Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay

Nicholas J McGlincy1,3, Lit-Yeen Tan1, Nicodeme Paul2, Mihaela Zavolan2, Kathryn S Lilley1, Christopher WJ Smith1*

Abstract

Background: In addition to acting as an RNA quality control pathway, nonsense-mediated mRNA decay (NMD) plays roles in regulating normal gene expression. In particular, the extent to which alternative splicing is coupled to NMD and the roles of NMD in regulating uORF containing transcripts have been a matter of debate.

Results: In order to achieve a greater understanding of NMD regulated gene expression we used 2D-DiGE proteomics technology to examine the changes in protein expression induced in HeLa cells by UPF1 knockdown. QPCR based validation of the corresponding mRNAs, in response to both UPF1 knockdown and cycloheximide treatment, identified 17 bona fide NMD targets. Most of these were associated with bioinformatically predicted NMD activating features, predominantly upstream open reading frames (uORFs). Strikingly, however, the majority of transcripts up-regulated by UPF1 knockdown were either insensitive to, or even down-regulated by, cycloheximide treatment. Furthermore, the mRNA abundance of several down-regulated proteins failed to change upon UPF1 knockdown, indicating that UPF1’s role in regulating mRNA and protein abundance is more complex than previously appreciated. Among the bona fide NMD targets, we identified a highly conserved AS-NMD event within the 3’ UTR of the HNRNPA2B1 gene. Overexpression of GFP tagged hnRNP A2 resulted in a decrease in endogenous hnRNP A2 and B1 mRNA with a concurrent increase in the NMD sensitive isoforms.

Conclusions: Despite the large number of changes in protein expression upon UPF1 knockdown, a relatively small fraction of them can be directly attributed to the action of NMD on the corresponding mRNA. From amongst these we have identified a conserved AS-NMD event within HNRNPA2B1 that appears to mediate autoregulation of HNRNPA2B1 expression levels.

Background

Nonsense-mediated mRNA decay (NMD) is one of a number of RNA surveillance pathways that help to ensure the fidelity of gene expression by degrading mRNAs that lack the proper arrangement of translational signals (reviewed in [1-4]). As the name suggests, NMD is responsible for recognizing and degrading mRNAs that contain premature termination codons (PTCs). In mammals, a termination codon is generally defined as premature by its spatial relationship to exon-junctions. The presence of one or more junctions at a distance of > 50-55 nucleotides downstream of the termination codon marks the mRNA for destruction [2,4]. The biochemical basis of this effect is an interaction between the exon junction complex (EJC), a large multi-protein complex that is deposited on the mRNA as a result of splicing, and the complex formed at the stop codon by the terminating ribosome during the first round of translation [1,2,4]. This interaction is mediated by the essential NMD factors UPF1, UPF2 and UPF3 (LIF-Frameshift suppressor, from their original identification in Saccharomyces cerevisiae [5]). Furthermore, in a number of metazoans, the phosphorylation state of UPF1 is regulated by the factors SMG1 and SMG5-7
(Suppressor with Morphological defects on the Genitalia, from their original identification in Caenorhabditis elegans [6]), which is required for NMD to take place [7]. Until recently the position of any downstream EJC’s was thought to be the primary determinant of a PTC in mammals. Recent studies, however, have shown that the distance from the PTC to various cues within the 3’ UTR (particularly the cytoplasmic poly-A binding protein PABP) can also play an important role in defining termination events as aberrant [3,8-13]. This is in a similar fashion to those organisms, such as S. cerevisiae, C. elegans and Drosophila melanogaster, where the EJC is absent or plays no role in NMD, and is thought to reflect a primordial mechanism of PTC recognition upon which, in mammals, the EJC has been superimposed [3,4,13].

Apart from its role in dealing with unintended errors in gene expression, NMD has a well-documented role in regulating the abundance of many physiological transcripts in all model organism examined to date [14-21]. Moreover, many NMD factors are now known to have additional functions extending beyond NMD (reviewed in [4]). The difference between species in those genes regulated by NMD is thought to be the cause of the differing phenotypes of animals in which Upf1 has been removed [16,19,22]. Amongst these, Mus musculus lacking Upf1 are embryonic lethal [23], indicating that Upf1, and presumably NMD, plays an important role in mammalian physiology and development. NMD-regulated transcripts can be divided into two broad categories. First those mRNAs that “normally” possess a PTC. These include transcripts that contain upstream open reading frames (uORF) within their 5’ UTR, or in which a PTC is introduced as the result of a regulated alternative splicing event (AS-NMD), including those transcripts with an intron more than 50-55 nt into the 3’ UTR such that the coding sequence (CDS) termination codon appears premature [15,16,24-26]. Secondly, those mRNAs where NMD is co-opted as a decay mechanism through the interaction of UPF1 with a protein that recognizes a specific set of mRNAs. Two examples are staufen-1 (STAU1) mediated decay (SMD) and the decay of certain replication-dependent histone mRNAs at the end of S-phase of the mammalian cell cycle [27-29]. In both cases a protein recognizes a specific cis-element within the 3’ UTR and also interacts with UPF1 [27-29]. mRNA degradation is then triggered in a fashion that is dependent on UPF1 and active translation, but independent of the other UPF proteins [27-29].

Previous large-scale investigations into the role of UPF1/NMD in regulating physiological gene expression in metazoans have focused on changes in mRNA abundance [15,20,21,29-34]. In this study we have sought to deepen our understanding of the role of UPF1 in regulating physiological gene expression by examining the changes in protein expression in response to siRNA mediated depletion of UPF1 in HeLa cells, using the expression proteomics technique 2D difference gel electrophoresis (2D-DiGE). We observed a large number of alterations in protein levels; both increases and decreases. By analyzing the levels of the corresponding mRNAs after treatment with either siRNAs against UPF1 or the translation inhibitor cycloheximide we were able to identify a small group of bona fide NMD targets; indicating that UPF1’s role in regulated gene expression may be more limited than previously thought. However, the majority of these bona fide NMD targets contained recognizable NMD-activating features, such as splicing dependent PTCS, introns in the 3’ UTR and uORFs. From among these we identified a highly conserved AS-NMD event with the 3’ UTR of the HNRNPA2B1 gene, which appears to be involved in the autoregulation of HNRNPA2B1 mRNA levels.

**Results**

**mRNA stabilized as a result of the inhibition of NMD is translated to yield protein**

Before embarking upon a proteomic analysis of the consequences of UPF1 knockdown we carried out a proof of principle experiment to demonstrate that mRNAs stabilized by inhibition of NMD could be translated to yield protein. To this end we constructed the pGFPint reporter plasmid, containing an efficiently spliced artificial intron (based on α-tropomyosin exons 2 and 3 and a 111nt intron from β-globin [35]) in the 3’ UTR. Splicing of the intron creates an exon-exon junction 105 nucleotides downstream of the GFP stop codon, making it appear premature and hence the mRNA NMD sensitive. A cell line stably expressing pGFPint was constructed and subjected to knockdown of NMD factors UPF1, UPF2 and SMG1 or transfection with a control siRNA, C2. Depletion of UPF1 and UPF2 protein was achieved to levels less than 12.5% of control, as estimated by western blot (Figure 1A). In the absence of an antibody against SMG1 the reduction in SMG1 mRNA was measured by QPCR. A reduction to levels approximately 35% of control was achieved (Figure 1B). Both GFP mRNA (Figure 1D, left) and protein (Figure 1C) showed a large increase in response to UPF1 knockdown, a much smaller increase in response to UPF2 knockdown and an intermediate response to SMG1 knockdown. Analysis of pGFPint mRNA by RT-PCR indicated that the artificial intron was entirely spliced (data not shown).

To further examine the differing response of pGFPint to the knockdown of different NMD factors, the levels of the NMD sensitive isoforms of SC35 (SFRS2) were examined by QPCR (1.6 and 1.7 kb [36], Figure 1D,
Figure 1 Knockdown of UPF1, UPF2 and SMG1 results in the production of GFP protein by cells stably expressing pGFPint. A Upper panel: western blot of 10 μg total cell protein for UPF1 and actin as a loading control. Lower panel: western blot of 30 μg total cell protein for UPF2 and actin as a loading control. In each case the C2 treated sample was diluted 1:2, 1:4 and 1:8 in RIPA buffer in order to better estimate the degree of knockdown in the knockdown sample. Representative samples of three biological repeats are shown. B Histogram comparing the fold change in SMG1 mRNA levels in response to treatment with C2 (black bar) or SMG1_A (light grey bar) siRNAs, as measured by QPCR on parallel RNA samples. The height of each bar represents the mean of three biological repeats, while error bars represent the standard error of the mean (SEM). C Western blot of 10 μg total cell protein for GFP and actin as a loading control. Representative samples of three biological repeats for each siRNA are shown. D Histogram comparing the fold change in mRNA levels for GFP and the NMD sensitive isoforms of SC35 (1.6 kb and 1.7 kb, [36]) in response to treatment with each siRNA, as measured by QPCR on parallel RNA samples. The height of each bar represents the mean of three biological repeats, while error bars represent SEM. The colour scheme is indicated in the side panel. E Schematic of AS-NMD events within the 3' UTR of SC35 (SFRS2). Dark boxes represent exons, and white boxes introns. Dashed lines denote alternative splicing patterns and arrows denote the QPCR primers used in D.
right. Schematics of events in Figure 1E). Interestingly, the pattern of changes in the SC35 isoforms mirrored that of pGFPint mRNA: UPF1 knockdown provoked the largest fold change followed by SMG1 and then UPF2. Given the high degree of knockdown achieved for all the factors examined, it appears that NMD of pGFPint and SC35 1.6 and 1.7 kb mRNA is differentially sensitive to the knockdown of UPF1, UPF2 or SMG1. This phenomenon has been described previously and is thought to reflect distinct branches of the NMD pathway with differential requirements for UPF2 [37]. Since the inception of this work two further reports of similar NMD reporter systems support our conclusion that mRNAs stabilized as a result of NMD inhibition are active substrates for translation [38,39]. Indeed, this idea is also borne out by published data from more physiological circumstances [40-42].

A multi-gel 2D-DiGE study of UPF1 knockdown in HeLa cells reveals numerous changes in protein expression

To identify global changes in protein expression in response to UPF1 knockdown we performed a multi-gel 2D-DiGE study of HeLa cells in which UPF1 had been depleted by RNA interference (RNAi). Samples were harvested 48 h after the second siRNA hit on the basis of the PCA, others were picked that had changed significantly in response to both siRNAs, although one change or the other may not achieve significance in the univariate sense. In order to identify which proteins were responsible for the observed increases in expression, and hence are candidate UPF1 targets, 85 of the protein spots that had increased in expression were excised for identification by mass spectrometry (Additional file 1). Spots were picked first from those that had changed significantly in response to both Upf1_A and Upf1_B. Then, on the basis of the PCA, others were picked that had changed significantly in response to either siRNA, starting with those that had narrowly escaped significance with the second siRNA. 17 down-regulated spots were also picked, from those that had decreased significantly in response to both siRNAs (Additional file 1). Of the 85 up-regulated spots, 58 yielded sufficient material to allow protein identification, whereas of the 17 down-regulated spots, 13 yielded sufficient material. For each spot, potential protein constituents were identified from liquid chromatography-tandem mass spectrometry of tryptic peptides produced by in-gel digestion, using the MASCOT search engine [47]. As a result, 128 unique proteins were identified from up-regulated spots and 21
Figure 2 Changes in protein spot expression in response to UPF1 knockdown. A. Representative western blot of 10 μg total cell protein for UPF1 and actin as a loading control for each sample. C2 treated sample was diluted 1:2, 1:4 and 1:8 in ASB14 buffer in order to estimate the degree of knockdown.

B. Representative U2AF35 RT-PCR assay on parallel RNA samples. The adjoining cartoon illustrates the identity of each band. The lower band results from alternative inclusion of one of a pair of normally mutually exclusive exons of equal size, termed E3 (yielding isoform U2AF35a) and EAb (yielding isoform U2AF35b) [45]. The upper band represents inclusion of both exons, which results in a frameshift that creates a PTC - making isoform U2AF35c NMD sensitive [45]. M: 1 kb plus marker (GE healthcare). RT-: addition of RT performed without reverse transcriptase. PCR-: PCR performed without template. Underlying numbers indicate the percentage of the signal from both bands represented by the upper, double-included NMD-sensitive, band for each siRNA treatment (mean of 6 biological replicates ± SEM).

C. Proportional Venn diagrams representing the number of protein spot changes unique and common to each siRNA against UPF1. The upper, red coloured, diagram details upward changes whereas the lower, blue coloured diagram details downward changes.

D. Principal component analysis (PCA) scores plot illustrating the similarity of the multi-gel study samples to each other by their relationships with the first two principal components (PCs) describing the whole multi-gel study dataset. Blue squares - C2 treated samples; red circles - Upf1_A treated samples, green triangles - Upf1_B treated samples. t[1] - score relating to PC1; t[2] - score relating to PC2.
from down-regulated spots. Some proteins were identified in more than one spot, including 6 that were found in both up- and down-regulated spots. This could be due to post-translational modification of the proteins, or the existence of isoforms that are different in size or pI, but are indistinguishable by their tryptic peptide pattern. Information on the identified proteins and peptide sequences, along with the change observed for each spot is detailed in Additional file 2.

Validation of NMD targeted mRNAs
In order to determine which of the identified proteins represented bona fide NMD targets, two rounds of QPCR validation were employed. In the first round we tested whether levels of the mRNAs corresponding to the identified proteins were affected by UPF1 knockdown mediated by Upf1_A siRNA. Since NMD and the other UPF1-dependent mRNA decay pathways (SMD and histone mRNA decay) are dependent on active translation [27,28,48], we next measured changes in mRNA levels following treatment of cells with the translation inhibitor cycloheximide. For each gene of interest QPCR primers were designed to mRNA regions not known to undergo alternative splicing, and eight replicate samples were used in order to achieve the same statistical power as the 2D-DiGE multi-gel study [49,50]. Only genes that showed statistically significant increases (p < 0.05, Student's t-test, one-tail) in response to both treatments were deemed to be genuine targets of NMD/UPF1-dependent mRNA decay.

The results of the QPCR validation are detailed in Additional file 2 and summarised in Figure 3. Of the 128 mRNAs corresponding to proteins from up-regulated spots, 47 (37%) increased in response to UPF1 knockdown, as expected of UPF1/NMD targets. A further 62 (48%) showed no significant change in expression, 12 decreased (9%), while 8 (6%) failed to produce an intelligible signal. Strikingly, from the 17 down-regulated protein spots none of the 21 corresponding mRNAs was down-regulated upon UPF1 knockdown. Eleven did not change significantly in expression, 4 failed to show an intelligible signal, while the remaining 6 actually increased in expression (Additional file 2). While 3 of these 6 proteins were among those also found in up-regulated spots, it is striking that not a single one of the genes corresponding to the protein constituents of down-regulated spots showed an accompanying decrease in mRNA expression. This suggests that UPF1 knockdown has a negative role in the translational efficiency of at least some of these proteins.

The 47 validated UPF1 target genes from up-regulated spots were then subjected to the second round of validation by cycloheximide treatment. Of the 47 UPF1 targets, 17 (34%) increased in expression, as expected of authentic NMD targets (IMMT, GOLGA4, ID12, TH1L, PLG, KIAA1529, CCT8, TUBA8, SFRS7, MAGEA1, CDC6, SERPINB1, NAPA, HNRNPA2B1, VAPB, RAB1A, RAB35). Of the remainder, one (CANX) failed to give an intelligible signal, 10 (21%) showed no significant change in expression, while 19 (40%) actually decreased in expression. The 30 genes that did not increase in expression in response to cycloheximide treatment may represent secondary effects of UPF1 knockdown or targets of UPF1 in processes other than NMD.

With reference to the original 2D DiGE analysis, of the 58 upregulated spots that yielded protein identifications, 34 (59%) contained at least one protein whose corresponding mRNA was validated as being upregulated by knockdown of UPF1, and in 26 cases (45%) the upregulated mRNA corresponded to the major protein constituent of the spot, as indicated by MASCOT scores (see Additional file 2). Of the 34 spots, 18 (i.e. 31% of the total) contained proteins that were also upregulated by cycloheximide treatment. In five cases (CCT8, SERPINB1, NAPA, HNRNPA2B1, RAB1A) the doubly validated NMD target corresponded to the major protein constituent of the protein spot, as judged by MASCOT scores, and so provides a clear explanation for the observed upregulation. A number of the other NMD targets were among the more abundant components of the spots in which they were identified (e.g. IMMT, VAPB, TUBA8, RAB35). In the remaining cases (e.g. GOLGA4, ID12, TH1L, PLG, SFRS7, MAGEA1, CDC6, KIAA1529), the doubly validated NMD targets represented relatively minor constituents of their resident spots. In four of these cases the spot also contained a more abundant validated UPF1 target that was not cycloheximide upregulated.

The remaining 24 spots (41% of the total) contained no protein whose mRNA was upregulated by UPF1 knockdown. This suggests that relatively few of the changes in protein expression upon UPF1 knockdown can be attributed directly to UPF1’s role in mRNA decay. This result is in broad agreement with the more extensively validated studies of UPF1’s role in physiological gene expression [20,21,34,51]. (discussed below)

Identification of NMD activating features
We next sought to identify possible NMD activating features within the doubly validated genes. To this end, maximum transcript alignments [52] of corresponding Unigene clusters [53] were generated using SPA [54] and examined for instances of: i) introns more than 50 nt downstream of the termination codon of the largest ORF within a transcript, as an indication of AS-NMD. ii) The presence of ORFs upstream of the largest ORF within a maximum transcript, as an indication of the
presence of uORFs within the 5′ UTR. Potential NMD sensitive maximum transcripts were then scored according to the number of peptides identified by mass spectrometry that were present within the protein sequence encoded by the largest ORF of the maximum transcript. Only transcripts encoding all of the identified peptides were considered as being potentially responsible for the observed upregulation of protein spots. For some genes, this step eliminated many potential NMD substrates. For example, the SFRS7 gene encodes the SR splicing regulatory protein 9G8, which has a well characterized AS-NMD event involving a “poison” cassette exon [55]. In this study however, the 9G8 peptides identified by mass spectrometry are not encoded by the alternatively spliced PTC-containing isoform. In contrast, many of the SFRS7 maximum transcripts contain one or two additional short uORFs upstream of the main protein coding ORF.
Of the 17 doubly validated genes, we found evidence of at least one NMD feature in 12 cases. Notably, all five doubly validated genes that constituted the major components of their resident spots (CCT8, SERPINB1, NAPA, HNRNPA2B1, RAB1A) had at least one NMD feature that was consistent with the peptide data. The most common features, found in 11 cases, were uORFs. In addition, there were three genes (TH1L, NAPA and HNRNPA2B1) with one or more alternative splicing events at a sufficient distance downstream of the main ORF to activate NMD.

**AS-NMD events in TH1L and NAPA**

We next tested each of the AS-NMD predictions by RT-PCR or QPCR. TH1L was a relatively minor constituent of spot 831, which was up-regulated 1.39 fold in response to UPF1 knockdown (Additional file 2). TH1L mRNA was up-regulated approximately 1.4 fold in response to UPF1 knockdown and to a similar extent by cycloheximide (Figure 4A). TH1L is predicted to contain three AS-NMD events: first, a 101 nt intron within its 3′ UTR, which when spliced causes the normal stop codon to appear premature (Figure 4B upper panel). Secondly, retention of the intron between exons 13 and 14. Thirdly, the use of an alternative 5′ splice site within exon 13, resulting in a frameshift that creates a PTC (Figure 4C upper panel). The first two events are consistent with the peptide data, while one of the peptides lay downstream of the PTC introduced by use of the internal 5′ splice site on exon 13, thus ruling out the latter event as a contributor to the upregulation of protein spot 831. RT-PCR was performed with primers flanking each AS-NMD event to determine whether the PTC containing isoform was stabilised as a result of UPF1 knockdown (Figure 4B and 4C, lower panel). The 3′UTR splicing event and the exon 13 alternative 5′ splice site event both showed a significant 2-3 fold increase in the proportion of the NMD sensitive isoform in response to UPF1 knockdown, to a level of approximately 11% (Figure 4B and 4C, lower panel). In contrast, no effect of UPF1 knockdown was seen upon the intron 13 retention event (data not shown). Taken together, the peptide and RT-PCR data therefore suggest that the 3′UTR intron is the NMD feature responsible for the observed upregulation of TH1L protein, but that both events contributed to upregulation of its mRNA upon UPF1 knockdown.

NAPA was the major component of spot 1997, which was upregulated ~1.3 fold by UPF1 knockdown (Additional file 2). Likewise, NAPA mRNA was upregulated 1.35 - 1.4 fold by UPF1 knockdown and cycloheximide (Figure 5A). NAPA has two predicted NMD features that are consistent with the 9 peptides that identified it. An alternatively spliced intron 53 nt into its 3′UTR is at the threshold distance for inducing NMD (Figure 5). In addition, the 5′UTR of a maximum transcript has overlapping 9 and 3 codon uORFs. QPCR was carried out to analyze levels of the isoforms in which the 3′UTR intron was retained or spliced. Levels of NAPA mRNA with the intron spliced out (denoted junction b) were elevated by ~2.5 fold after UPF1 knockdown, whereas transcripts with the intron retained were not significantly affected (Figure 5B), suggesting that the 3′UTR intron is the feature responsible for the observed upregulation of the NAPA containing spot.

**AS-NMD mediated autoregulation of HNRNPA2B1**

HNRNPA2B1 was the most abundant protein in spot 2105, which was upregulated 1.3 - 1.7 fold by the two UPF1 siRNAs (Additional file 2). HNRNPA2B1 mRNA showed an approximate two-fold increase upon UPF1 knockdown and a small but significant increase upon cycloheximide treatment (Figure 6A and Additional file 2). Bioinformatic analysis indicated that HNRNPA2B1 contains extensive alternative splicing within its 3′ UTR that would cause the normal stop codon to appear premature (Figure 6B). As an NMD feature this would be consistent with the 5 peptides that identified HNRNPA2B1. The predicted UTR structure is, however, in conflict with the Refseq annotation. In order to confirm the existence of the predicted NMD sensitive isoforms of the 3′ UTR, we performed 3′ RACE (Rapid Amplification of cDNA Ends) for HNRNPA2B1 on mRNA taken from the UPF1 knockdown RNA samples used for validation. The resulting sequences were aligned to the genome using BLAT [56]. The RACE tags clearly support the 3′ UTR structure predicted from the unigene cluster; comprising three additional exons after that in which the HNRNPA2B1 CDS ends (Figure 6B, Additional file 3). Splicing of the final intron or inclusion of the first additional exon would create an exon-exon junction at sufficient distance downstream to make the normal stop codon appear premature. Splicing of the final intron was examined by exon-junction specific QPCR and found to be up-regulated by 3.5-fold upon UPF1 knockdown (junction b, Figure 6B and 6C), whereas a junction in the Refseq mRNA expected not to be NMD sensitive (denoted junction a, Figure 6B) was not upregulated (Figure 6C). The whole area of the 3′ UTR is very highly conserved (Figure 6B), suggesting that AS-NMD might be important in the regulation of HNRNPA2B1 levels. Indeed full-length mRNAs and ESTs representing similar UTR structures can also be observed in the orangutan (Pongo pygmaeus abelii), other mammals (M. musculus and Bos taurus) and even the chicken (Gallus gallus). While 3′ UTRs have many roles in regulating gene expression, any of which may
Figure 4 AS-NMD within the TH1L gene. A Histogram representing QPCR validation results for TH1L. Bars represent mean fold change in mRNA levels in response to either UPF1 knockdown (left panel, N = 8) or cycloheximide treatment (right panel, N = 8) ± SEM. P-value summary (Student's t test, one tail): * p < 0.05, ** p < 0.01, *** p < 0.001. QPCR primers were located upstream of the schematic shown in B, in a region expected to be unaffected by alternative splicing. B Upper panel: schematic of the AS-NMD event within 3' UTR of TH1L produced using the UCSC genome browser [114], running 5' left to 3' right. Boxes represent exons while chevroned lines represent introns. The upper blue cartoon indicates the Refseq annotated 3' UTR structure, the thinning of the box indicating the end of the protein coding sequence. The underlying cartoons are Genbank mRNAs illustrating the AS isoforms predicted by our analysis, the red mRNA represents the NMD sensitive isoform - a retained intron is spliced to make the normal stop codon appear premature. The lower blue histogram represents conservation across 17 vertebrate species as calculated by [115]. Lower panel: RT-PCR of the AS-NMD event, illustrating the effect of UPF1 knockdown. M: 1 kb plus marker (GE healthcare). RT-: addition of RT performed without reverse transcriptase. PCR-: PCR performed without template. Underlying numbers indicate the mean (± SEM, N = 3) percentage of the total signal from both bands represented by the NMD-sensitive for each siRNA treatment.

C UCSC genome browser schematic and RT-PCR of a second predicted event within TH1L. Use of an alternative 5' splice site results in a frameshift creating a downstream PTC. The arrangement of elements is the same as for section B.
result in the observed conservation [57], many other genes of the hnRNP and SR protein families have been identified as containing such highly conserved AS-NMD events [32,33,55,58-60]. Indeed, murine Hnrnpa2b1 was also identified as containing AS-NMD by Ni et al. [32], although the event was not the same as reported here (M. Ares, personal communication). Many such conserved AS-NMD events have been shown to be subject to autoregulation by their own protein levels [32,36,58,60-66]. We examined whether this is also the case for HNRNPA2B1. FLAG tagged HNRNPA2 isoform or an empty expression vector was transfected into HeLa cells in combination with a GFP expressing plasmid. GFP expressing cells were then isolated by flow cytometry in order to enrich for cells possessing the cotransfected plasmid. Expression of FLAG-HNRNPA2 in the sorted cells was confirmed by western blot (Figure 6D). The expression and splicing of HNRNPA2B1 was examined in parallel RNA samples, using primers specific for the endogenous transcripts. Overexpression of FLAG-HNRNPA2 reduced HNRNPA2 and HNRNPB1 mRNA levels to 75-80% of control (Figure 6E), demonstrating that HNRNPA2 protein can regulate HNRNPA2B1 mRNA levels. Examining the effect of overexpression on UTR junctions a and b indicated that non-NMD sensitive junction a was decreased in expression to approximately 60% of control, while NMD sensitive junction b was increasing in expression by 20% over

**Figure 5 AS-NMD within the NAPA gene.** A. Histogram representing QPCR validation results for NAPA. Bars represent mean fold change in mRNA levels in response to either UPF1 knockdown (left panel, N = 8) or cycloheximide treatment (right panel, N = 8) ± SEM. p-value summary (Student’s t test, one tail): * p < 0.05, ** p < 0.01, *** p < 0.001. QPCR primers were located upstream of the schematic shown in B, in a region expected to be unaffected by alternative splicing. B. Schematic of the AS-NMD event within 3’ UTR of NAPA produced using the UCSC genome browser [114]. Boxes represent exons while lines with chevrons represent introns. The upper blue cartoon indicates the Refseq annotated 3’ UTR structure, the thinning of the box indicating the end of the protein coding sequence. The underlying cartoons are Genbank mRNAs illustrating the AS isoforms predicted by our analysis, the red mRNA represents the NMD sensitive isoform - a retained intron is spliced to make the normal stop codon appear premature. The lower blue histogram represents conservation across 17 vertebrate species as calculated by [115]. C. Histogram representing QPCR validation results of the predicted AS-NMD. Bars represent mean fold change in of the detailed exon junction in response to either UPF1 knockdown (± SEM, N = 3). p-value summary (Student’s t test, one tail): * p < 0.05, ** p < 0.01, *** p < 0.001.
Figure 6  AS-NMD within the HNRNPA2B1 gene.  

A. Histogram representing QPCR validation results for HNRNPA2B1. Bars represent mean fold change in mRNA levels in response to either UPF1 knockdown (left panel, N = 3 in this instance) or cycloheximide treatment (right panel, N = 8) ± SEM. p-value summary (Student’s t test, one tail): * p < 0.05, ** p < 0.01, *** p < 0.001. QPCR primers were located upstream of the schematic shown in B, in a region expected to be unaffected by alternative splicing.  

B. Schematic of the 3’ UTR of HNRNPA2B1 produced using the UCSC genome browser [114], running 5’ right to 3’ left as indicated by the upper arrow. The upper black cartoon represents the prediction of the exonic structure (boxes) predicted by our bioinformatic analysis. The underlying red cartoons represent the highest scoring BLAT alignments [56] of several of the RACE sequences obtained in our analysis. Underlying these are cartoons representing the UCSC prediction of 3’ UTR and the Refseq annotated UTR. The lower blue histogram represents conservation across 17 vertebrate species as calculated by [115]. The small UTR intron in the Refseq mRNA is too close to the stop codon for its splicing to make the stop codon appear premature (junction a, denoted by the green line). Splicing of the final intron in the UTR (junction b, denoted by the green line) would be expected to render all incumbent isoforms NMD sensitive.  

C. Histogram representing QPCR results for exon junctions a and b. Bars represent mean fold change (N = 3, ± SEM) in mRNA levels in response to UPF1 knockdown.  

D. Western blot of GFP positive HeLa cells sorted by flow cytometry. HeLa cells were co-transfected with a GFP expressing plasmid and either FLAG-tagged hnRNP A2 or an empty vector control. ERK1 was used as a loading control.  

E. Histogram representing QPCR results from parallel RNA samples to D. Bars represent mean fold change (N = 3, SEM) of HNRNPA2 and B1 in response to hnRNPA2 overexpression. p-value summary (Student’s t test, two tails): * p < 0.05, ** p < 0.01, *** p < 0.001.  

F. Histogram representing QPCR results from parallel RNA samples to D. Bars represent mean fold change (N = 3, SEM) of UTR junctions a and b in response to hnRNPA2 overexpression. p-value summary (Student’s t test, two tails): * p < 0.05, ** p < 0.01, *** p < 0.001.
control levels (Figure 6F). These results are consistent with HNRNP2B1 regulating the abundance of its own mRNA by altering splicing of the 3’ UTR to promote the production of isoforms degraded by NMD.

Discussion

The role of UPF1-dependent mRNA decay in regulating physiological gene expression

We have demonstrated that the knockdown of UPF1 in HeLa cells results in a large number of changes in protein expression but that only a minority of these can be attributed to UPF1’s characterized roles in NMD and other translation dependent mRNA decay pathways. This result is in broad agreement with the more extensively validated studies of UPF1’s role in physiological gene expression [20,21,34,51], suggesting that NMD plays a more restricted role in regulating gene expression than previously claimed [67,68]. For example, in order to measure the efficiency of NMD in cell lines expressing different amounts of RNPS1, Viegas and colleagues sought to identify genuine NMD target genes from genes identified by microarray as upregulated upon UPF1 knockdown [51]. They examined both the abundance of the fully processed mRNA and pre-mRNA, finding that in the majority of cases (14/16) both were upregulated. This indicates that the increase in mRNA levels was likely to be a transcriptional effect rather than due to the direct action of NMD [51].

The majority of the 17 confirmed NMD targets (defined as UPF1 and translation-dependent) have not been previously identified by microarray studies of UPF1 dependent gene expression in mammals [15,29-32,51,69], indicating the complementarity of proteomic and transcriptomic analyses. However, given the bias of 2D-DiGE towards higher abundance proteins, it may be that this mechanism is more pertinent for genes with lower absolute expression levels. Recent findings have indicated that the distance between the stop codon and poly(A) tail can also be an important determinant of NMD (reviewed in [3]). However, we were unable to find an over-representation of greater 3’ UTR lengths within our set of doubly validated genes (data not shown). We believe that it is relevant that all of the identified NMD activating features within the 17 doubly validated NMD targets were either uORFs or introns positioned in the 3’UTR or towards the 3’ end of the transcript. As a result the detected proteins would all represent regular full-length isoforms, and would be expected to be stable. The protein products produced from other stabilized AS-NMD substrate mRNAs might be inherently unstable due to their truncated nature. This would argue against the functional roles often proposed for these products [70,71]. In support of this, we observed that the potential truncated PTB isoform, encoded by the AS-NMD targeted mRNA in which exon 11 is skipped [58], is not detectably expressed from cDNA expression vectors (in which there is no NMD activating feature) unless proteasome activity is inhibited by the proteasomal inhibitor MG132 (NJM and CWJS, unpublished observation). This suggests that the inherent instability of some protein products might preclude their identification as NMD targets by proteomic approaches. Nevertheless, a pilot DiGE experiment did not reveal a noticeable increase in upregulated spots when UPF1 knockdown was analyzed under conditions of proteasome inhibition by MG-132 (NJM and CWJS, unpublished observation), suggesting that our current investigation has not suffered substantially from this potentially confounding effect.

In addition to identifying genuine NMD targets, our validation strategy also highlighted a number of other interesting groups. First, those proteins whose mRNA was not up-regulated by UPF1 knockdown may represent targets of nonsense mediated translational repression (NMTR), wherein PTC containing mRNAs apparently escape NMD but do not produce detectable levels of protein [72,73]. Secondly, those genes that were upregulated by UPF1 knockdown but not upregulated by cycloheximide treatment. An interesting feature of these 19 unusual UPF1 targets is an enrichment of heat shock proteins of the hsp90, hsp70 and hsp60 classes (HSP90AA1, HSP90AB1, HSPA1A, HSPD1, AHSA1). In S. cerevisiae a specialized mRNA decay pathway, termed initiation dependent decay, operates under conditions where translation initiation is down-regulated but not abolished [74,75]. Of particular interest, initiation dependent decay targets a variety of heat shock proteins, including hsp70 and 90, and is dependent upon UPF1 and UPF2. While our experiments were not carried out under conditions where translation initiation is expected to be impaired, they suggest that initiation dependent decay may be relevant to mammalian systems. However, this observation could also be the result of a shared transcriptional activator of these heat-shock proteins being NMD sensitive, but generally translationally down-regulated by cyclohexamide. Thirdly, analysis of the spots that were down-regulated upon UPF1 knockdown indicate yet another potentially interesting group of UPF1 targets. None of the mRNAs corresponding to the protein constituents of these spots were down-regulated by UPF1 knockdown; indeed some were actually upregulated. This suggests that UPF1 may play a role in the synthesis or degradation of these proteins. Given UPF1’s known roles, it seems more likely that it acts at the translational level and that for these proteins UPF1 has a positive influence on their translation.
uORFs as NMD features

The most prevalent NMD activating feature predicted for the doubly validated NMD targets was the presence of uORFs. Indeed, two of our confirmed cases that can be explained by uORFs have previously been identified as AS-NMD targets. SFRS7 (also known as 9G8) is already a well-characterised example of AS-NMD [55,62,76], our finding that some isoforms may also possess a uORF indicate that SFRS7′s post-transcriptional regulation may be more complex than previously appreciated, as recently described for SFRS1 (SF2/ASF) [66]. Furthermore, a CCT8 alternative isoform was reported to be sensitive to the translation inhibitor emetine by AS-sensitive microarray [32].

A priori it might be expected that all uORF containing mRNAs would be NMD sensitive. The termination codon of the uORF(s) would be premature with regard to the exon-exon junctions within the CDS, and would likely be at great distance from signals in the 3′ UTR determining proper translation termination. This, however, does not appear to be the case as uORFs often mediate translational repression of the protein coding ORF without an accompanying decrease in mRNA abundance [26,77]. Moreover, inhibition of NMD in C. elegans results in the upregulation of only some transcripts bearing uORFs [21]. So what characteristics make a uORF baring transcript NMD sensitive? uORF length has been shown to be important; short open reading frames or those that have been translated more quickly favour translation reinitiation downstream [11,78,79], possibly by the remaining association of translation initiation factors with the ribosome [80,81]. Reinitiation then inhibits NMD [11,78,79]. So NMD sensitive uORFs might be expected to be longer than average, or composed of sequence that results in slow translation. Ramani et al. (2009) found a weak correlation between the Kozak consensus at the start codon of the protein coding ORF and the extent to which the uORF containing transcript was up-regulated in NMD deficient C. elegans [21]. Calvo and colleagues examined the effect of various uORF characteristics on reduction in protein expression. They found that uORF containing 5′ UTRs, when examined in a heterologous system, generally exerted a greater effect on protein abundance (average 58% decrease) than mRNA abundance (average 5% decrease). Similar trends were observed in published datasets they examined. The decreasing protein expression correlated with stronger context at the uAUG, a greater cap-uORF distance (both in the case of a single uORF) and, to a lesser extent, an increasing number of uORFs. In the case of a single uORF, uORF length and uORF-CDS distance did not correlate with the extent of protein repression. The absolute amplitude of changes in mRNA abundance are similar to those observed for our predicted uORF carrying genes. Unfortunately the relationship of these variables with decreasing mRNA abundance, which we might presume to be due to NMD, was not examined [82].

We have been unable to find any characteristics that distinguish the uORFs in our doubly validated NMD targets from others (data not shown). However, uORF mediated regulation is often complex, involving not only the uORFs themselves but also interactions with other conserved sequences and trans-acting factors [26,83,84]. Indeed, the relatively small fold changes in mRNA level observed during validation indicate that in each case only a proportion of the mRNA is NMD-sensitive, suggesting that some sort of probabilistic event is responsible. Thus, it is not clear that identification of a uORF is currently a powerful predictor of NMD sensitivity. A larger dataset and broader scope of analysis may prove more fruitful in finding some association between specific uORF characteristics, or groups thereof, and NMD-sensitivity.

Autoregulation of HNRNPA2B1 via AS-NMD

We have identified a highly conserved example of autoregulatory AS-NMD within the HNRNPA2B1 gene. The hnRNP family of proteins plays many roles in RNA metabolism [85,86]. HNRNPA2B1 itself has been shown to regulate both alternative splicing [87] and mRNA stability [88]. Our data is consistent with HNRNPA2B1 activating splicing within it’s own UTR to produce NMD sensitive forms that account for the decrease in total gene expression: over-expression of the HNRNPA2 isoform resulted in down-regulation of A2 and B1 isoform expression and the up-regulation of one of the NMD-sensitive exon junctions within the 3′ UTR (Figure 6). For simplicity we measured two UTR exon-exon junctions that would always be expected to be NMD (in)sensitive, but the splicing patterns involved are almost certainly more complex. While hnRNP A/B proteins were initially characterised as splicing repressors [86,89,90], it has recently been shown that intronic binding of hnRNP A1/A2 proteins can activate splicing, particularly of elongated introns [87,91]. This is thought to proceed by homophilic interactions between A2/A1 proteins bound at separate sites (or indeed heterophilic interactions between A1/A2 proteins and hnRNP F/H proteins) causing the looping out of portions of intron, which in turn results in promoting the splicing of said intron [87,92]. We have not examined whether HNRNPA2B1 directly binds to its own UTR, but motifs thought to represent bindings sites for HNRNPA1 and HNRNPA2B1 (taken from [91]) are clearly present in the HNRNPA2B1 3′ UTR introns we identified (Additional file 3). This also raises the possibility that HNRNPA1 may also regulate these splicing
events. Indeed increased expression of HNRNPA1 has been observed to correlate with decreased HNRNPA2/B1 in some cancer cell lines [93].

AS-NMD within hnRNP and SR protein genes has been shown to mediate quantitative regulation by repressing protein expression at inappropriate times [63,94-96] or providing homeostatic regulation of protein levels through autoregulatory negative feedback [32,36,58,60-66]. Our data is consistent with the later case. However, these two modes of regulation are not mutually exclusive, and it will be interesting to see whether biological circumstances can be identified where this AS-NMD event is used to repress expression of HNRNPA2B1. Indeed, up-regulation of HNRNPA2B1 levels has pathological associations. Increased HNRNPA2 expression has also been observed in pancreas and breast cancer [97,98], and in the foetal brain of Down’s syndrome patients [99]. Underscoring the functional relevance of these increases, increased expression of HNRNPA2B1 and PTBP1 has been shown to be responsible for the predominance of the PKM2 isoform that is the hallmark of many types of cancer, promoting the aerobic glycolysis that is important for cell growth [100,101]. Furthermore, HNRNPA2 was recently shown to be responsible for splicing events that promote invasive migration of cancer cells in three-dimensional matrices [102].

**Conclusions**

Despite the large number of changes in protein expression upon UPF1 knockdown, our two-stage validation shows that a relatively small fraction of them can be directly attributed to the action of NMD on the corresponding mRNA. This indicates that the role of NMD in directly regulating gene expression may be less prominent than previously suggested. The majority of the doubly-validated mRNAs contain computationally predicted uORFs, confirming this feature as an indicator of NMD sensitivity. We have also identified three examples of AS-NMD, including a highly conserved AS-NMD event that appears to mediate autoregulation of HNRNPA2B1 expression levels. This extends the observation that many RNA binding proteins auto-regulate their own expression through highly conserved elements. Consideration of this autoregulation will be important when examining biological situations, such as several types of cancer, where HNRNPA2B1 levels are deregulated.

**Methods**

**Cloning**

*Construction of pEGFPint*: an efficient artificial intron with associated exonic sequence was amplified from plasmid pY7 [35] using primers PY7INTF (CTCTAGAGCTCAATTCGCTGGCAGGCA) and PY7INTR2 (CTCTAGAGTCCAATTCGCTGGCAGGCA) and pfu high fidelity polymerase. The pY7 intron is based on a β-globin intron while the exons are exons 2 and 3 from the α-tropomyosin [35]. The resulting PCR product was cloned into pGEM T-easy (Invitrogen) and its identity confirmed by sequencing with T7 and SP6 primers. pEGFP-N1 (Promega) contains unique NotI and MfeI sites between the GFP stop codon and the SV40 polyA signals. The insert was liberated from pGEM T-easy by sequential digestion of the NotI and MfeI sites within PY7INTF and PY7INTR2 respectively and ligated with the corresponding fragment of pEGFP-N1. Cloning of FLAG-hnRNPA2: hnRNPA2 sequence was amplified from HeLa cell cDNA using primers hnRNPA2_F TACAGAATTCTGAAAGAGAAGAAGGAAC and hnRNPA2B1_R TCAGGTGCACTATCGGCTCCTCC- CACC. PCR was performed using 1.25 U Stratagene Native Pfu DNA Polymerase, 200 uM dNTPs and 400 uM primers. Cycling parameters: 95°C 2 min, [95°C 30 sec, 48°C 30 sec, 72°C 2 min]35, 72°C 5 min. The resulting PCR product was digested with EcoRI and Sall and ligated into pCI-NLS-FLAG [103], which allows expression of hnRNPA2 protein with N-terminal Flag tag and NLS, and 13 amino acid C-terminal tag. This cloning was confirmed by sequencing with T7 primer.

**Cell culture and transfections**

HeLa cells were cultured under standard conditions in DMEM medium with glutamax (Invitrogen) and 10% fetal bovine serum. HeLa cells stably expressing pGFPInt were generated by lipofectAMINE (Invitrogen) transfection of approximately 6 µg of pGFPInt linearised at the Apa LI site. Transformant cell lines were then selected through growth in medium supplemented with 1 mg/mL G418 (Sigma) and isolated through ring cloning. Cell lines were then constantly maintained in growth medium containing G418 with the exception of when they were being used in an experiment. siRNA transfection of HeLa cells was performed using lipofectAMINE 2000 (Invitrogen) according to the “two-hit” protocol previously described [58]. The 19-mer sense target sequence and associated details of each siRNA are: control C2, 5′-GGUCGCGCCCGCCCGAUG-3′, 120 pmol/transfection or pGFPInt experiments, 2.5 pmol/transfection 2D-DIGE experiments [104]; Upf1_A, 5′-GAUCGAGUCCGCGUCCAUUU-3′, 120 pmol/transfection or pGFPInt experiments, 2.5 pmol/transfection 2D-DIGE experiments [105]; Upf1_B, 5′-GCUCUCCAC- CUGGUGACUGA-3′, 2.5 pmol/transfection; Upf2, 5′-GCGUUUGUGCCGACGAUC-3′, 120 pmol/transfection; SMG1_A, 5′-GUGAAGAUGUUGCCUAUGA-3′, 120 pmol/transfection, Dharmaco siGENOME duplex D-005033-01-0050. Unless otherwise noted siRNA were designed and purchased from Dharmaco Inc.
were harvested for analysis 48 hours after the second transfection. In each case prior analysis had confirmed a high degree of knockdown at this time-point. This time-point also represented a suitable trade-off between the need to allow sufficient accumulation of proteomic alterations directly resulting from UPF1 knockdown, while minimizing secondary effects. Such secondary effects might include false positives, which could arise if primary targets included, for example, transcription or translation factors, as well as false negatives due to compensatory mechanisms. Inhibition of NMD by cycloheximide was achieved by dosing HeLa cells to a final concentration of 10 μg/mL, or with an equivalent volume of DMSO, for 8 hours, as described [55].

Analysis of protein expression by western blot
Extracts of total protein were obtained from tissue culture cells using RIPA buffer or ASB14 buffer. The concentration of these extracts was determined by Bradford assay [106]. Protein extracts were separated on 15% SDS-PAGE gels, transferred to PVDF membrane and then detected by a standard immunoblotting procedure followed by enhanced chemiluminescence detection. Primary antibodies used: rabbit anti-UPF1 [105], goat anti-UPF2 (Santa Cruz Biotechnology inc.), rabbit anti-GFP (Molecular Probes), rabbit anti-actin (Sigma), rabbit anti-ERK1 (invitrogen) Primary antibodies were detected by donkey anti-rabbit and donkey anti-goat antibodies conjugated to horseradish peroxidase.

Analysis of mRNA expression by RT-PCR and QPCR
Total cellular RNA was harvested using TRI reagent (Sigma) according to the manufacturers instructions. 1 μg total RNA was treated with DNase I (Ambion) before oligo-dT reverse transcription using 200 U Superscript II (Sigma) according to the manufacturers instructions. 1 μl of total RNA was used for RT reactions. After RT, a 1/20th aliquot of the RT reaction was used for 38 cycles of amplification for the AS-NMD event within U2AF35 (Invitrogen). PCR for the AS-NMD event within U2AF35 oligo-dT reverse transcription using 200 U Superscript II (Invitrogen) was performed on 1/20th of the RT reaction using the primers U2AF35_F: 5′-TAGCCAG-3′ and Hs.517148_1_R 5′-GCCACAATAAAACCGACGT'TT TAGCCAG-3′, and U2AF35_R: 5′-TGGATCGGCTGTGC- CATAAAACCAAC-3′ for 30 cycles with an annealing temperature of 59°C. AS-NMD events within the TH1L gene were examined using primers: Hs.517148_1_F 5′-GGAGGAGGTGATGACTTC-3′ and Hs.517148_1_R 5′-GGTCACGGCTGGAAAGGAGTT-3′ (intron retention, Figure 4B) for 38 cycles at 60°C annealing; Hs.517148_2_F 5′-ACTGCTGGACAGGATGGTTC-3′ and Hs.517148_2_R 5′-TACCTGCGATGCTGTCATTTC-3′ for 40 cycles at 60°C annealing (alternative 5′ splice site, Figure 4C). Electronic images of gels were captured using a MultiDoc-It Imaging System (UVP) and band intensities were analysed using the associated Doc-It(r) LS Image Analysis Software (UVP). Quantitative PCR (QPCR) was performed on a Rotor-GeneTM 6000 (QIAGEN) using a SYBR green master mix (Applied Biosystems). QPCR data was analysed using the comparative concentration module of the Rotor-Gene software, which is based on the second derivative maximum method described by Tichopad et al. [107]. Signal for the gene of interest (GOI) was normalized to GAPDH or HPRT levels then fold change in mRNA levels was calculated relative to the control sample. Gene specific QPCR primers were generally obtained from Primerbank [108] or designed using Primer3 plus [109], for sequences see Additional file 4. For each primer pair the formation of a single product was confirmed by melt curve analysis [110].

Proteomics 2D-DiGE multi-gel study
Proteomics work was performed at the Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, University of Cambridge. Extracts of total protein for analysis by 2D-DiGE were obtained from tissue culture cells using ASB14 lysis buffer and their concentrations quantified by DC Bradford protein assay (Biorad). For each sample 100 μg total protein was used for analysis. Protein separation was performed using a pH range of pH3-10 (non-linear strip) and a 12.5% SDS-PAGE gel. CyDye labeling, 2D protein separation, gel imaging, and analysis were performed as described previously using systems and software primarily obtained from GE healthcare [46,49,63,111]. PCA was performed using the software package SIMCA (Umetrics). Spots present in < 75% of the experimental samples (1061) were excluded, leaving 2021 for analysis. The first eight principal components (PCs) describing the data were calculated and the first two, PC1 and PC2, identified as significant. PC1 and PC2 result in a model with R² = 0.55 (goodness of fit) and Q² = 0.34 (goodness of prediction). Protein spots were excised both manually from gels stained with colloidal Coomassie, and automatically from fluorescently labelled gels using a CyProt-Picker robotics system (GE Healthcare). The protein constituents of the spots were then identified by LC-MS/MS sequencing of the tryptic peptides produced by in-gel digestion of the spots with trypsin. MS/MS fragmentation data were used to search the NCBI primary sequence database using MASCOT search engine [47].

Bioinformatics
Computational prediction of AS-NMD was performed using a computational pipeline based on that described previously [112,113]. To begin with the Unigene cluster(s) [53] corresponding to each gene of interest was aligned to the genome sequence using SPA [54]. The resulting clusters of alignments were then processed by PASA [52]. PASA acts to subsume equivalent alignments to form a number of maximal transcript assemblies (termed maximum transcripts) that represent
alternative mRNA isoforms consistent with the data from the alignment cluster. For each maximum transcript the largest ORF was defined and tested to determine whether the stop codon lay greater than 50 nucleotides upstream of an exon-exon junction. If this was the case for one or more of the maximum transcripts corresponding to a particular gene, then the gene was designated as possessing AS-NMD.

uORF prediction was performed on all the RefSeq transcripts and Ensembl annotated 5′ UTRs corresponding to the genes of interest. In the case of the Refseq transcripts the largest ORF was defined and then the region directly upstream of this taken to be the 5′ UTR. Then, for both sets of UTRs, ORFs beginning with ATG were sought for in the forward three reading frames.

Finally, potential NMD sensitive maximum transcripts were then scored according to the number of peptides identified by mass spectrometry that were present within the protein sequence encoded by the largest ORF of the maximum transcript. Only transcripts encoding all of the identified peptides were considered as being potentially responsible for the observed upregulation of protein spots.

3′ Race
1 μg of DNAase I treated total RNA from HeLa cells having undergone UPF1 knockdown was subject to reverse transcription by superscript II (Invitrogen) using an oligo-dT primer with the 5′ adaptor sequence: 5′-GGACGC- TAAAGCTTGTTCGAC-3′. PCR was then performed using a primer with the adaptor sequence and primers within both the terminal coding exon of hnRNPA2/B1 (A2B15′: TTTTGTTGTTACGAGGAACAT, A2B15′2: TGGAG- GAAACTATGGTCCAG) and within predicted portions of the 3′ UTR (A2B15′3: TTGGTTCCTTCTCGTGGTT, A2B15′4: TGCTGCCACAAAGACTGTAA). Sequences from these reactions were cloned into pGEM T-easy (Invitrogen) and sequenced.

Sorting of cells by flow cytometry
2 × 10^5 HeLa cells/well of a 6 well plate were co-transfected with 1 μg pCI-NLS-FLAG hnRNPA2 (or the empty pCI-NLS-FLAG vector) and an equivalent molar amount of pEGFP-N1 (927.94 ng) using Lipofectamine 2000 (Invitrogen). 48 hours later cells were harvested for flow cytometry by trypsinisation. Cells from four wells were pooled for each replicate and resuspended in 1 mL DMEM + 2% FCS. For each replicate 1 × 10^6 GFP positive cells were collected using a MoFlo high-speed cell sorter (Beckman Coulter). GFP florescence was detected using a 530/30 filter and live/dead cells discriminated with To-Pro-3 staining, detected using a 670/30 filter. Cells were gated based on forward and side scatter to eliminate debris and then doublet discrimination was carried out to ensure only single cells were sorted (Additional file 5).

Additional material

**Additional file 1:** Protein spots of interest excised for identification by mass spectrometry. Cy2 image of one gel from the 2D-DIGE multi-gel study. The 85 up-regulated spots excised for identification by mass spectrometry are circled in red and the 17 down-regulated spots excised are circled in blue. Each spot is labelled with its number from Additional file 2.

**Additional file 2:** XLS file containing the collated data from the 2D-DIGE multi-gel study, peptide sequence data and protein IDs produced by mass spectrometry, computational prediction of NMD activating features and QPCR based validation of changes in mRNA expression.

**Additional file 3:** A. Novel 3′ UTR sequence of HNRNPA2B1 constructed from RACE tags. Exon sequence is capitalised. HNRNPA1 and A2/B1 motifs from [91] are emboldened and underlined. The small UTR intron present in the Refseq UTR is outlined in black. B. Sequences of the 3′ RACE tags illustrated in Figure 6. C. 3′ Race PCR products illustrating the predicted alternative mRNA isoforms consistent with the data produced by mass spectrometry, computational prediction of NMD activating features and QPCR based validation of changes in mRNA expression. D. Sorting of GFP positive cells by flow cytometry. Representative plots of the cell sorting used in Figure 6. Cell events are denoted by dots. Increasingly “hot” colours represent increasing density of cell events. Each panel, the cells selected for further sorting are outlined by a polygon (termed a gate), and the inside number indicates the percentage of cells at that stage within the gate. A Cells were gated based on forward and side scatter to eliminate debris. B. Doublet discrimination was carried out to ensure only single cells were sorted. C Live/dead cells were discriminated with To-Pro-3 staining, detected using a 670/30 filter. D. GFP fluorescence was detected using a 530/30 filter.

**Acknowledgements**

We would like to thank members of the Cambridge Centre for Proteomics for their invaluable help and advice in the execution of this study. We thank P. Berninger for his help with programming in Perl, J. Mendell for the kind gift of antibodies against UPF1. Work in CWJS laboratory is funded by the Wellcome Trust (077877) and in part by the EC grant EURASNET-LSHG-CT-2005-518238. Collaboration with MZ’s group was also supported by EURASNET. NJM was supported by a studentship from the MRC. LYT was supported by a studentship from the Cambridge Australia Trust.

**Author details**

1Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge, CB2 1QW, UK. 2Biozentrum, University of Basel, Klingelbergstr. 50-70, CH 4056 Basel, Switzerland. 3Current address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 0QH, UK.

**Authors’ contributions**

NJM, KL & CWJS designed the study, NJM performed the experiments and analysed the data, LYT performed experiments pertaining to HNRNPA2B1 autoregulation, NP & MZ performed bioinformatic prediction of AS and NMD activating features, KL assisted with proteomics data analysis, NJM & CWJS wrote the paper. All authors read and approved the final manuscript.

**Received:** 2 June 2010  **Accepted:** 14 October 2010  **Published:** 14 October 2010

**References**

1. Shyu AB, Wilkinson MF, van Hoof A. Messenger RNA regulation: to translate or to degrade. *Embo J* 2008, 27(3):471-481.
2. Isken O, Maquat LE. The multiple lives of NMD factors: balancing roles in gene and genome regulation. *Nat Rev Genet* 2008.
3. Brogna S, Wen J. Nonsense-mediated mRNA decay (NMD) mechanisms. Nat Struct Mol Biol 2009, 16(2):107-113.

4. Nicholoson P, Yepiskoposyan H, Metze S, Zamudio Orozco R, Kleinschmidt N, Muhlemann O. Nonsense-mediated mRNA decay in human cells: mechanistic insights, functions beyond quality control and the double-life of NMD factors. Cell Mol Life Sci 2010, 67(5):677-700.

5. Culbertson JM, Underbrink KM, Fink GR. Frameshift suppression Saccharomyces cerevisiae. II. Genetic properties of group II suppressors. Genetics 1980, 95(4):833-853.

6. Hodgkin J, Papp A, Pulak R, Ambros V, Anderson P. A new kind of informational suppression in the nematode Caenorhabditis elegans. Genetics 1989, 123(2):301-313.

7. Yamashita A, Kashima I, Ohno S. The role of SMG-1 in nonsense-mediated mRNA decay. Biochim Biophys Acta 2005, 1754(1-2):305-315.

8. Behm-Anastas I, Gatfield D, Rehwinkel J, Hilgers V, Izaurralde E. A conserved role for cytoplasmic poly(A)-binding protein 1 (PABPC1) in nonsense-mediated mRNA decay. Embo J 2007, 26(6):1591-1601.

9. Ivanov PV, Gehring NH, Kunz JB, Hentze MW, Kulozek AE. Interactions between UPF1, eRF3, PABP and the exon junction complex suggest an integrated model for mammalian NMD pathways. Embo J 2008, 27(5):736-747.

10. Eberle AB, Stalder L, Wathys H, Orozco RZ, Muhlemann O. Posttranscriptional nonsense mRNA decay by recombination of the 3′ untranslated region. PLoS Biol 2008, 6(4):e92.

11. Silva AL, Ribeiro P, Inacio A, Liebhaber SA, Romao L. Proximity of the poly(A)-binding protein to a premature termination codon inhibits mammalian nonsense-mediated mRNA decay. RNA 2008, 14(3):563-576.

12. Singh G, Rebbapragada L, Lykke-Andersen J. A competition between stimulators and antagonists of Upf complex recruitment governs human nonsense-mediated mRNA decay. PLoS Biol 2008, 6(4):e111.

13. Rebbapragada L, Lykke-Andersen J. Execution of nonsense-mediated mRNA decay: what defines a substrate? Curr Opin Cell Biol 2009, 21(3):394-402.

14. Morrison M, Harris KS, Roth MB. Interactions A with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 2007, 21(8):708-718.

15. Saltzman AL, Kim YK, Pan G, Fagnani MM, Maquat LE, Blencowe BJ. Regulation of multiple core splicingosom proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol Cell Biol 2008.

16. Rehwinkel J, Letunic I, Ries J, Bork P, Izaurralde E. Nonsense-mediated mRNA decay factors act in concert to regulate common mRNA targets. Nucleic Acids Res 1997, 25(9):9782-9785.

17. Mendell JT, Sharif NA, Meyers JL, Martinez-Murillo F, Dietz HC. Nonsense-mediated mRNA decay effectors are essential for mammalian nonsense-mediated mRNA decay. Mol Biol Cell 2005, 16(11):5218.

18. Rehwinkel J, Letunic I, Ries J, Bork P, Izaurralde E. Nonsense-mediated mRNA decay targets of mammalian transcripts and mutes genomic noise. Nat Genet 2004, 36(10):1073-1078.

19. Neu-Yilik G, Gehring NH, Hentze MW, Kulozek AE. Nonsense-mediated mRNA decay: from vacuum cleaner to Swiss army knife. Genome Biol 2004, 5(4):218.

20. Mongodin MM, Roth MB. smg mutants affect the expression of alternatively spliced SR protein mRNAs in Caenorhabditis elegans. Proc Natl Acad Sci USA 2004, 101(18):6750-6755.

21. Wittmann J, Hoh EM, Jack HM. NUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated RNA decay. Mol Cell Biol 2006, 26(4):1272-1287.

22. Ni JZ, Grate L, Donohue JP, Preston C, Nobida N, O’Brien G, Shiue L, Clark TA, Blume JE, Ares M Jr. Ultraspliced elements are associated with homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. Genes Dev 2007, 21(8):708-718.

23. Saltzman AL, Kim YK, Pan G, Fagnani MM, Maquat LE, Blencowe BJ. Regulation of multiple core splicingosom proteins by alternative splicing-coupled nonsense-mediated mRNA decay. Mol Cell Biol 2008.

24. Barberan-Soler S, Lambert NJ, Zaher AM. Global analysis of alternative splicing uncovers developmental regulation of nonsense-mediated decay in C. elegans. RNA 2009, 15(9):1652-1660.

25. Scadden ADJ, Smith CW. Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. Embo J 1995, 14(13):3236-3246.

26. Sachs MS, Geballe AP. Downstream control of upstrem open reading frames. Genes Dev 2006, 20(8):915-921.

27. Kaygın H, Marzluff WF. Regulated degradation of replication-dependent histone mRNAs requires both ATR and Upf1. Nat Struct Mol Biol 2005, 12(9):794-800.

28. Kim YK, Furic L, Desgroselliers L, Maquat LE. Mammalian Staufen1 recruits Upf1 to specific mRNA 3′UTR so as to elicit mRNA decay. Cell 2005, 120(2):195-208.

29. Kim YK, Furic L, Parisien M, Major F, Desgroselliers L, Maquat LE. Staufen1 regulates diverse classes of mammalian transcripts. Embo J 2007, 26(11):760-7628.

30. Pan Q, Saltzman AL, Kim YK, Misquitta C, Shai O, Maquat LE, Frey BJ, Blencowe BJ. Quantitative microarray profiling provides evidence against widespread coupling of alternative splicing with nonsense-mediated mRNA decay to control gene expression. Genes Dev 2006, 20(2):153-158.

31. Wittmann J, Hoh EM, Jack HM. NUPF2 silencing identifies physiologic substrates of mammalian nonsense-mediated RNA decay. Mol Cell Biol 2005, 26(4):1272-1287.
48. Chang YF, Imam JS, Wilkinson MF: The nonsense-mediated decay RNA surveillance pathway. Annu Rev Biochem 2007, 76:51-74.

49. Karp NA, Lilley KS: Maximising sensitivity for detecting changes in protein expression: experimental design using minimal CyDyes. Proteomics 2005, 5(12):3105-3115.

50. Lenth RV: Java Applets for Power and Sample Size. 2006.

51. Viegas MH, Gehring NH, Brett S, Hentze MW, Kulozik AE: The abundance of RNP51, a protein component of the exon junction complex, can determine the variability in efficiency of the Nonsense Mediated Decay pathway. Nucleic Acids Res 2007.

52. Haas BJ, Dekker AL, Mount SM, Worman JR, Smith RK, Jr, Hancock LJ, Maiti R, Ronning CM, Ruch DB, Town CD, et al: Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res 2003, 31(19):5654-5666.

53. Pontius JU, Wagner L, Schuler GD: UniGene: a unified view of the human transcriptionalome. Nat Genet 2003, 31(4):310-314.

54. Kuersten S, Goodwin EB: Alternative splicing of intron 3 of the epsilonRIbeta gene. J Biol Chem 2003, 278(11):10755-10764.

55. Lareau LF, Inada M, Green RE, Wengrod JC, Brenner SE: Maximal transcript alignment assemblies. Nucleic Acids Res 2003, 31(19):5654-5666.

56. Kent WJ: BLAT—the BLAST-like alignment tool. Genome Res 2002, 12(4):656-664.

57. Kuersten S, Goodwin EB: The power of the 3′ UTR: translational control and development. Nat Rev Genet 2003, 4(8):626-637.

58. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW: Translation reinitiation abrogates nonsense-mediated RNA decay in mammalian cells. Embo J 1997, 16(4):826-833.

59. Popielarz M, Cavalaro Y, Mattei MG, Gattioni R, Stevenin J: The gene encoding human splicing factor 9G8. Structure, chromosomal localization, and expression of alternatively processed transcripts. J Biol Chem 1995, 270(30):17830-17835.

60. Rossbach O, Hung LH, Schreiner S, Grishina I, Heiner M, Hui J, Bindereif A: SF2/ASF autoregulation involves multiple layers of post-transcriptional and translational control. J Cell Biol 2004, 165(3):420-434.

61. Hoover DA, Apel K, Eckstein L: Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen expression by influencing alternative splicing of its own pre-mRNA. PLoS Biol 2007, 5(2):e42-e51.

62. Karp NA, Lilley KS: Maximising sensitivity for detecting changes in protein expression: experimental design using minimal CyDyes. Proteomics 2005, 5(12):3105-3115.

63. Kuersten S, Goodwin EB: Alternative splicing of intron 3 of the epsilonRIbeta gene. J Biol Chem 2003, 278(11):10755-10764.

64. Staiger D, Zecca L, Wieczorek Kirk DA, Apel K, Eckstein L: RNA binding protein Pub1 modulates alternative splicing. Mol Cell Biol 2002, 22(22):8093-8103.

65. Stoilov P, Daoud R, Nayler O, Stamm S: Reckoned RBP1 and its paralogs nPTB and ROD1. Nat Rev Genet 2007, 8(8):509-524.

66. Sun S, Zhang Z, Sinha R, Karni R, Krainer AR: Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA 2009, 106(18):7507-7512.

67. Morris DR, Geballe AP: Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 2000, 20(23):8653-8662.

68. Ruiz-Echavarria M, Peltz SW: The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell 2000, 101(7):741-751.

69. Kreic AJ, Swanson MS: hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol 2009, 19(3):363-371.

70. Matlin AJ, Clark F, Smith CW: Understanding alternative splicing: toward a cellular code. Nat Rev Mol Cell Biol 2005, 6(5):386-398.

71. Martinez-Cantelar R, Fisette JF, Stahl M, Warneke F-U, Maddren R, Cordeau J, Janse C, Bacht D: Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. PloS Biology 2006, 4(2):e21.

72. Lee HC, Oh N, Cho H, Choe J, Kim YK: Nonsense-mediated translational repression involves exon junction complex downstream of premature translation termination codon. FEBS Lett 2010, 584(4):795-800.

73. You KT, Li LS, Kim NG, Kang HJ, Koh KH, Chiwa YJ, Kim RM, Kim YK, Park SM, Jang SK, et al: Selective translational repression of truncated proteins from frameshift mutation-derived mRNAs in tumors. Proc Natl Acad Sci USA 2007, 104(5):1809-1814.

74. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D: An unappreciated role for RNA in Arabidopsis thaliana. Proc Natl Acad Sci USA 1999, 96(18):10020-10025.

75. Pontius JU, Wagner L, Schuler GD: UniGene: a unified view of the human transcriptionalome. Nat Genet 2003, 31(19):5654-5666.

76. Wollerton MC, Gooding C, Wagner EJ, Garcia-Blanco MA, Smith CW: Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay, Mol Cell 2004, 13(1):91-100.

77. Roy B, Vaughn JN, Kim BH, Zhou F, Gilchrist MA, Von Amim AG: The h subunit of eIF3 promotes reinitiation competence during translation of mRNAs harboring upstream open reading frames. RNA 2010, 16(4):748-761.

78. Valtorta L, Paglierini DJ, Nooth V: Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci USA 2009, 106(18):7507-7512.

79. Morris DR, Geballe AP: Upstream open reading frames as regulators of mRNA translation. Mol Cell Biol 2000, 20(23):8653-8662.

80. Ruiz-Echavarria M, Peltz SW: The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell 2000, 101(7):741-751.

81. Kreic AJ, Swanson MS: hnRNP complexes: composition, structure, and function. Curr Opin Cell Biol 2009, 19(3):363-371.

82. Matlin AJ, Clark F, Smith CW: Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 2005, 6(5):386-398.

83. Martinez-Cantelar R, Fisette JF, Stahl M, Warneke F-U, Maddren R, Cordeau J, Janse C, Bacht D: Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. PloS Biology 2006, 4(2):e21.

84. Fahling M, Mrowka R, Steege A, Martinka P, Persson PB, Thiele BJ: Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability. J Biol Chem 2007, 282(14):9279-9286.

85. Fisette JF, Toutant J, Dugre-Brisson S, Desgroseillers L, Chabot B: The RNA binding protein Pub1 modulates the stability of transcripts containing upstream open reading frames. Cell 2000, 101(7):741-751.

86. Matlin AJ, Clark F, Smith CW: Understanding alternative splicing: towards a cellular code. Nat Rev Mol Cell Biol 2005, 6(5):386-398.

87. Martinez-Cantelar R, Fisette JF, Stahl M, Warneke F-U, Maddren R, Cordeau J, Janse C, Bacht D: Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing. PloS Biology 2006, 4(2):e21.

88. Fahling M, Mrowka R, Steege A, Martinka P, Persson PB, Thiele BJ: Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability. J Biol Chem 2007, 282(14):9279-9286.

89. Black DJ: Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 2003, 72:291-336.

90. Goina E, Skoko N, Pagani F: Binding of DAZAP1 and hnRNP A1/A2 to an exonic splicing silencer in a natural BRCA1 exon 18 mutant. Mol Cell Biochem 2008, 314(1-2):385-390.

91. Venables JP, Koh CS, Froehlich U, Lapointe E, Couture S, Inkel L, Bramard A, Calvo SE, Pagliarini DJ, Mootha VK: Heterogeneous expression of transcripts harboring upstream open reading frames. J Cell Biol 2008, 18(4):91-100.

92. Fisette JF, Toutant J, Dugre-Brisson S, Desgroseillers L, Chabot B: Heterogeneous nuclear ribonucleoprotein-A2/B1 modulate collagen prolyl 4-hydroxylase, alpha (I) mRNA stability. J Biol Chem 2007, 282(14):9279-9286.
elements in vertebrate, insect, worm, and yeast genomes. Genome Res 2005, 15(8):1034-1050.

doi:10.1186/1471-2164-11-565

Cite this article as: McGlinchy et al.: Expression proteomics of UPF1 knockdown in HeLa cells reveals autoregulation of hnRNP A2/B1 mediated by alternative splicing resulting in nonsense-mediated mRNA decay. BMC Genomics 2010 11:565.