (Lei et al., 2011, 2013; Wang et al., 2018; Guang et al., 2005; Huang and Carmichael, 1997; Huang et al., 1999).

Studies of endogenous and viral RNAs in various species have identified the core components of the nuclear RNA export pathway (Carmody and Wente, 2009). Cases of selective export of RNA subsets have also been identified (Wickramasinghe and Laskey, 2015). Importantly, many of the canonical members of the export pathway were implicated in the export of polyadenylated RNAs using fluorescence *in situ* hybridization (FISH) with oligo-d(T) probes, and so whereas these factors are likely required for export of a substantial portion of polyadenylated transcripts, it is unclear what fraction of genes rely on them for export.

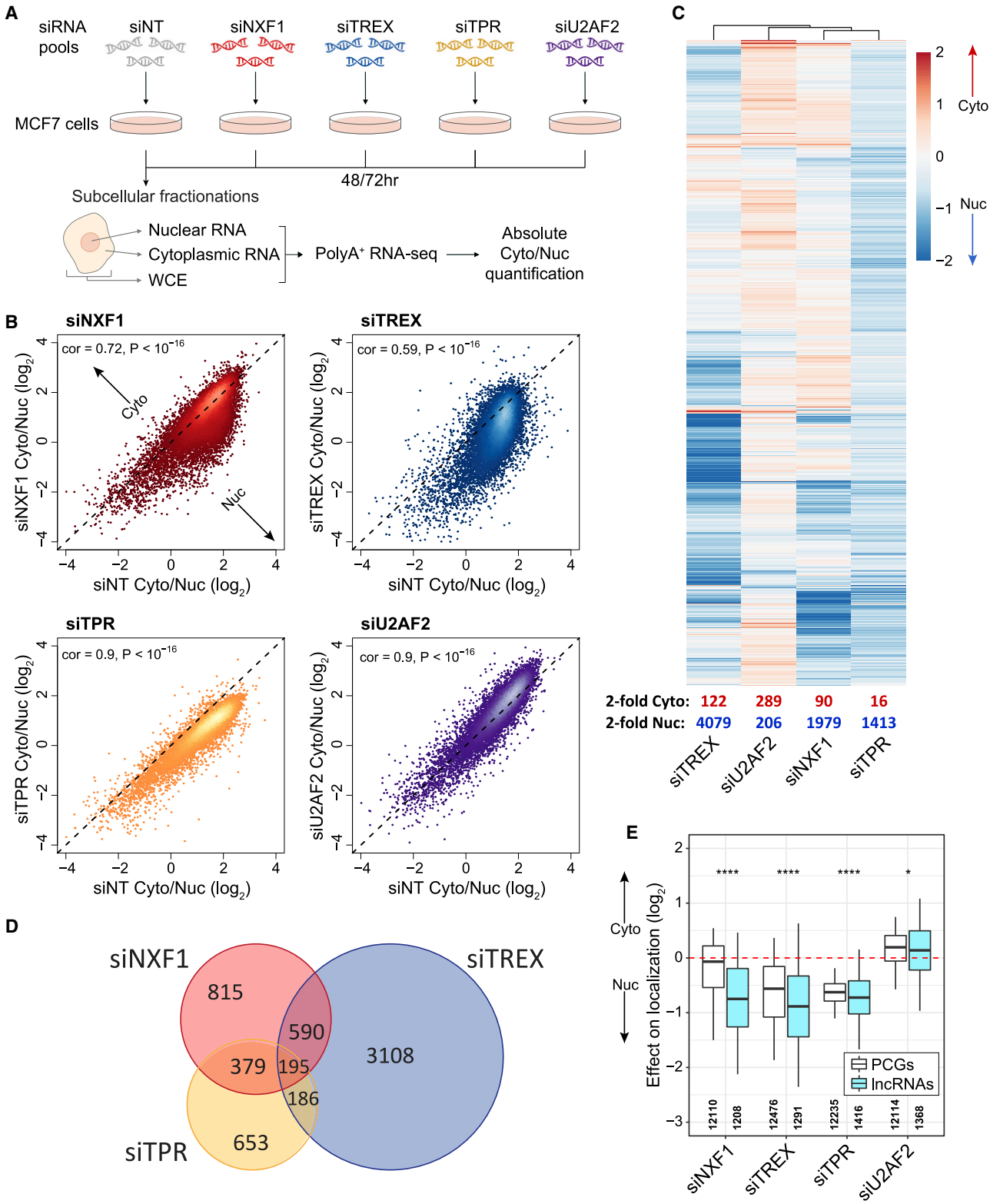


Figure 1. Depletion of Export Factors Has a Selective Effect on Subcellular Localization of Coding and Non-coding Genes

(A) Experimental design of export factor depletion, cytoplasmic and nuclear fractionation, RNA-seq, and quantification of absolute subcellular localization for each gene (see STAR Methods).

(legend continued on next page)

Early transcriptome-wide studies in fly S2 cells have shown that depletion of NXF1 (Tap), NXT1 (p15), and UAP56 (*DDX39B*, one of the core components of the TREX complex) affected the Cyto/Nuc distribution and expression levels of the vast majority of mRNAs in a similar way, leading to a conclusion that these proteins all act in the same pathway and export most mRNAs (Herold et al., 2003). In contrast, differences between RNAs associated with ALY (*ALYREF*, another TREX component) and NXF1 homologs were observed in yeast (Hieronymus and Silver, 2003). ALY depletion has been reported to have a minimal effect on export of polyadenylated RNA in fly (Gatfield and Izaurralde, 2002) but reduced export in human cells (Katahira et al., 2009; Silla et al., 2018). Similarly, UAP56 depletion was shown to cause strong nuclear accumulation of polyadenylated RNA in nematodes, flies, and human cells (Fan et al., 2018; Gatfield et al., 2001; MacMorris et al., 2003), suggesting that core TREX components are essential for nuclear export of all processed transcripts. Later studies proposed a three-step mechanism of export-competent messenger ribonucleoprotein (mRNP) complex formation (Carmody and Wente, 2009; Hautbergue et al., 2008): (1) UAP56 is recruited to mRNAs cotranscriptionally, (2) UAP56 recruits ALY to mRNAs in ATP-dependent manner, and (3) NXF1-NXT1 dimer is recruited by and eventually replaces ALY and then facilitates the transition of the RNA through the nuclear pore.

NXF1 is a key component of the canonical export pathway that is thought to be required for export of virtually all long RNAs (Carmody and Wente, 2009). NXF1 by itself has weak affinity for RNA (Katahira et al., 1999), and cross-linking immunoprecipitation (CLIP) studies of NXF1 did not identify any clear sequence specificity (Viphakone et al., 2019). It has been suggested that various adaptor proteins, including ALY (Rodrigues et al., 2001) and SR proteins (Huang et al., 2003, 2004; Müller-McNicoll et al., 2016; Wang et al., 2018), recruit NXF1 to long RNAs and that TREX binding leads to a change in NXF1 conformation that increases its RNA affinity (Hautbergue et al., 2008; Viphakone et al., 2012). The exon junction complex (EJC), which binds G/C-rich sequences upstream of excised introns (Singh et al., 2012), was also reported to recruit NXF1 (Le Hir et al., 2001), although mRNAs can clearly recruit NXF1 without splicing, likely to their 5′ end (Cheng et al., 2006). A recent study also found that co-transcriptional recruitment of NXF1 regulates alternative polyadenylation, as NXF1 depletion caused increased use of proximal cleavage and polyadenylation sites in last exons of many genes (Chen et al., 2019). The currently accepted model is that the export of the vast majority of RNAs relies on both TREX and NXF1.

TPR is a nuclear basket component that has also been proposed in multiple studies to play regulatory roles in RNA export (Lee et al., 2019; Umlauf et al., 2013). Using reporters, it was shown that loss of TPR facilitates export of incompletely spliced RNAs through the NXF1 pathway (Coyle et al., 2011), although more recent studies found no evidence for requirement for TPR for retention of unspliced transcripts (Lee et al., 2019). TPR was also recently shown to restrict the number of nuclear pores per nucleus (McCloskey et al., 2018). Another proposed regulator of the export of incompletely spliced RNAs is the 65 kD subunit of U2AF, U2AF2 (U2AF^{65}), which was shown to mediate the nuclear retention of an intron-containing reporter (Takemura et al., 2011). Interestingly, U2AF2 is also required for nuclear export of some intronless transcripts, pointing to its dual role in splicing-dependent and splicing-independent export (Lei et al., 2013). Yet detailed export regulation mechanisms of incompletely spliced RNAs and their global effect on the transcriptome are unknown.

Here, we used RNAi-based perturbations followed by sequencing of RNAs from different compartments and found that in human cells, depletion of core TREX components and of NXF1 affected different subsets of long noncoding RNAs (lncRNAs) and mRNAs, which differ in their gene architecture, sequence composition, and protein binding partners. In contrast, depletion of TPR and U2AF2 had a much less selective effect. We further focused on NXF1, which is preferentially required for export of long transcripts with few exons or multiexon but A/U-rich transcripts. In selected genes with one or two exons, we identified specific, structured, 5′-biased, and G/C-rich RNA regions, enriched with N^6-methyladenosine (m^{6}A) modifications, that drive NXF1-dependent nuclear export. We also studied the crosstalk between NXF1 dependence and splicing efficiency. Last, we show that NXF1 selectivity is related to selectivity of RNA export block induced by viral proteins.

RESULTS

Non-uniform Export Sensitivity upon Depletion of RNA Export Factors

To explore the selectivity of canonical RNA export factors, we used RNAi to deplete NXF1, TPR, U2AF2, and the TREX components ALY and UAP56 in human breast cancer MCF7 cells and examined the subcellular localization of polyadenylated transcripts by Cyto/Nuc fractionation followed by RNA sequencing (RNA-seq) (Figures 1A, S1A, and S1B). We used the whole-cell extract (WCE) samples to normalize the Cyto/Nuc ratios and obtain absolute RNA localization values, as suggested by Carlevaro-Fita and Johnson (2019) (Figure S1C; Table S1; see STAR Methods). In order to avoid splicing-related bias and complexity, we focused the analysis on the gene rather than on the isoform

(B) Genome-wide effects of export factor depletions on subcellular localization (normalized Cyto/Nuc values in export factor depletion samples versus control [siNT]; see STAR Methods). Color intensity indicates local point density. Spearman's R and p values are indicated. Dots, genes; black dashed line, X = Y.
(C) Subcellular localization changes upon export factor depletions. Color indicates the difference between knockdown and control (siNT) $\log_2$-transformed Cyto/Nuc ratios; cytoplasmic shift is colored red, while nuclear shift is colored blue. Number of genes with at least 2-fold localization difference is indicated below the heatmap for each sample.
(D) Venn diagram indicating number of genes in different categories exhibiting at least 2-fold increase of nuclear enrichment upon export factor depletion.
(E) Effects of export factor depletions on subcellular localization of PCGs and lncRNAs. Plot indicates the median, quartiles, and 5th and 95th percentiles. Differences between medians of PCGs and lncRNAs are 0.68 for siNXF1, 0.32 for siTREX, 0.10 for siTPR, and 0.06 for siU2AF2. *p < 0.05 and ****p < 0.0001 (Wilcoxon rank-sum test).
See Figure S1.

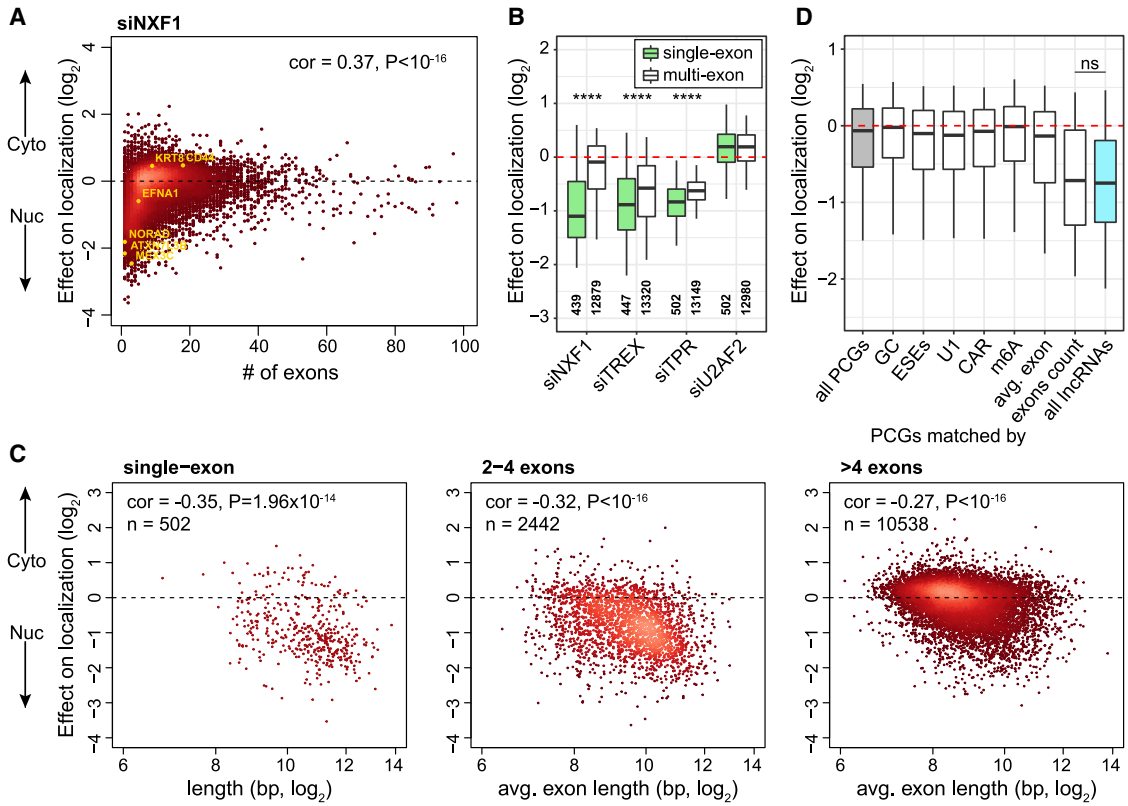


Figure 2. Genes with Few or No Introns Are Particularly Sensitive to NXF1 Depletion

(A) Correlation between exon count and effect of NXF1 depletion on localization. Genes that were selected for validation by smFISH (see Figure 3) are highlighted. Sixteen genes with more than 100 exons were omitted from the plot.

(B) Effect of export factor depletions on localization of single-exon and multi-exon transcripts. Number of genes represented by each box is indicated at the bottom. ****p < 0.0001 (Wilcoxon rank-sum test).

(C) Correlation between average exon length and effect of NXF1 depletion on localization of single-exon (left), two- to four-exon (middle), and multi-exon (more than four; right) genes.

(D) Effect of NXF1 depletion on localization of all PCGs (gray, n = 12,110), all lncRNAs (cyan, n = 1,208), and PCGs matched to lncRNAs using the indicated sequence feature (white, n = 1,208 for each group of sampled PCGs; see STAR Methods). Wilcoxon rank-sum test p values for all comparisons with lncRNAs are $<10^{-15}$, except the last group (PCGs matched using exon counts), for which p > 0.05.

Color intensity indicates local point density. Spearman's R and p values are indicated. See Figure S2.

level. Depletions of NXF1, TPR, and ALY+UAP56 (siTREX) each led to nuclear enrichment of thousands of transcripts, which outnumbered the transcripts that became more cytoplasmic (possibly through indirect effects) by >20-fold (Figures 1B and 1C). Surprisingly, depletion of NXF1, as well as siTREX, affected distinct gene groups, while depletion of TPR and U2AF2 had a smaller and less selective effect on localization (Figures 1C and 1D). TPR depletion led to changes that were significantly similar to those of NXF1 depletion (Figures 1C and 1D; Spearman's R = 0.40, $p < 10^{-20}$), suggesting that its role in regulating nuclear export is related to that of NXF1. Also surprisingly, the effect of export factor depletion on subcellular localization exhibited rather mild correlations with differential expression (Figure S1D), suggesting that whereas some transcripts that are not properly exported are downregulated, that is not the case for most transcripts.

To identify possible secondary and feedback effects on other known export factors, we scrutinized changes in localization and expression of genes associated with "RNA export from nucleus" (taken from Gene Ontology; Figure S1E). Expression levels were mostly unaffected in all samples, with the notable exception of RBM15, which was strongly induced by NXF1 depletion, possibly to compensate for its strong nuclear retention in these conditions, and UAP56 paralog DDX39A (URH49), which was induced by both NXF1 and TREX depletions, possibly as part of feedback regulation to enhance export.

Transcripts with Few Exons Are Selectively Retained in the Nucleus upon NXF1 Depletion

When compared with protein coding genes (PCGs), lncRNAs were more sensitive to NXF1 depletion and to a lesser extent to depletion of TREX components (Figure 1E). We therefore sought features that may explain this differential sensitivity. LncRNAs have fewer and shorter exons compared with PCGs (Derrien et al., 2012). NXF1 depletion caused nuclear enrichment of transcripts with few exons (Figures 2A and 2B). Single-exon

genes were also more affected than multi-exon ones by TREX and TPR depletions, though to a lesser extent (Figure 2B), and an association between exon count and export factor sensitivity was observed for TPR but not for TREX depletion (Figure S2A).

Genes with longer exons were significantly more sensitive to NXF1 depletion (and, to a lesser extent, of other factors), regardless of exon number (Figures 2C and S2B, left). Expression level correlations with exon length did not follow the pattern of localization changes (Figure S2B, right), supporting the hypothesis that at least some localization changes are independent of changes in gene expression. To test whether gene architecture or other sequence features can explain the differences between PCGs and lncRNAs, we randomly sampled PCGs matching to lncRNA by several features, which included gene architecture (exon count and average exonic length), G/C content, and enrichment with sequence features associated by previous studies with nuclear RNA export (see STAR Methods and below), and examined their NXF1 sensitivity (Figure 2D). PCGs matched by number of exons to lncRNAs were affected by NXF1 depletion similarly to lncRNAs, while other groups of sampled PCGs did not resemble lncRNAs (Figure 2D). When subpopulations of lncRNAs and PCGs binned by exon count were compared, the most significant difference was for genes with at least seven exons, in which PCGs were mostly unaffected by NXF1 depletion, whereas some lncRNAs exhibited sensitivity (Figure S2C). We conclude that differences in number of exons explains most of the difference in NXF1 sensitivity between PCGs and lncRNAs.

As cell fractionation followed by sequencing can be noisy, we examined three NXF1-sensitive genes using single-molecule FISH (smFISH): *NORAD*, an abundant intronless and mostly cytoplasmic lncRNA; *ATXN7L3B*, an intronless PCG; and *MEX3C*, a PCG with one efficient intron. We also examined three efficiently spliced multi-exon and apparently NXF1-insensitive genes (*CD44*, *EFNA1*, and *KRT8*) and used an oligo-dT probe to examine subcellular distributions of all PolyA$^+$ RNA. Upon NXF1 depletion, the three NXF1-sensitive genes exhibited strong nuclear enrichment and PolyA$^+$ RNA accumulated in the nucleus, whereas the localization and abundance of transcripts from the control genes was largely unaffected and NXF1 depletion did not increase their nuclear presence (Figures 3A, 3B, S3B, and S3C). When small interfering RNA (siRNA) treatment was followed with NXF1 staining, cells with greater knockdown efficiency exhibited a more substantial nuclear accumulation of *NORAD* (Figures 3C and 3D). Abundance of intronless transcripts *NORAD* and *ATXN7L3B* was also strongly reduced by NXF1 depletion, while abundance of *MEX3C* was only moderately affected (Figure S3C). TREX depletion had also some effect on the localization and/or expression levels of the three NXF1 targets, as well as on *KRT8*, but it did not resemble the strong and highly selective effect of NXF1 depletion (Figures 3B and S3A–S3C). Imaging results therefore supported the conclusions drawn from the RNA-seq analysis.

Subnuclear Distribution of RNA following Export Factor Depletion

Depletion of TREX components promotes accumulation of PolyA$^+$ RNA in nuclear speckles (Chi et al., 2013; Dias et al., 2010), and speckles were shown to be important for export of intronless transcripts (Wang et al., 2018). We used smFISH to examine *MALAT1*, a marker of nuclear speckles (Hutchinson et al., 2007). PolyA$^+$ RNA accumulated in nuclear speckles upon NXF1 depletion, though to a lesser extent than upon TREX depletion (Figure 3E, left; Figure S3A), consistent with previous studies (Wickramasinghe et al., 2013). Three-dimensional (3D) object-based colocalization analysis (see STAR Methods) showed that *NORAD* colocalization with *MALAT1* was mildly but significantly elevated in NXF1-depleted cells but not in TREX-depleted cells (Wilcoxon rank-sum test $p = 3.95 \times 10^{-4}$ and 0.09, respectively; Figure 3E, right; Figures 3F and S3A), despite the significant nuclear retention of *NORAD* in both conditions (Figure 3B). These results suggest that export factor depletion also differentially affects subnuclear RNA distribution and that the response of at least some genes is distinct from that of bulk PolyA$^+$ RNA.

NXF1 Sensitivity Is Distinct from Sensitivity to Additional Export Factors

To test whether NXF1-sensitive transcripts are affected by inhibition of additional export pathways, we used RNAi and fractionations followed by qRT-PCR for the three model NXF1-sensitive transcripts. We examined the sensitivity of export and expression to depletion of additional factors, including THO complex member *THOC2* (Strässer et al., 2002), TREX-2 member *MCM3AP* (GANP) (Wickramasinghe et al., 2010), nuclear pore component *NUP153* (Ullman et al., 1999), and *RBM15*, which has been implicated in NXF1-dependent export (Lindtner et al., 2006). Among all tested factors, NXF1 depletion was unique in causing a strong nuclear enrichment of *NORAD*, *ATXN7L3B*, and *MEX3C*, which was accompanied by a corresponding reduction in their expression (Figures 3G, S3D, and S3E). We also tested combinations of knockdowns of NXF1 together with several other factors, and found no substantial synergistic effects on either localization or expression (Figures S3D and S3F).

Sequence Composition and Short Sequence Motifs Are Additional Determinants of Transcript Sensitivity to Export Factor Depletion

We next examined the relationship between overall sequence composition and sensitivity of transcripts to depletion of each of the export factors. Strikingly, we observed a strong correlation between G/C content and nuclear enrichment upon TREX depletion (Figure 4A, bottom; Table S1). When considering G and C content separately, the correlation was particularly strong for C content (Spearman's R = –0.55 between change in localization and C content, compared with R = –0.39 for G content; Table S1). This correlation was much stronger for genes with many exons but was significant also for single-exon genes (Figure 4A, bottom). In contrast, NXF1 dependence was correlated with A/U content, but only for multi-exon genes (Figure 4A, top). Among these transcripts, those that are more A/U rich became more nuclear upon NXF1 depletion regardless of average exon length (Spearman's R = 0.34 for multi-exon genes with exons < 1 kb on average versus R = 0.37 with exons > 1 kb). This suggests three groups of transcripts that are particularly dependent on NXF1 for their export: single-exon transcripts, transcripts with

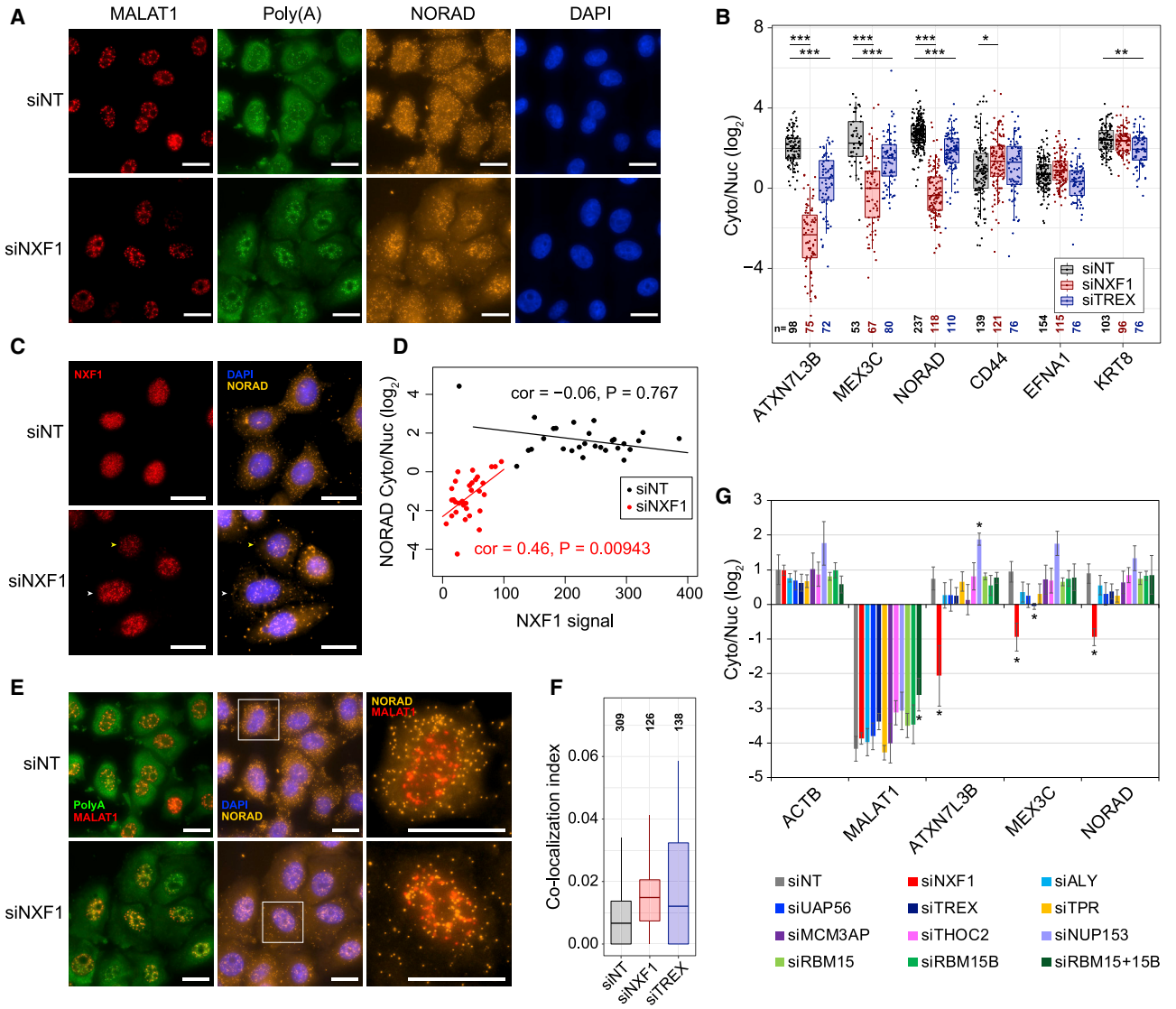


Figure 3. NXF1 Depletion Uniquely Prevents Cytoplasmic Accumulation of Intronless Genes

(A) Representative smFISH images of PolyA$^+$ RNA (green) and the lncRNA *NORAD* (orange) in control and NXF1-depleted cells. DAPI staining (blue) and *MALAT1* (red) were used as markers for the nucleus and nuclear speckles, respectively.

(B) Quantification of Cyto/Nuc ratios of indicated genes as measured by smFISH (see STAR Methods). Each dot is a cell. *p < 0.05, **p < 0.005, and ***p < 0.0005 (Wilcoxon rank-sum test).

(C) Left: staining for NXF1 protein in control and NXF1-depleted cells. Right: *NORAD* smFISH signal and DAPI staining in the same cells. Yellow and white arrows point at cells with strong and weak reduction in NXF1 levels, respectively, and corresponding effect on *NORAD* localization.

(D) Quantification of NXF1 staining intensity and *NORAD* subcellular localization in control (black) and NXF1-depleted (red) cells. Each dot is a cell. Linear regression lines corresponding to control and NXF1 depletion data and corresponding Spearman's R and p values are shown.

(E) Same images as in (A), merged by PolyA$^+$ and *MALAT1* signals (left) or by DAPI and *NORAD* signals (middle). Right: enlarged image of one cell with merged *MALAT1* and *NORAD* signals. White square in the middle image shows the enlarged area in the right panels.

(F) Quantification of *NORAD* and *MALAT1* colocalization (see STAR Methods).

(G) Effect of export factor depletions on localization of the NXF1 targets *ATXN7L3B*, *MEX3C*, and *NORAD* as measured using qRT-PCR. Localization of *ACTB* and *MALAT1* was used to estimate fractionation efficiencies. Knockdown efficiencies are shown in Figure S3D. n ≥ 3. Error bars represent SEM. *p < 0.05 (t test, compared with siNT).

Scale bars: 25 μM. See Figure S3.

long exons, and A/U-rich multi-exon transcripts. We note that a potential common denominator of these groups is low density of EJCs, which are deposited at structured, purine-rich sequences upstream of spliced introns (Saulière et al., 2012). Genes that became more nuclear upon TPR depletion had low G/C content, resembling NXF1 depletion (Figure S4A, left), consistent with the general correlation between TPR and NXF1 sensitivity (Figures 1C and 1D). Differential expression upon knockdown of NXF1

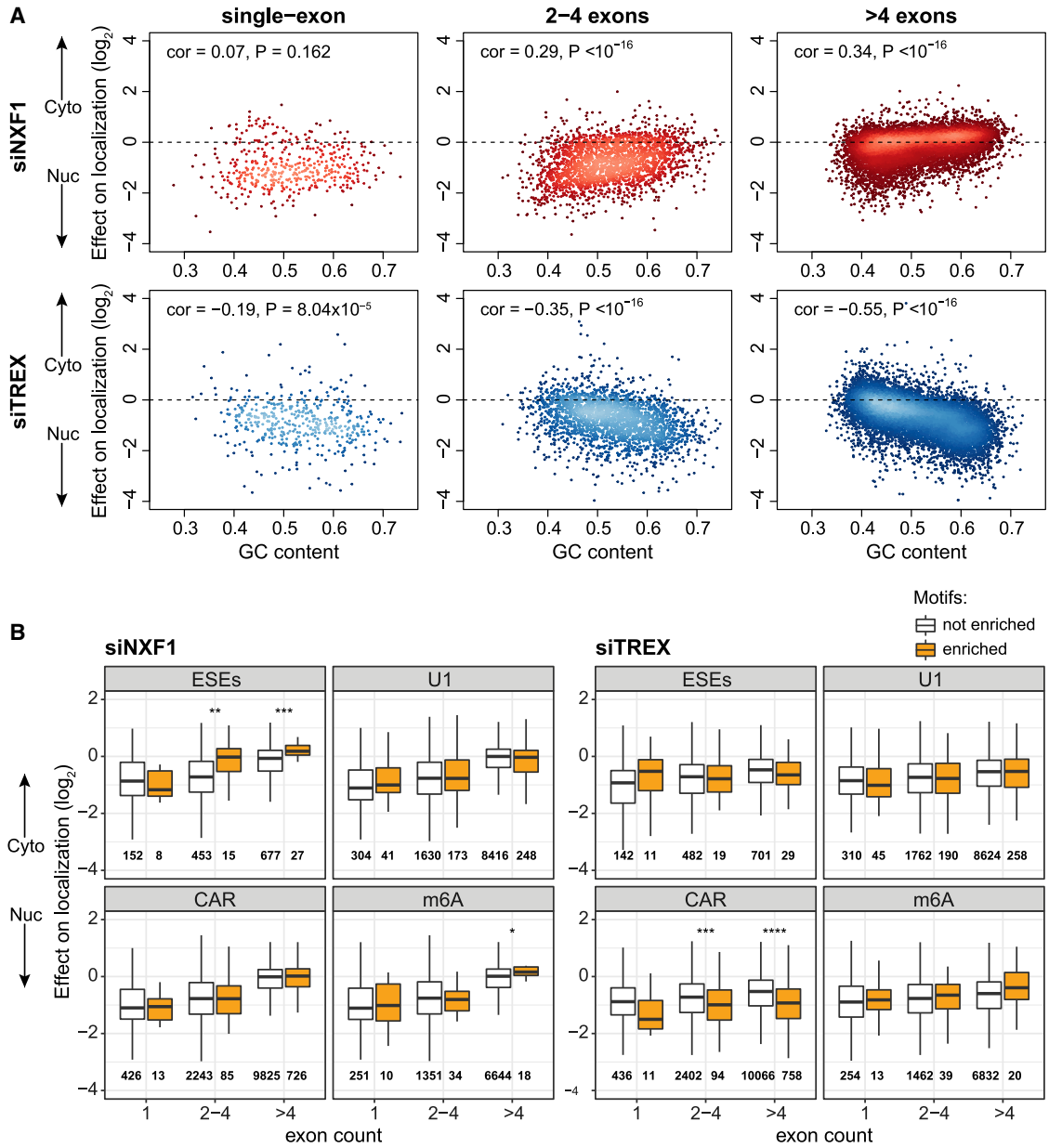


Figure 4. G/C Content and Short Sequence Elements Are Associated with Export Factor Sensitivity

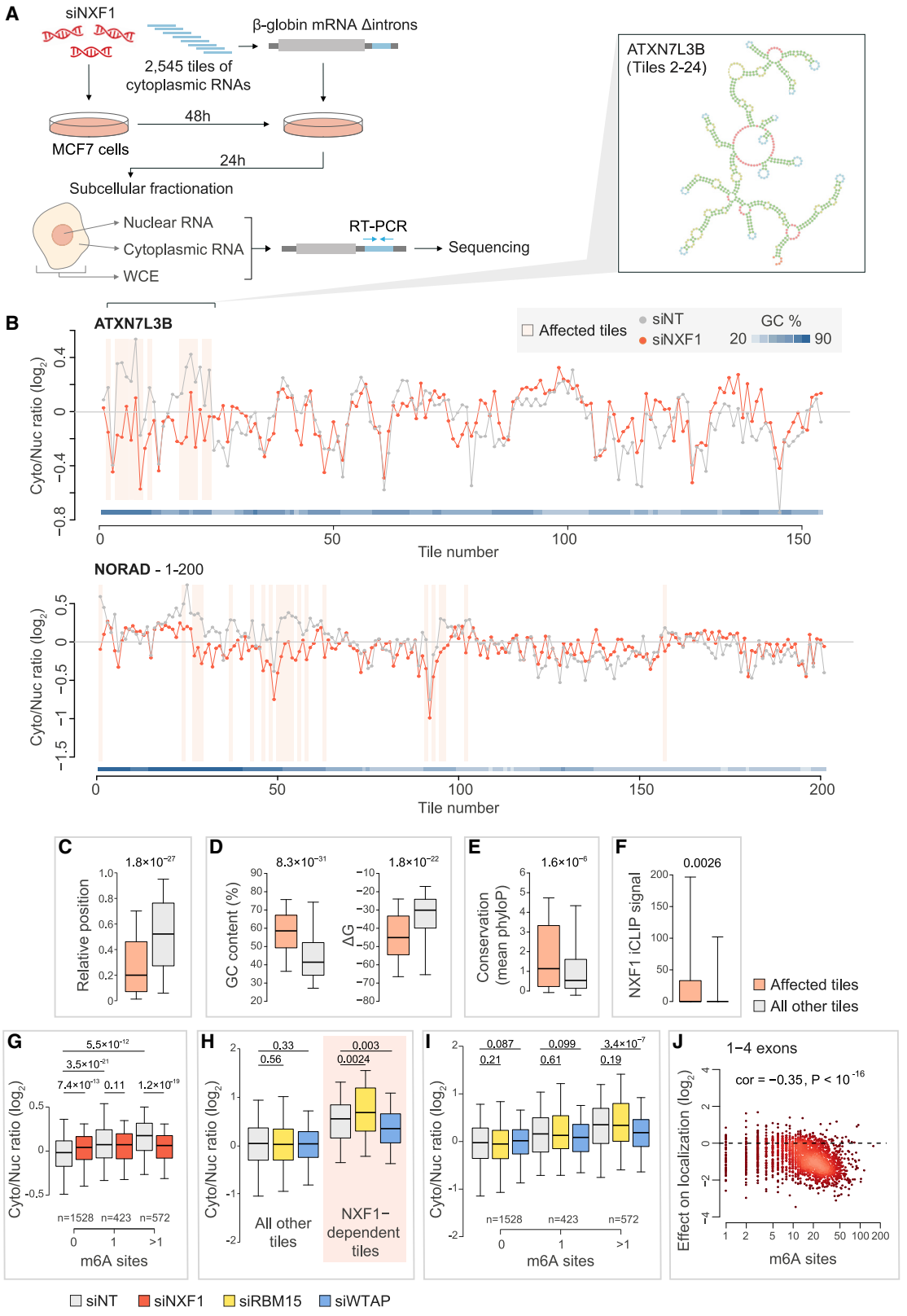
(A) Correlations between G/C content and effects of NXF1 (top) and TREX (bottom) depletion on localization. Correlations are shown separately for single-exon (left), two- to four-exon (middle), and multi-exon (more than four; right) genes. Color intensity indicates local point density. Spearman's R and p values are indicated.

(B) Effect of export factor depletions on localization of transcripts enriched (>2 enrichment value, orange; see STAR Methods) with indicated sequence elements, grouped by exon count. As control, we examined the effect on localization of all genes with no enrichment (≤1 enrichment value, white). Numbers indicate the amount of genes in each group.

See Figure S4.

and U2AF2 was mostly unrelated to G/C content, while effects of TREX and TPR depletions were associated with high G/C content regardless of exon count (Figure S4A, right).

Sequence composition can affect nuclear retention and export through short sequence elements that might recruit RBPs and actively promote or restrict nuclear export (Palazzo and Lee, 2018). We therefore calculated the enrichment of several such elements (Table S1; see STAR Methods) and examined correlations with export factor depletion sensitivity. Multi-exon transcripts enriched with exonic splicing enhancers (ESEs) exhibited resistance to NXF1 depletion, whereas sensitivity of single-exon transcripts was not related to ESE enrichment (Figure 4B, left).



(legend on next page)

This suggests that the generally protective effect of the increasing number of introns toward NXF1 depletion is mediated at least in part through recruitment of splicing-related factors. Other motifs, which were previously reported to have strong effects on RNA localization, such as U1 binding site (Azam et al., 2019; Lee et al., 2015; Yin et al., 2020; Roundtree et al., 2017) had minor and marginally significant effects (Figure 4B, left). Enrichment of cytoplasmic accumulation region (CAR), another sequence element reported to be important for nuclear export of specific intronless genes (Lei et al., 2011, 2013), had no effect on NXF1 sensitivity but was, interestingly, associated with enhanced TREX sensitivity and reduced TPR sensitivity regardless of exon number (Figures 4B and S4B). This suggests that this motif is likely indeed relevant to nuclear export, but not necessarily specifically to export of single-exon transcripts.

Short sequence elements thus modulate the sensitivity of transcripts to depletion of different export factors in a manner that is dependent on gene architecture and in a manner that is typically different from the models based on the study of individual genes.

Structured and Conserved Regions within Genes with Few Exons Drive NXF1-Dependent Nuclear Export

In order to identify sequences that may promote NXF1-dependent export of intronless RNAs, we used a massively parallel RNA assay (Lubelsky and Ulitsky, 2018; Shukla et al., 2018) (Figures 5A; Figure S5A). We designed short oligos tiled across the sequences of *NORAD*, *ATXN7L3B*, *MEX3C*, and eight additional single-exon, cytoplasmic, and NXF1-sensitive human genes, including six PCGs and two lncRNAs (Figure S5B). As a control, we also included the *JPX* lncRNA and a fragment of the *MLXIPL* gene, which we studied previously (Lubelsky and Ulitsky, 2018). Most of the transcripts were tiled with 140 nt sequences with offsets of 20 nt (10 nt for NORAD and 25 nt for JPX). Overall, 2,545 sequences (collectively called CytoLib) were cloned into the 3′ UTR of an intronless variant of the β-globin gene (βΔ1,2), which is relatively inefficiently exported and was previously used as a model sequence for study of elements affecting nuclear export (Akef et al., 2015; Brown and Steitz, 2016). We transfected MCF7 cells, first with siRNAs targeting NXF1 or non-targeting controls (siNT) and then with the CytoLib plasmids, fractionated them, and prepared sequencing libraries from CytoLib fragments using input plasmids, WCE, and Cyto/Nuc fractions. We quantified the effects of 2,473 sequences on subcellular localization (Cyto/Nuc ratios) and on the expression levels of the reporter (WCE/plasmid ratios) (other sequences were not efficiently cloned or particularly poorly expressed). When we combined five biological replicates, 257 sequences originating from all 13 genes exhibited significant sensitivity to depletion of NXF1 ($p < 0.05$ and $|\Delta\log_2[\text{Cyto/Nuc}]| > 0.3$), with 181 tiles associated with nuclear enrichment in siNXF1-treated cells compared with siNT and 76 tiles associated with cytoplasmic enrichment (Figure 5B; Table S2).

Inspection of the positions of the 181 tiles driving NXF1-dependent export (Figure 5B; Figure S5C) showed that these sequences tended to be clustered within their host genes and often found near the 5′ of the transcript (Figure 5C). This clustering is expected in part because consecutive tiles share most of their sequences, though in several cases tiles close to each other but without any sequence overlap had similar effects (Figure 5B).

Tiles driving NXF1-dependent export were significantly more G/C rich and were predicted to form more stable secondary structures than other tiles (Figures 5B, 5D, and S5D). They were also more conserved in vertebrate evolution than other tiles, supporting their functional importance (Figure 5E). We hypothesized that these sequences mediated NXF1-dependent export because they could recruit NXF1, and indeed, the NXF1 individual-nucleotide CLIP (iCLIP) read density in these tiles was significantly higher than in the other tiles in CytoLib (Figure 5F; data from Fan et al., 2019). We conclude that efficiently exported long RNAs with one or few exons harbor focal regions that can drive NXF1-dependent export. There is no obvious sequence homology between these regions in different genes, but they tended to be highly structured, features that appear to closely mirror these observed in viral sequences known to drive NXF1-dependent nuclear export (see Discussion).

Figure 5. A Screen for Sequence Elements Enabling NXF1-Dependent Nuclear Export Reveals Roles for RBM15 and RNA Methylation in Nuclear Export

(A) Experimental design of the CytoLib massively parallel assay.

(B) Cyto/Nuc ratios for individual tiles in the CytoLib library for the indicated regions and condition. Each point shows the median of five replicates. Regions of tiles with $p < 0.05$ and $\Delta\log_2(\text{Cyto/Nuc}) < -0.3$ between siNXF1 and siNT conditions are shaded. The squares at the bottom represent G/C content within the tile. Inset shows structure of the indicated region predicted by RNAfold server (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) using default parameters.

(C) Comparison of the relative position of the tile within the gene between the tiles sensitive to loss of NXF1 ($p < 0.05$ and $\Delta\log_2[\text{Cyto/Nuc}] < -0.3$) and all other tiles.

(D) As in (C), comparing G/C content (left) and ΔG predicted by RNAfold within the tiles.

(E) As in (C), comparing average PhyloP scores (Pollard et al., 2010) across the tiles (taken from the University of California, Santa Cruz [UCSC], Genome Browser).

(F) As in (C), comparing the total number of iCLIP reads in NXF1 iCLIP data from (Fan et al., 2019).

(G) Cyto/Nuc ratios for tiles overlapping the indicated number of m^6A sites in their source RNAs (from Liu et al., 2018). p values computed using Wilcoxon rank-sum test.

(H) Cyto/Nuc ratios for tiles affected by NXF1 knockdown (from C) and all other tiles following the indicated treatment. p values computed using Wilcoxon rank-sum test.

(I) Cyto/Nuc ratios for tiles with the indicated number of m^6A sites (as in H) and all other tiles following the indicated treatment. p values computed using Wilcoxon rank-sum test.

(J) Correlation between number of m^6A sites and effect of NXF1 depletion on endogenous genes with one to four exons. Color intensity indicates local point density. Spearman's R and p values are indicated.

See Figure S5.

Sequences that Promote NXF1-Dependent Export Depend on WTAP for Their Activity

In order to identify proteins whose binding to the export-promoting sequences might be consequential, we intersected the 181 tiles associated with nuclear enrichment in siNXF1-treated cells with enhanced CLIP (eCLIP) data from the ENCODE project (see STAR Methods). This analysis highlighted RBM15, which has been previously implicated in NXF1-dependent export (Lindtner et al., 2006; Meyer and Jaffrey, 2017) (Table S2; STAR Methods). RBM15 has also been implicated in m^6A modifications (Knuckles et al., 2018; Lence et al., 2016; Patil et al., 2016), and indeed, we found that tiles that overlap experimentally determined m^6A sites significantly increased export in an NXF1-dependent manner (Figure 5G; 181 NXF1-affected tiles had on average 1.55 bases reported as m^6A modified compared with 0.69 in other tiles, $p = 1.1 \times 10^{-19}$).

In order to test the potential importance of RBM15 binding and m^6A in nuclear export, we knocked down RBM15 alongside its paralog RBM15B, and WTAP, a core member of the m^6A writer complex (Schwartz et al., 2014), and examined localization of CytoLib tiles (Figure S5E; Table S2). RBM15+RBM15B knockdown led to increased export of NXF1-dependent tiles, whereas WTAP knockdown decreased it (Figure 5H). WTAP depletion caused nuclear enrichment of tiles that contained sequences with m^6A modifications and those that overlap RBM15 eCLIP peaks (Figures 5I and S5F). These results suggest that m^6A plays an important role in the context of specific structured regions within long RNAs that are dependent on NXF1 for their export. Indeed, when considering endogenous transcripts, particularly those with few exons, the number of experimentally determined m^6A sites was significantly correlated with increased nuclear enrichment upon NXF1 depletion (Spearman's $R = -0.35$ and -0.11 for genes with one to four exons and genes with more than four exons, respectively; Figure 5J).

Interestingly, sequences that supported NXF1-dependent export were also associated with significant and quite uniform reduction in β-globin reporter levels in siNXF1-treated cells (Figure S5G). This reduction, which happens in a setting in which all the sequences share the same promoter and thus there is no transcriptional compensation, supports the notion that inability to leave the nucleus through the NXF1 pathway results in nuclear degradation.

Splicing Efficiency Influence on Export Is Largely NXF1 Independent

As efficient splicing has been implicated as a key regulator of nuclear export, we next systematically examined the relationship between splicing efficiency and NXF1 sensitivity. We estimated gene-level splicing efficiency (Zuckerman and Ulitsky, 2019) from RNA-seq data from control and factor-depleted cells. All factor depletions had very limited effect on splicing efficiency in the cytosolic fraction (Figure 6A; Table S1), and we found no evidence for systematic "leakage" of unspliced transcripts to the cytoplasm. In contrast, depletion of NXF1 and TPR led to an increase in splicing efficiency in the nuclear fraction (Figure 6A), possibly because transcripts spent more time in the nucleus and thus had an additional opportunity to undergo post-transcriptional splicing. We then examined the correlation between baseline splicing efficiency (in control cells) and effects of export factor depletion on localization. Transcripts with retained introns (splicing efficiency < 60%) exhibited significant association between inefficient splicing and nuclear retention upon depletion of NXF1, especially if they had few exons (Figure 6B), and with no correlation with the retained intron length (Figure S6A). In contrast, TREX depletion exhibited a uniform effect on transcripts with retained introns and a broader association between completed splicing and resistance to TREX depletion (Figure S6B). These results thus suggest that splicing efficiency has a complex and non-linear relationship with export and a differential crosstalk with NXF1- and TREX-dependent export pathways.

To examine the relationship between NXF1 sensitivity and splicing in a better controlled setting, we used a library of thousands of reporters carrying splicing events based on a set of 38 native introns with various systematic splicing-related sequence changes and hence variable splicing efficiencies stably integrated into K562 cells (Mikl et al., 2019). We perturbed NXF1 with RNAi in these cells and measured subcellular localization and splicing efficiency of the reporter mRNAs (Figures 6C, S6C, and S6D; Table S3). In this system, NXF1 depletion had a negligible effect on splicing efficiency in both cytoplasmic and nuclear fractions (Figure S6E). Consistently with the observations on endogenous genes, NXF1 knockdown did not have a substantial effect on localization of unspliced RNAs, and no major leakage of such RNAs into the cytoplasm was observed (Figure 6D, left). In contrast, NXF1 depletion substantially affected localization of spliced transcripts for some of the introns but not others (Figure 6D, right; Figure S6F). When considering all the different sequence variants together, NXF1 depletion led to an increased correlation between splicing efficiency and Cyto/Nuc ratios (Figure 6E), suggesting that efficient splicing of individual exons enhances nuclear export in a mostly NXF1-independent manner, with some transcripts becoming dependent on NXF1 when their splicing is particularly inefficient, possibly to overcome sequences in their introns that cause nuclear retention.

NXF1 Depletion Does Not Affect RNA Stability in the Cytoplasm

The observed effects of export factor depletion on steady-state subcellular localization might result from either nuclear export block or from changes in cytoplasmic decay rates. To test this, we used actinomycin D (ActD) to inhibit transcription and measured expression of the NXF1-sensitive and NXF1-insensitive transcripts in the cytosolic and nuclear fractions (Figures S7A–S7C). With the possible exception of *ATXNL7B*, which was slightly destabilized in the cytoplasm upon NXF1 loss, we observed no substantial effects on RNA stability in either the nucleus or the cytoplasm upon NXF1 depletion. To test the effects of NXF1 depletion on cytoplasmic stability more broadly, we combined Cyto/Nuc fractionations with SLAM-seq (Herzog et al., 2017) and focused our analysis on cytoplasmic fraction, which was free of any detectable nuclear contamination, and so the "new" and "old" RNA could be reliably quantified (Figures S7D-S7E). We found no substantial changes in half-lives upon NXF1 depletion (Figure S7F), as well as no correlation between

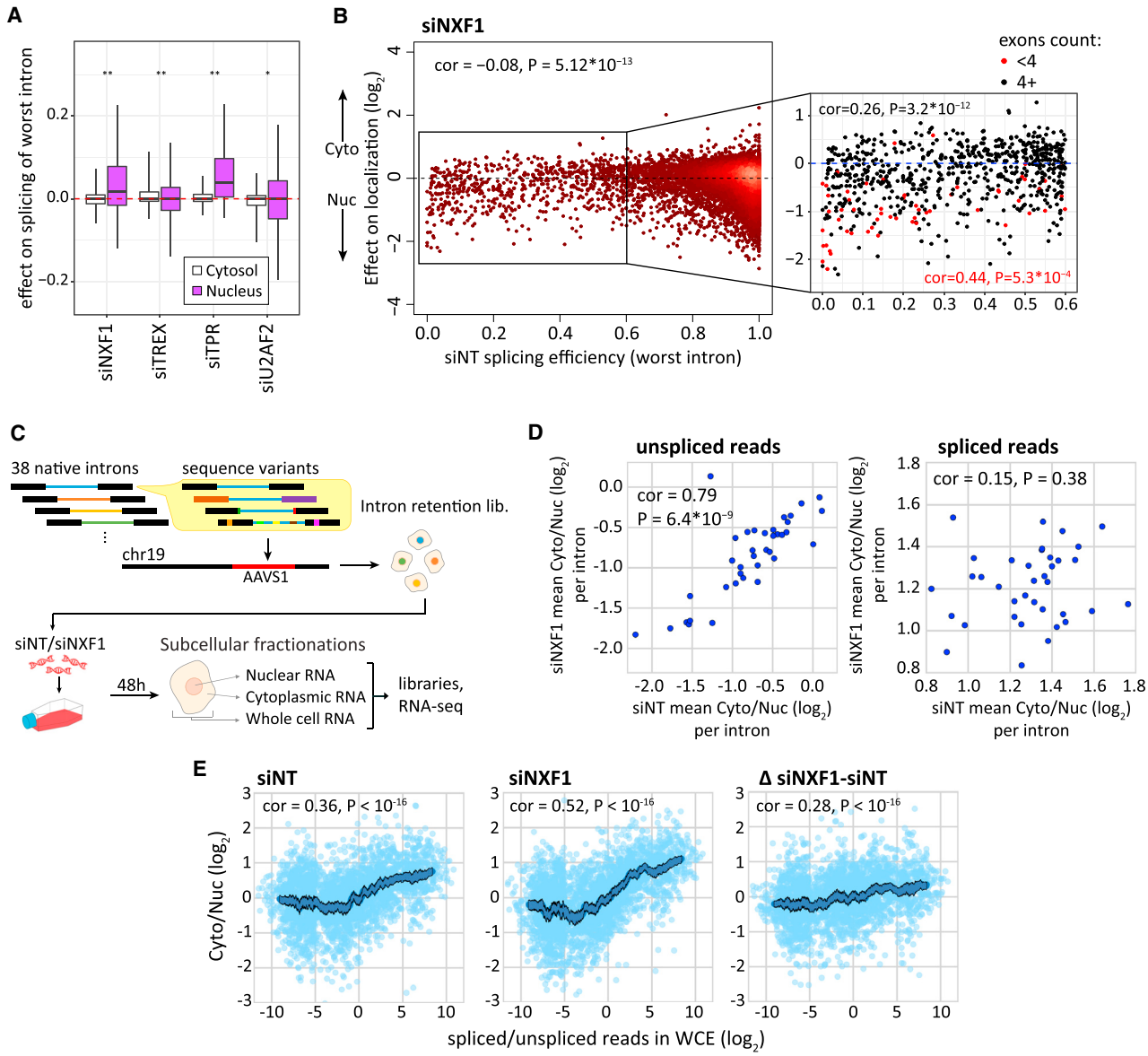


Figure 6. Splicing Effects on Subcellular Localization Are Largely NXF1 Independent

(A) Effect of NXF1 depletion on splicing efficiency of the most inefficient intron in each gene in cytosolic and nuclear fractions. Splicing efficiency effect is defined as the difference in splicing efficiency between control (siNT) and knockdown of the indicated factor. *p < 0.0005 and **p < 10^{-10} (two-sided Wilcoxon rank-sum test).

(B) Correlation between splicing efficiency in siNT WCE sample and effect of NXF1 depletion on localization. Color intensity indicates local point density. Spearman's R and p value are indicated. Subset: enlarged plot of inefficiently spliced genes (efficiency < 0.6), showing the correlation separately for genes with many (at least four, black) or few (fewer than four, red) exons.

(C) Experimental design of massively parallel reporter assay employing intron retention library to examine the effect of NXF1 depletion on thousands of sequence variants in a fixed genomic context (see STAR Methods).

(D) Average Cyto/Nuc ratios of all variants for each sequence context (dots) in control and in NXF1-depleted samples. Unspliced (left) and spliced (right) reads were quantified separately. Spearman's R and p values are indicated.

(E) Correlations between siNT WCE spliced/unspliced ratios and Cyto/Nuc values in control (left) and NXF1-depleted (middle) cells, and the effect on localization (siNXF1 - siNT, right). Dots, sequence variants; blue line, rolling average (mean of 100 data points) and 95% confidence interval. Spearman's R and p values are indicated.

See Figures S6 and S7.

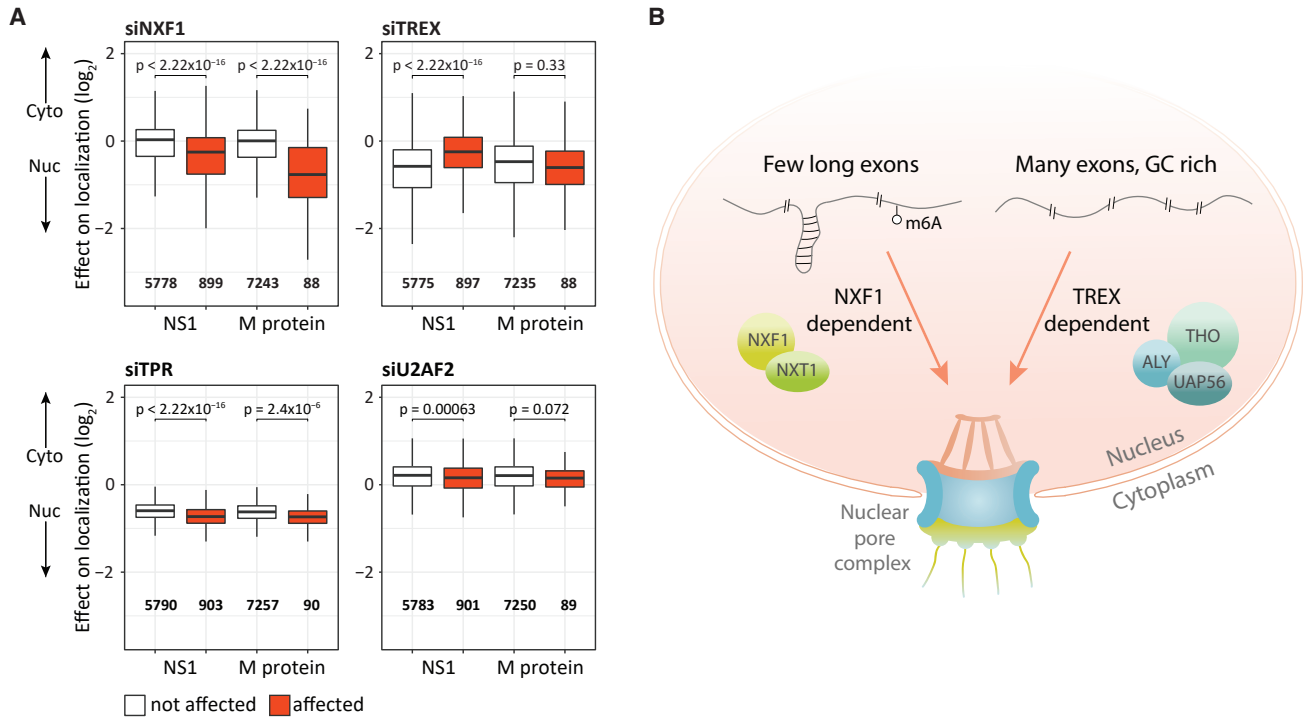


Figure 7. Viral Protein Overexpression Preferentially Affects Localization of NXF1-Sensitive Genes

(A) Effects of export factor depletions on genes affected or unaffected by overexpression of either NS1 or M protein. "Affected" genes are those with Cyto/Nuc ratios (from Zhang et al., 2012) reduced by at least 2-fold compared with controls, and "unaffected" are genes with Cyto/Nuc ratio reduced by less than 25%. (B) Features associated with NXF1- and TREX-dependent nuclear export.

changes in the estimated half-lives and exon counts (Figure S7G). These results suggest that NXF1 depletion predominantly affects nuclear export rather than RNA stability in the cytoplasm.

Viral Proteins that Block RNA Export Preferentially Affect NXF1-Sensitive Genes

In order to examine the physiological relevance of the differential sensitivity of different genes to NXF1-dependent export, we focused on export block induced by the influenza NS1 protein, which specifically acts on the NXF1 protein (Satterly et al., 2007; Zhang et al., 2019), and the vesicular stomatitis virus (VSV) M protein that blocks mRNA export through RAE1 that forms a complex with NXF1 (Blevins et al., 2003; Faria et al., 2005). Selective export block was previously reported for both viral mRNAs (Larsen et al., 2014; Read and Digard, 2010) and host mRNAs in infected cells (Zhang et al., 2012), and we wondered whether the differential sensitivity of different RNAs to viral export block was correlated with our observations of differential sensitivity to depletion of NXF1. Indeed, in data from HeLa cells (Zhang et al., 2012), transcripts whose export was inhibited by expression of the M protein (and to a lesser extent NS1) were selectively and significantly enriched in the nucleus upon NXF1 depletion, and, with a smaller effect size, upon TPR depletion (Figure 7A), despite the differences in cell lines and in the technology used to measure gene expression (RNA-seq versus microarray). Notably, the genes that were unaffected by NS1 expression and exhibited reduced NXF1 sensitivity were significantly more sensitive to TREX depletion (Figure 7A, top right). Interestingly, many of the immediate-early genes acting in response to viral infection, such as interferons, have few exons or a single exon. The selective sensitivity to NXF1 loss thus has consequences for the fates of RNAs during viral infection.

DISCUSSION

The canonical model of RNA export is based on a large number of studies from various systems, that mostly used viral or reporter RNAs and/or FISH with oligo-d(T) probes in human cells. These studies, alongside genome-wide studies in yeast and fly cells, have been instrumental for implicating many factors in the mRNA export pathway, but they had limited resolution toward the fates of human transcripts, that have on average seven introns and vary quite extensively in their exon architecture and sequence composition. Here, we linked two prominent characteristics that vary extensively among human genes, namely, exon architecture and G/C content, to preferential sensitivity to two different parts of the export pathway, TREX and NXF1, which have been proposed to tightly cooperate in the export of the typical mRNA (Figure 7B). Transcripts having one or few long exons, or many exons but high A/U content, preferentially require NXF1. At least the single-exon transcripts appear to recruit NXF1 through focal (100–200 nt), structured and relatively G/C-rich elements, with contribution from m^6A modifications, whereas their overall G/C content

appears to be less important. There are no obvious shared sequences or structural characteristics between these patches in different genes, resembling the situation among the viral elements driving NXF1-dependent export, that are also highly structured, and for which structure was shown to be important (Ernst et al., 1997; Lindtner et al., 2002; Pasquinelli et al., 1997; Sakuma et al., 2014; Smulevitch et al., 2005). We note that we do not presently know the m^6A stoichiometry in these NXF1-sensitive regions, and it can be quantified in future studies using genome-wide methods that recently became available (Garcia-Campos et al., 2019), but it is likely substantial as effects of WTAP depletion were comparable with those of NXF1 depletion.

Transcripts with high C content (and to a lesser extent G content) preferentially require ALY/UAP56. Interestingly, C-rich transcripts typically have lower export efficiencies in control cells (Lubelsky and Ulitsky, 2018), and indeed, TREX depletion preferentially affects transcripts relatively enriched in the nucleus at baseline (Figure 1B). Interestingly, we recently implicated HNRNPK, an abundant nuclear protein in inhibition of nuclear export, in part through binding to C-rich fragments in internal exons and in Alu repeats (Lubelsky and Ulitsky, 2018). This suggests a possible model in which HNRNPK might compete with TREX components for shared binding sites. Indeed, we observe a strong anti-correlation between the effects of HNRNPK and TREX depletion on localization in MCF7 cells (Spearman's $R = -0.39$, $p < 10^{-15}$ for multi-exon transcripts and much more modest $R = -0.14$, $p = 0.007$ for single-exon ones).

One limitation of our approach (as well as most previous studies of the export pathway) is the use of RNAi to target individual export factors or their combinations, which implies that the studied cells were not completely null for the targeted protein. We note that, on the basis of CRISPR screens, most of the genes we studied appear universally essential in human cells: NXF1, TPR, and U2AF2 are essential in the vast majority of 625 human cell lines profiled in CRISPR screens (623 for NXF1, 622 for TPR, and 624 for U2AF2; data from Cancer Dependency Map [https://www.broadinstitute.org/cancer/cancer-dependency-map]), and ALY is individually essential in >85% of the cell lines (542 lines). UAP56 is essential in ~45% of the cell lines, possibly because of its partial redundancy with URH49. This implies that it is not feasible to obtain and compare knockout cells for these factors, and our siRNA-treated cells indeed still contain residual amounts of each of the proteins (Figure S1A). Importantly, in the conditions we used, we did not observe any significant induction of a transcriptional stress response following the knockdowns. Other potential limitations of our approach are that only one or two factors were targeted in each experiment, that some factors may have additional functions beyond nuclear export, and that some of the effects on export might be indirect.

The differential sensitivity to depletion of NXF1 and TREX can have variable underlying reasons. It is possible that TREX and NXF1 cooperate in export of most polyadenylated transcripts, yet different transcripts interact with them at different stoichiometries. For example, it is possible that transcripts with many ESEs and/or excised introns are decorated with a large number of EJCs or SR proteins, which may allow them to compete favorably with transcripts with few or no exons for NXF1 binding when NXF1 amounts are limiting. It is also possible, however, that transcripts with a sufficient number of exons or extensive SR protein binding do not require NXF1 for export at all, as has been demonstrated for other mRNA subsets (Culjkovic et al., 2006; Wickramasinghe and Laskey, 2015). The co-existence of distinct pathways for export of large numbers of transcripts may facilitate further sorting of the long RNAs once they leave the nucleus. For example, using HIV-1 genome and different viral sequences, it was shown that different viral export sequences (working through NXF1 or the CRM1 pathways) can lead to a different cytoplasmic distribution of the exported cargo, with an NXF1-dependent constitutive transport element (CTE) driving association of RNAs with the cytoskeleton in the cytoplasm (Pocock et al., 2016).

Our findings relate to those of several recent studies of nuclear export factors. A recent study reported that RNAi-mediated depletion of NXF1 in HeLa cells leads to preferential expression of transcripts with shorter last exons (Chen et al., 2019). The authors showed that changes in Pol2 elongation can account for some of these changes. Our results further suggest that the preferential requirement of NXF1 for export of transcripts with long exons could also underlie some of these differences. Supporting this idea, NXF1 depletion does not result in preferential use of intronic polyadenylation sites, which typically produce transcripts with short 3′ UTRs, but rather leads to avoidance of distal poly(A) sites in the long terminal exons (Chen et al., 2019). Our results are consistent with those described in a recent preprint that describes RNAi-mediated depletion of TPR in U2OS cells (Lee et al., 2019), in which it was found that loss of TPR does not increase export of poorly spliced RNAs, and leads to a global reduction in export of $PolyA^+$ RNAs, with preferential effect on transcripts with relatively few introns, including *NORAD*.

mRNAs and lncRNAs differ in their primary cellular functions and show wide, overlapping, and significantly different distributions of various characteristics, such as expression levels, splicing efficiency, and subcellular localization (Derrien et al., 2012). Here we report that they also differ in the export pathways that they are tunneled through, with lncRNAs showing preferential sensitivity to loss of NXF1, which can be explained by lncRNAs typically having substantially fewer exons (two or three) compared to mRNAs (seven or eight on average). This difference may expose lncRNAs and mRNAs to differential regulation upon conditions that specifically inhibit NXF1-dependent export, such as viral infection. Furthermore, as NXF1-sensitive and predominantly cytoplasmic transcripts typically harbor quite long (>100 nt) structured regions that drive efficient export, we suggest that evolutionary young lncRNAs, which are typically derived from previously non-transcribed RNA (Ulitsky, 2016), and typically contain few exons, are often inefficiently exported. The introns in such transcripts will also typically contain suboptimal splice sites, and inefficient splicing also correlates with inefficient export, although to a lesser extent in lncRNAs than in mRNAs (Zuckerman and Ulitsky, 2019). Our findings thus help explain the overall differences in RNA distribution among lncRNAs and mRNAs (Derrien et al., 2012).

The ability to measure with high precision, and in a single experiment, the sensitivity of all expressed genes to loss of individual factors, and to measure the effects of thousands of rationally designed RNA elements on export, provides an opportunity

to test and revise the textbook models of how the post-transcriptional fate of long RNAs is determined. The development of proximity-based approaches for RNA labeling (Fazal et al., 2019; Padrón et al., 2019) is expected to yield many further developments beyond the nuclear and cytoplasmic divide. The emerging picture is that diverse aspects of RNA production and the RNA product, namely, transcription, processing, sequence elements, and secondary and tertiary structures, all come together to dictate delivery of long RNAs at the right time to the right place inside the cell and outside of it.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
- METHOD DETAILS
  - RNAi treatments
  - Extraction of cytoplasmic and nuclear RNA
  - RNA-seq and data analysis
  - Gene architecture analysis
  - Transcription inhibition by actinomycin D
  - Metabolic labeling and SLAM-seq
  - Single-molecule FISH and immunofluorescence
  - Colocalization analysis
  - Fractionation and library construction for massively parallel splicing reporter assay
  - Massively parallel splicing reporter assay data analysis
  - CytoLib library plasmid construction
  - CytoLib sequencing library generation
  - CytoLib data analysis
  - qRT-PCR
  - Western blot
  - Motif enrichment analysis
  - Additional datasets
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.molcel.2020.05.013.

ACKNOWLEDGMENTS

We thank Schragi Schwartz, Shalev Itzkovitz, Yaron Shav-Tal, and members of the Ulitsky lab for comments on the manuscript and helpful discussions. We thank Shalev Itzkovitz for the kind gift of smFISH probes and help with the smFISH protocol. We thank Joan Steitz for the kind gift of the beta-globin Δintrons plasmid. We also thank Raya Eilam for the help with the smFISH protocol and imaging and Ofra Golani for help with image analysis. We thank Florian Eddhard for the support in performing GRAND-SLAM analysis and Florian Mueller for the support in using FishQuant. This research was supported by grants to I.U. from the Israeli Science Foundation (ISF) (grant 852/19), the ISF-Natural Science Foundation of China (NSFC) joint research program (grant 2406/18), the Germany-Israeli Foundation for Scientific Research and Development (grant I-144-417.5-2017), the Israeli Ministry of Health as part of the ERA-NET localMND, and the Azrieli Institute for Systems Biology. I.U. is incumbent of the Sygnet Career Development Chair for Bioinformatics.

AUTHOR CONTRIBUTIONS

B.Z. and I.U. conceived the study. B.Z. conducted and designed experiments and analyzed data. M.R. performed CytoLib experiments and analyzed data. M.M. and E.S. generated the intron retention library and analyzed data. B.Z. and I.U. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 4, 2019
Revised: March 23, 2020
Accepted: May 11, 2020
Published: June 5, 2020

REFERENCES

Akef, A., Lee, E.S., and Palazzo, A.F. (2015). Splicing promotes the nuclear export of β-globin mRNA by overcoming nuclear retention elements. RNA *21*, 1908–1920.

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.

Azam, S., Hou, S., Zhu, B., Wang, W., Hao, T., Bu, X., Khan, M., and Lei, H. (2019). Nuclear retention element recruits U1 snRNP components to restrain spliced lncRNAs in the nucleus. RNA Biol. *16*, 1001–1009.

Bahar Halpern, K., and Itzkovitz, S. (2016). Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. Methods *98*, 134–142.

Bahar Halpern, K., Caspi, I., Lemze, D., Levy, M., Landen, S., Elinav, E., Ulitsky, I., and Itzkovitz, S. (2015). Nuclear retention of mRNA in mammalian tissues. Cell Rep. *13*, 2653–2662.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. Cell *163*, 1596–1610.

Blevins, M.B., Smith, A.M., Phillips, E.M., and Powers, M.A. (2003). Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. J. Biol. Chem. *278*, 20979–20988.

Brown, J.A., and Steitz, J.A. (2016). Intronless β-globin reporter: a tool for studying nuclear RNA stability elements. Methods Mol. Biol. *1428*, 77–92.

Carlevaro-Fita, J., and Johnson, R. (2019). Global positioning system: understanding long noncoding RNAs through subcellular localization. Mol. Cell *73*, 869–883.

Carlevaro-Fita, J., Polidori, T., Das, M., Navarro, C., Zoller, T.I., and Johnson, R. (2019). Ancient exapted transposable elements promote nuclear enrichment of human long noncoding RNAs. Genome Res. *29*, 208–222.

Carmody, S.R., and Wente, S.R. (2009). mRNA nuclear export at a glance. J. Cell Sci. *122*, 1933–1937.

Chen, S., Wang, R., Zheng, D., Zhang, H., Chang, X., Wang, K., Li, W., Fan, J., Tian, B., and Cheng, H. (2019). The mRNA export receptor NXF1 coordinates transcriptional dynamics, alternative polyadenylation, and mRNA export. Mol. Cell *74*, 118–131.e7.

Cheng, H., Dufu, K., Lee, C.-S., Hsu, J.L., Dias, A., and Reed, R. (2006). Human mRNA export machinery recruited to the 5′ end of mRNA. Cell *127*, 1389–1400.

Chi, B., Wang, Q., Wu, G., Tan, M., Wang, L., Shi, M., Chang, X., and Cheng, H. (2013). Aly and THO are required for assembly of the human TREX complex and association of TREX components with the spliced mRNA. Nucleic Acids Res. *41*, 1294–1306.

Coyle, J.H., Bor, Y.-C., Rekosh, D., and Hammarskjold, M.-L. (2011). The Tpr protein regulates export of mRNAs with retained introns that traffic through the Nxf1 pathway. RNA *17*, 1344–1356.

Culjkovic, B., Topisirovic, I., Skrabanek, L., Ruiz-Gutierrez, M., and Borden, K.L.B. (2006). eIF4E is a central node of an RNA regulon that governs cellular proliferation. J. Cell Biol. *175*, 415–426.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Dias, A.P., Dufu, K., Lei, H., and Reed, R. (2010). A role for TREX components in the release of spliced mRNA from nuclear speckle domains. Nat. Commun. *1*, 97.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m^6A RNA methylomes revealed by m^6A-seq. Nature *485*, 201–206.

Elbarbary, R.A., and Maquat, L.E. (2016). Coupling pre-mRNA splicing and 3′ end formation to mRNA export: alternative ways to punch the nuclear export clock. Genes Dev. *30*, 487–488.

Ernst, R.K., Bray, M., Rekosh, D., and Hammarskjöld, M.L. (1997). A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. Mol. Cell. Biol. *17*, 135–144.

Fan, J., Kuai, B., Wang, K., Wang, L., Wang, Y., Wu, X., Chi, B., Li, G., and Cheng, H. (2018). mRNAs are sorted for export or degradation before passing through nuclear speckles. Nucleic Acids Res. *46*, 8404–8416.

Fan, J., Wang, K., Du, X., Wang, J., Chen, S., Wang, Y., Shi, M., Zhang, L., Wu, X., Zheng, D., et al. (2019). ALYREF links 3′-end processing to nuclear export of non-polyadenylated mRNAs. EMBO J. *38*, e99910.

Faria, P.A., Chakraborty, P., Levay, A., Barber, G.N., Ezelle, H.J., Enninga, J., Arana, C., van Deursen, J., and Fontoura, B.M.A. (2005). VSV disrupts the Rae1/mrnp41 mRNA nuclear export pathway. Mol. Cell *17*, 93–102.

Fazal, F.M., Han, S., Parker, K.R., Kaewsapsak, P., Xu, J., Boettiger, A.N., Chang, H.Y., and Ting, A.Y. (2019). Atlas of subcellular RNA localization revealed by APEX-seq. Cell *178*, 473–490.e26.

Garcia-Campos, M.A., Edelheit, S., Toth, U., Safra, M., Shachar, R., Viukov, S., Winkler, R., Nir, R., Lasman, L., Brandis, A., et al. (2019). Deciphering the "M^6 A Code" via Antibody-Independent Quantitative Profiling. Cell *178*, 731–747.

Gatfield, D., and Izaurralde, E. (2002). REF1/Aly and the additional exon junction complex proteins are dispensable for nuclear mRNA export. J. Cell Biol. *159*, 579–588.

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Köcher, T., Wilm, M., and Izaurralde, E. (2001). The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in Drosophila. Curr. Biol. *11*, 1716–1721.

Guang, S., Felthauser, A.M., and Mertz, J.E. (2005). Binding of hnRNP L to the pre-mRNA processing enhancer of the herpes simplex virus thymidine kinase gene enhances both polyadenylation and nucleocytoplasmic export of intronless mRNAs. Mol. Cell. Biol. *25*, 6303–6313.

Hansen, M.M.K., Desai, R.V., Simpson, M.L., and Weinberger, L.S. (2018). Cytoplasmic amplification of transcriptional noise generates substantial cell-to-cell variability. Cell Syst. *7*, 384–397.e6.

Hautbergue, G.M., Hung, M.-L., Golovanov, A.P., Lian, L.-Y., and Wilson, S.A. (2008). Mutually exclusive interactions drive handover of mRNA from export adaptors to TAP. Proc. Natl. Acad. Sci. U S A *105*, 5154–5159.

Herold, A., Teixeira, L., and Izaurralde, E. (2003). Genome-wide analysis of nuclear mRNA export pathways in Drosophila. EMBO J. *22*, 2472–2483.

Herzog, V.A., Reichholf, B., Neumann, T., Rescheneder, P., Bhat, P., Burkard, T.R., Wlotzka, W., von Haeseler, A., Zuber, J., and Ameres, S.L. (2017). Thiol-linked alkylation of RNA to assess expression dynamics. Nat. Methods *14*, 1198–1204.

Hieronymus, H., and Silver, P.A. (2003). Genome-wide analysis of RNA-protein interactions illustrates specificity of the mRNA export machinery. Nat. Genet. *33*, 155–161.

Hocine, S., Singer, R.H., and Grünwald, D. (2010). RNA processing and export. Cold Spring Harb. Perspect. Biol. *2*, a000752.

Huang, Y., and Carmichael, G.G. (1997). The mouse histone H2a gene contains a small element that facilitates cytoplasmic accumulation of intronless gene transcripts and of unspliced HIV-1-related mRNAs. Proc. Natl. Acad. Sci. U S A *94*, 10104–10109.

Huang, Y., Wimler, K.M., and Carmichael, G.G. (1999). Intronless mRNA transport elements may affect multiple steps of pre-mRNA processing. EMBO J. *18*, 1642–1652.

Huang, Y., Gattoni, R., Stévenin, J., and Steitz, J.A. (2003). SR splicing factors serve as adapter proteins for TAP-dependent mRNA export. Mol. Cell *11*, 837–843.

Huang, Y., Yario, T.A., and Steitz, J.A. (2004). A molecular link between SR protein dephosphorylation and mRNA export. Proc. Natl. Acad. Sci. U S A *101*, 9666–9670.

Hutchinson, J.N., Ensminger, A.W., Clemson, C.M., Lynch, C.R., Lawrence, J.B., and Chess, A. (2007). A screen for nuclear transcripts identifies two linked noncoding RNAs associated with SC35 splicing domains. BMC Genomics *8*, 39.

Jürges, C., Dölken, L., and Erhard, F. (2018). Dissecting newly transcribed and old RNA using GRAND-SLAM. Bioinformatics *34*, i218–i226.

Katahira, J., Strässer, K., Podtelejnikov, A., Mann, M., Jung, J.U., and Hurt, E. (1999). The Mex67p-mediated nuclear mRNA export pathway is conserved from yeast to human. EMBO J. *18*, 2593–2609.

Katahira, J., Inoue, H., Hurt, E., and Yoneda, Y. (2009). Adaptor Aly and co-adaptor Thoc5 function in the Tap-p15-mediated nuclear export of HSP70 mRNA. EMBO J. *28*, 556–567.

Knuckles, P., Lence, T., Haussmann, I.U., Jacob, D., Kreim, N., Carl, S.H., Masiello, I., Hares, T., Villaseñor, R., Hess, D., et al. (2018). Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m^6A machinery component Wtap/Fl(2)d. Genes Dev. *32*, 415–429.

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

Larsen, S., Bui, S., Perez, V., Mohammad, A., Medina-Ramirez, H., and Newcomb, L.L. (2014). Influenza polymerase encoding mRNAs utilize atypical mRNA nuclear export. Virol. J. *11*, 154.

Le Hir, H., Gatfield, D., Izaurralde, E., and Moore, M.J. (2001). The exon-exon junction complex provides a binding platform for factors involved in mRNA export and nonsense-mediated mRNA decay. EMBO J. *20*, 4987–4997.

Lee, E.S., Akef, A., Mahadevan, K., and Palazzo, A.F. (2015). The consensus 5′ splice site motif inhibits mRNA nuclear export. PLoS ONE *10*, e0122743.

Lee, E.S., Wolf, E.J., Smith, H.W., Emili, A., and Palazzo, A.F. (2019). TPR is required for the nuclear export of mRNAs and lncRNAs from intronless and intron-poor genes. bioRxiv. https://doi.org/10.1101/740498.

Lei, H., Dias, A.P., and Reed, R. (2011). Export and stability of naturally intronless mRNAs require specific coding region sequences and the TREX mRNA export complex. Proc. Natl. Acad. Sci. U S A *108*, 17985–17990.

Lei, H., Zhai, B., Yin, S., Gygi, S., and Reed, R. (2013). Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. Nucleic Acids Res. *41*, 2517–2525.

Lence, T., Akhtar, J., Bayer, M., Schmid, K., Spindler, L., Ho, C.H., Kreim, N., Andrade-Navarro, M.A., Poeck, B., Helm, M., and Roignant, J.Y. (2016). m^6A modulates neuronal functions and sex determination in Drosophila. Nature *540*, 242–247.

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

Lindtner, S., Felber, B.K., and Kjems, J. (2002). An element in the 3′ untranslated region of human LINE-1 retrotransposon mRNA binds NXF1(TAP) and can function as a nuclear export element. RNA *8*, 345–356.

Lindtner, S., Zolotukhin, A.S., Uranishi, H., Bear, J., Kulkarni, V., Smulevitch, S., Samiotaki, M., Panayotou, G., Felber, B.K., and Pavlakis, G.N. (2006). RNA-binding motif protein 15 binds to the RNA transport element RTE and provides a direct link to the NXF1 export pathway. J. Biol. Chem. *281*, 36915–36928.

Liu, H., Wang, H., Wei, Z., Zhang, S., Hua, G., Zhang, S.-W., Zhang, L., Gao, S.-J., Meng, J., Chen, X., and Huang, Y. (2018). MeT-DB V2.0: elucidating context-specific functions of N6-methyl-adenosine methyltranscriptome. Nucleic Acids Res. *46* (D1), D281–D287.

Love, M., Anders, S., and Huber, W. (2014). Differential analysis of count data–the DESeq2 package. Genome Biol. *15*, 550.

Lubelsky, Y., and Ulitsky, I. (2018). Sequences enriched in Alu repeats drive nuclear localization of long RNAs in human cells. Nature *555*, 107–111.

Lyubimova, A., Itzkovitz, S., Junker, J.P., Fan, Z.P., Wu, X., and van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. Nat. Protoc. *8*, 1743–1758.

MacMorris, M., Brocker, C., and Blumenthal, T. (2003). UAP56 levels affect viability and mRNA export in Caenorhabditis elegans. RNA *9*, 847–857.

Mauger, O., Lemoine, F., and Scheiffele, P. (2016). Targeted intron retention and excision for rapid gene regulation in response to neuronal activity. Neuron *92*, 1266–1278.

McCloskey, A., Ibarra, A., and Hetzer, M.W. (2018). Tpr regulates the total number of nuclear pore complexes per cell nucleus. Genes Dev. *32*, 1321–1331.

McQuin, C., Goodman, A., Chernyshev, V., Kamentsky, L., Cimini, B.A., Karhohs, K.W., Doan, M., Ding, L., Rafelski, S.M., Thirstrup, D., et al. (2018). CellProfiler 3.0: next-generation image processing for biology. PLoS Biol. *16*, e2005970.

Meyer, K.D., and Jaffrey, S.R. (2017). Rethinking m^6A readers, writers, and erasers. Annu. Rev. Cell Dev. Biol. *33*, 319–342.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. Cell *149*, 1635–1646.

Mikl, M., Hamburg, A., Pilpel, Y., and Segal, E. (2019). Dissecting splicing decisions and cell-to-cell variability with designed sequence libraries. Nat. Commun. *10*, 4572.

Mueller, F., Senecal, A., Tantale, K., Marie-Nelly, H., Ly, N., Collin, O., Basyuk, E., Bertrand, E., Darzacq, X., and Zimmer, C. (2013). FISH-quant: automatic counting of transcripts in 3D FISH images. Nat. Methods *10*, 277–278.

Müller-McNicoll, M., Botti, V., de Jesus Domingues, A.M., Brandl, H., Schwich, O.D., Steiner, M.C., Curk, T., Poser, I., Zarnack, K., and Neugebauer, K.M. (2016). SR proteins are NXF1 adaptors that link alternative RNA processing to mRNA export. Genes Dev. *30*, 553–566.

Ni, T., Yang, W., Han, M., Zhang, Y., Shen, T., Nie, H., Zhou, Z., Dai, Y., Yang, Y., Liu, P., et al. (2016). Global intron retention mediated gene regulation during CD4+ T cell activation. Nucleic Acids Res. *44*, 6817–6829.

Padrón, A., Iwasaki, S., and Ingolia, N.T. (2019). Proximity RNA labeling by APEX-seq reveals the organization of translation initiation complexes and repressive RNA granules. Mol. Cell *75*, 875–887.e5.

Palazzo, A.F., and Lee, E.S. (2018). Sequence determinants for nuclear retention and cytoplasmic export of mRNAs and lncRNAs. Front. Genet. *9*, 440.

Pasquinelli, A.E., Ernst, R.K., Lund, E., Grimm, C., Zapp, M.L., Rekosh, D., Hammarskjöld, M.L., and Dahlberg, J.E. (1997). The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. EMBO J. *16*, 7500–7510.

Patil, D.P., Chen, C.-K., Pickering, B.F., Chow, A., Jackson, C., Guttman, M., and Jaffrey, S.R. (2016). m(6)A RNA methylation promotes XIST-mediated transcriptional repression. Nature *537*, 369–373.

Pocock, G.M., Becker, J.T., Swanson, C.M., Ahlquist, P., and Sherer, N.M. (2016). HIV-1 and M-PMV RNA nuclear export elements program viral genomes for distinct cytoplasmic trafficking behaviors. PLoS Pathog. *12*, e1005565.

Pollard, K.S., Hubisz, M.J., Rosenbloom, K.R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. *20*, 110–121.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Read, E.K.C., and Digard, P. (2010). Individual influenza A virus mRNAs show differential dependence on cellular NXF1/TAP for their nuclear export. J. Gen. Virol. *91*, 1290–1301.

Reed, R., and Hurt, E. (2002). A conserved mRNA export machinery coupled to pre-mRNA splicing. Cell *108*, 523–531.

Rodrigues, J.P., Rode, M., Gatfield, D., Blencowe, B.J., Carmo-Fonseca, M., and Izaurralde, E. (2001). REF proteins mediate the export of spliced and unspliced mRNAs from the nucleus. Proc. Natl. Acad. Sci. U S A *98*, 1030–1035.

Roundtree, I.A., Luo, G.-Z., Zhang, Z., Wang, X., Zhou, T., Cui, Y., Sha, J., Huang, X., Guerrero, L., Xie, P., et al. (2017). YTHDC1 mediates nuclear export of N^6-methyladenosine methylated mRNAs. eLife *6*, e31311.

Sakuma, T., Davila, J.I., Malcolm, J.A., Kocher, J.-P.A., Tonne, J.M., and Ikeda, Y. (2014). Murine leukemia virus uses NXF1 for nuclear export of spliced and unspliced viral transcripts. J. Virol. *88*, 4069–4082.

Satterly, N., Tsai, P.-L., van Deursen, J., Nussenzveig, D.R., Wang, Y., Faria, P.A., Levay, A., Levy, D.E., and Fontoura, B.M.A. (2007). Influenza virus targets the mRNA export machinery and the nuclear pore complex. Proc. Natl. Acad. Sci. U S A *104*, 1853–1858.

Saulière, J., Murigneux, V., Wang, Z., Marquenet, E., Barbosa, I., Le Tonquèze, O., Audic, Y., Paillard, L., Roest Crollius, H., and Le Hir, H. (2012). CLIP-seq of eIF4AIII reveals transcriptome-wide mapping of the human exon junction complex. Nat. Struct. Mol. Biol. *19*, 1124–1131.

Savisaar, R., and Hurst, L.D. (2016). Purifying selection on exonic splice enhancers in intronless genes. Mol. Biol. Evol. *33*, 1396–1418.

Schwartz, S., Mumbach, M.R., Jovanovic, M., Wang, T., Maciag, K., Bushkin, G.G., Mertins, P., Ter-Ovanesyan, D., Habib, N., Cacchiarelli, D., et al. (2014). Perturbation of m^6A writers reveals two distinct classes of mRNA methylation at internal and 5′ sites. Cell Rep. *8*, 284–296.

Shukla, C.J., McCorkindale, A.L., Gerhardinger, C., Korthauer, K.D., Cabili, M.N., Shechner, D.M., Irizarry, R.A., Maass, P.G., and Rinn, J.L. (2018). High-throughput identification of RNA nuclear enrichment sequences. EMBO J. *37*, e98452.

Silla, T., Karadoulama, E., Mąkosa, D., Lubas, M., and Jensen, T.H. (2018). The RNA exosome adaptor ZFC3H1 functionally competes with nuclear export activity to retain target transcripts. Cell Rep. *23*, 2199–2210.

Singh, G., Kucukural, A., Cenik, C., Leszyk, J.D., Shaffer, S.A., Weng, Z., and Moore, M.J. (2012). The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. Cell *151*, 915–916.

Smulevitch, S., Michalowski, D., Zolotukhin, A.S., Schneider, R., Bear, J., Roth, P., Pavlakis, G.N., and Felber, B.K. (2005). Structural and functional analysis of the RNA transport element, a member of an extensive family present in the mouse genome. J. Virol. *79*, 2356–2365.

Strässer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., and Hurt, E. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304–308.

Takemura, R., Takeiwa, T., Taniguchi, I., McCloskey, A., and Ohno, M. (2011). Multiple factors in the early splicing complex are involved in the nuclear retention of pre-mRNAs in mammalian cells. Genes Cells *16*, 1035–1049.

Ulitsky, I. (2016). Evolution to the rescue: using comparative genomics to understand long non-coding RNAs. Nat. Rev. Genet. *17*, 601–614.

Ullman, K.S., Shah, S., Powers, M.A., and Forbes, D.J. (1999). The nucleoporin nup153 plays a critical role in multiple types of nuclear export. Mol. Biol. Cell *10*, 649–664.

Umlauf, D., Bonnet, J., Waharte, F., Fournier, M., Stierle, M., Fischer, B., Brino, L., Devys, D., and Tora, L. (2013). The human TREX-2 complex is stably associated with the nuclear pore basket. J. Cell Sci. *126*, 2656–2667.

Valencia, P., Dias, A.P., and Reed, R. (2008). Splicing promotes rapid and efficient mRNA export in mammalian cells. Proc. Natl. Acad. Sci. U S A *105*, 3386–3391.

Van Nostrand, E.L., Freese, P., Pratt, G.A., Wang, X., Wei, X., Xiao, R., Blue, S.M., Dominguez, D., Cody, N.A.L., Olson, S., et al. (2018). A large-scale binding and functional map of human RNA binding proteins. bioRxiv. https://doi.org/10.1101/179648.

Viphakone, N., Hautbergue, G.M., Walsh, M., Chang, C.-T., Holland, A., Folco, E.G., Reed, R., and Wilson, S.A. (2012). TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. Nat. Commun. *3*, 1006.

Viphakone, N., Sudbery, I., Griffith, L., Heath, C.G., Sims, D., and Wilson, S.A. (2019). Co-transcriptional loading of RNA export factors shapes the human transcriptome. Mol. Cell *75*, 310–323.e8.

Wang, K., Wang, L., Wang, J., Chen, S., Shi, M., and Cheng, H. (2018). Intronless mRNAs transit through nuclear speckles to gain export competence. J. Cell Biol. *217*, 3912–3929.

Wickramasinghe, V.O., and Laskey, R.A. (2015). Control of mammalian gene expression by selective mRNA export. Nat. Rev. Mol. Cell Biol. *16*, 431–442.

Wickramasinghe, V.O., McMurtrie, P.I.A., Mills, A.D., Takei, Y., Penrhyn-Lowe, S., Amagase, Y., Main, S., Marr, J., Stewart, M., and Laskey, R.A. (2010). mRNA export from mammalian cell nuclei is dependent on GANP. Curr. Biol. *20*, 25–31.

Wickramasinghe, V.O., Savill, J.M., Chavali, S., Jonsdottir, A.B., Rajendra, E., Grüner, T., Laskey, R.A., Babu, M.M., and Venkitaraman, A.R. (2013). Human inositol polyphosphate multikinase regulates transcript-selective nuclear mRNA export to preserve genome integrity. Mol. Cell *51*, 737–750.

Yap, K., Lim, Z.Q., Khandelia, P., Friedman, B., and Makeyev, E.V. (2012). Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. Genes Dev. *26*, 1209–1223.

Yin, Y., Lu, J.Y., Zhang, X., Shao, W., Xu, Y., Li, P., Hong, Y., Cui, L., Shan, G., Tian, B., et al. (2020). U1 snRNP regulates chromatin retention of noncoding RNAs. Nature *580*, 147–150.

Zhang, L., Das, P., Schmolke, M., Manicassamy, B., Wang, Y., Deng, X., Cai, L., Tu, B.P., Forst, C.V., Roth, M.G., et al. (2012). Inhibition of pyrimidine synthesis reverses viral virulence factor-mediated block of mRNA nuclear export. J. Cell Biol. *196*, 315–326.

Zhang, K., Xie, Y., Muñoz-Moreno, R., Wang, J., Zhang, L., Esparza, M., García-Sastre, A., Fontoura, B.M.A., and Ren, Y. (2019). Structural basis for influenza virus NS1 protein block of mRNA nuclear export. Nat. Microbiol. *4*, 1671–1679.

Zhou, H., Bulek, K., Li, X., Herjan, T., Yu, M., Qian, W., Wang, H., Zhou, G., Chen, X., Yang, H., et al. (2017). IRAK2 directs stimulus-dependent nuclear export of inflammatory mRNAs. eLife *6*, e29630.

Zuckerman, B., and Ulitsky, I. (2019). Predictive models of subcellular localization of long RNAs. RNA *25*, 557–572.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-U2AF65	Sigma-Aldrich	Cat#U4758; RRID:AB_262122
Mouse monoclonal anti-TPR	Abcam	Cat#ab58344; RRID:AB_2287600
Mouse monoclonal anti-NXF1	Abcam	Cat#ab50609; RRID:AB_881770
Rabbit monoclonal anti-GAPDH	Cell Signaling	Cat#C-2118S
Mouse monoclonal anti-ALY	Abcam	Cat#ab6141; RRID:AB_305312
Rabbit polyclonal anti-UAP56	Abcam	Cat#ab47955; RRID:AB_2258813
AzureSpectra 700 Goat-anti-rabbit secondary antibody	Azure biosystems	Cat#AC2128
AzureSpectra 800 Goat-anti-mouse secondary antibody	Azure biosystems	Cat#AC2135
Cy5-conjugated donkey anti-rabbit	Jackson ImmunoResearch	Cat# 711-175-152; RRID:AB_2340607
Chemicals, Peptides, and Recombinant Proteins		
4-Thiouridine	Sigma	T4509
Iodoacetamide	Sigma	A3221
Critical Commercial Assays		
SENSE mRNA-Seq Library Prep Kit	Lexogen	001
NextSeq 500 High Output v2 Kit (75 cycles)	Illumina	FC-404-2005
Deposited Data		
RNA-seq dataset	This paper	GEO: GSE139151
smFISH images	This paper	Mendeley data, https://doi.org/10.17632/4x52dtt4hj.2, 10.17632/w9p9nj53hk.2, 10.17632/7hxjypvbps.2, 10.17632/fd36zzr37r.2, 10.17632/d6m6jnsbfk.2, 10.17632/8p45hryxnh.2, 10.17632/fdtzbsw8v6.2, 10.17632/6hsf4fyhsn.3, 10.17632/9s9m4wytfw.2, 10.17632/bggz8pzxcf.2, 10.17632/yp7z58pjy9.2, 10.17632/cv7n2bbcb4.2.
Western blots	This paper	Mendeley data, https://doi.org/10.17632/6p8tdn5vmm.2
Experimental Models: Cell Lines		
Human: MCF7 cells	ATCC	HTB-22, RRID:CVCL_0031
Human: K562 cells with integrated intron retention library	Mikl et al., 2019	N/A
Oligonucleotides		
Primers for real-time PCR assays	This study	See Table S5
ON-TARGETplus siRNAs	Dharmacon	siNT (Non Targeting): D-001320-10-05; siNXF1: L-013680-01-0005;siALY: L-012078-00-0005; siUAP56: L-003805-00-0005; siTPR: L-010548-00-0005; siU2AF2: L-012380-02-0005; siMCM3AP: L-017379-00-0005; siTHOC2: L-025006-01-0005; siNUP153: L-005283-00-0005; siRBM15: L-010854-00-0005; siRBM15B: L-018823-01-0005; siWTAP: L-017323-00-0005.

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Primers for gene-specific RT and library generation of intron retention library	This study	See Table S5
Software and Algorithms		
Fiji (ImageJ) analysis software	Fiji project	https://fiji.sc/
FishQuant V3	Mueller et al., 2013	https://bitbucket.org/muellerflorian/fish_quant/src/master/
CellProfiler	McQuin et al., 2018	https://cellprofiler.org/
GRAND-SLAM	Jürges et al., 2018	https://github.com/erhard-lab/gedi/wiki/GRAND-SLAM
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/
RSEM	Li and Dewey, 2011	N/A
STAR	Dobin et al., 2013	N/A
DESeq2	Love et al., 2014	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
Other		
Lipofectamine 3000 Transfection Reagent	Life	L3000008
Stellaris RNA FISH probes	Biosearch Technologies	See Table S4

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by Igor Ulitsky (igor.ulitsky@weizmann.ac.il).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

All sequencing data are available in the GEO database under accession GEO: GSE139151.

All smFISH raw images (for Figures 3 and S3) are available on Mendeley data: http://dx.doi.org/10.17632/6hsf4fyhsn.2; http://dx.doi.org/10.17632/9s9m4wytfw.1; http://dx.doi.org/10.17632/cv7n2bbcb4.1; http://dx.doi.org/10.17632/bggz8pzxcf.1; http://dx.doi.org/10.17632/yp7z58pjy9.1; http://dx.doi.org/10.17632/4x52dtt4hj.1; http://dx.doi.org/10.17632/7hxjypvbps.1; http://dx.doi.org/10.17632/w9p9nj53hk.1; http://dx.doi.org/10.17632/d6m6jnsbfk.1; http://dx.doi.org/10.17632/8p45hryxnh.1; http://dx.doi.org/10.17632/fdtzbsw8v6.1; http://dx.doi.org/10.17632/fd36zzr37r.1

All western blots shown in supplemental figures are available on Mendeley data repository with the following http://dx.doi.org/10.17632/6p8tdn5vmm.2

The code supporting the current study is available from the corresponding author on request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

MCF7 cells (ATCC, RRID:CVCL_0031) were grown in DMEM (GIBCO) containing 10% fetal bovine serum and Penicillin-Streptomycin mixture (1%) at 37°C in a humidified incubator with 5% CO_2.

METHOD DETAILS

RNAi treatments

For export factor knockdown, cells were transfected with 10nM siRNA pool targeting indicated genes or with a control pool (Dharmacon) using Lipofectamin 3000 reagent (L3000001, Thermo Fisher). For double knockdowns, 5nM of each pool were used for a total mix of 10nM. Cells were collected for cytosolic and nuclear fractionation 72h post-transfection for all export factor depletion assays, except for siNXF1 and siTREX, where we collected the cells 48h post-transfection. For CytoLib experiments, MCF7 cells were transfected with a 10 nM siRNA pool targeting NXF1 or with a control pool. 28 hr after transfection, cells were washed and transfected with CytoLib

plasmid pool. After an additional 24 hr cells were fractionated and RNA was extracted from cytoplasmic, nuclear, and whole cells extracts.

Extraction of cytoplasmic and nuclear RNA

Cells were washed in cold PBS and detached from plates by 10mM EDTA, a fraction was transferred to a new tube and RNA was extracted with TRIREAGENT (MRC) to obtain WCE. Remaining cells were washed in cold PBS, resuspended in 180μl RLN buffer (50mM Tris·Cl pH8, 140mM NaCl, 1.5mM MgCl2, 10mM EDTA, 1mM DTT, 0.5% NP-40, 10U/ml RNase inhibitor), and incubated on ice for 5 min. The extract was centrifuged for 5 min at 300 g in a cold centrifuge, the supernatant was transferred to a new tube and centrifuged again for 1 min at 500 g in a cold centrifuge. The supernatant (cytoplasmic fraction) was transferred to a new tube and RNA was extracted using TRIREAGENT. The nuclear pellet was washed once in 180μl RLN buffer, resuspended in 1ml of buffer S1 (250mM Sucrose, 10mM MgCl2, 10U/ml RNase inhibitor), layered over 3ml of buffer S3 (880mM Sucrose, 0.5mM MgCl2, 10U/ml RNase inhibitor), and centrifuged for 10 min at 2800 g in a cold centrifuge. The supernatant was removed and RNA was extracted from the nuclear pellet using TRIREAGENT.

RNA-seq and data analysis

Knockdown efficiency of all samples was checked by qRT-PCR and western blot (Figure S1A), and fractionation quality was validated with qRT-PCR using primers for ACTB and MALAT1, which are expected to be enriched in the cytosolic and the nuclear fractions, respectively (Figure S1B). WCE, cytosolic and nuclear fractions obtained after fractionation (from 3 biological replicates) were used to generate cDNA libraries using the SENSE mRNA-Seq Library prep kit (Lexogen) according to manufacturer's protocol and sequenced on a NextSeq 500 machine to obtain 75 nt single-end reads. RNA-seq reads were mapped to the human genome (hg19 assembly) with STAR (Dobin et al., 2013) and gene expression levels were quantified using RSEM (Li and Dewey, 2011) and Bowtie2. GENCODE v26 annotations were used for this and all subsequent analysis, and gene type classification was used as described (Zuckerman and Ulitsky, 2019). At this point, one of the three siTREX samples was excluded from further analysis due to poor sequencing quality. Differential expression in WCE samples was computed using DESeq2 (Love et al., 2014) with default parameters.

Splicing efficiency analysis was performed as described (Zuckerman and Ulitsky, 2019) (Table S1). To estimate splicing efficiency at gene level, we pre-selected a unique and non-overlapping set of introns for each gene, confidently supported by the entire RNA-seq dataset. We then quantified splicing of: (a) all introns together, and (b) the least effectively spliced intron in each gene ("worst intron"). We used only splice-site junction spanning reads for quantification, and defined splicing efficiency as the ratio between exon-exon reads and all reads (exon-exon plus exon-intron) mapped to junctions.

To quantify subcellular localization, we performed normalization of gene-level RSEM output using a method that employs the WCE samples to fit normalization coefficients for Cyto and Nuc samples (Carlevaro-Fita and Johnson, 2019). For each sample and fraction, we used genes whose expression levels were between the 50th and 90th percentiles of all annotated genes to estimate the normalization coefficients using a linear regression analysis. These coefficients were used to normalize the Cyto and Nuc data and obtain absolute localization values. The accuracy of our coefficients was validated by comparing the predicted and actual WCE values for all expressed genes ($r > 0.98$ for all samples, Figure S1C).

Gene-level length parameters, expression levels in WCE, cytosol and nucleus and normalized Cyto/Nuc ratios are listed in Table S1.

Gene architecture analysis

For gene length parameters and exon counts we used the maximal value among GENCODE transcripts per gene. For G/C content analysis, we applied the *alphabetFrequency* function from the Biostrings R package to all GENCODE transcripts and divided the frequency of either G, C or both by the total frequency of all bases per transcript. Mean value across transcripts was used as the gene-level G/C content. GC, C and G content values are listed in Table S1.

Transcription inhibition by actinomycin D

MCF7 cells were treated with siRNA pools targeting NXF1 or with a control pool as described above. 48h post-transfection growth medium was replaced and actinomycin D (A9415, Sigma) was added to a final concentration of 5 μg/ml. DMSO (MP Biomedicals, 1:400) was used as control. Cytosol-nucleus fractionations were performed as described above for non-treated cells (0h), after 4h and after 8h of actinomycin D or DMSO treatment. Extracted RNA was analyzed with qRT-PCR. Delta-Ct values were used to estimate relative abundance across samples. Knockdown efficiency (in the cytoplasmic fraction) and fractionation quality were analyzed by qRT-PCR and/or western blot.

Metabolic labeling and SLAM-seq

For cytoplasmic half life estimation, MCF7 cells were treated with siRNA pools targeting NXF1 or with a control pool as described above. 48h post-transfection growth medium was replaced and 4-Thiouridine (T4509, Sigma) was added to a final concentration of 500 μM. In 8h samples, growth medium was replaced after 4h of labeling with a fresh medium supplemented with 4-Thiouridine. Cells were collected for cytosol-nucleus fractionation before and at 2h, 4h and 8h post-labeling. RNA extraction and Iodoacetamide (A3221, Sigma) treatment were performed as described in (Herzog et al., 2017). Knockdown efficiency (in WCE) was analyzed by

qRT-PCR and western blot. 300 ng of cytosolic RNA per sample (from 2 biological replicates) were used to generate cDNA libraries as described above, and sequenced on a NextSeq 500 machine to obtain 150 nt single-end reads, with a yield of 10-25 million reads per sample. Estimation of cytosolic RNA half-life was performed using GRAND-SLAM (Jürges et al., 2018). Estimated half lives are listed in Table S1. Notably, this estimation is assuming exponential decay in the cytoplasm, and valid for the cytoplasmic fraction assuming that the export rate is much higher than the decay rate.

Single-molecule FISH and immunofluorescence

Probe libraries were designed according to Stellaris guidelines and synthetized by Stellaris (Stellaris RNA FISH probes, Biosearch Technologies) as described in Raj et al. (2008). Libraries targeting NORAD consisted of 96 probes, libraries targeting all other genes consisted of 48 probes (Table S4, labeling fluorophore is indicated for each gene). An oligo-dT(50-mer) probe labeled with FAM was used to examine Poly(A)$^+$ RNA distribution. Hybridization conditions and imaging were as described previously (Bahar Halpern and Itzkovitz, 2016; Lyubimova et al., 2013). Hybridizations were done overnight at 30 °C with probes at a final concentration of 0.1 ng/μl for all probes and 0.3 ng/μl for oligo-dT probe. For immunofluorescence, NXF1 antibody was diluted in glucose oxidase (GLOX) buffer (1:1000) and applied to cells for 1 hour at room temperature. Secondary antibody Cy5-conjugated donkey anti-mouse (1:500) was added to GLOX buffer for 1 hour at room temperature. For nuclear staining, 1.25 μg/ml Hoechst 33342 (H3570, Thermo Fisher) was added during the washes. Images were taken with a Nikon Eclipse Ti2-E inverted fluorescence microscope equipped with a × 100 oil-immersion objective and an iXon 888 EMCCD camera using NIS-Elements Advanced Research software. The image-plane pixel dimension was 0.13 μm and distance between Z stacks was 0.3 μm. Quantification was done with FishQuant V3 (Mueller et al., 2013). We performed automatic 2D projections as suggested in FishQuant documentation, followed by automatic cell segmentation using CellProfiler (McQuin et al., 2018). Hoechst signal was used to segment nuclei and either the oligo-dT signal or Alexa Fluor 488 Phalloidin (A12379, Thermo Fisher) were used to segment cell bodies. Following batch analysis, we manually examined segmentation and removed incorrectly segmented cells from further analysis using Fiji (ImageJ) software. Quantification of cytoplasmic and nuclear signals was performed with default parameters and recommended filters of FishQuant. Same analysis pipeline was used to quantify the NXF1 signal.

Colocalization analysis

To quantify colocalization of NORAD with nuclear speckles, we performed deconvolution of our images using AutoQuant X3.0.4 (Media cybernetics) with default parameters. Then, we used Imaris 9.3.1 (Bitplane) to perform manual 3D segmentation of nuclei and nuclear speckles using Hoechst and MALAT1 signals, respectively, and NORAD signal was used to define 'spots'. We filtered nuclei by distance from image border and from each other to avoid quantification of incomplete or clumped nuclei, and applied the "Find Spots Close To Surface" extension with distance parameter set to 0 to identify all NORAD spots inside nuclei. These spots were then used to identify NORAD spots within nuclear speckles using the same extension. We then calculated colocalization indices by dividing the total amount of nuclear speckle-overlapping spots by the total amount of nuclear spots for each image.

Fractionation and library construction for massively parallel splicing reporter assay

We employed the existing collection of designed and genomically integrated intron retention library, composed of ~8500 variants of a retained intron flanked by exonic sequences as described in Mikl et al.(2019). K562 cells with integrated library were grown in IMDM (GIBCO) containing 10% fetal bovine serum and Penicillin-Streptomycin mixture (1%) at 37°C in a humidified incubator with 5% CO_2. For NXF1 knockdown, 5 × 10^6 cells were transfected with 10nM siRNA pool targeting NXF1 or with a control pool (Dharmacon) using Lipofectamine RNAiMAX reagent (13778-075, Thermo Fisher). Cells were collected for cytosolic and nuclear fractionation 48h post-transfection. Fractionation was carried out as described above for MCF7 cells. WCE, cytosolic and nuclear fractions from 2 biological replicates were used to generate cDNA libraries as follows: 1 μg of RNA per sample was used for cDNA production using the qScript Flex cDNA synthesis kit (95049, Quanta) and a gene specific primer containing part of the Illumina RD2 region (see Table S5). The entire cDNA reaction was diluted into 100μl second strand reaction with a mix of 4 primers introducing a unique molecular identifier (UMI) and a shift as well as part of the Illumina RD1 region. The second strand reaction was carried for a single cycle using Phusion Hot Start Flex DNA Polymerase (NEB, M0535), purified using AMpure beads (A63881, Beckman) at 1.5:1 beads:sample ratio, and eluted in 20μl of ddH_2O. 15μl of the second strand reaction were used for amplification with barcoded primers, the amplified libraries were purified by two-sided AMpure purification first with 0.5:1 beads to sample ratio followed by a 0.8:1 ratio. Libraries were sequenced on a NextSeq 500 machine to obtain 150 nt single-end reads.

Massively parallel splicing reporter assay data analysis

To unambiguously identify the variant of origin of reads from spliced and unspliced isoforms, a unique 12-mer barcode sequence was placed at the 5′ end of each variable region in an exonic region. Reads were first assigned according to their barcode and subsequently the exact position of splicing (or lack thereof) was mapped using the full length read and assigned to either of the splice variants or, in the case of even a single mismatch or usage of a cryptic splice site, discarded. Both steps were performed using custom-made Python scripts.

For all variants with at least 30 reads mapped we computed the fraction of spliced over all reads (spliced/(spliced+unspliced), "PSI") and the log2 ratio of spliced/unspliced reads for each fraction (cytosolic, nuclear or total) separately. In cases where more

than 30 reads were mapped to a given variant, but all of them represented the same isoform, we added one pseudo count to either isoform in order to calculate the log ratio for these variants. For each isoform (spliced and unspliced) we also computed the $\log_2$ ratio of Cyto/Nuc RNA reads. All spliced/unspliced ratios, PSI values and Cyto/Nuc ratios are listed in Table S3.

CytoLib library plasmid construction

The oligonucleotide pool was amplified by PCR (10 ng template in 250μl reaction, divided into 5 50 μL reactions) and purified using AMpure beads (A63881, Beckman) at a 2:1 beads:sample ratio according to the manufacturer's protocol. For adding flanking regions of target plasmid, PCR product was amplified using primers containing parts of the beta-globin Δintrons construct (AATGCCCTGGCCCACAAGTATCACTAAGCC-TAGGAGGCCTCATCTGACTG and TAGGTGACACTATAGAATAGGGCCCTCTAG-GCAGACTAGAGCGGTCGGTA) (200 ng template in 2.4ml reaction, divided into 48 50 μL reactions), concentrated using Amicon tubes (UFC503096, Millipore) and purified using AMpure at a 2:1 beads:sample ratio. 500 ng of the PCR product was used in a PCR with 100 ng beta-globin Δintrons construct and X1 KAPA HiFi HotStart ReadyMixPCR (Kapa biosciences), in total volume of 125μl (divided into 5 25 μL reactions). Reactions were pooled and treated with 10μl DpnI for 1h at 37°C to digest the methylated parental plasmid, and purified using AMpure beads at a 1:1 beads:sample ratio. RF product was transformed into *Escherichia coli* electrocompetent bacteria (60117-2, Lucigen), and plated on 12 × 15-cm LB/Amp agar plates. Colonies were scraped off the plates and DNA was extracted using a plasmid midi-prep kit (12163, QIAGEN).

CytoLib sequencing library generation

One microgram of RNA was used for cDNA production using the qScript Flex cDNA synthesis kit (95049, Quanta) and a gene specific primer containing part of the Illumina RD2 region. The entire cDNA reaction was diluted into 100 μL second strand reaction with a primer containing a unique molecular identifier (UMI) and part of the Illumina RD1 region. The second strand reaction was carried for a single cycle using Phusion Hot Start Flex DNA Polymerase (NEB, M0535), purified using AMpure beads at a 1.2:1 beads:sample ratio and eluted in 20 μL ddH2O. 20μl of the second strand reaction was used for amplification with barcoded primers, and the amplified libraries were purified by two-sided AMpure purification; First with a 0.5:1 beads:sample ratio followed by a 0.7:1 ratio.

CytoLib data analysis

The sequenced reads were used to count individual library tiles using a custom Java script as in Lubelsky and Ulitsky (2018). We considered only R1 reads that contained the TAGGAGGCCTCATCTGACTG adaptor sequence, and extracted the unique molecular identifier (UMI) sequence preceding the adaptor. Each read was then matched to the sequences in the library, without allowing indels. The matching allowed mismatches only at positions with Illumina sequencing quality of at least 35 and we allowed up to two mismatches in the first 15 nt ('seed'), and no more than four overall mismatches. If a read matched more than one library sequence, the sequence with the fewest mismatches was selected, and if the read matched more than one library sequence with the same number of mismatches, it was discarded. The output from this step was a table of counts of reads mapping to each library sequence. Only fragments with at least 20 reads on average in the WCE samples were used in subsequent analysis. We then used these to compute nuclear/cytoplasmic and WCE/plasmid ratios after adding a pseudocount of 0.5 to each UPM. Median values from five biological replicates (siNXF1) and two biological replicates (siRBM15 and siWTAP) were used for subsequent analysis (Table S2). For comparison with the eCLIP data we used "bedtools intersect" to compare the significant tiles ($p < 0.05$ and $\Delta\log_2(\text{Cyto/Nuc}) < -0.3$) with eCLIP peaks (obtained from the ENCODE project (Van Nostrand et al., 2018), with > 2-fold enrichment and $p < 0.01$ based on ENCODE criteria). For each ENCODE experiment, we computed the % of significant tiles that overlapped peaks with the % of other tiles that overlapped the peaks, and compared the two (Table S2).

qRT-PCR

For analysis of gene expression and RNA subcellular localization, extracted RNA was subjected to reverse transcription reaction using qScript Flex cDNA synthesis kit with random primers, according to manufacturer's protocol. Real-time PCR was performed using Fast SYBR qPCR mix (Life Technologies) and analyzed on ViiA 7 real-time PCR machine (Applied Biosystems). Primers used for qRT-PCR are listed in Table S5. For differential expression analysis in WCE samples, expression was normalized by ACTB as the reference gene. Cyto/Nuc ratios were computed without normalization using ΔCt values, fractionation quality was validated by primers targeting the nuclear MALAT1 gene, and the cytoplasmic ACTB and GAPDH mRNAs.

Western blot

Protein was extracted from either TRIREAGENT lysates according to manufacturer's protocol or directly from cells using RIPA Lysis Buffer. Protein samples were resolved on 10% SDS-PAGE, transferred to nitrocellulose membrane and incubated with primary antibodies overnight. AzureSpectra fluorescent 700 anti-rabbit and 800 anti-mouse (Azure biosystems) were used as secondary antibodies for fluorescent quantification of western blots. Blots were imaged on an Azure c600 system.

Motif enrichment analysis

To examine enrichment of known sequence motifs, we pre-selected one isoform per gene that was the most supported isoform across all our samples using our splicing analysis pipeline (Zuckerman and Ulitsky 2019). We then counted sequence motifs'

occurrences in that isoform and compared them with the average counts in 100 control sequences obtained by random di-nucleotide distribution preserving sequences. To control also for the local A/U content, the shuffling was done separately for every 100 bases within the sequence. The examined motifs were: U1 - GGTAAG / GGTGAG / GTGAGT (Almada et al., 2013); m6A - RRACT (Dominissini et al., 2012; Meyer et al., 2012); CAR elements - CCAGTTCCTG / CCAGATCCTG / CCAGCTCCTG (Lei et al., 2013); ESEs - we used the INT3 ESE set from Savisaar and Hurst (2016), composed of 84 sequences. All enrichment values are listed in Table S1 (“ESEs,” “U1,” “CAR” and “m6A” columns).

Additional datasets

We reanalyzed our published dataset of siHNRNPK (Lubelsky and Ulitsky, 2018) using the same pipeline as for the other knockdowns in this study. For NS1 and M protein expression in HeLa cells, we used the processed Cyto/Nuc values in Zhang et al. (2012). Experimental m6A sites were downloaded from Met-DB V2.0 database (table “Single Based m6A site,” id 12 in http://compgenomics.utsa.edu/MeTDB/) (Liu et al., 2018). We pre-selected one isoform per gene as in motif enrichment analysis (see above) to estimate gene-level experimental m6A levels (Table S1, “m6A_data_sb” column).

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical details for individual experiments can be found in the figure legends, the Results section, or in the STAR Methods. This includes number of observations, number of replicates, statistical tests used, and significance level. All boxplots presented in this study indicate the median, quartiles, and 5th and 95th percentiles of the corresponding data.

Molecular Cell, Volume *79*

Supplemental Information

Gene Architecture and Sequence Composition Underpin Selective Dependency of Nuclear Export of Long RNAs on NXF1 and the TREX Complex

Binyamin Zuckerman, Maya Ron, Martin Mikl, Eran Segal, and Igor Ulitsky

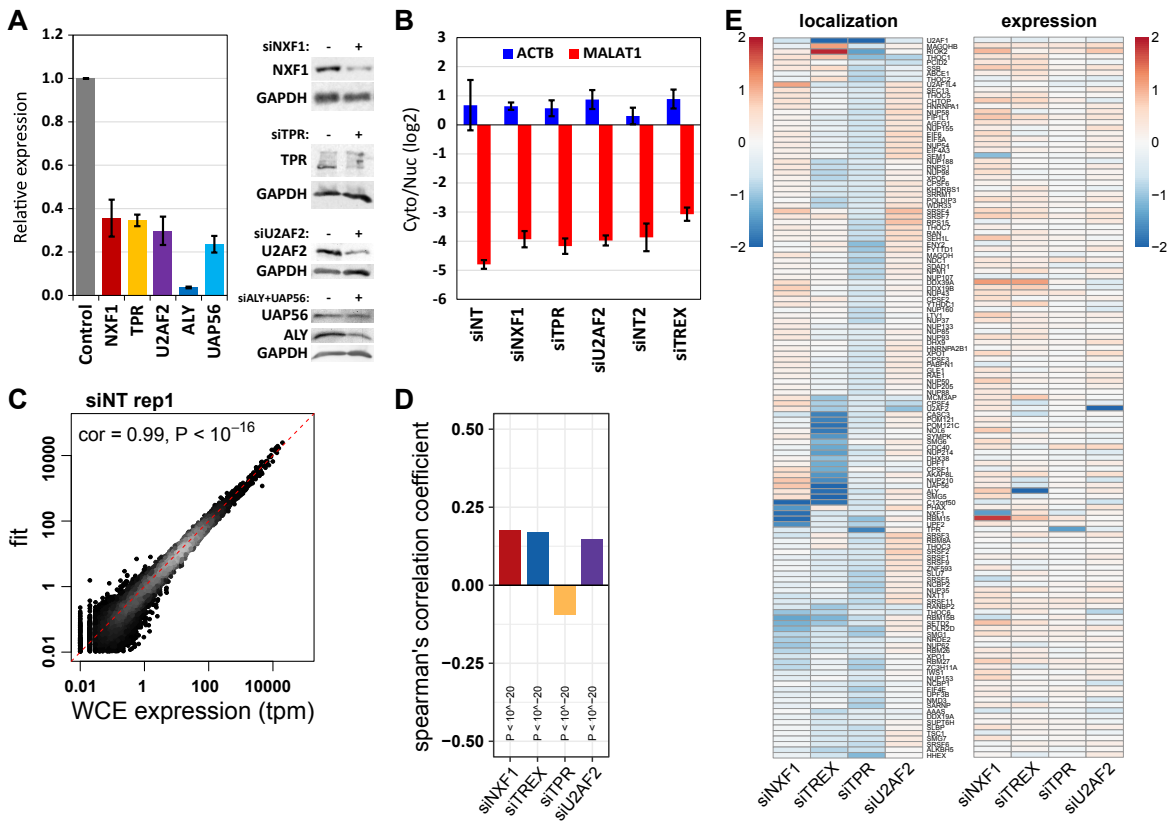


Figure S1. Knockdown and fractionation efficiency and correlations between effects on localization and expression upon export factor depletions. Related to **Figure 1**.

(A) Knockdown efficiency of export factors measured by qRT-PCR (normalized to ACTB expression; n=3, left) and western blot (right).

(B) Fractionation efficiency validation using Cyto/Nuc ratios of cytoplasmic and nuclear markers (*ACTB* and *MALAT1*, respectively) measured by qRT-PCR (unnormalized ΔCt). n=3. Error bars in (A) and (B) represent SEM.

(C) Estimation of accuracy of Cyto/Nuc normalization coefficients (see STAR Methods) by comparing actual WCE expression levels (X axis) vs. predictions by the fit (Y axis). Shown are results for the first replicate of siNT, similar results were observed for other samples and replicates (not shown).

(D) Correlations between changes in expression (in WCE samples) and in localization (Cyto/Nuc) upon export factor depletions. Positive correlation represents association of increased expression with cytoplasmic shift in localization. Spearman's R and P values are indicated.

(E) Effects of export factor depletions on subcellular localization (change in Cyto/Nuc values, left) and expression levels (in WCE, right) of transcripts encoding export-related proteins. Shown are all 147 genes included in the Gene Ontology category "RNA export from nucleus", and, in addition, *ALY* and *RBM15*, which are known to be implicated in nuclear RNA export. Color for localization is the same as in **Figure 1C**, color for expression indicates log2-fold change values relative to control (siNT).

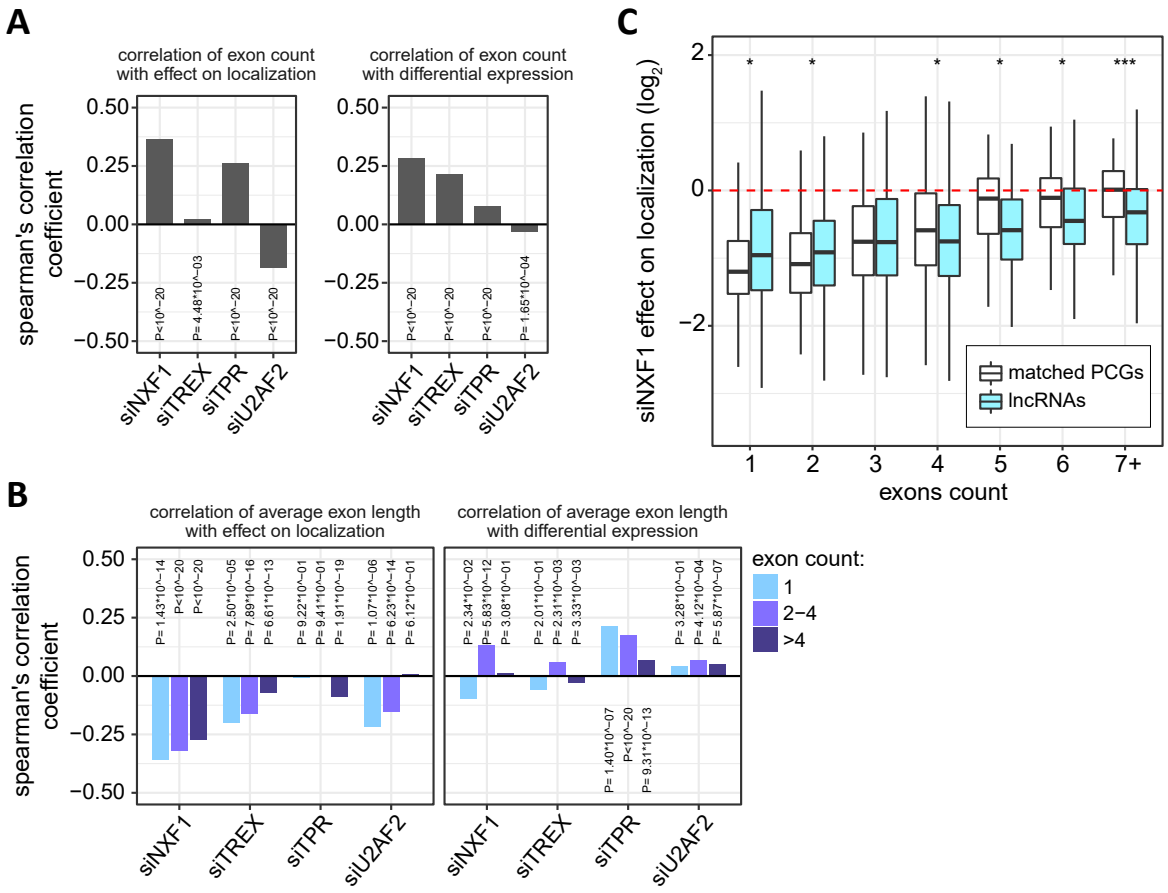


Figure S2. Effects of gene architecture on localization and expression changes upon depletion of export factors. Related to **Figure 2**.

(A) Correlations between exon count and effects of export factor depletions on localization (left) and expression (right). Positive correlation represents association of higher exon count with cytoplasmic shift in localization or with increased expression, respectively. Spearman's R and P values are indicated.

(B) Correlations between average exon length and effects of export factor depletions on localization (left) and expression (right). Positive correlation represents association of longer exons with cytoplasmic shift in localization or with increased expression, respectively. Spearman's R and P values are indicated for gene sub-groups divided by exon counts.

(C) Effect of NXF1 depletion on localization of lncRNAs and matched sample of PCGs binned by exon count. *, P < 0.05; ***, P < 0.001 (Wilcoxon rank-sum test).

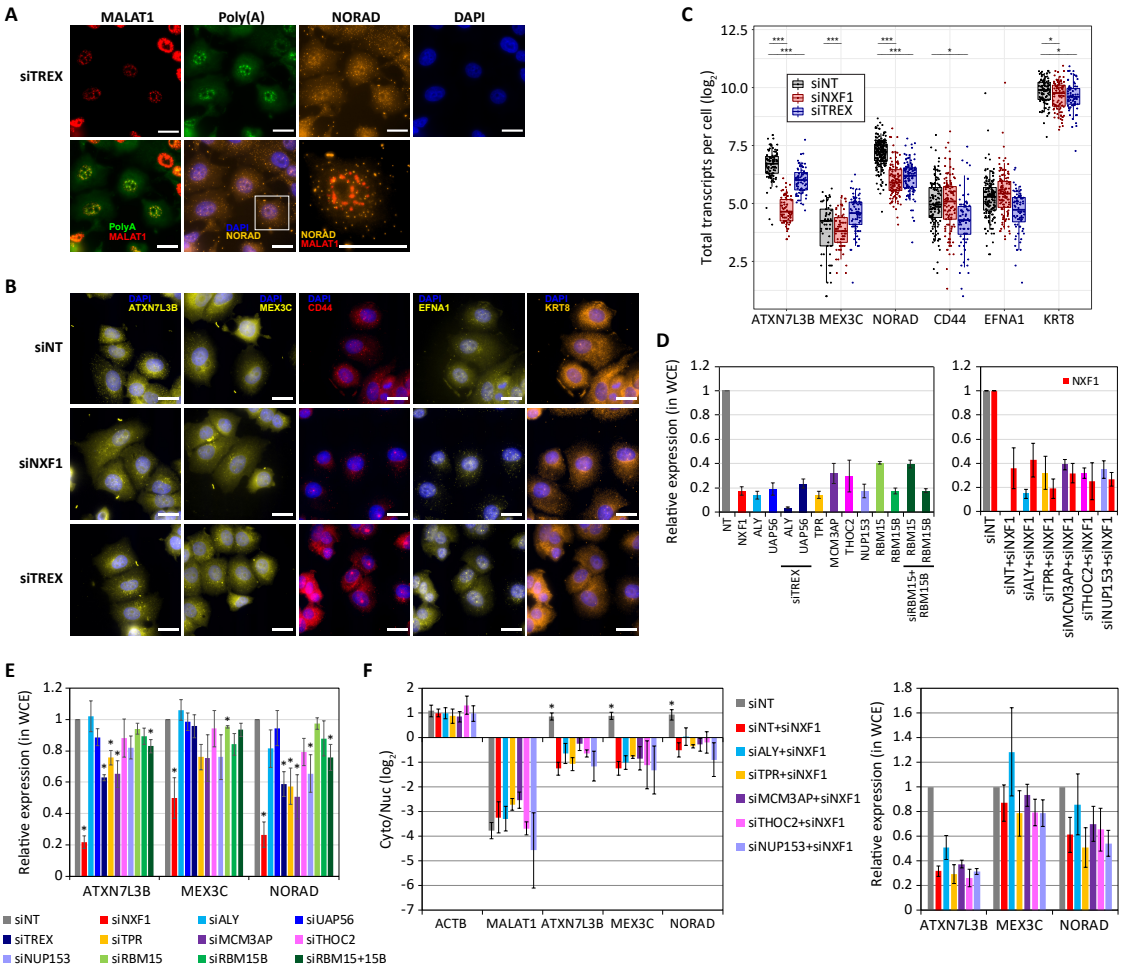


Figure S3. Effect of NXF1 and TREX depletions on other NXF1-sensitive and insensitive genes. Related to **Figure 3**.

(A) Top: Same as **Figure 3A**, for TREX depletion. Bottom: Same as **Figure 3E**, for TREX depletion.

(B) Representative smFISH images of other NXF1 targets and NXF1-insensitive genes upon NXF1 and TREX depletions. Shown are merged images of smFISH signal and DAPI nuclear staining (blue).

(C) Quantification of total abundance of indicated genes as measured by smFISH (see STAR Methods). Each dot is a cell. *, $p < 0.05$; **, $p < 0.005$; ***, $p < 0.0005$ (Wilcoxon rank-sum test).

(D) Export factors' knock-down efficiencies, related to **Figures 3G** (left) and **S3F** (NXF1 co-depletion, right), as measured by qRT-PCR. $n \geq 3$.

(E) Effect of export factor depletions on expression levels (in WCE) of the NXF1 targets *ATXN7L3B*, *MEX3C* and *NORAD* as measured by qRT-PCR. $n \geq 3$. *, $p < 0.05$ (Student's t-test, compared to siNT).

(F) Effect of co-depletions of NXF1 and other export factors on localization (left) and expression levels (right) of the NXF1 targets *ATXN7L3B*, *MEX3C* and *NORAD* as measured by qRT-PCR. Localization of *ACTB* and *MALAT1* was used to estimate fractionation efficiencies. Knockdown efficiencies are shown in **Figure S3D**. $n \geq 3$. No significant differences were observed between siNT+siNXF1 and the other double knock-downs for either localization or expression (Student's t-test).

Scale bars: 25 μM. Expression levels were normalized by *ACTB* expression. Error bars represent SEM.

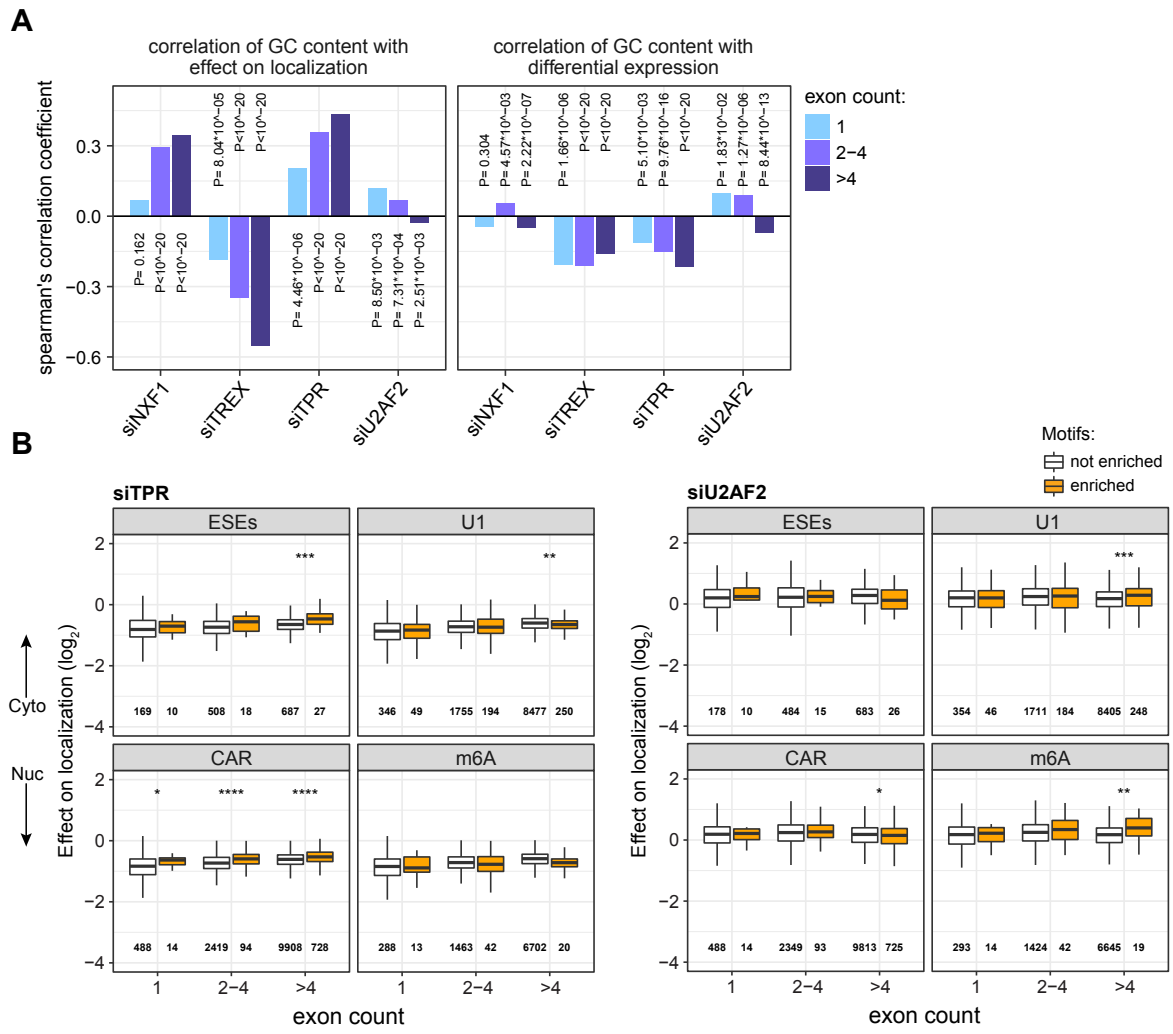


Figure S4. Correlations of G/C content with export factor depletions' effect on localization and expression. Related to **Figure 4**.

(A) Correlations between G/C content and effects of export factor depletions on localization (left) and expression (right). Positive correlation represents association of higher GC content with cytoplasmic shift in localization or with increased expression, respectively. Spearman's R and P values are indicated for gene sub-groups divided by exon counts.

(B) As in **Figure 4B**, but for siTPR and siU2AF2.

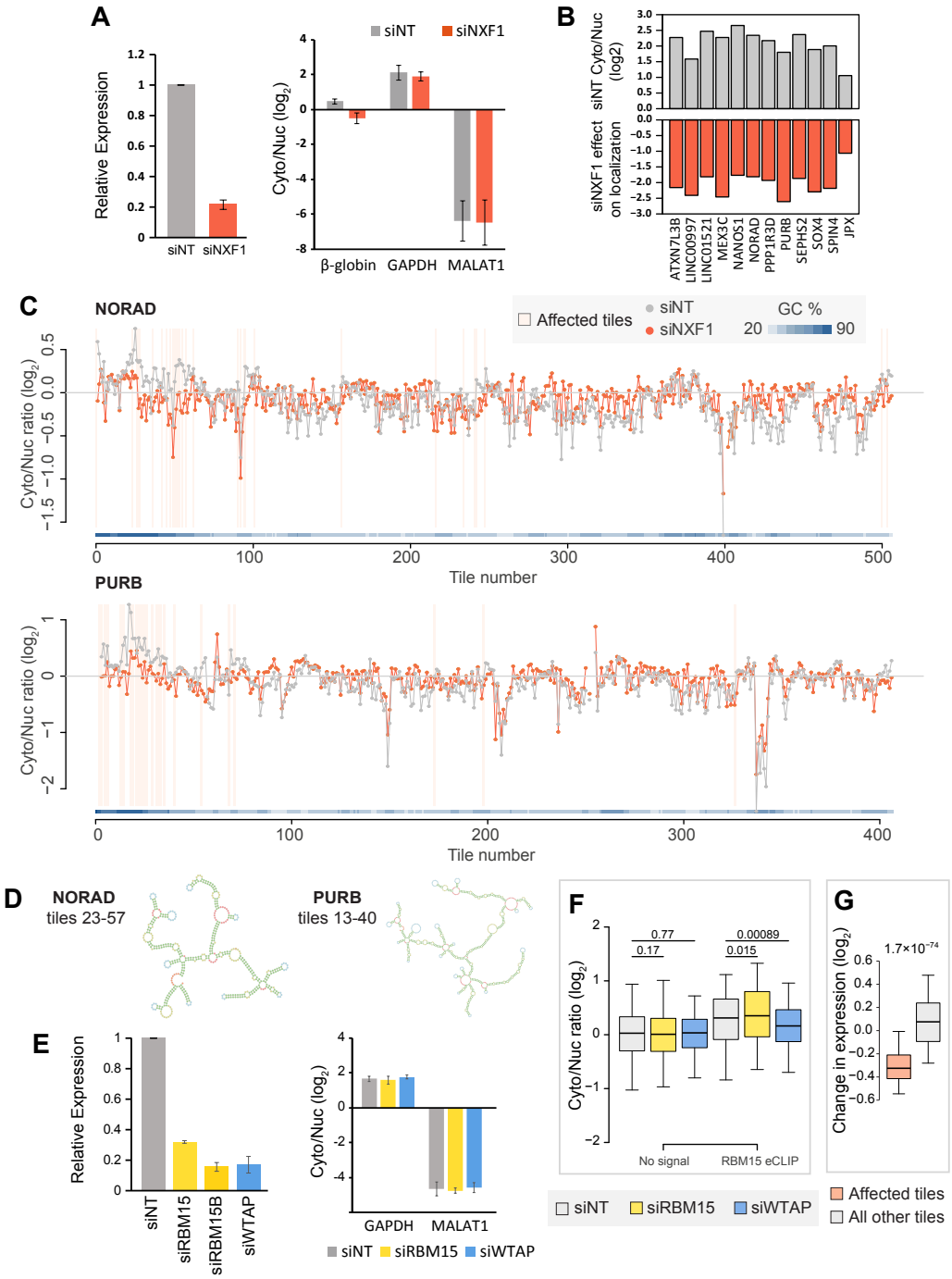


Figure S5. Related to **Figure 5**.

(A) NXF1 knockdown and fractionation efficiency in CytoLib experiments. qRT-PCR measuring localization (Cyto/Nuc, ΔCt) of *GAPDH* and *MALAT1* was used to estimate fractionation efficiency in intron retention library samples. Error bars represent SEM.

(B) Cyto/Nuc ratios (top) and sensitivity to NXF1 depletion (bottom) of the genes tiled in CytoLib.

(C) As in **Figure 5B**, for all the tiles in *NORAD* and in *PURB*.

(D) Structures of the indicated regions in *PURB* and *NORAD*, as predicted by RNAFold.

(E) As in A, for RBM15, RBM15B, and WTAP KD CytoLib experiments.

(F) As in **Figure 5I**, for tiles overlapping RBM15 eCLIP peaks, and all other tiles.

(G) Changes in expression (siNT/plasmid) following NXF1 KD, for tiles affected by NXF1 knockdown and all other tiles.

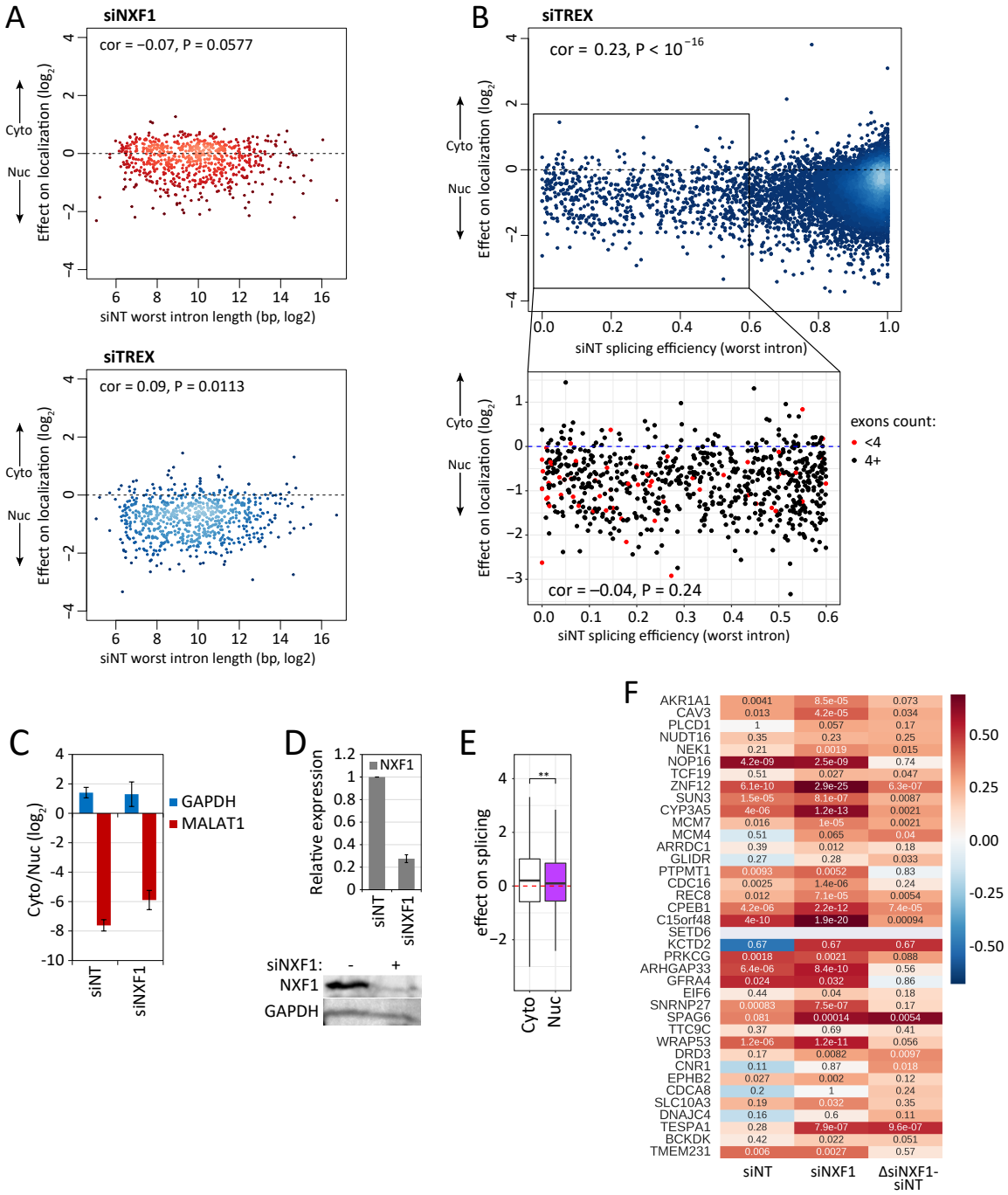


Figure S6. Association between splicing and export factor depletion sensitivity, and effects of NXF1 depletion on intron retention library. Related to **Figure 6**.

(A) Correlations between length of most inefficient intron and effect of NXF1 (top) or TREX (bottom) depletions on localization. Color intensity indicates local point density. Spearman's R and P value are indicated.

(B) As **Figure 6B**, for TREX depletion.

(C) qRT-PCR measuring localization (Cyto/Nuc, ΔCt) of *GAPDH* and *MALAT1* was used to estimate fractionation efficiency in intron retention library samples. Error bars represent SEM.

(D) Knockdown efficiency of NXF1 measured by qRT-PCR (normalized to *ACTB* expression, top) and Western blot (bottom). Error bars represent SEM.

(E) Effect of NXF1 depletion on spliced/unspliced $\log_2$ ratios of intron retention library in Cyto/Nuc fractions. **, $p < 0.01$ (Wilcoxon rank-sum test).

(F) Correlations between siNT WCE spliced/unspliced ratios and Cyto/Nuc values in control and NXF1-depleted cells, and the effect on localization (siNXF1 - siNT) as in **Figure 6E**, but for each indicated sequence context separately. Color indicates Spearman's R, and text is the corresponding P values.

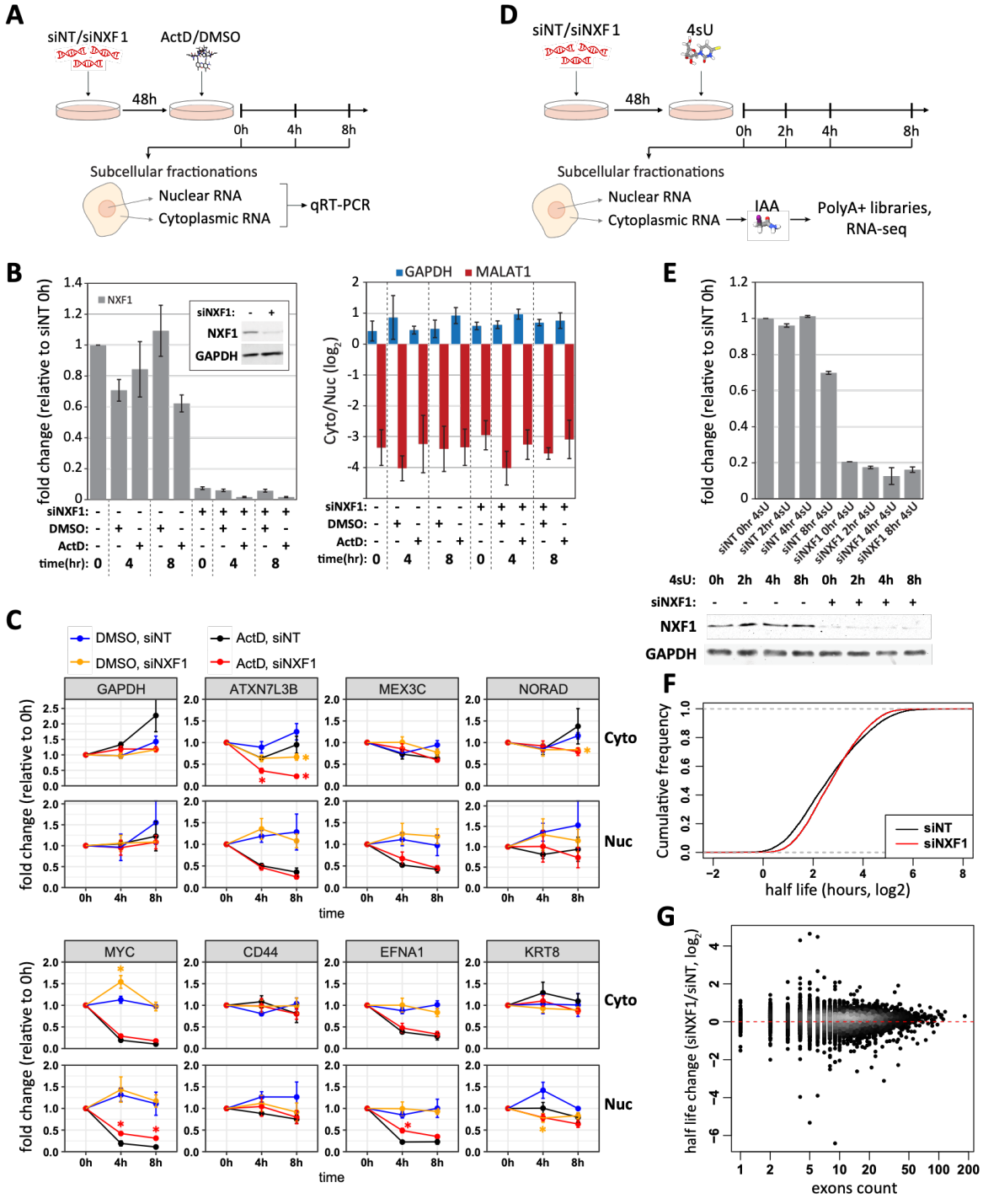


Figure S7. NXF1 depletion does not affect RNA stability. Related to **Figure 6**.

(A) Experimental design of ActD assay (see STAR Methods).

(B) Knockdown (left) and fractionation (right) efficiencies of all samples in ActD assay measured by qRT-PCR. Expression levels were normalized by *ACTB* expression. Localization (Cyto/Nuc, ΔCt) of *GAPDH* and *MALAT1* was used to estimate fractionation efficiencies. Subset, protein levels of NXF1 at t=0 as measured by Western blot.

(C) Relative abundance of indicated genes in ActD- or DMSO-treated samples pre-treated with siRNAs targeting NXF1 or non-targeting control siRNAs, during an 8-

hours time course. Relative abundance was measured in cytoplasmic and nuclear fractions separately. *GAPDH* and *MYC* were used as controls for stable and unstable genes, respectively. n ≥ 3 for all 4h samples, n ≥ 4 for all 0h and 8h samples. *, $p < 0.05$ (Student's t-test, compared to corresponding siNT).

(D) Experimental design of the SLAM-seq assay (see STAR Methods).

(E) Knockdown efficiency of NXF1 in SLAM-seq samples. RNA (top) and protein (bottom) levels are measured by qRT-PCR and Western blot, respectively. RNA levels were measured in cytoplasmic fractions and normalized by *ACTB* expression. Protein levels were measured in WCE samples.

(F) The cumulative distribution of half-lives of all genes in NXF1-depleted and control samples.

(G) Correlation between exon number and half-life change upon NXF1 depletion. Color intensity indicates local point density.

Error bars represent SEM.