Global analyses of UPF1 binding and function reveal expanded scope of nonsense-mediated mRNA decay

Jessica A. Hurt, Alex D. Robertson, and Christopher B. Burge[1]

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

UPF1 is a DNA/RNA helicase with essential roles in nonsense-mediated mRNA decay (NMD) and embryonic development. How UPF1 regulates target abundance and the relationship between NMD and embryogenesis are not well understood. To explore how NMD shapes the embryonic transcriptome, we integrated genome-wide analyses of UPF1 binding locations, NMD-regulated gene expression, and translation in murine embryonic stem cells (mESCs). We identified over 200 direct UPF1 binding targets using crosslinking/immunoprecipitation-sequencing (CLIP-seq) and revealed a repression pathway that involves 3′ UTR binding by UPF1 and translation but is independent of canonical targeting features involving 3′ UTR length and stop codon placement. Interestingly, NMD targeting of this set of mRNAs occurs in other mouse tissues and is conserved in human. We also show, using ribosome footprint profiling, that actively translated upstream open reading frames (uORFs) are enriched in transcription factor mRNAs and predict mRNA repression by NMD, while poorly translated mRNAs escape repression. Together, our results identify novel NMD determinants and targets and provide context for understanding the impact of UPF1 and NMD on the mESC transcriptome.

[Supplemental material is available for this article.]

The multistep nature of eukaryotic gene expression and RNA processing enables multiple layers of regulation but also introduces more opportunities for error. Nonsense-mediated mRNA decay (NMD) is a highly conserved RNA surveillance pathway that oversees mRNA translation and targets those mRNAs harboring premature termination codons (PTCs) for decay, preventing the cell from producing potentially deleterious truncated proteins. As a translation-dependent process, NMD is triggered when a ribosome stalls at the termination codon (TC) of a target RNA and recruits the RNA helicase UPF1 (for review, see Kervestin and Jacobson 2012). UPF1 is conserved in all studied eukaryotes and strictly required for NMD activity (Leeds et al. 1991; for review, see Conti and Izaurralde 2005). The NMD pathway has important implications in human disease, as ~11% of disease-causing mutations result in the production of nonsense-containing mRNAs (Mort et al. 2008) and frequently result in haploinsufficiency phenotypes (for review, see Kuzmiak and Maquat 2006).

Interestingly, while NMD is traditionally considered to be required to prevent the translation of aberrant mRNAs that harbor mutations or result from errors in transcription or splicing, this pathway is also implicated in regulating the expression of many normal ("wild-type") genes and mRNAs (for review, see Schweingruber et al. 2013). These include mRNAs harboring upstream open reading frames (uORFs), selenocysteine codons, long 3′ UTRs, or alternative splicing events that generate isoforms with PTCs. While this last mode is used to regulate the levels of specific factors, particularly splicing factors (Lareau et al. 2007; Ni et al. 2007), in general, the regulation of and importance of this pathway's effects on wild-type gene expression remains poorly understood. A large fraction of the mammalian genome appears to be regulated by NMD; two recent studies have estimated that between one sixth and one quarter of mammalian genes are affected by this pathway (McIlwain et al. 2010; Weischenfeldt et al. 2012). Mice homozygous null for key NMD

factors die during embryogenesis (Medghalchi et al. 2001; Weischenfeldt et al. 2008; McIlwain et al. 2010), suggesting that aberrant expression of NMD target mRNAs may contribute to these phenotypes. However, distinguishing primary from secondary effects of inhibition of the NMD pathway remains challenging, and how NMD activity regulates the mammalian transcriptome early in development is not well understood.

In mammals, targeting of UPF1 to mRNAs that harbor a PTC is primarily thought to occur via its specific interactions with additional, strategically positioned NMD factors. Pre-mRNA splicing results in the deposition of a multiprotein complex, known as the exon junction complex (EJC), ~20–24 nt upstream of the exon-exon junction. The EJC can recruit many different factors that affect mRNA metabolism, including the NMD factors UPF2 and UPF3 (Le Hir et al. 2001). When recruited to the EJC and sufficiently downstream from a TC, UPF2 and UPF3 can stabilize UPF1 interactions at the terminating ribosome and stimulate both its ATPase and its helicase activity (Chamieh et al. 2008; Chakrabarti et al. 2011) as well as its phosphorylation by the kinase SMG1 (Yamashita et al. 2001). These activities, in turn, trigger a cascade of events resulting in degradation of the target mRNA. Exon-exon junctions positioned >50 nt 3′ from the TC (downstream exon-exon junctions or dEJs) trigger NMD of the host mRNA (Cheng and Maquat 1993), a distance likely reflecting the sizes of the terminating ribosome and EJC. Since EJCs are normally displaced by a transiting ribosome during the first or "pioneer" round of translation (Lejeune et al. 2002), typical mammalian mRNAs lacking dEJs (Nagy and Maquat 1998; Giorgi et al. 2007) will be cleared of EJCs in this process and will, therefore, fail to recruit UPF1 and will escape from NMD.

An additional feature of mRNAs that enhances NMD susceptibility is extended 3′ UTR length (Buhler et al. 2006). Factors that associate with poly(A) tails (mainly poly[A] binding protein, cytoplasmic 1, PABPC1) can compete with UPF1 for binding to the terminating ribosome (Behm-Ansmant et al. 2007; Singh et al. 2008), and modulation of the PABPC1-TC distance alters message stability (Amrani et al. 2004; Eberle et al. 2008). Recent studies demonstrated that UPF1 can associate with 3′ UTRs of some

mRNAs, including an endogenous, long 3′ UTR previously shown to be sufficient for decay (Hogg and Goff 2010; Kurosaki and Maquat 2013). However, specificity of UPF1 for particular UTRs is not understood, and the transcriptome-wide binding profile of UPF1 remains largely unknown. Furthermore, the relative contributions of dEJs and 3′ UTR length to NMD of endogenous mRNAs have not been assessed genome-wide.

Despite progress in understanding NMD mechanisms, the canonical determinants of NMD—3′ UTR length and presence of a dEJ—do not fully explain the observed impact of NMD on the transcriptome. For example, many mRNAs that appear as NMD targets in genome-wide studies lack these canonical features, and many transcripts that harbor these traits are not repressed, suggesting that they possess features that enable full or partial escape from degradation. Genome-wide, presence of an upstream open reading frame (uORF) in a gene's 5′ UTR has been associated with NMD (Mendell et al. 2004; Ramani et al. 2009; Yepiskoposyan et al. 2011). However, detailed analysis of specific uORF-containing mRNAs has revealed that only a fraction is actually targeted by this pathway (Linz et al. 1997; Stockklausner et al. 2006; Zhao et al. 2010). Genes with longer than average 3′ UTRs have been associated globally with decay (Mendell et al. 2004; Hansen et al. 2009; Ramani et al. 2009; Yepiskoposyan et al. 2011), but only a few specific UTRs have been shown to confer this activity (Singh et al. 2008; Yepiskoposyan et al. 2011). Similarly, direct binding of mRNAs by UPF1 has been associated with NMD for only a handful of metazoan messages (Hogg and Goff 2010; Hwang et al. 2010). Thus, large-scale identification of direct NMD targets remains challenging and the transcriptome-wide binding and activity of UPF1 insufficiently characterized.

Here, we sought to define the role of UPF1 in gene expression of an early developmental system, murine embryonic stem cells (mESCs), by identifying UPF1 binding locations within the transcriptome and globally measuring the changes in mRNA abundance and translation following perturbations to the NMD pathway. We associate uORF translation with NMD susceptibility and identify a class of UPF1-bound mRNAs that undergo repression by NMD in the absence of canonical NMD features. Interestingly, the set of messages bound by UPF1 in mESCs is repressed by NMD in other mouse cells/tissues, and NMD-dependent repression of this group of mRNAs is conserved in humans. Our results enabled us to describe additional features associated with NMD, to quantify the contributions of these and canonical NMD-triggering features to the decay of endogenous mRNAs, and to better understand the role of NMD in embryonic cells.

Results

Hundreds of mRNAs with dEJs and long 3′ UTRs are derepressed by UPF1 depletion and translational inhibition in mESCs

To identify NMD-regulated genes and isoforms in an early developmental system, we performed RNA-seq analysis of mESCs (v6.5) depleted of UPF1 or treated with cycloheximide (CHX). CHX is a potent translation elongation inhibitor, and relatively short treatment of cells with this drug results in the stabilization of NMD-targeted mRNAs (Carter et al. 1995). We reasoned that use of multiple methods to inhibit NMD, including a translational inhibitor, would increase our ability to identify authentic NMD target mRNAs and that RNA-seq analysis would enable isoform-specific as well as gene-level quantitation. Stable mESC lines were derived

using two independent shRNA sequences targeting *Upf1* (denoted Upf1-1 and Upf1-2) or a control shRNA targeting GFP. In cells infected with *Upf1*-specific shRNAs, UPF1 protein and mRNA levels were reduced to 21%–37% and 14%–15%, respectively, of those in control cells (Supplemental Fig. S1A,B). POU5F1 (also known as OCT4) levels and alkaline phosphatase staining remained similar between UPF1- and control-depleted cells, supporting that ESC state is maintained in the knockdowns (Supplemental Fig. S1C,D). Translational inhibition using CHX was performed on wild-type mESCs for 2 h, a duration that caused a significant increase in abundance of known NMD target mRNAs without causing visible cytotoxicity.

RNA-seq reads were mapped to the mouse genome and exon-exon junctions, and both gene- and isoform-specific abundances were calculated (Trapnell et al. 2010). Pairwise comparisons of gene and mRNA expression values for each experiment were made relative to controls, i.e., v6.5 CHX to v6.5, Upf1-1 to GFP, and Upf1-2 to GFP, following normalization (see Methods). Expression changes were more similar between the two RNAi experiments than between either of these and CHX treatment, as expected (Fig. 1A). The overlap between the sets of mRNA isoforms whose expression increased or decreased by more than 1.1-fold in the three NMD inhibition treatments was twice that expected by chance, and the extent of this overlap rose with increasing fold change, indicating consistency in the response across the treatments (Supplemental Fig. S1E). The extent of overlap above background was greater for mRNAs that increased in abundance after treatment than for those that decreased, consistent with NMD's function as a decay pathway (Supplemental Fig. S1E).

We next sought to assess the relative importance in the mESC NMD pathway of canonical targeting features by analyzing mRNAs with one or more downstream exon-exon junctions (dEJs) or with varying 3′ UTR lengths. We defined a dEJ as an exon-exon junction located ≥50 nt 3′ from an annotated TC (Nagy and Maquat 1998), a classification that includes both mRNAs that harbor a PTC (e.g., as a result of alternative splicing) and mRNAs with introns in their 3′ UTRs. While not a universal rule (Sauliere et al. 2012; Singh et al. 2012), these mRNAs are likely to have an EJC between the TC and dEJ, potentially stimulating UPF1 activity. As expected from previous studies in other cell types and organisms, messages harboring a dEJ increased significantly in abundance following *Upf1* knockdown relative to mRNAs without dEJs (Fig. 1B). Since mRNAs whose expression changed similarly in the three NMD inhibitory treatments are likely enriched for authentic NMD targets, we developed a consistency criterion (see Methods) and identified mRNAs and genes that consistently increased, consistently decreased, or did not change across the three experiments, yielding ~3900 mRNAs and ~4500 genes designated as consistent (Supplemental Table S2). Indeed, the consistent subset of dEJ-containing mRNAs showed stronger derepression (median fold change ~1.19, $P < 1 \times 10^{-7}$) upon *Upf1* knockdown than the full set of dEJ-containing messages (~1.12-fold, $P < 1 \times 10^{-13}$) relative to non-dEJ mRNAs, supporting the enrichment for authentic NMD targets in this set (Fig. 1B,C; and not shown). Similar comparisons between two control clones expressing a GFP-targeting hairpin yielded much smaller fold changes of ±4% (NS, not shown).

Increasing mRNA 3′ UTR length was correlated with increasing derepression following NMD inhibition for each UTR length bin considered (Fig. 1D). The difference in derepression between mRNAs with long 3′ UTRs (>1500 nt) and mRNAs with the shortest 3′ UTRs (50–350 nt) was somewhat greater (on average 1.26-fold) for the *Upf1* knockdowns than for the CHX treatments
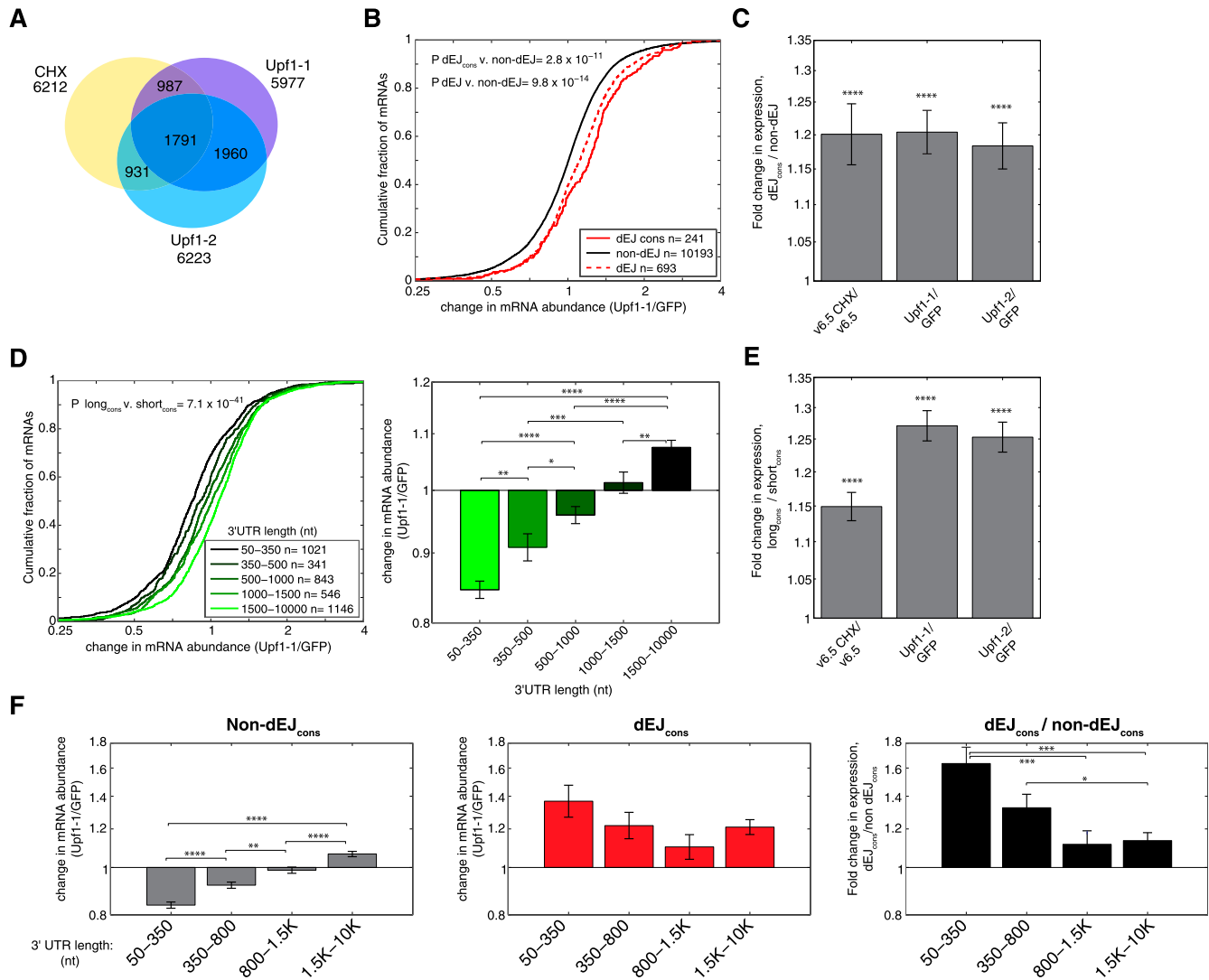
Figure 1. Consistent derepression of hundreds of mRNAs with and without canonical NMD features occurs following UPF1 depletion and translational inhibition. (A) Overlap of mRNAs that changed expression by more than 1.1-fold in the same direction in each of three NMD inhibition experiments (shRNA Upf1-1, shRNA Upf1-2, and CHX treatment). (B) Cumulative distribution functions (CDFs) of changes in mRNA abundance following UPF1 depletion (shRNA Upf1-1) for all dEJ mRNAs (dashed red line), consistently changing dEJ mRNAs (solid red line), or mRNAs without an annotated dEJ (black line). P-value determined by Wilcoxon rank sum test. (C) Ratios of median fold expression change following NMD inhibition of dEJ_{cons} to non-dEJ isoforms. Error bar represents standard error of the two populations compared. P-values determined as in B. (D) As in B for isoforms behaving consistently with different annotated 3′ UTR lengths (different green lines). Median expression changes and standard error are shown at right. See also Supplemental Figure S1F. (E) As in C for mRNAs with long (1500–10k nt) versus short (50–350 nt) 3′ UTRs. (F) Interaction between dEJ and 3′ UTR length. Median fold change in expression following UPF1 depletion of consistent mRNAs with different 3′ UTR lengths without (left) and with (middle) an annotated dEJ. Ratios of median expression change following UPF1 depletion between mRNAs with and without an annotated dEJ of a given 3′ UTR length (right). Significance of differences between different 3′ UTR length bins was determined by permutation test (n = 2000). This trend was also observed when comparing expression changes between each dEJ isoform and non-dEJ isoforms with equivalent 3′ UTR lengths (±10%) following NMD inhibition (Supplemental Fig. S1H). Results were similar for other NMD inhibitory treatments. All fold change values and ratios are plotted on a log_2 scale. P-values: (*) $P < 0.05$, (**) $P < 0.01$, (***) $P < 0.001$, (****) $P < 0.0001$.

(1.15-fold, all $P < 2 \times 10^{-23}$) (Fig. 1E). In all cases, the magnitude of the 3′ UTR length effect seen in the experimental treatments exceeded that observed between controls (Methods). The derepression associated with 3′ UTR length was not dependent on the presence of a dEJ (Fig. 1F), as mRNAs with constitutive long 3′ UTRs lacking dEJs exhibited similar derepression upon NMD inhibition (Supplemental Fig. S1F). Similar results were observed using different minimum expression cutoffs (Supplemental Fig. S1G).

In summary, dEJs and long 3′ UTRs were associated with a similar extent of NMD activity in mESCs. While median expression changes were moderate, some mRNAs changed much more than this. For example, 24 consistent dEJ messages and 33 consistent long 3′ UTR messages were derepressed more than twofold following NMD inhibition (Supplemental Table S2). The observed fold changes almost certainly underestimate the magnitude of NMD's effects, since the ~65%–80% knockdown of *Upf1* achieved likely does not completely abolish NMD activity. Some messages that possessed either a dEJ or a long 3′ UTR were not derepressed by NMD inhibition, either because they remain repressed by residual UPF1 activity or perhaps because they have additional features that

enable escape from NMD. We observed reduced repression of mRNAs harboring long 3′ UTRs containing A-rich segments and a weaker effect for T-rich segments (not shown). Poly(A) stretches internal to a long 3′ UTR and proximal to a TC might recruit PABPC1, thus shortening the apparent 3′ UTR and potentially inhibiting NMD of the message.

Repression associated with a dEJ is strongest in the context of a short 3′ UTR

While addition of a dEJ to an mRNA with a long 3′ UTR can enhance NMD-associated repression (Singh et al. 2008), the relationship between 3′ UTR length and presence of a dEJ as NMD determinants has not been assessed genome-wide in mammalian cells. To address this question, we compared the changes in abundance of dEJ and non-dEJ subsets of mRNAs in different 3′ UTR length classes. In general, presence of a dEJ was associated with increased derepression, irrespective of 3′ UTR length class, following NMD inhibition (Fig. 1F; Supplemental Fig. S1H; data not shown). However, after correction for the derepression associated with UTR length, the relative increase in expression associated with presence of a dEJ was much higher (1.63-fold) for mRNAs with short 3′ UTRs (50–350 nt) than for those with longer UTRs (1.14-fold for UTRs longer than 800 nt) (Fig. 1F). This finding suggests that NMD triggered by a downstream EJC is most active for transcripts with short 3′ UTRs and that transcripts with longer UTRs are less sensitive to the presence of a dEJ.

Genes derepressed following NMD inhibition are enriched for transcription factors

Analysis of the biological functions of derepressed genes revealed expected results as well as some surprises. Several known NMD-targeted isoforms increased in abundance upon NMD inhibition in mESCs, including isoforms of genes involved in pre-mRNA splicing and NMD itself (Supplemental Table S2). In addition, one of the largest and most strongly enriched categories among derepressed genes was transcriptional regulators, including many DNA binding transcription factors (GO:0045449, regulation of transcription $P = 1.5 \times 10^{-11}$, Benjamini-corrected $P = 4.4 \times 10^{-9}$) (Supplemental Table S3). While some changes might be indirect (Dahlseid et al. 2003), repression of transcription factors (TFs) by NMD has also been previously observed in mouse embryonic fibroblasts (MEFs) and HeLa cells (McIlwain et al. 2010; Wang et al. 2011).

Translated but not untranslated uORFs are associated with NMD

mRNAs harboring upstream open reading frames (uORFs) may be susceptible to NMD. If an uORF is translated prior to initial translation of the main ORF in a gene with typical intron distribution, downstream EJCs will be present when the ribosome terminates. Additionally, the typically large distance from the uORF to the poly-A tail could trigger NMD. Under this model, however, decay is triggered only if the uORF is translated.

A previous integrative study reported that genes with uORFs tend to produce ~10%–40% less protein than those without uORFs, with less significant effects on mRNA levels (Calvo et al. 2009). Furthermore, several cases of uORFs that seemingly escape NMD have been described (Stockklausner et al. 2006), leaving the question open as to the degree that uORFs globally affect mRNA stability. Only recently has the translational status of uORFs been assessed genome-wide (Ingolia et al. 2009, 2011). Here we sought

to identify uORFs that are actively translated and to assess their contribution to NMD in the mESC transcriptome.

Ribosome footprint profiling (Ingolia et al. 2009) was performed using UPF1-depleted and control-depleted mESCs, and ribosome locations were mapped within mRNAs and to assess the translational status of uORFs. The density of footprint reads was used to distinguish actively translated uORFs ("tuORFs") from nontranslated uORFs ("ntuORFs") in each cell line (Supplemental Table S4). In our classification scheme, we only considered uORFs located completely upstream of the annotated translation start site in order to cleanly distinguish footprint reads belonging to the uORF from those of the main ORF. Overall, the density of footprint reads in uORFs was well correlated between experiments (Spearman ρ = 0.86 between Upf1-1 and control cells) (Supplemental Fig. S2A). We defined a tuORF as a uORF that had ribosome footprint coverage at least fivefold greater than that of surrounding regions and defined an ntuORF as a uORF that had footprint coverage no greater than the coverage of surrounding regions (see Methods). These definitions are conservative, enabling determination of translation status when the evidence is fairly strong, but leaving some uORFs unclassified. Genes were then classified by the presence and translation status of their uORFs. Ribosome footprint data for a typical tuORF-containing gene, the transcription factor *Dmtf1*, and a ntuORF-containing gene, *Armc1*, are shown in Figure 2A. Using these definitions, we identified 392 and 464 tuORF genes in control and *Upf1* knockdown cells, respectively, with most (347) in common between the two sets. Conversely, we identified 237 and 204 ntuORF genes in control and *Upf1* knockdown cells, respectively. Most of the ntuORFs identified in each cell line also had low uORF to background footprint density ratios in the other cell line (Supplemental Fig. S2B), and <1% of all uORFs that were confidently classified changed classification between cell lines. For downstream analyses, we used uORF classifications derived from *Upf1* knockdown cells, as we reasoned that this condition would enhance our opportunity to observe isoforms that are actively targeted by NMD.

Notably, tuORF genes were modestly but significantly derepressed relative to ntuORF genes following UPF1 depletion ($P < 0.001$ for both hairpins) (Fig. 2B,C). While the degree of tuORF-associated derepression was strongest for the subset of consistent genes, it was also significant for all tuORF-containing genes (Fig. 2B). Furthermore, tuORF genes were significantly derepressed upon NMD inhibition compared to uORF-containing genes overall (Fig. 2B; Supplemental Fig. S2C). Similar trends with smaller magnitudes were observed following translational inhibition (Fig. 2C; Supplemental Fig. S2C). Together, these results suggest that regulated uORF translation can often modulate mRNA stability via NMD.

Interestingly, tuORF-containing genes were enriched for transcriptional regulators compared to all expressed genes (GO: 0045449, regulation of transcription, $P = 3.7 \times 10^{-5}$, Benjamini-corrected $P < 0.05$). Furthermore, we observed that genes encoding transcriptional regulators were enriched 1.5-fold for tuORFs compared to all expressed genes ($P = 4.1 \times 10^{-6}$), and this enrichment increased to twofold for consistently derepressed messages (Fig. 2D). Together, these findings suggest that NMD triggered by uORF translation is an important mechanism of gene expression regulation in mESCs and particularly for modulators of transcription.

Identification of hundreds of mRNAs bound by UPF1, mostly in 3′ UTRs

One challenge facing study of mammalian NMD and of UPF1, in particular, is the identification of direct regulatory targets. While
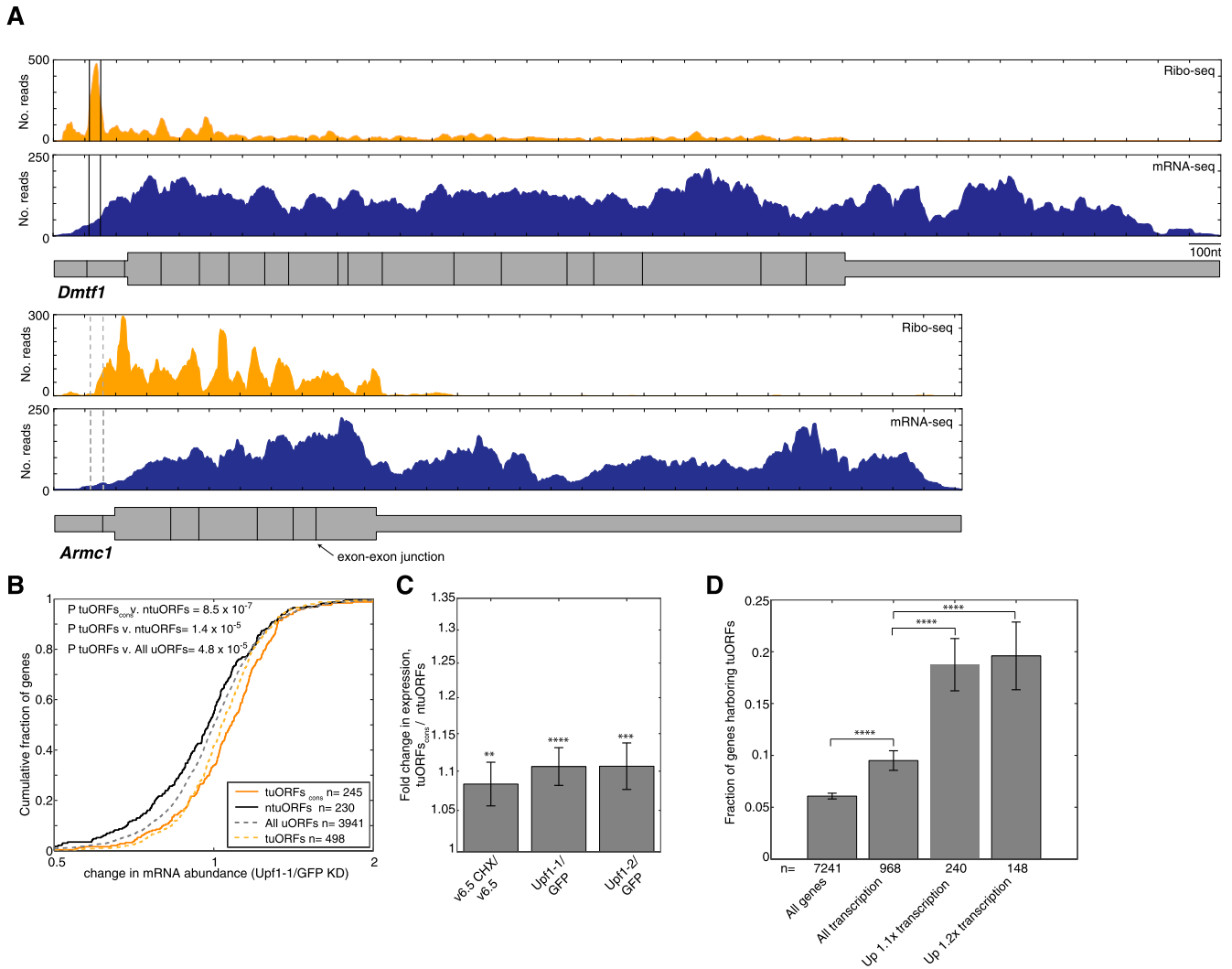
A

Figure 2. Translation of uORFs is associated with UPF1-mediated repression. (*A*) mRNA-seq (blue) and ribosome footprint (orange) reads from UPF1 depleted (shRNA Upf1-1) cells mapping to *Dmtf1* and *Armc1* mRNAs. *Dmtf1* (*top*) contains a tuORF (outlined with dark gray lines), while *Armc1* contains a single ntuORF (dashed lines). (*B*) CDFs of changes in mRNA abundance following UPF1 depletion (shRNA Upf1-1) for all genes with a tuORF (dashed yellow line), consistent genes with a tuORF (orange line), genes with a uORF (dashed gray line), and genes with an ntuORF (black line). *P*-values determined by Wilcoxon rank sum test. (*C*) Ratios of median fold expression change following NMD inhibition of consistent genes with a tuORF to genes with an ntuORF. Error bars represent standard error of the two populations compared. *P*-values determined as in *B*. (*D*) Fraction of genes harboring a tuORF for all expressed genes, all expressed transcriptional regulators (GO:0045449), and transcriptional regulators derepressed by more than 1.1- or 1.2-fold in at least two out of three NMD inhibition experiments. *P*-value of enrichment determined by hypergeometric test. Numbers of genes in each category are indicated. Error bars indicate binomial standard deviations. Fold change values and ratios shown in *B* and *C* are plotted on log$_2$ scale. Asterisks as in Figure 1.

UPF1-bound mRNAs have been associated with NMD in yeast genome-wide (Johansson et al. 2007), metazoan studies have mostly inferred UPF1 targets using indirect evidence such as changes in gene expression following UPF1 depletion. Here, we identified binding targets of UPF1 in mESCs using CLIP-seq. Wild-type mESCs were UV-crosslinked, and the resulting RNA-UPF1 complexes were immunoprecipitated using antibodies against endogenous UPF1 after limited RNase digestion. Since the RNase used can impact CLIP-seq results (Kishore et al. 2011), we prepared libraries using both RNase A (two libraries: Upf1.A1 and Upf1.A2) and RNase I (one library: Upf1.I) to enhance the robustness of the analysis. Small RNA fragments that coprecipitated with UPF1 were isolated, amplified, and sequenced. Anti-rabbit IgG precipitates harvested in parallel contained little or no crosslinked RNA, indicating low levels of intact background RNA-protein complexes remaining after stringent washing during the CLIP procedure (Supplemental Fig. S3A). After mapping the resulting CLIP-seq reads to the mouse genome and transcriptome and subtracting background read density, we determined the fraction of reads mapping to different genic regions (Supplemental Fig. S3B). The density of CLIP reads per nucleotide was ~10- to 30-fold higher in exons than introns in all samples, consistent with the expectation that UPF1 interacts predominantly with mature mRNAs in the cytoplasm (Supplemental Fig. S3C).

Based on the standard model of NMD activity, we had initially hypothesized that the majority of binding events would reside in close proximity to PTCs and/or dEJs. Instead, we observed a pronounced bias for UPF1 binding to occur in mRNA 3′ UTRs, which was consistent in all three CLIP libraries (Fig. 3A). When combining
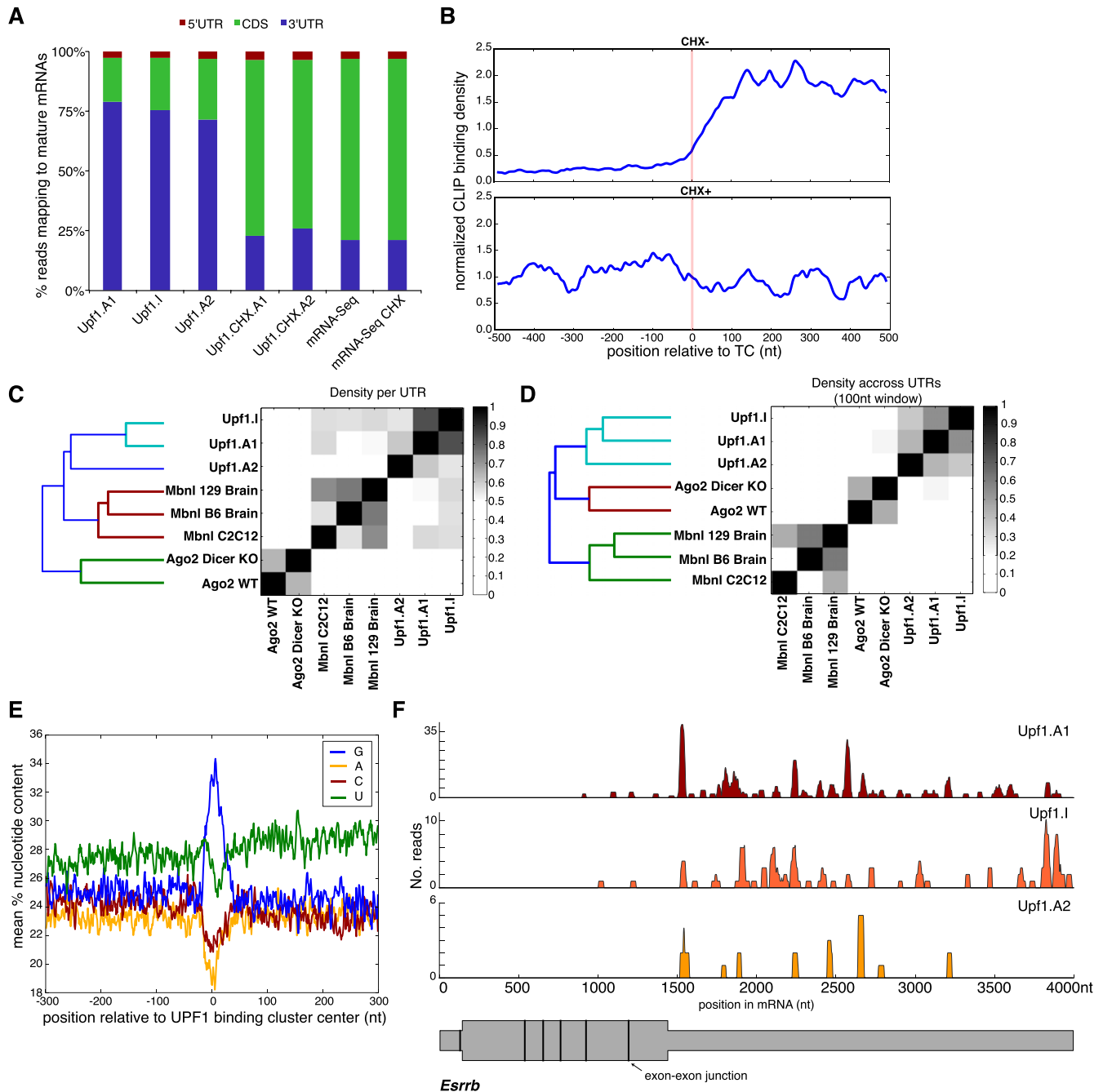
Figure 3. UPF1 interacts predominantly with 3′ UTRs of mature mRNAs. (*A*) Distribution of CLIP-seq reads mapping to 5′ UTRs, coding sequences (CDS), and 3′ UTRs using RNase A ("A") or RNase I ("I") and RNA-seq reads, from control and CHX+ treatments. (*B*) Metagene plot of average UPF1 CLIP tag density per gene in 500-nt regions flanking the TC (red line) under normal (*top*) and CHX (*bottom*) conditions. Density was smoothed using a Gaussian with standard deviation of 10 nt. (*C,D*) Correlation of UPF1 CLIP samples binding in 3′ UTRs of genes with minimum FPKM (fragments per kilobase per million mapped reads) of 50. Correlations of MBNL1 CLIP data in mouse C2C12 cells and two mouse brain samples (Wang et al. 2012) and of AGO2 CLIP data in wild-type and *Dicer* null mESCs (Leung et al. 2011) are shown for comparison. In *C*, correlation coefficients were calculated between densities of CLIP binding over all UTRs. In *D*, correlation coefficients were calculated between densities over 100-nt windows across all UTRs. Grayscale emphasizes higher values in *C*. (*E*) Mean nucleotide content flanking UPF1 binding sites within 3′ UTRs. Lines indicate moving average of values for each nucleotide using a 3-bp sliding window. (*F*) UPF1 binding to the *Esrrb* mRNA in three CLIP experiments in untreated cells. Schematic of *Esrrb* mature mRNA displayed *below*. Width of gray bar indicates CDS and UTR regions, and vertical black bars indicate exon-exon junctions.

the data for all genes in a metagene analysis, the density of UPF1 binding increased rapidly to ~10 times that seen in coding regions just downstream from the TC and remained high throughout the 3′ UTR (Fig. 3B). Preferential binding to the 3′ UTRs of specific mRNAs

was observed (controlling for gene expression) (Supplemental Fig. S3D), and these preferences were strongly correlated across replicate UPF1 CLIP samples, indicating the gene-specific nature of the UPF1 binding signal (Fig. 3C). UPF1 also exhibited preferential binding to

specific locations within 3′ UTRs (Fig. 3D). The positions of binding along 3′ UTRs were correlated between replicates and clustered separately from CLIP-seq locations obtained for other RNA binding proteins—AGO2 and MBNL1—in two recent studies of mouse cells (Leung et al. 2011; Wang et al. 2012), indicating the specificity of the interactions identified. Correlations involving sample Upf1.A2 were less strong than those between samples Upf1.I and Upf1.A1, likely reflecting the lower complexity of the Upf1.A2 library. Comparison of UPF1 and AGO2 binding sites revealed some significant overlaps (Supplemental Fig. S3E). While overlap by itself does not imply a functional relationship, a previous study showed that AGO2 can inhibit NMD (Choe et al. 2010).

While analysis of UPF1 binding sites in 3′ UTRs did not reveal a clear sequence motif, we did find that UPF1-bound regions are enriched for guanosine residues ($P < 0.0001$, χ^2 test) (Fig. 3E). Given UPF1's function as an RNA helicase, we also analyzed RNA structural features. We observed that UPF1 binding sites had higher propensity to form secondary structures (more negative $\Delta G_{folding}$) than surrounding areas (Supplemental Fig. 3SF), an effect that was significant overall but could be attributed to increased GC content (not shown). Thus, our data suggest that UPF1's residence within a 3′ UTR is biased toward primary sequences rich in G nucleotides or toward structures produced by G-rich RNA. Furthermore, analysis of the two CLIP libraries that had deeper coverage (Upf1.A1 and Upf1.I) (Supplemental Table S1) revealed that the extent of UPF1 binding in the upstream half of 3′ UTRs was correlated with the extent of binding to the downstream half of the same UTR (Spearman $\rho = 0.3$ to 0.4, $P = 0.018$ and 0.0013, respectively, in the two libraries). This observation might result from sliding (translocation) of UPF1 along the 3′ UTRs of some mRNAs (Melero et al. 2012). Together, the binding data paint a picture of a factor with a moderate degree of specificity for particular mRNAs and locations within their 3′ UTRs.

Translation displaces UPF1 from ORFs

To ask whether the process of translation influences UPF1 binding locations, we performed CLIP-seq analysis of UPF1 after 2 h of CHX treatment. Under these conditions, UPF1 CLIP tags were enriched in mature mRNAs, as in control conditions (Supplemental Fig. S3C). However, CHX treatment also caused a dramatic redistribution of UPF1 binding within mRNAs, resulting in much higher levels of binding to coding regions (Fig. 3A), with similar densities of binding upstream of and downstream from TCs overall (Fig. 3B). The redistribution of UPF1 binding locations following a 2-h inhibition of translation indicates that UPF1 binding to RNAs is fairly dynamic and suggests that translating ribosomes normally displace UPF1 from ORFs, as likely occurs for other RNA binding factors (Grimson et al. 2007).

We identified significantly UPF1-bound mRNAs in control conditions by comparing the number of UPF1-bound positions within mRNAs relative to what would be expected if binding were random (controlling for gene length and expression level). Given that UPF1 is an RNA helicase, likely interacting transiently with RNA, we adopted a method of identifying high confidence targets within specific gene regions (3′ UTRs or coding regions) rather than specific positions (Methods). After filtering for significant binding in replicate CLIP samples, we identified just over 200 high confidence target mRNAs with significant UPF1 binding in their 3′ UTRs and 17 genes with significant binding to coding regions (Supplemental Table S5). As a control, reads sampled randomly from the RNA-seq data at comparable 3′ UTR depths as the CLIP

reads yielded very few significantly enriched genes (Supplemental Fig. S3G). Unbound mRNAs were defined as those displaying no UPF1 binding in any CLIP experiment. Analyzing genes encoding UPF1-bound mRNAs by Gene Ontology analysis did not yield significant biases, but we noted that some bound mRNAs encoded proteins involved in a cell cycle (thioredoxin interacting protein, TXNIP and cell division cycle 25A, CDC25A), ESC pluripotency (estrogen related receptor, beta, ESRRB) (Fig. 3F; Zhang et al. 2008), and NMD (SMG6 and SMG7). Interestingly, several NMD factors, including SMG6 and SMG7, participate in auto-regulatory feedback circuits to regulate their own levels (Huang et al. 2011). In the case of SMG7, at least, the 3′ UTR appears to mediate this regulation (Yepiskoposyan et al. 2011). Our data raise the possibility that this regulation involves direct binding of UPF1 to this mRNA's 3′ UTR.

UPF1 binding in 3′ UTRs is associated with repression

We next assessed whether UPF1 binding is associated with UPF1 activity by measuring the abundance of mRNAs bound by UPF1 following UPF1 depletion and translational inhibition. Given that UPF1 binding occurred predominantly to the 3′ UTRs of mRNAs, we chose to focus on messages bound in this region for further analysis. mRNAs with significant UPF1 binding in their 3′ UTRs were derepressed compared to unbound mRNAs following all of the NMD-inhibitory treatments (between 1.16- and 1.20-fold, $P < 1 \times 10^{-7}$), implicating UPF1 binding in regulation of their mRNA levels (Fig. 4A,B). No significant change was observed between two control lines (<1.01-fold; not shown). Notably, despite the modest numbers of genes identified, we also observed that genes bound by UPF1 in their CDS were, on average, derepressed following NMD inhibition (Supplemental Fig. S4A).

UPF1 also plays a role in other cellular processes in addition to NMD, including Staufen-mediated mRNA decay (SMD). SMD is a translation- and UPF1-dependent, but UPF2-independent, decay mechanism in which UPF1 is recruited to mRNA 3′ UTRs via Staufen binding (Kim et al. 2005). In order to further characterize the regulation of the UPF1-bound mRNAs in this study, we took advantage of two recently published mouse RNA-seq data sets investigating the role of UPF2 and the UPF1 kinase SMG1 in gene expression regulation (McIlwain et al. 2010; Weischenfeldt et al. 2012). The extent to which UPF1 binds to similar targets in different cell types has not been examined comprehensively. However, we observed significant derepression of mRNAs bound by UPF1 in mESCs in data from *Upf2* knockout liver and *Smg1* knockout MEFs (1.28- and 1.20-fold, respectively, both $P < 1 \times 10^{-4}$), further supporting a connection between UPF1 3′ UTR binding and NMD (Fig. 4B,C; Supplemental Fig. S4B).

The 3′ UTR binding that we observed does not appear to reflect the canonical dEJ-based NMD pathway, as genes encoding 3′ UTR-bound mRNAs were not enriched for expression of dEJ-containing isoforms (Fig. 4D). However, we did observe that UPF1-bound 3′ UTRs were, on average, 2145 nt in length, 70% longer than the average 3′ UTR length of 1262 nt (difference of all versus bound 3′ UTRs, $P = 3.5 \times 10^{-30}$) (Fig. 4E). Given that extended 3′ UTR length is itself an NMD-triggering feature, we asked whether increased 3′ UTR length could explain all of the derepression associated with UPF1-bound mRNAs following NMD inhibition. We next analyzed derepression of genes bound by UPF1, controlling for 3′ UTR length (Supplemental Fig. S4D) or for both 3′ UTR length and expression level (Fig. 4F). Interestingly, UPF1-bound mRNAs were also derepressed relative to these control gene sets
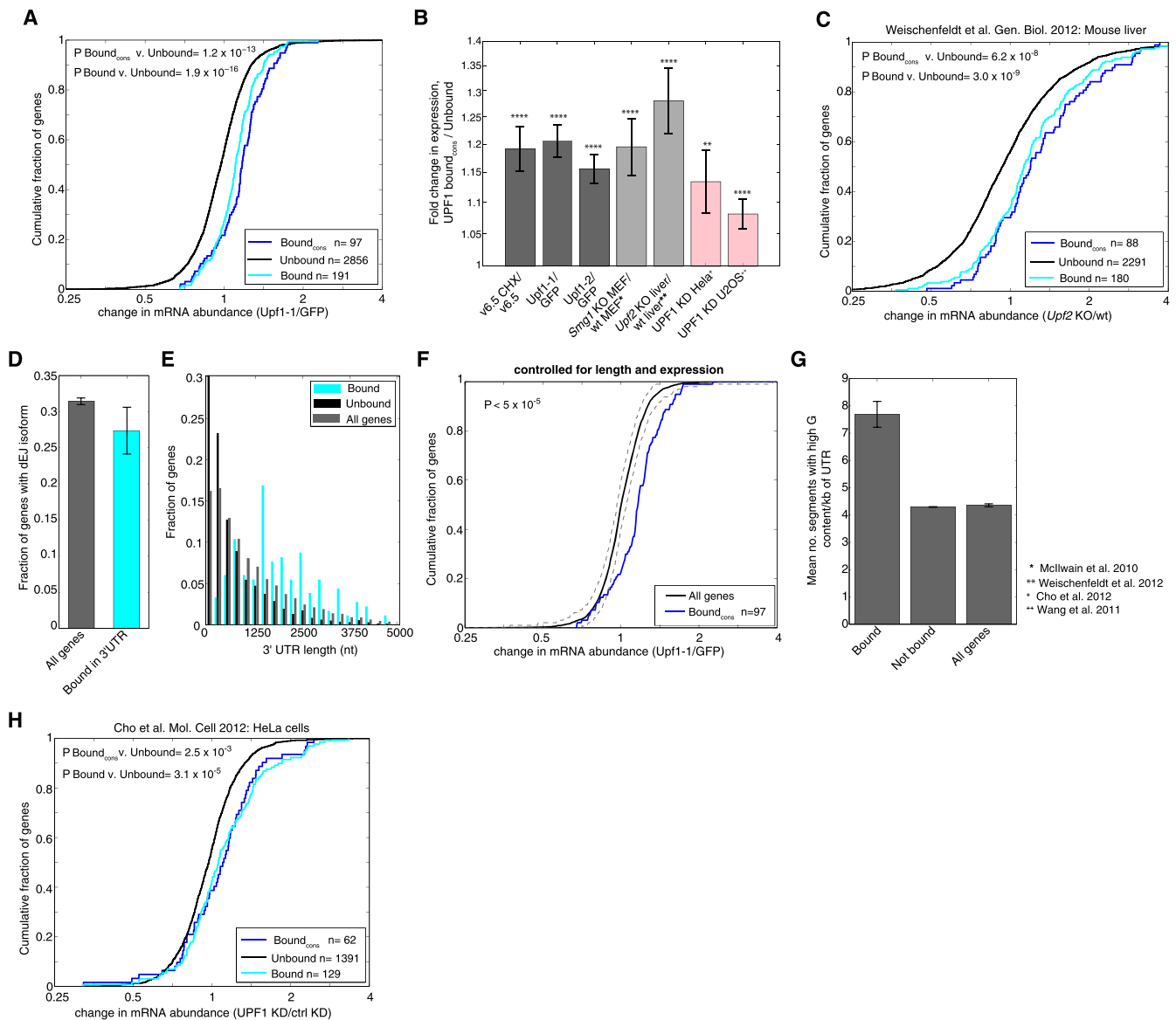
Figure 4. 3′ UTR binding by UPF1 is associated with mRNA repression. (*A*) CDFs of gene expression changes following UPF1 depletion (shRNA Upf1-1) for consistently behaving genes bound by UPF1 in the 3′ UTR (blue line), all genes bound by UPF1 in the 3′ UTR (cyan line), and unbound genes (black line). (*B*) Ratios of median fold expression change following NMD inhibition between consistently behaving genes bound by UPF1 in the 3′ UTR and unbound genes in mESCs (dark gray), as well as for *Smg1* KO in MEFs (McIlwain et al. 2010), and *Upf2* KO in mouse liver (Weischenfeldt et al. 2012) (light gray), and for human homologs of these genes following UPF1 depletion in HeLa (Cho et al. 2012) and U2OS cells (Wang et al. 2011) (pink). Error bars represent standard error of the two populations compared. (*C*) As in *A*, except mRNA abundance measurements were made in wild-type and *Upf2* KO mouse liver (Weischenfeldt et al. 2012). (*D*) Fraction of all genes and genes bound by UPF1 in the 3′ UTR that have an annotated isoform harboring a dEJ. Error bars indicate binomial standard deviation. (*E*) Distribution of 3′ UTR lengths of genes bound by UPF1 in their 3′ UTRs. Lengths were assigned based on the best-annotated isoform for each gene. (*F*) As in *A* for consistently behaving genes bound in the 3′ UTR by UPF1 and genes sampled with replacement to match the distribution of expression levels and 3′ UTR lengths. Significance was calculated by bootstrapping ($n = 20,000$). Mean and 95% confidence intervals of subsampled populations are shown in black and gray lines. Median fold change in expression between UTR groups shown is 1.16, and results were similar using shRNA Upf1-2 and CHX treatment (median fold changes 1.10 and 1.15, respectively) (data not shown). (*G*) Mean number of 50-nt regions with high G content (95th percentile) per kb of 3′ UTR for genes with UPF1-bound 3′ UTRs, all genes, and genes not bound by UPF1 (controlled for expression level and 3′ UTR length). Error bars represent standard error of the mean (Bound or All genes) or standard error of the means of sampled populations (Not bound). (*H*) As in *A*, except gene names were identified by homology with mouse genes either bound or unbound by UPF1 in their 3′ UTR, and mRNA abundance measurements were made in control- and UPF1-depleted HeLa cells (Cho et al. 2012). All expression fold change values and ratios are plotted on a $\log_2$ scale. Asterisks as in Figure 1. See also Supplemental Figure S4.

following both *Upf1* knockdown and translational inhibition (Supplemental Fig. S4D; Fig. 4F). While long UTRs were more likely to exhibit binding, a small subset of mRNAs with 3′ UTRs <800 nt were also bound, and this set of genes was also derepressed upon NMD inhibition, indicating that binding is associated with decay regardless of 3′ UTR length (Supplemental Fig. S4E). These observations suggest that, while UPF1 binding may contribute to the association between 3′ UTR length and NMD, UPF1 binding contributes directly to mRNA decay independent of 3′ UTR length.

Given the high G content of UPF1-binding sites, we also asked whether UPF1-bound 3′ UTRs were enriched for G-rich regions. We defined G-rich regions as 50-bp segments with G content within the top 5% of all such segments in 3′ UTRs. Indeed, we found that UPF1-bound UTRs have, on average, nearly twice the number of G-rich regions per kb than do unbound UTRs with similar lengths and expression levels (Fig. 4G), supporting a link between UPF1 association and G content.

In order to determine whether UPF1-dependent repression of mouse mRNAs bound via their 3′ UTR is conserved, we assessed whether human homologs of genes encoding mRNAs bound by UPF1 in mESCs were similarly repressed. Interestingly we observed that human homologs of mouse UPF1 targets were significantly derepressed following UPF1 depletion in two human cell lines, HeLa and U2OS cells (both $P < 0.01$) (Fig. 4B,H; Supplemental Fig. S4C; Wang et al. 2011; Cho et al. 2012). Together, these findings provide evidence that these genes are similarly repressed in other mouse tissues and cells and in humans.

Genes with low translational efficiency escape NMD

As NMD is a translation-dependent process, we asked whether the level of translational activity influenced susceptibility to NMD. For this purpose, we calculated the average ribosome density—also referred to as "translational efficiency" (TE) (Ingolia et al. 2009)—of each message by dividing the ribosome footprint density of the ORF by the RNA-seq read density of this same region. We then analyzed

the effects on mRNA stability of different NMD-associated features as a function of TE values. When comparing TE to gene expression changes, we calculated each measure using data from separate experiments to avoid an established source of false positives (Larsson et al. 2010) and used only consistently behaving mRNAs to enrich for NMD-related effects.

Overall, TE was positively correlated with derepression following *Upf1* knockdown, consistent with the known translation-dependence of NMD (Fig. 5A). Similar results were observed for the other NMD inhibition treatments (not shown). Furthermore, transcripts harboring dEJs derived from genes with very low TE failed to exhibit significant derepression following UPF1 depletion (Fig. 5B). Similarly, UPF1 binding in the 3′ UTR was not associated with significant derepression following UPF1 depletion for genes with very low TE (Fig. 5C). Low TE genes are likely to encode mRNAs that are translated infrequently, reducing their sensitivity to NMD. For example, such mRNAs might undergo signal-induced translation but otherwise be held in a nontranslating (but stable) state. Interestingly, little difference in 3′ UTR length-dependent derepression was observed between genes grouped by TE in the UPF1 depletion experiments (Fig. 5D). Thus, 3′ UTR length-dependent NMD may be less reliant on robust translation than regulation based on UPF1 binding or presence of a dEJ. While these patterns were somewhat more variable following CHX treatment (not shown), genes with higher TE values were more derepressed overall after all treatments. Together, these data provide systematic evidence for modulation of NMD activity by translational efficiency.
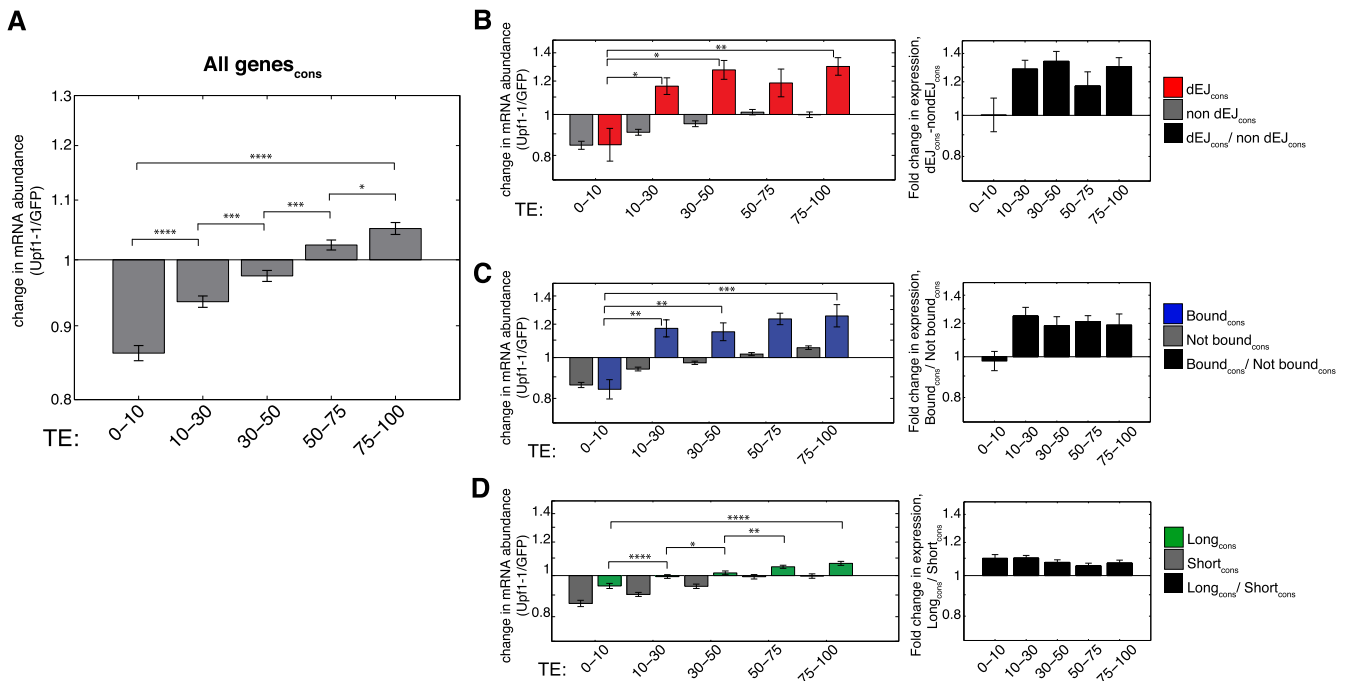

Figure 5. Relationship between TE and NMD-triggering gene features. (*A*) Median fold change of mRNA abundance following UPF1 depletion (shRNA Upf1-1) for consistently behaving genes grouped by percentile rank of translational efficiency (TE). Results were similar using shRNA Upf1-2 and CHX treatment (not shown). (*B*) As in *A*, except for non-dEJ (gray) and dEJ (red) mRNAs. Ratios of median fold change in expression following UPF1 depletion between dEJ and non-dEJ mRNAs with given TE is shown at *right* (black). For dEJ calculation, expression changes were calculated on an isoform level and TE was assigned based on the TE of the full ORF. (*C,D*) As in *B*, except for UPF1 binding in 3′ UTRs (*C*, blue) and 3′ UTR length (*D*, green). Long and short 3′ UTRs were defined as 1500–10,000 nt and 50–350 nt, respectively. Significance of differences in expression changes between TE bins of non-feature-containing genes (non-dEJ, not bound, and short 3′ UTR) were similar to those of all genes in *A*. All expression fold changes and ratios are plotted on a log$_2$ scale. *P*-values calculated using Wilcoxon rank sum test. Asterisks as in Figure 1.

Discussion

Features predictive of mRNA repression

One goal of this study was to identify features that predict an mRNA's susceptibility to NMD. In mESCs, we observed that established NMD-triggering features—long 3′ UTR, presence of a dEJ, and presence of a uORF—were predictive of derepression following NMD inhibition (Fig. 6A). We also observed that presence of a uORF with ribosome footprint coverage indicative of active translation was more predictive of regulation by NMD than mere occurrence of a uORF (Fig. 6A), emphasizing the potential for regulated uORF translation to impact message levels. Most notably, detection of UPF1 binding in a gene's 3′ UTR was as predictive of NMD regulation as was presence of a long 3′ UTR or dEJ, whether analyzed at the gene level (Fig. 6A) or at the level of individual mRNA isoforms (Supplemental Fig. S5). Furthermore, messages with longer 3′ UTRs were more likely to be bound, but binding was predictive of NMD regulation independent of 3′ UTR length (Fig. 4F). In total, ~30% of genes in the mESC transcriptome contain at least one of the four features characterized here (dEJ, long 3′ UTR, translated uORF, or 3′ UTR binding by UPF1). Overall, these genes were 1.7-fold more likely to be repressed by NMD than are genes lacking these features (Fig. 6A), and just requiring presence of UPF1 binding was associated with a 2.5-fold increase in repression potential. Additional mRNA properties that modify the efficacy of these features in triggering NMD undoubtedly exist.

We observed that presence of a dEJ was associated with increased mRNA repression regardless of 3′ UTR length class (Fig. 1F; Singh et al. 2008). The extent of repression associated with presence of a dEJ was greatest for mRNAs with short UTRs (Fig. 1F), suggesting that decay triggered by presence of a long 3′ UTR might reduce the scope of repression achievable by addition of a dEJ.

UPF1 binds extensively in the 3′ UTRs of a cohort of mRNAs

This study provides the first genome-wide identification of UPF1 binding sites within mRNAs. Given models of UPF1 recruitment release factors and/or components of the EJC (Kurosaki and Maquat 2013), we initially hypothesized that the majority of CLIP sites would reside near PTCs and their downstream EJCs. While we did observe a modest number of UPF1 CLIP reads near PTCs of dEJ-containing mRNAs (Supplemental Fig. S3H), most such sites were not detected consistently between CLIP libraries and thus were not emphasized here. mRNAs with dEJs may be degraded more quickly than other classes of NMD targets, making detection by CLIP more difficult. Furthermore, the small sizes of most alternative exons make it challenging to detect isoform-specific binding by CLIP-seq.

Most UPF1 binding locations were distributed along the 3′ UTRs of hundreds of mouse mRNAs (Fig. 3). This finding extends recent reports that UPF1 can associate with several, mostly exogenous 3′ UTRs (Hogg and Goff 2010; Kurosaki and Maquat 2013). Importantly, our genome-wide approach enabled us to find that UPF1 binding sites are not randomly distributed but are concentrated in the UTRs of specific mRNAs (Fig. 3C,D; Supplemental Fig. S3D). A prior study proposed a 3′ UTR length-sensing function for UPF1 (Hogg and Goff 2010), and we show that UPF1-bound 3′ UTRs tend to be longer than average (Fig. 4E). However, we also show that UPF1 binding is associated with NMD-dependent repression in excess of that predicted by 3′ UTR length (Fig. 4F), that UPF1 has specificity for certain UTRs, and that binding events that occur within short 3′ UTRs are associated with NMD. Thus, our data indicate that message-specific features beyond 3′ UTR length

and dEJ presence determine UPF1 binding and associated mRNA decay.

UPF1's interactions with RNA are dynamic and regulated by both ATP binding and interaction with auxiliary factors. The CLIP assay likely captures multiple distinct states, including UPF1 that is stably loaded onto the RNA before interaction with UPF2 or ATP (Chakrabarti et al. 2011), helicase-active UPF1 (post-UPF2 interaction) (Melero et al. 2012), and UPF1 that is actively involved in disassembly of mRNPs as degradation progresses (Franks et al. 2010). RNA helicases often transit through or rearrange mRNP complexes associated with a wide variety of RNA sequences or bind at specific mRNA locations without recognizable primary sequence motifs (Bohnsack et al. 2009; Sievers et al. 2012). UPF1 bound locations lacked a detectable sequence motif but displayed increased G content (Fig. 3E). Interestingly, earlier studies using purified human UPF1 found that its ATPase activity is more than fourfold less active when in the presence of poly(rG) than for any other ribohomopolymer tested (Bhattacharya et al. 2000), suggesting that UPF1 may preferentially pause at G-rich regions. Together, our data indicate that UPF1 resides in particular mRNAs, often in G-rich regions, and localizes mostly to 3′ UTRs except when translation is inhibited.

The degradation of messages targeted by UPF1 binding to the 3′ UTR is not only UPF1-dependent but also translation-dependent (Figs. 4B, 5C; Kurosaki and Maquat 2013). The strong association that we observed between UPF1 binding and derepression in Upf2 knockout tissue (Fig. 4B,C) suggests that our findings are predominantly relevant to NMD rather than SMD. How might additional NMD factors, in particular UPF2, be recruited to typical UPF1-bound 3′ UTRs devoid of exon-exon junctions? One possibility is that, in addition to the canonical mode (Fig. 6B, left), NMD can also operate in a 3′ UTR targeting mode, wherein UPF1 associates with 3′ UTRs of messages and interacts with soluble UPF2 and/or UPF3 or with these factors bound to 3′ UTRs in the absence of an EJC (Fig. 6B). A recent study reported that the degree of UPF1 association with mature mRNA was not affected by depletion of EIF4A3, the primary RNA-binding component of the EJC, suggesting that UPF1's association with mRNA may not require fully assembled EJCs (Singh et al. 2012). Interestingly, the proposed 3′ UTR targeting mode of regulation could occur after the initial rounds of translation of the message, raising the possibility that NMD could be used to deplete bulk pools of post-pioneer messages.

Our CLIP-seq analysis of UPF1 binding appears to be far from saturating (Supplemental Fig. S3G), indicating that UPF1 may bind several hundred or even thousands of mRNAs, potentially accounting for additional NMD-repressed mRNAs (Supplemental Table S3). Notably, we observed that mRNAs bound in ESCs appear to be similarly regulated in MEFs and in mouse liver (Fig. 4B,C; Supplemental Fig. S4B), suggesting that binding to the same or similar sets of mRNAs occurs in other cellular contexts. Intriguingly, human homologs of genes bound by UPF1 in mouse were similarly regulated by UPF1 in human cells (Fig. 4B,H; Supplemental Fig. S4C), indicating conserved UPF1-dependent repression of this set of genes. Together these findings posit that association of UPF1 with the 3′ UTRs of many mESC mRNAs plays a widespread role in gene expression and that this mode of regulation is likely conserved in many cell types and organisms.

NMD regulation via translation of uORFs

We also identified translated uORFs genome-wide and observed an association with UPF1-dependent mRNA repression. Translation of
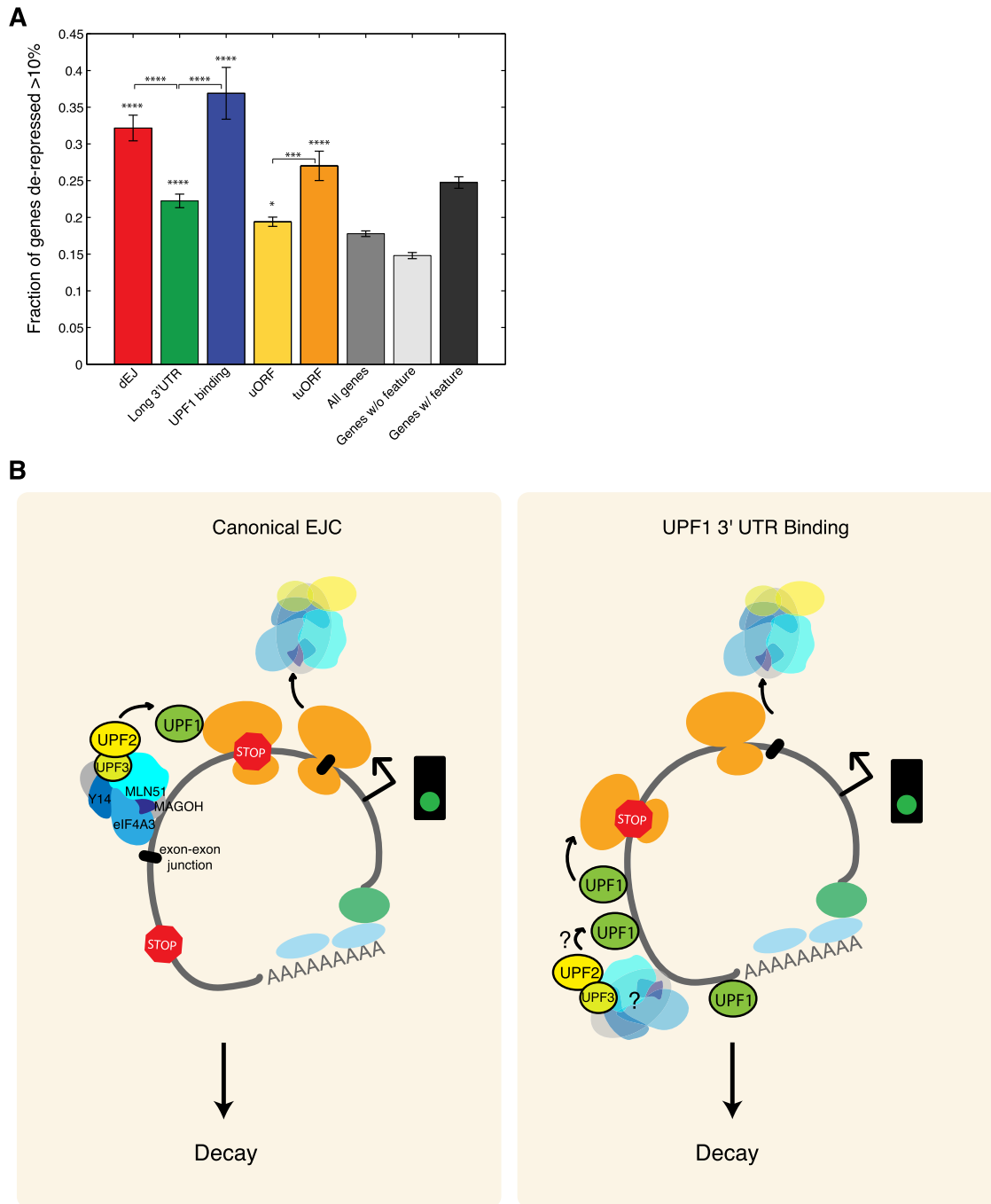
Figure 6. NMD features and models of UPF1-dependent mRNA repression. (*A*) Predictive capacity of mRNA features for NMD-regulation. Fractions of expressed genes harboring a dEJ (red), long 3′ UTR (green), UPF1 3′ UTR binding (blue), uORF (light orange), and/or tuORF (orange) that were derepressed consistently are shown. Shown for comparison are the fractions of genes derepressed consistently regardless of feature content (medium gray), without any NMD-inducing feature (light gray), and with at least one feature (dark gray). *P*-values *above* each feature indicate significance relative to all expressed genes; brackets indicate significant comparisons between features (hypergeometric test). See also Supplemental Figure S5A. Asterisks as in Figure 1. (*B*) (*Left*) Canonical dEJ-mediated regulation. EJC components (blue and gray) are deposited as a consequence of splicing in the nucleus ~24 nt upstream of an exon-exon junction (black bar). Members of the EJC, including UPF2 and UPF3, help to stabilize the transient UPF1-ribosome interaction as well as to stimulate UPF1's phosphorylation and helicase activity, ultimately leading to decay of the message. (*Right*) 3′ UTR UPF1 binding-mediated regulation. UPF1 binds to mRNA 3′ UTRs independent of the presence of an exon-exon junction. At some frequency, UPF1 is activated by interaction with cytoplasmic EJC components. These factors may either be recently released from mRNAs due to translation or perhaps stably associated with message 3′ UTRs independent of an exon-exon junction.

a uORF prior to initial translation of the main ORF could trigger EJC-dependent NMD. Translation of a uORF after main ORF translation could also contribute to NMD by inhibiting translation of the main ORF and thereby increasing the accessibility of the coding region to UPF1 (as seen following CHX treatment) and/or extending the effective length of the 3′ UTR. Regulation of main ORF translation via regulated uORF translation controls expression of key transcription factors involved in cellular stress responses (Gaba et al. 2001; Vattem and Wek 2004). Our data suggest that regulated uORF translation could commonly trigger NMD to reinforce repression at the translational level (Calvo et al. 2009). As a group, transcriptional regulators are derepressed in response to NMD inhibition, and this group is also enriched for tuORFs (Supplemental Table S3; Fig. 2D), suggesting that tuORF-dependent NMD is a common regulator of the expression levels of TFs.

Role of NMD in the mESC transcriptional program

Efficient depletion or elimination of several NMD components results in multisystemic developmental abnormalities and eventual embryonic lethality in worms, zebrafish, flies, and mice, which may be attributable to NMD or non-NMD functions (for review, see Hwang and Maquat 2011; Varsally and Brogna 2012). Notably, examination of available ChIP data revealed that 116 genes that were derepressed following NMD-inhibition have been previously identified as targets of the POU5F1 (also known as OCT4) TF, a master regulator of pluripotency ($P < 0.005$ by hypergeometric test) (Supplemental Table S3; Kim et al. 2008). Several mRNAs encoding developmentally relevant TFs—some of which contain tuORFs—were derepressed following NMD inhibition, including *Klf9* (Martin et al. 2001), *Ncor1* (Jepsen et al. 2000), and *Tbx3* (Ivanova et al. 2006; Lu et al. 2011) (Supplemental Table S3). Misregulation of these or other TFs in NMD-compromised mice might contribute to developmental irregularities.

Our findings, including the recognition of UPF1 binding to 3′ UTRs as a widespread NMD targeting determinant, the identification of hundreds of direct NMD targets, and delineation of the relationships between mRNA translation and NMD susceptibility, provide a context for understanding the role of UPF1 and NMD in development and transcriptome control.

Methods

Cell culture and stable knockdown of Upf1

Mouse v6.5 (129SvJae × C57BL/6) ESCs were cultured on irradiated DR4 mouse embryonic fibroblasts (MEFs) (Applied Stem Cell) in KO DMEM (Gibco), Pen Strep, L-Glutamine, nonessential amino acids, LIF, and either 10% FBS (HyClone) (for wild type) or 15% FBS (for knockdown lines) in gelatinized culture dishes. Puromycin was added to media (1.5 μg/mL) during selection as well as during routine culture of knockdown cells. Puromycin was removed from media for 48 h prior to performing any analysis of knockdown lines. For translational inhibition, 100 μg/mL cycloheximide was added to culture media 2 h prior to harvest. For stable knockdown of *Upf1*, mESCs (20,000 cells) were plated off of MEFs for 24 h prior to infection with 40 μl of ~1.37×10^8 titer virus particles (RNAi Consortium shRNA Library). shRNA sequences are listed in Supplemental Methods. After 24 h, media was changed on all infections, and after 48 h, cells were replated with MEFs using media containing puromycin. Clonal populations were isolated and tested for *Upf1* KD as well as expression of pluripotency factor *Pou5f1* by RT-PCR. Clones with minimal *Pou5f1* expression variation

from wild-type cells but significant change in *Upf1* expression were chosen for further analysis; Upf1-1 KD (4.4, 4.7), Upf1-2 KD (5.2, 5.7), GFP-1,2 KD (2.4, 2.6).

RNA isolation and library preparation

mESCs were trypsinized and preplated on gelatinized dishes for 30 min to remove MEFs prior to harvest in TRIzol reagent (Invitrogen). Total RNA was further purified following isopropanol precipitation, using RNeasy columns and on-column DNase digestion (Qiagen). Twice Poly-T-selected RNA was isolated from 10 μg of total RNA and used as starting material in paired-end, strand-specific dUTP (Parkhomchuk et al. 2009) library prep using the SPRIworks Fragment library system (Beckman Coulter). Final libraries were amplified using 14 PCR cycles, size selected by agarose gel for 290-bp fragments, and sequenced using either 2×80-nt (for knockdown cells) or 2×40-nt (for v6.5 and CHX v6.5 cells) reads on an Illumina HiSeq 2000. To maximize power of detection of lowly expressed isoforms, sequencing data from clones of same hairpins were merged prior to calculation of gene expression values.

UPF1 CLIP-seq

mESCs were plated off of MEFs for 24 h prior to 254 nm UV irradiation (400 mJ/cm^2) in 15-cm plates, trypsinized, washed, and snap-frozen. CLIP-seq was performed similar to Wang et al. (2009, 2012), using antibodies against endogenous UPF1. Details of CLIP procedure are described in Supplemental Methods.

Ribosome footprinting

Footprinting was performed essentially as described in Ingolia et al. (2011), and RNA subtraction as described in Wang et al. (2012), using snap-freezing and no cycloheximide treatment. Additional modifications are described in Supplemental Methods.

Computational analysis

Computational analyses were performed using custom scripts in Python, Perl, MATLAB, or Matplotlib.

Gene expression analysis

For expression quantification, a custom gene annotation database was used consisting of combined 2011 Ensembl (Flicek et al. 2011) and UCSC (Fujita et al. 2011) annotations, with duplicate transcripts removed, and a handful of well-documented PTC+ isoforms that were not in these database releases (Bradley et al. 2012). The RNA-seq and ribosome footprinting reads were mapped to the mouse genome (mm9) and a list of junctions enumerated from our annotation database using TopHat v.1.4.0 (Trapnell et al. 2009), and gene and isoform expression levels were quantitated using Cufflinks v.1.3.0 (Trapnell et al. 2010) as described in Supplemental Methods.

Overlap analyses and consistency filter

For overlap analyses, genes or isoforms changing in expression above or below a given threshold in all three NMD-inhibiting experiments (CHX, Upf1-1, and Upf1-2) were identified (Fig. 1A; Supplemental Fig. S1E). For all other expression analyses, we developed a consistency criterion. For an isoform or gene to pass this criterion, it must have either (1) increased consistently (more than 1.1-fold increase in two or more experiments, and not decreased more than 1.1-fold in the third), (2) decreased consistently (more than 1.1-fold decrease in two or more experiments, and not increased

more than 1.1-fold in the third), or (3) remained consistently unchanged (not changed more than 1.1-fold in either direction in all experiments) (Supplemental Table S2). Isoform and gene groups filtered to pass this filter are designated "cons" in all figures.

Analysis of UPF1 binding

Unique UPF1 CLIP sequences were first trimmed for adapter sequence and 5′ randomized barcode (2 nt) to yield fragments 22–38 nt in length and then mapped uniquely to the mouse genome and splice junction database allowing 2-nt mismatches (-m 1–best –strata) using Bowtie (v.0.12.6) (Langmead et al. 2009). Sequencing reads from the IgG libraries were similarly processed, after which read counts at a given position were amplified by an order of magnitude and "subtracted" from the respective UPF1 library by iteratively canceling out any reads that overlapped. Regional distribution (exons, introns, intergenic/other) of unique UPF1 CLIP (IgG-subtracted) sequences and unique RNA-seq reads were calculated by determining if a read mapped within any known coding region, any UTR, and finally within any intron. Remaining reads were assigned to the intergenic/other category. Lengths of all these regions were calculated based on the best isoform for each coding gene. A P-value for binding to a given region of a gene or isoform was calculated for each CLIP library using a Poisson distribution as described in the Supplemental Methods.

UPF1 binding correlation

Correlation coefficients between binding of UPF1, MBNL1 (Wang et al. 2012), and AGO2 (Leung et al. 2011) were calculated as the correlation of CLIP densities in each 3′ UTR (Fig. 3C) or in 100-nt windows within each 3′ UTR (Fig. 3D) for highly expressed genes (FPKM ≥ 50) similar to Wang et al. (2012).

Comparison with previously published data

Upf2 KO and control mouse liver data sets were downloaded from Gene Expression Omnibus (GSE26561) (Weischenfeldt et al. 2012), and *Smg1* KO and control MEF data were obtained from the lab of Benjamin Blencowe (University of Toronto) (McIlwain et al. 2010). All RNA-seq data were processed as described for data generated in this study. Homologs of UPF1 bound and unbound genes were determined using the BioMart tool (Vilella et al. 2009). To assess gene expression changes for homologs of bound genes, microarray data were downloaded from Gene Expression Omnibus, and the following parameters were used: For GSE30499, a minimum expression threshold of 64 was required of the RMA processed data for inclusion in analyses (Wang et al. 2011); for GSE26781, a minimum expression threshold of 8 of the log transformed, quantile normalized data was required for inclusion in analyses (Cho et al. 2012). For cases in which a gene was represented by multiple probe sets, the mean value of all the corresponding probe sets was used.

Genome-wide survey of NMD

Gene annotations and expression analysis in v6.5 mESCs were used to calculate the number of expressed genes (FPKM > 1) harboring different NMD-inducing features. dEJ genes were defined as those that harbor at least one expressed annotated dEJ isoform (expression level > 1 FPKM and accounting for at least 10% of that of the entire gene). Long 3′ UTR genes were defined as those whose primary annotation harbors a 3′ UTR > 2000 nt. uORF genes were defined as those with at least one annotated uORF. tuORF genes and CLIP genes were defined as described in tuORF methods and

UPF1 3′ UTR binding methods. Predictive capacity of NMD features was determined by calculating the fraction of genes harboring a given feature that were up-regulated by a certain threshold in at least two of three NMD inhibition experiments (without significant down-regulation in the third) compared to all expressed genes in the transcriptome that harbored this feature (significance calculated by two-tailed Fisher's exact test). In order to avoid double-counting of genes, features were called for genes using the following hierarchy: (1) UPF1 3′ UTR binding; (2) presence of dEJ; (3) presence of a long 3′ UTR; and (4) presence of a tuORF.

uORF classification

For classification of uORF translation status, ribosome footprinting reads were reduced to the codon occupied by the A site of the ribosome calibrated based on the pile up at known stop codons, similar to Ingolia et al. (2011). For each uORF, densities of mapped A sites were calculated within the uORF and the background untranslated region consisting of the 10 codons upstream of and downstream from the uORF. uORFs were required to be covered by at least one RNA-seq read to ensure they were spliced into the message. uORFs were called as translated when there was at least a fivefold greater density above the higher of either the footprint density in the flanking 60 nt or a minimum threshold coverage of two-thirds. If an uORF was not called as translated, it could be called as confidently untranslated (an ntuORF) if the footprint density within the uORF was less than the higher of either the footprint density in the flanking 60 nt or one read per 60 nt. Genes were classified as tuORF-containing if they harbored a transcript with one or more tuORFs or as ntuORF-containing if they harbored one or more ntuORFs and no tuORFs as called in data from clone 4.7 (shRNA Upf1-1).

Additional procedural details are described in Supplemental Methods.

Data access

High throughput sequencing data generated in this study has been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE41785.

References

Amrani N, Ganesan R, Kervestin S, Mangus DA, Ghosh S, Jacobson A. 2004. A faux 3′-UTR promotes aberrant termination and triggers nonsense-mediated mRNA decay. *Nature* **432:** 112–118.

Behm-Ansmant I, Gatfield D, Rehwinkel J, Hilgers V, Izaurralde E. 2007. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. *EMBO J* **26:** 1591–1601.

Bhattacharya A, Czaplinski K, Trifillis P, He F, Jacobson A, Peltz SW. 2000. Characterization of the biochemical properties of the human Upf1 gene product that is involved in nonsense-mediated mRNA decay. *RNA* **6:** 1226–1235.

Bohnsack MT, Martin R, Granneman S, Ruprecht M, Schleiff E, Tollervey D. 2009. Prp43 bound at different sites on the pre-rRNA performs distinct functions in ribosome synthesis. *Mol Cell* **36:** 583–592.

Bradley RK, Merkin J, Lambert NJ, Burge CB. 2012. Alternative splicing of RNA triplets is often regulated and accelerates proteome evolution. *PLoS Biol* **10:** e1001229.

Buhler M, Steiner S, Mohn F, Paillusson A, Muhlemann O. 2006. EJC-independent degradation of nonsense immunoglobulin-µ mRNA depends on 3′ UTR length. *Nat Struct Mol Biol* **13:** 462–464.

Calvo SE, Pagliarini DJ, Mootha VK. 2009. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *Proc Natl Acad Sci* **106:** 7507–7512.

Carter MS, Doskow J, Morris P, Li S, Nhim RP, Sandstedt S, Wilkinson MF. 1995. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. *J Biol Chem* **270:** 28995–29003.

Chakrabarti S, Jayachandran U, Bonneau F, Fiorini F, Basquin C, Domcke S, Le Hir H, Conti E. 2011. Molecular mechanisms for the RNA-dependent ATPase activity of Upf1 and its regulation by Upf2. *Mol Cell* **41:** 693–703.

Chamieh H, Ballut L, Bonneau F, Le Hir H. 2008. NMD factors UPF2 and UPF3 bridge UPF1 to the exon junction complex and stimulate its RNA helicase activity. *Nat Struct Mol Biol* **15:** 85–93.

Cheng J, Maquat LE. 1993. Nonsense codons can reduce the abundance of nuclear mRNA without affecting the abundance of pre-mRNA or the half-life of cytoplasmic mRNA. *Mol Cell Biol* **13:** 1892–1902.

Cho H, Kim KM, Han S, Choe J, Park SG, Choi SS, Kim YK. 2012. Staufen1-mediated mRNA decay functions in adipogenesis. *Mol Cell* **46:** 495–506.

Choe J, Cho H, Lee HC, Kim YK. 2010. microRNA/Argonaute 2 regulates nonsense-mediated messenger RNA decay. *EMBO Rep* **11:** 380–386.

Conti E, Izaurralde E. 2005. Nonsense-mediated mRNA decay: Molecular insights and mechanistic variations across species. *Curr Opin Cell Biol* **17:** 316–325.

Dahlseid JN, Lew-Smith J, Lelivelt MJ, Enomoto S, Ford A, Desruisseaux M, McClellan M, Lue N, Culbertson MR, Berman J. 2003. mRNAs encoding telomerase components and regulators are controlled by UPF genes in *Saccharomyces cerevisiae*. *Eukaryot Cell* **2:** 134–142.

Eberle AB, Stalder L, Mathys H, Orozco RZ, Muhlemann O. 2008. Posttranscriptional gene regulation by spatial rearrangement of the 3′ untranslated region. *PLoS Biol* **6:** e92.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, et al. 2011. Ensembl 2011. *Nucleic Acids Res* **39:** D800–D806.

Franks TM, Singh G, Lykke-Andersen J. 2010. Upf1 ATPase-dependent mRNP disassembly is required for completion of nonsense-mediated mRNA decay. *Cell* **143:** 938–950.

Fujita PA, Rhead B, Zweig AS, Hinrichs AS, Karolchik D, Cline MS, Goldman M, Barber GP, Clawson H, Coelho A, et al. 2011. The UCSC Genome Browser database: Update 2011. *Nucleic Acids Res* **39:** D876–D882.

Gaba A, Wang Z, Krishnamoorthy T, Hinnebusch AG, Sachs MS. 2001. Physical evidence for distinct mechanisms of translational control by upstream open reading frames. *EMBO J* **20:** 6453–6463.

Giorgi C, Yeo GW, Stone ME, Katz DB, Burge C, Turrigiano G, Moore MJ. 2007. The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell* **130:** 179–191.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: Determinants beyond seed pairing. *Mol Cell* **27:** 91–105.

Hansen KD, Lareau LF, Blanchette M, Green RE, Meng Q, Rehwinkel J, Gallusser FL, Izaurralde E, Rio DC, Dudoit S, et al. 2009. Genome-wide identification of alternative splice forms down-regulated by nonsense-mediated mRNA decay in *Drosophila*. *PLoS Genet* **5:** e1000525.

Hogg JR, Goff SP. 2010. Upf1 senses 3′UTR length to potentiate mRNA decay. *Cell* **143:** 379–389.

Huang L, Lou CH, Chan W, Shum EY, Shao A, Stone E, Karam R, Song HW, Wilkinson MF. 2011. RNA homeostasis governed by cell type-specific and branched feedback loops acting on NMD. *Mol Cell* **43:** 950–961.

Hwang J, Maquat LE. 2011. Nonsense-mediated mRNA decay (NMD) in animal embryogenesis: To die or not to die, that is the question. *Curr Opin Genet Dev* **21:** 422–430.

Hwang J, Sato H, Tang Y, Matsuda D, Maquat LE. 2010. UPF1 association with the cap-binding protein, CBP80, promotes nonsense-mediated mRNA decay at two distinct steps. *Mol Cell* **39:** 396–409.

Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324:** 218–223.

Ingolia NT, Lareau LF, Weissman JS. 2011. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* **147:** 789–802.

Ivanova N, Dobrin R, Lu R, Kotenko I, Levorse J, DeCoste C, Schafer X, Lun Y, Lemischka IR. 2006. Dissecting self-renewal in stem cells with RNA interference. *Nature* **442:** 533–538.

Jepsen K, Hermanson O, Onami TM, Gleiberman AS, Lunyak V, McEvilly RJ, Kurokawa R, Kumar V, Liu F, Seto E, et al. 2000. Combinatorial roles of the nuclear receptor corepressor in transcription and development. *Cell* **102:** 753–763.

Johansson MJ, He F, Spatrick P, Li C, Jacobson A. 2007. Association of yeast Upf1p with direct substrates of the NMD pathway. *Proc Natl Acad Sci* **104:** 20872–20877.

Kervestin S, Jacobson A. 2012. NMD: A multifaceted response to premature translational termination. *Nat Rev Mol Cell Biol* **13:** 700–712.

Kim YK, Furic L, Desgroseillers L, Maquat LE. 2005. Mammalian Staufen1 recruits Upf1 to specific mRNA 3′UTRs so as to elicit mRNA decay. *Cell* **120:** 195–208.

Kim J, Chu J, Shen X, Wang J, Orkin SH. 2008. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132:** 1049–1061.

Kishore S, Jaskiewicz L, Burger L, Hausser J, Khorshid M, Zavolan M. 2011. A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins. *Nat Methods* **8:** 559–564.

Kurosaki T, Maquat LE. 2013. Rules that govern UPF1 binding to mRNA 3′ UTRs. *Proc Natl Acad Sci* **110:** 3357–3362.

Kuzmiak HA, Maquat LE. 2006. Applying nonsense-mediated mRNA decay research to the clinic: Progress and challenges. *Trends Mol Med* **12:** 306–316.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25.

Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE. 2007. Unproductive splicing of SR genes associated with highly conserved and ultraconserved DNA elements. *Nature* **446:** 926–929.

Larsson O, Sonenberg N, Nadon R. 2010. Identification of differential translation in genome wide studies. *Proc Natl Acad Sci* **107:** 21487–21492.

Le Hir H, Gatfield D, Izaurralde E, Moore MJ. 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.

Leeds P, Peltz SW, Jacobson A, Culbertson MR. 1991. The product of the yeast UPF1 gene is required for rapid turnover of mRNAs containing a premature translational termination codon. *Genes Dev* **5:** 2303–2314.

Lejeune F, Ishigaki Y, Li X, Maquat LE. 2002. The exon junction complex is detected on CBP80-bound but not eIF4E-bound mRNA in mammalian cells: Dynamics of mRNP remodeling. *EMBO J* **21:** 3536–3545.

Leung AK, Young AG, Bhutkar A, Zheng GX, Bosson AD, Nielsen CB, Sharp PA. 2011. Genome-wide identification of Ago2 binding sites from mouse embryonic stem cells with and without mature microRNAs. *Nat Struct Mol Biol* **18:** 237–244.

Linz B, Koloteva N, Vasilescu S, McCarthy JE. 1997. Disruption of ribosomal scanning on the 5′-untranslated region, and not restriction of translational initiation per se, modulates the stability of nonaberrant mRNAs in the yeast *Saccharomyces cerevisiae*. *J Biol Chem* **272:** 9131–9140.

Lu R, Yang A, Jin Y. 2011. Dual functions of T-box 3 (Tbx3) in the control of self-renewal and extraembryonic endoderm differentiation in mouse embryonic stem cells. *J Biol Chem* **286:** 8425–8436.

Martin KM, Metcalfe JC, Kemp PR. 2001. Expression of Klf9 and Klf13 in mouse development. *Mech Dev* **103:** 149–151.

McIlwain DR, Pan Q, Reilly PT, Elia AJ, McCracken S, Wakeham AC, Itie-Youten A, Blencowe BJ, Mak TW. 2010. Smg1 is required for embryogenesis and regulates diverse genes via alternative splicing coupled to nonsense-mediated mRNA decay. *Proc Natl Acad Sci* **107:** 12186–12191.

Medghalchi SM, Frischmeyer PA, Mendell JT, Kelly AG, Lawler AM, Dietz HC. 2001. Rent1, a *trans*-effector of nonsense-mediated mRNA decay, is essential for mammalian embryonic viability. *Hum Mol Genet* **10:** 99–105.

Melero R, Buchwald G, Castano R, Raabe M, Gil D, Lazaro M, Urlaub H, Conti E, Llorca O. 2012. The cryo-EM structure of the UPF-EJC complex shows UPF1 poised toward the RNA 3′ end. *Nat Struct Mol Biol* **19:** 498–505.

Mendell JT, Sharifi NA, Meyers JL, Martinez-Murillo F, Dietz HC. 2004. Nonsense surveillance regulates expression of diverse classes of mammalian transcripts and mutes genomic noise. *Nat Genet* **36:** 1073–1078.

Mort M, Ivanov D, Cooper DN, Chuzhanova NA. 2008. A meta-analysis of nonsense mutations causing human genetic disease. *Hum Mutat* **29:** 1037–1047.

Nagy E, Maquat LE. 1998. A rule for termination-codon position within intron-containing genes: When nonsense affects RNA abundance. *Trends Biochem Sci* **23:** 198–199.

Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O'Brien G, Shiue L, Clark TA, Blume JE, Ares M Jr. 2007. Ultraconserved elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev* **21:** 708–718.

Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37:** e123.

Ramani AK, Nelson AC, Kapranov P, Bell I, Gingeras TR, Fraser AG. 2009. High resolution transcriptome maps for wild-type and nonsense-mediated decay-defective *Caenorhabditis elegans*. *Genome Biol* **10:** R101.

Sauliere J, Murigneux V, Wang Z, Marquenet E, Barbosa I, Le Tonqueze O, Audic Y, Paillard L, Roest Crollius H, 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.

Schweingruber C, Rufener SC, Zund D, Yamashita A, Muhlemann O. 2013. Nonsense-mediated mRNA decay—mechanisms of substrate mRNA recognition and degradation in mammalian cells. *Biochim Biophys Acta* **1829:** 612–623.

Sievers C, Schlumpf T, Sawarkar R, Comoglio F, Paro R. 2012. Mixture models and wavelet transforms reveal high confidence RNA-protein interaction sites in MOV10 PAR-CLIP data. *Nucleic Acids Res* **40:** e160.

Singh G, Rebbapragada I, Lykke-Andersen J. 2008. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. *PLoS Biol* **6:** e111.

Singh G, Kucukural A, Cenik C, Leszyk JD, Shaffer SA, Weng Z, Moore MJ. 2012. The cellular EJC interactome reveals higher-order mRNP structure and an EJC-SR protein nexus. *Cell* **151:** 750–764.

Stockklausner C, Breit S, Neu-Yilik G, Echner N, Hentze MW, Kulozik AE, Gehring NH. 2006. The uORF-containing thrombopoietin mRNA escapes nonsense-mediated decay (NMD). *Nucleic Acids Res* **34:** 2355–2363.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **28:** 511–515.

Varsally W, Brogna S. 2012. UPF1 involvement in nuclear functions. *Biochem Soc Trans* **40:** 778–783.

Vattem KM, Wek RC. 2004. Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci* **101:** 11269–11274.

Vilella AJ, Severin J, Ureta-Vidal A, Heng L, Durbin R, Birney E. 2009. EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. *Genome Res* **19:** 327–335.

Wang Z, Tollervey J, Briese M, Turner D, Ule J. 2009. CLIP: Construction of cDNA libraries for high-throughput sequencing from RNAs cross-linked to proteins in vivo. *Methods* **48:** 287–293.

Wang D, Zavadil J, Martin L, Parisi F, Friedman E, Levy D, Harding H, Ron D, Gardner LB. 2011. Inhibition of nonsense-mediated RNA decay by the tumor microenvironment promotes tumorigenesis. *Mol Cell Biol* **31:** 3670–3680.

Wang ET, Cody NA, Jog S, Biancolella M, Wang TT, Treacy DJ, Luo S, Schroth GP, Housman DE, Reddy S, et al. 2012. Transcriptome-wide regulation of pre-mRNA splicing and mRNA localization by muscleblind proteins. *Cell* **150:** 710–724.

Weischenfeldt J, Damgaard I, Bryder D, Theilgaard-Monch K, Thoren LA, Nielsen FC, Jacobsen SE, Nerlov C, Porse BT. 2008. NMD is essential for hematopoietic stem and progenitor cells and for eliminating by-products of programmed DNA rearrangements. *Genes Dev* **22:** 1381–1396.

Weischenfeldt J, Waage J, Tian G, Zhao J, Damgaard I, Jakobsen JS, Kristiansen K, Krogh A, Wang J, Porse BT. 2012. Mammalian tissues defective in nonsense-mediated mRNA decay display highly aberrant splicing patterns. *Genome Biol* **13:** R35.

Yamashita A, Ohnishi T, Kashima I, Taya Y, Ohno S. 2001. Human SMG-1, a novel phosphatidylinositol 3-kinase-related protein kinase, associates with components of the mRNA surveillance complex and is involved in the regulation of nonsense-mediated mRNA decay. *Genes Dev* **15:** 2215–2228.

Yepiskoposyan H, Aeschimann F, Nilsson D, Okoniewski M, Muhlemann O. 2011. Autoregulation of the nonsense-mediated mRNA decay pathway in human cells. *RNA* **17:** 2108–2118.

Zhang X, Zhang J, Wang T, Esteban MA, Pei D. 2008. *Esrrb* activates *Oct4* transcription and sustains self-renewal and pluripotency in embryonic stem cells. *J Biol Chem* **283:** 35825–35833.

Zhao C, Datta S, Mandal P, Xu S, Hamilton T. 2010. Stress-sensitive regulation of IFRD1 mRNA decay is mediated by an upstream open reading frame. *J Biol Chem* **285:** 8552–8562.