Table of contents

1. Supplementary Methods

Method S1: Gene Ontology analysis of NMD targets

We use the GO Term Finder http://www.yeastgenome.org/cgi-bin/GO/goTermFinder.pl to detect functional bias for our NMD targets. We have 732 NMD targets as defined in text, and use all the examined 4823 ORFs as background. We tested for "Process", "Function" and "Component" ontologies. None of them showed significantly biased GO terms.

Method S2: UTR information for ORFs

The information on UTRs for each ORF was downloaded from a recent large-scale sequencing study (Nagalakshmi et al, 2008). The authors sequenced all the mRNAs after fragmenting them, and then mapped the reads to genomic sequence. The UTR regions were determined by comparing the read-mapped regions and the CDS annotation (Nagalakshmi et al, 2008).

2. Supplementary Results

Result S1: NMD targets have longer UTRs

Previous studies in mammals show that NMD targets have shorter coding sequence (CDS) length but longer 5' and 3' untranslated regions (UTRs) (Zhang et al, 2009). Longer 3' UTR can promote NMD in budding yeast (Kebaara & Atkin, 2009). We tested whether the sequence features of the NMD targets identified on the genome-wide scale for budding yeast match expectations. We compared the lengths of coded proteins, 5' UTRs and 3' UTRs. As shown in Figure S2 (top panel), the 'Inter' set shows a significantly shorter protein length (Wilcoxon rank sum test: P = 0.0229), consistent with the finding in mammals (Zhang et al, 2009). However, when we compared the 'Union' set against non-NMD ORFs, there is no significant difference for protein length (or CDS) between NMD and non-NMD targets (Wilcoxon rank sum test, p = 0.8807). Meanwhile, both sets of NMD targets showed longer 5' UTRs than non-NMD ORFs (Figure S2, middle panel, p = 0.0017 and 0.0027, respectively), and the 'Inter' set shows the larger difference. For 3' UTRs, similar patterns are observed, but the 'Inter' set shows only marginal significance (Figure S2, bottom panel, p = 0.0876 and 8.73E-07 for 'Inter' and 'Union' sets, respectively), likely caused by small sample size. In sum, the results for UTR are consistent with those observed in mammals (Zhang et al, 2009). The longer UTRs in NMD targets is consistent with possible role of UTR in the suppression of NMD targets (Kebaara & Atkin, 2009).

Result S2: NMD targets show lower ribosome occupancy and ribosome density

To confirm the low TE of NMD targets and to see if the low TE is caused by lower ribosome occupancy (for a given gene, the fraction of mRNA copies associated with ribosomes (Arava et al, 2003)), we employed another ribosome dataset based on polysomal separation on sucrose gradients (Arava et al, 2003). As shown in Figure S5A, NMD targets indeed have lower ribosome occupancy (Kolmogorov-Smirnov tests, p = 1.72E-08). This is consistent with NMD mRNAs being less likely to bind ribosomes.

What about the situation when only considering NMD-target mRNAs that are associated with ribosomes? Is it possible that ribosomes on NMD-target mRNA molecules are more spaced out, as expected were the time interval between docking events longer? Consistent with this idea, when only ribosome-associated mRNAs are considered, NMD targets have a smaller number of ribosomes on each mRNA (Figure S5B and S5C, wilcoxon rank sum test, p = 6.92E-07). As there is no difference in CDS length between NMD and non-NMD targets in the 'Union' set, as then expected, the ribosome density (scaled by CDS length) is also lower for NMD targets (Figure S5D, p = 2.99E-07). This result is conservative because we did not use the length including the UTR regions to calculate the ribosome density, which would result in an even lower density in NMD targets due to their longer UTRs (Figure S2).

Result S3: High codon bias is an unlikely explanation for low TE

We checked several relationships which suggest high codon bias is an unlikely explanation for low TE. First, AUG context optimality and codon usage bias are positively correlated in *Saccharomyces cerevisiae* (Miyasaka, 1999) (see also Table S4). As NMD transcripts have low initiation ability (Figure 2C), they may also have low

codon bias and hence slow ribosome movement. Second, codon bias is high in highly expressed (abundant transcripts) genes (as we also see Figure S10) and NMD transcripts tend to be lowly expressed. Third, selection on unwanted transcripts should act to slow translation rates by slowing the ribosome. Finally, we observed a positive correlation between codon usage bias index and TE (Figure S13). These prior results suggest NMD targets may have lower codon usage bias, which is unlikely the reason for the low TE of NMD targets.

Result S4: Lower initiation rates is not the sole reason for low ribosome densities

We have shown that NMD targets initiate poorly and translate poorly. How do they contribute to the ribosome density? For example, if the successful initiation rate is very low, the ribosome density in CDS region may not affected too much by codon usage, because they have enough time for tRNA charging. By contrast, codon usage may slow down ribosomes and increase the ribosome densities.

To examine this, we mapped all the mRNA and ribosome reads onto the CDS and UTR regions based on the reads' genomic coordinates. Then the mean number of reads from both ribosome footprint and mRNA fragments on each codon was calculated. As shown in Figure S12A, there is a great peak around the start codon for the ribosomes, confirming previous observations (Kudla et al, 2009). This is true for both NMD and nonNMD targets, but compared with nonNMD genes, the peak is much smaller for NMD targets. Following the start codon, the number of ribosomes on each codon gradually decreases with increasing distance from start codon. In all of 150 codons after the start codon, NMD targets show greatly lower ribosome density than the nonNMD targets. However, the ribosome densities in the 5'UTR regions are similar between NMD and

nonNMD targets, except for the region near start codon.

The analysis using the ribosome reads alone may have a sequence sample bias because NMD targets are usually lowly expressed (Figure S12B). Therefore, we compared the ribosome read densities between NMD and nonNMD genes accounting for the mRNA level. As shown in Figure S12C, the ratio of ribosome to mRNA read densities at each codon for NMD targets are consistently lower in CDS regions and near the 5'UTR region. As we observed a positive correlation between the free folding energy near the start codon and codon usage (Spearman's rank correlation rho = 0.135, P = 1.450e-13), and likewise for initiation codon optimality and codon usage (Spearman's rank correlation rho = 0.229, P < 2.2e-16), it is possible that the lower densities at CDS regions observed in NMD targets are only caused by lower translation initiation rates and are unconnected to the elongation rate. To check this, we normalized the ribosome densities by the mean of values near the start codon (from -15 to +3 relative to start codon). As shown in Figure S12D, we can still see a generally lower ribosome density in the CDS region for NMD targets. This suggests that a lower initiation rate is not the sole reason for lower densities. Alternatively, there might also be higher rates of stalling at the ATG followed by abortion of translation or rapid movement of ribosomes without translation.

Result S5: NMD activity and the ratio of mRNA read density in UTR to that in CDS, and mRNA half-lives

To check if the low TE of NMD targets is just the preamble of mRNA decay, we designed two tests. First, if the low TE of NMD targets is just a prestep of mRNA decay, we expect that NMD is active in the mRNA sample, that is, there will be more sequences

being degraded from both 5' and 3' ends of an target mRNA (Guan et al, 2006; He et al, 2003; Mitchell & Tollervey, 2003; Takahashi et al, 2003). This speedup will result in fewer sequence reads at the two ends of NMD target mRNAs than non-NMD ORFs after normalizing with the reads in the middle. Consider two genes with the same length, one is an NMD target and the other is not. At the beginning, either gene has 100 mRNA copies. Because of NMD, the NMD target has more copies entering a decay pathway in an unit time, say, 10 copies per unit time, while non-NMD target has fewer, say, 5 copies in one unit time. Meanwhile, it costs time to degrade the whole mRNA and the time depends on the length of mRNA and degradation speed. For simplicity, let's set it to be the same at this moment. Here, because the mRNAs are degraded from two ends, the end point for degradation is the middle of the mRNA. Consider if the time to decay the whole mRNA is *t*, say, 5 units of time. So when the first mRNA is degraded, there are 50 and 25 copies of mRNA being degraded at the ends for NMD and non-NMD, respectively. The number for the middle part is 10 and 5 for NMD and non-NMD, respectively. Therefore, when someone sequences the mRNA samples, 50 and 75 copies can be sequenced at two ends, and 90 and 95 in the middle for NMD and non-NMD, respectively. The ratio of the end to the middle is 50/90 and 75/95 for NMD and non-NMD, respectively. Obviously, the NMD gene shows a lower ratio. To reduce the noise, we used the UTR to represent the end and CDS to represent middle part. If NMD is effective and our assumptions are correct, we should observe a lower ratio for NMD. We use this to support the postulate that the sequence sample used in Ingolia NT, *et al*, 2009 is not under NMD regulation.

As shown in Figure 3A and 3B, no significant difference for the ratio in 5' (Wilcoxon rank sum test, P = 0.1467) and 3' UTR (P = 0.2468) are observed between

NMD and nonNMD targets. This is also true when only the NMD target set containing direct targets is considered (P = 0.82 and 0.77 for 5'UTR and 3'UTR, respectively). This result suggests that the sequence sample was not under NMD suppression. Consistent with this, mRNA samples were poly-A selected (Ingolia et al, 2009), suggesting these mRNAs are likely free of NMD degradation. Therefore the reduced TE observed is for functional mRNAs, but not just a by-product of the NMD mRNA decay pathway.

Second, if NMD is active in mRNA sample, we should observe a shorter half-life for NMD targets. NMD degradation is very effective and rapid. Indeed, when we compared the half-lives of representative ORFs in Table S7 of (Guan et al, 2006) with the half-lives in our dataset. We found that most NMD representatives have shorter half-lives (5 ~ 18 mins) than the median value in our dataset. Thus we expect shorter half-lives for NMD targets when NMD is active. However, using the mRNA degradation half-lives from an independent study (Wang et al, 2002), the mRNA decay half-lives for NMD targets are significantly longer, with more NMD targets being longer than 25 minutes (Figure 3C, Kolmogorov-Smirnov tests, P = 5.72E-9). This suggests NMD should not operate on the samples. In sum, the translation repression observed here is not just an intermediate step of the NMD pathway and occurs in the absence of NMD.

3. Supplementary Figures

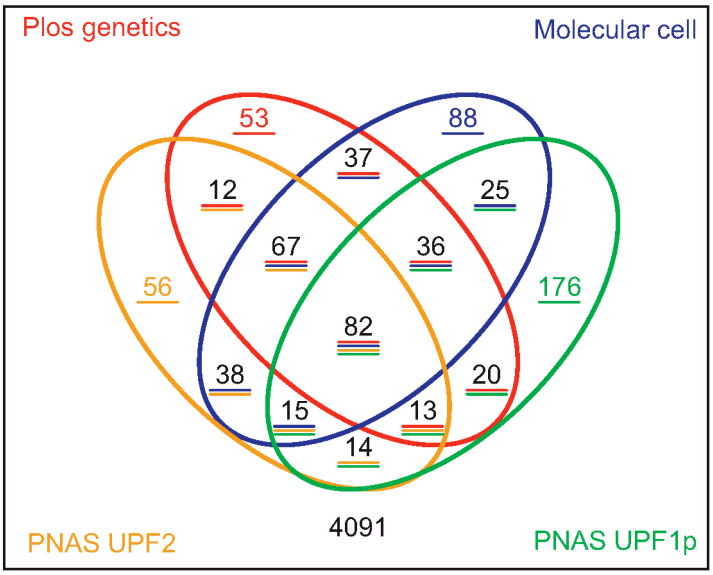

Figure S1. Venn diagram for the NMD gene lists from previous studies.

There are four lists: Plos Genetics, the targets identified by mRNA expression and decay rate change after NMD inhibition (Guan et al, 2006). Molecular Cell, the targets identified by expression change after NMD inhibition (He et al, 2003). PNAS UPF2, the mRNAs down-regulated after UPF2 activation (Johansson et al, 2007). PNAS UPF1p, the mRNAs bound by Upf1p (Johansson et al, 2007). Each set is represented by a different color, and the number of overlapping ORFs shown in the corresponding spot.

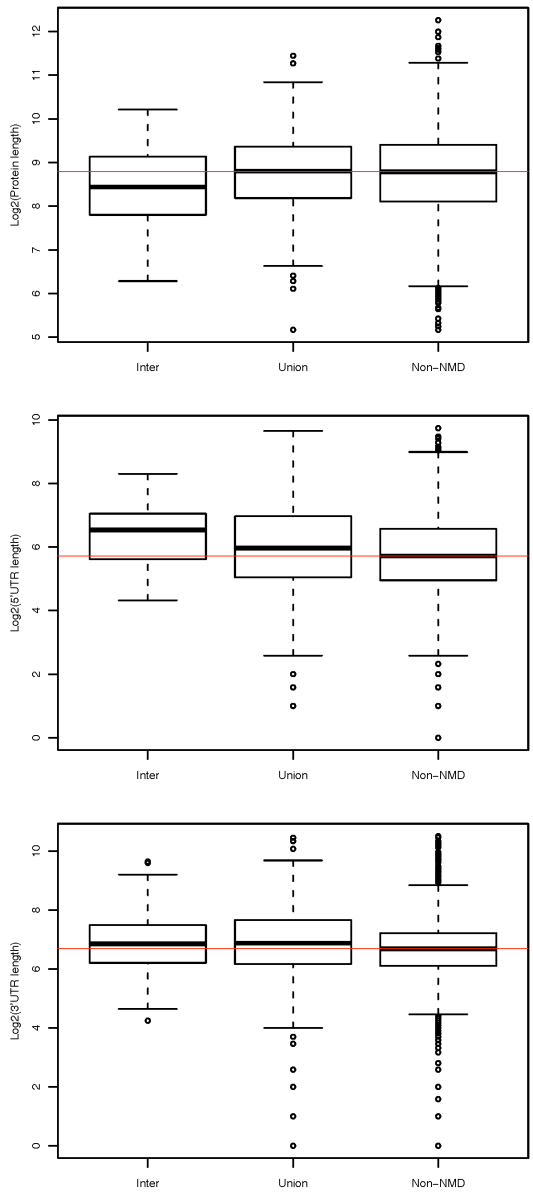

Figure S2. Comparison of the lengths of protein, 5' and 3' UTR sequences between NMD and non-NMD ORFs.

Protein sequences in the NMD 'Inter' set but not in the 'Union' set show a significant shorter length (WRST, P = 0.0229). 5' UTRs in NMD targets are significantly longer than those in non-NMD ORFs (P = 0.0017 and 0.0027 for 'Inter' and 'Union' sets, respectively). For 3'UTR, NMD 'Union' set shows a weak but significantly longer length (P = 8.73E-7). 'Inter' set is marginally significant (P = 0.0876).

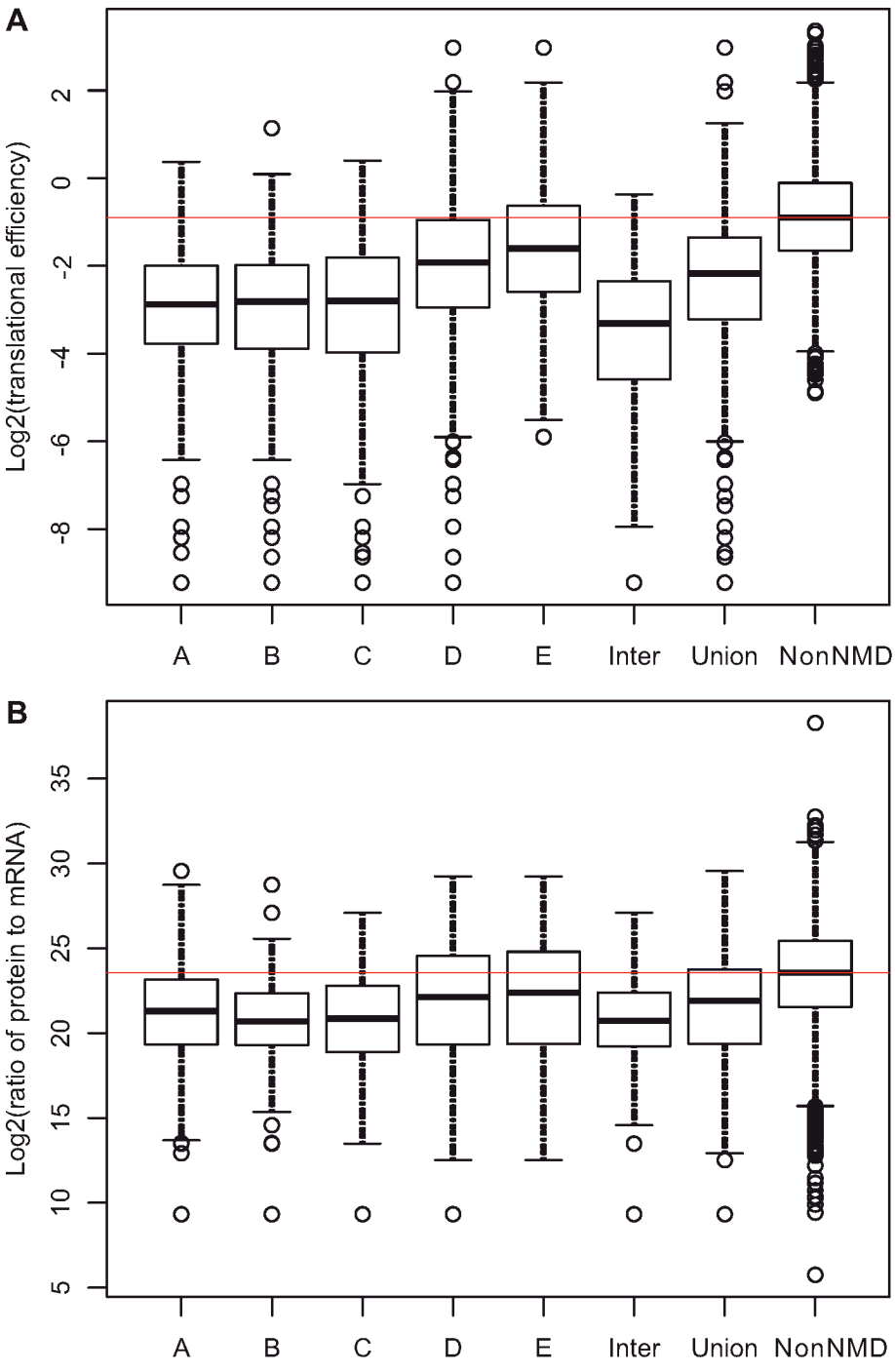

Figure S3. Low translational efficiency in NMD targets.

Comparison was done in all the defined NMD target sets and non-NMD ORFs. Labels for different sets: A: Molecular Cell (He et al, 2003, *Molecular Cell*); B: Plos_Genetics

(Guan et al, 2006, *Plos Genetics*); C: PNAS UPF2 (Johansson et al, 2007, *PNAS*), gene set down-regulated when UPF2 was activated; D: PNAS UPF1P (Johansson et al, 2007, *PNAS*) , gene set bound by UPF1p; E: ORFs bound by UPF1P but no response to UPF2 reactivation; Inter: intersection of A, B, C and D; Union: all the ORFs which is at least affected in one of A, B, C and D; Non-NMD: the genes which do not exist in 'Union' set. (A) TE is estimated using ribosome density on each mRNA. (B) The ratio of protein to mRNA abundance is used to represent TE.

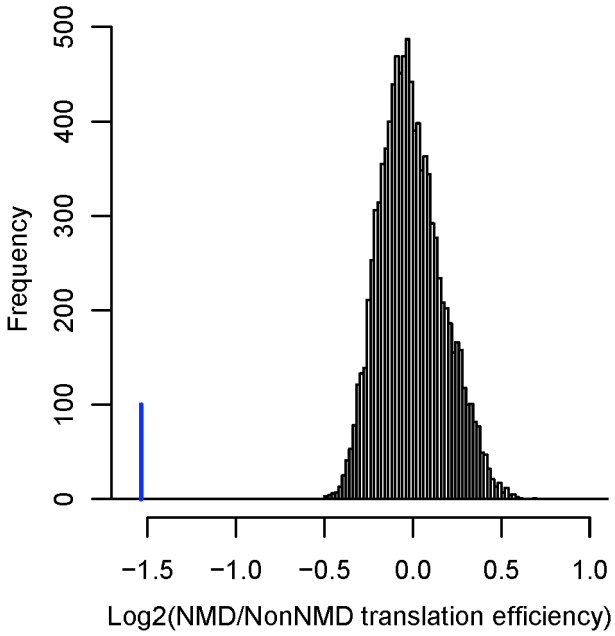

Figure S4. Permutation test for the difference in translational efficiency between NMD and non-NMD ORFs.

The actual ratio of mean translational efficiency in NMD targets to that in non-NMD ORFs is calculated and marked on the x-axis with a blue bar. The null distribution for the ratio of NMD to non-NMD translational efficiency is generated by 10000 random samplings. In each sample, each ORF is labeled as NMD or non-NMD randomly with the sample containing the same number of NMD targets as the original set. The ratio of translational efficiency is calculated for each sample and the distribution is plotted.

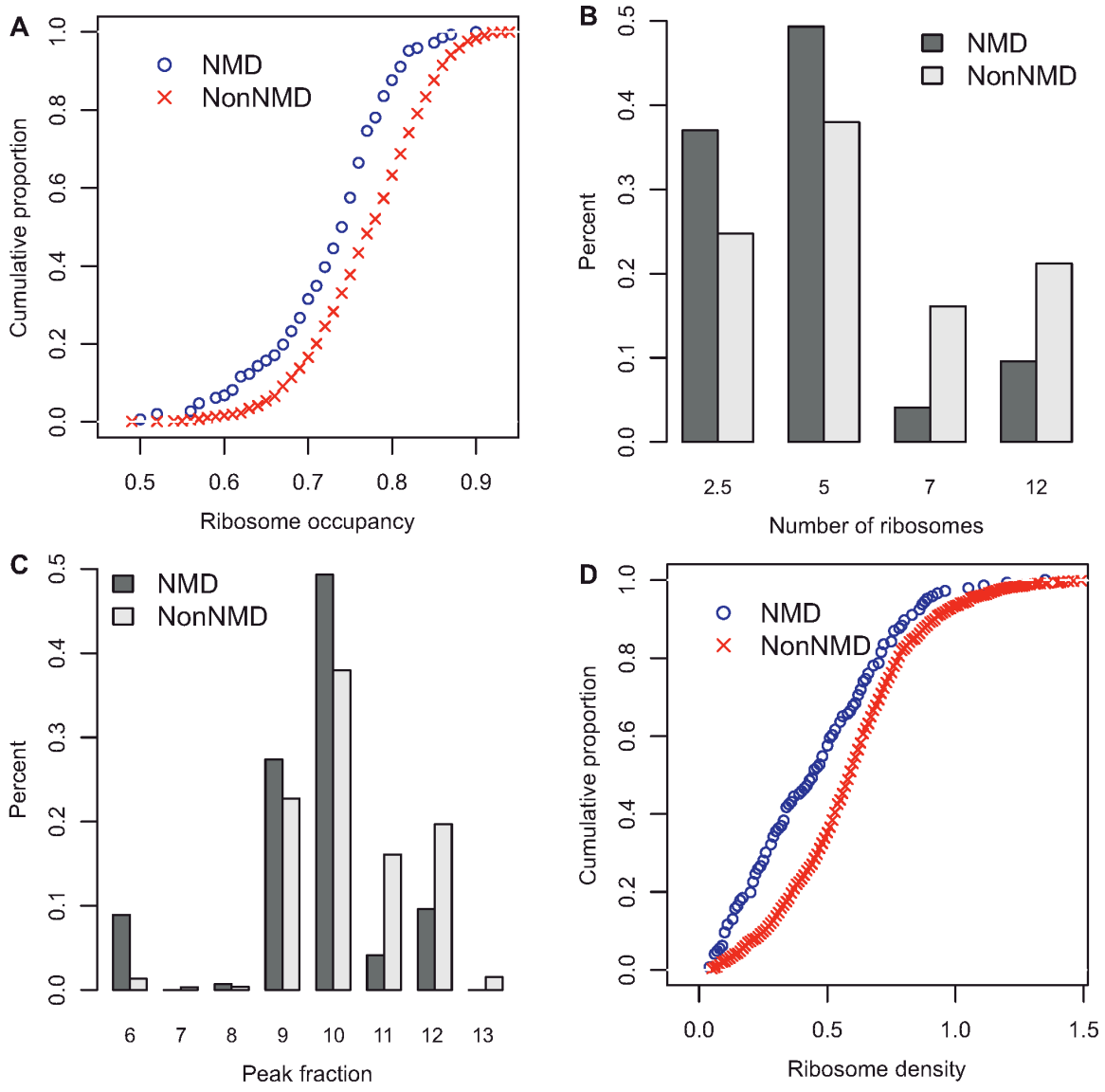

Figure S5. Comparison of ribosome density between NMD and non-NMD targets when controlling ribosome occupancy.

(A) NMD targets show a lower ribosome occupancy (Kolmogorov-Smirnov tests, P = 1.72E-08). (B) and (C) NMD targets have fewer associated ribosomes on each mRNA (P = 6.92E-07). (D) NMD targets have a lower ribosome density when only ribosome-associated mRNA molecules are considered (P = 2.99E-07).

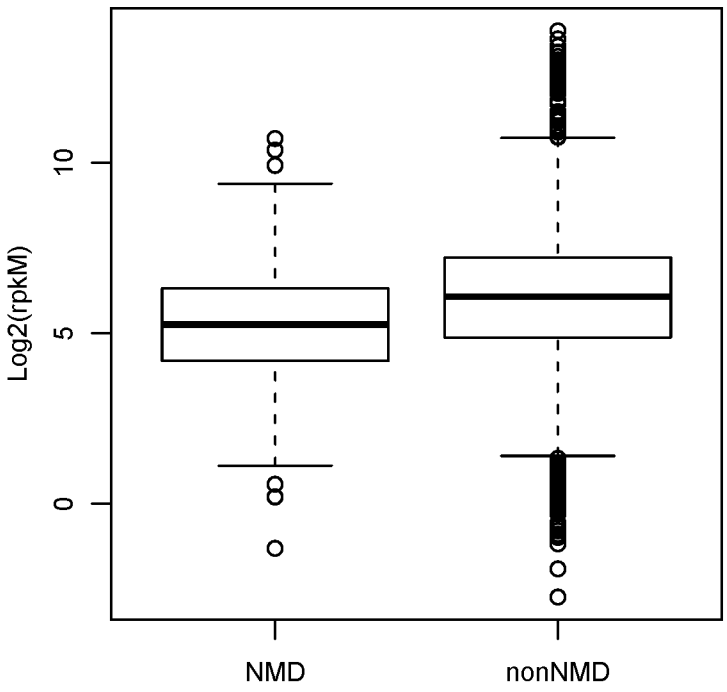

Figure S6. NMD targets show significantly lower mRNA expression.
The mRNA read density (rpkM) is used to estimate the expression level. Wilcoxon rank sum test, $p < 2.2e-16$.

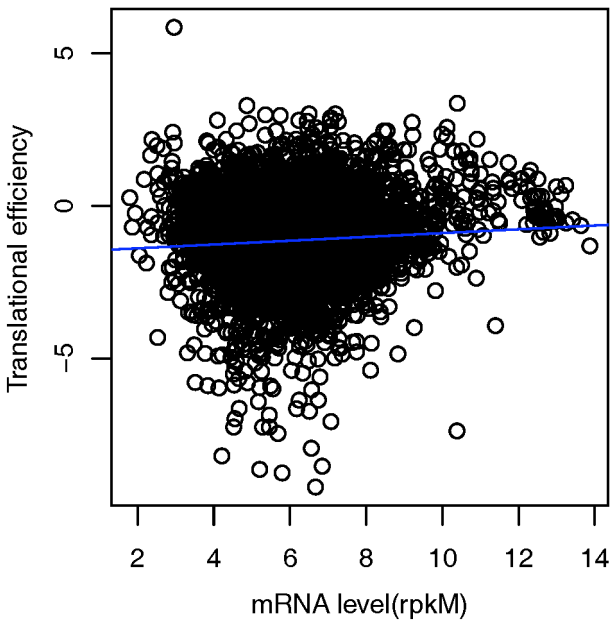

Figure S7. Translational efficiency and mRNA expression level are weakly positively correlated.

Spearman rank correlation, Rho=0.0215, P = 0.1683

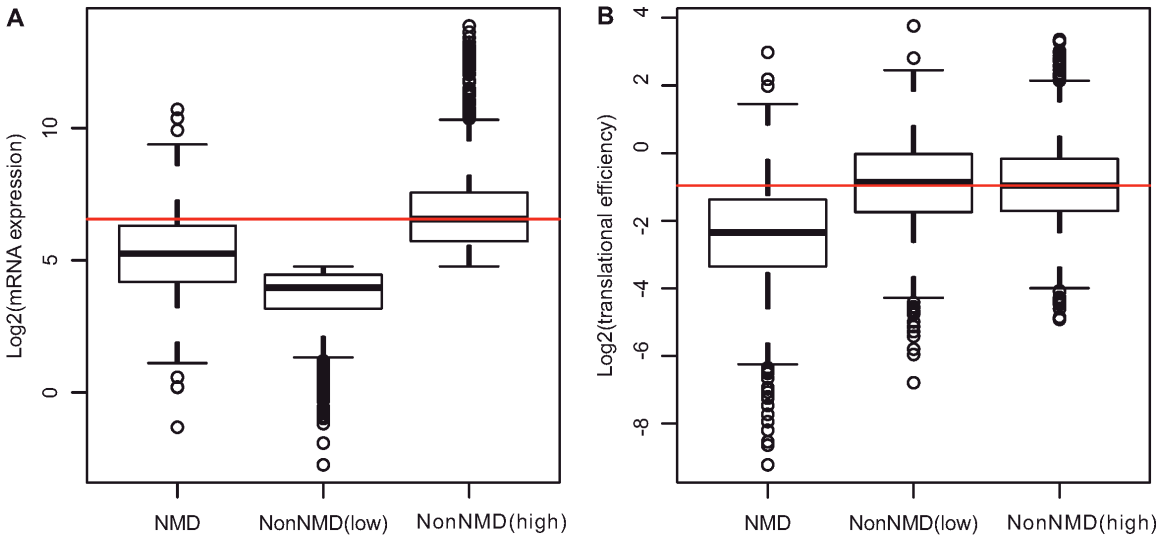

Figure S8. NMD targets show lower TE than lowest expressed ORFs.

The non-NMD ORFs are divided into two groups. The 'low' group are ORFs with lowest expression in the first quartile, while 'high' represents the rest of non-NMD ORFs. (A) NMD targets have significantly higher mRNA expression than non-NMD 'low' group (Wilcoxon rank sum test, $P < 2.2e-16$), but (B) they have lower TE (Wilcoxon rank sum test, $P < 2.2e-16$).

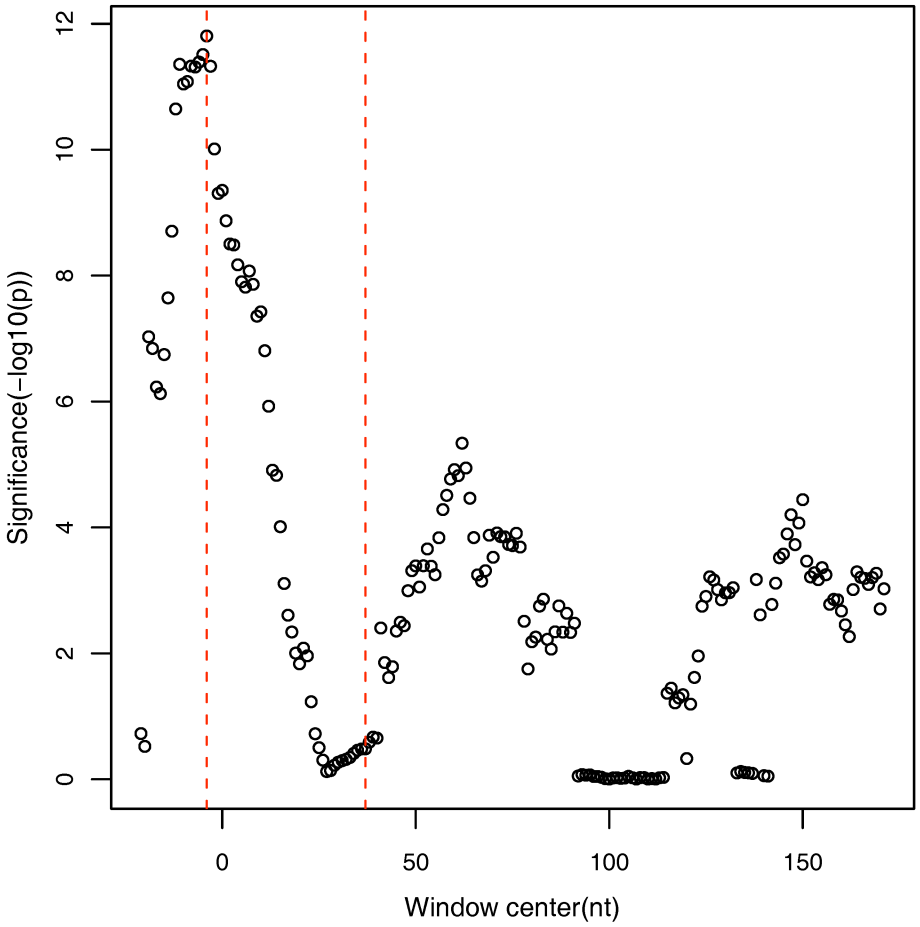

Figure S9. The folding energy of 42-nt window centered at -4 position relative start codon (the first position of start codon is 0) gives the most significant correlation with translational efficiency.

The Y-axis is the $-\log_{10}(p)$, where p is the p-value of spearman correlation test at each window.

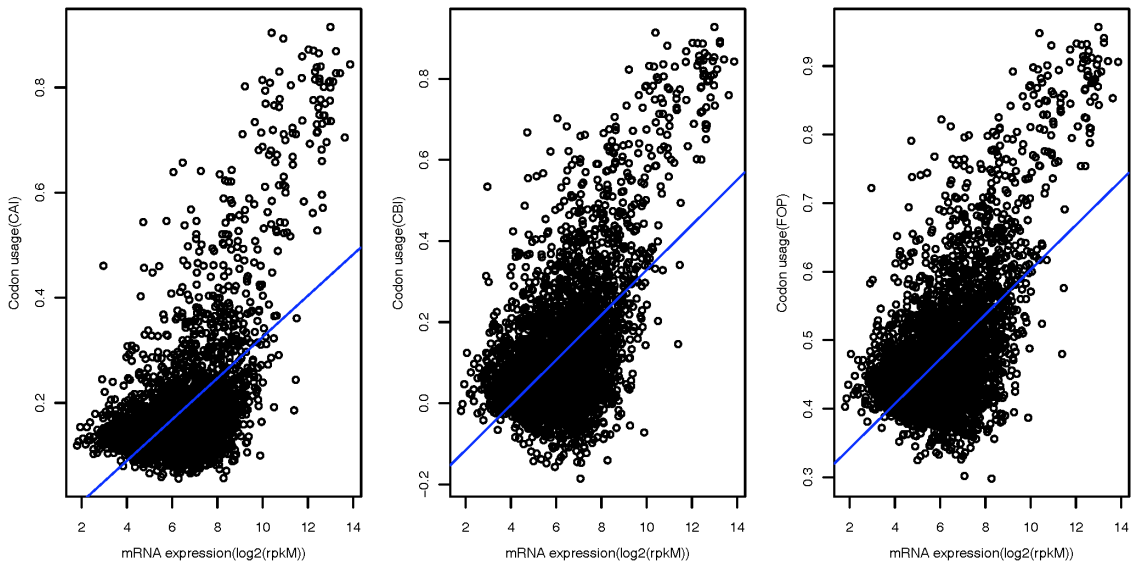

Figure S10. Codon usage score is positively correlated to mRNA expression level.
Spearman rank correlation, P < 2.2e-16.

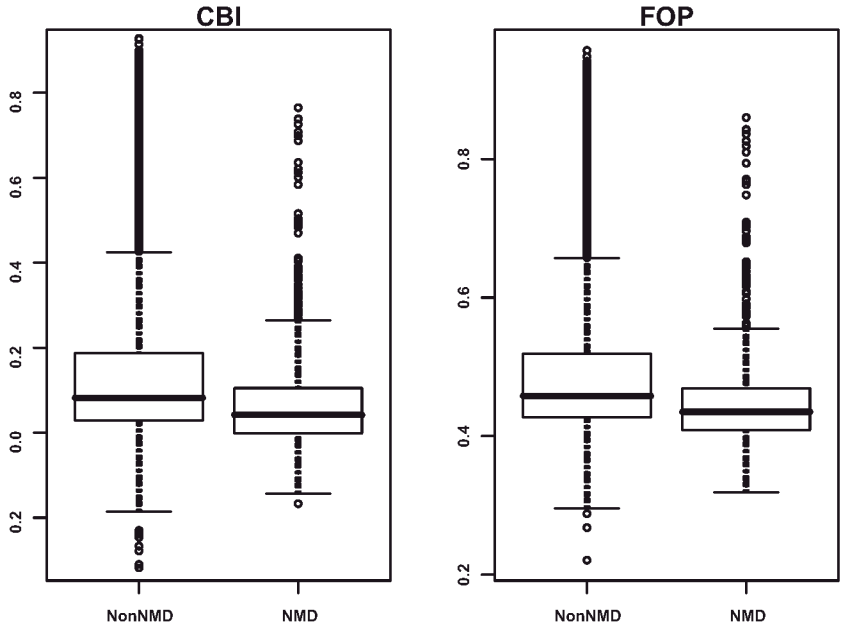

Figure S11. NMD targets show significantly lower CBI and FOP values.
Wilcoxon rank sum test, P < 2.2e-16.

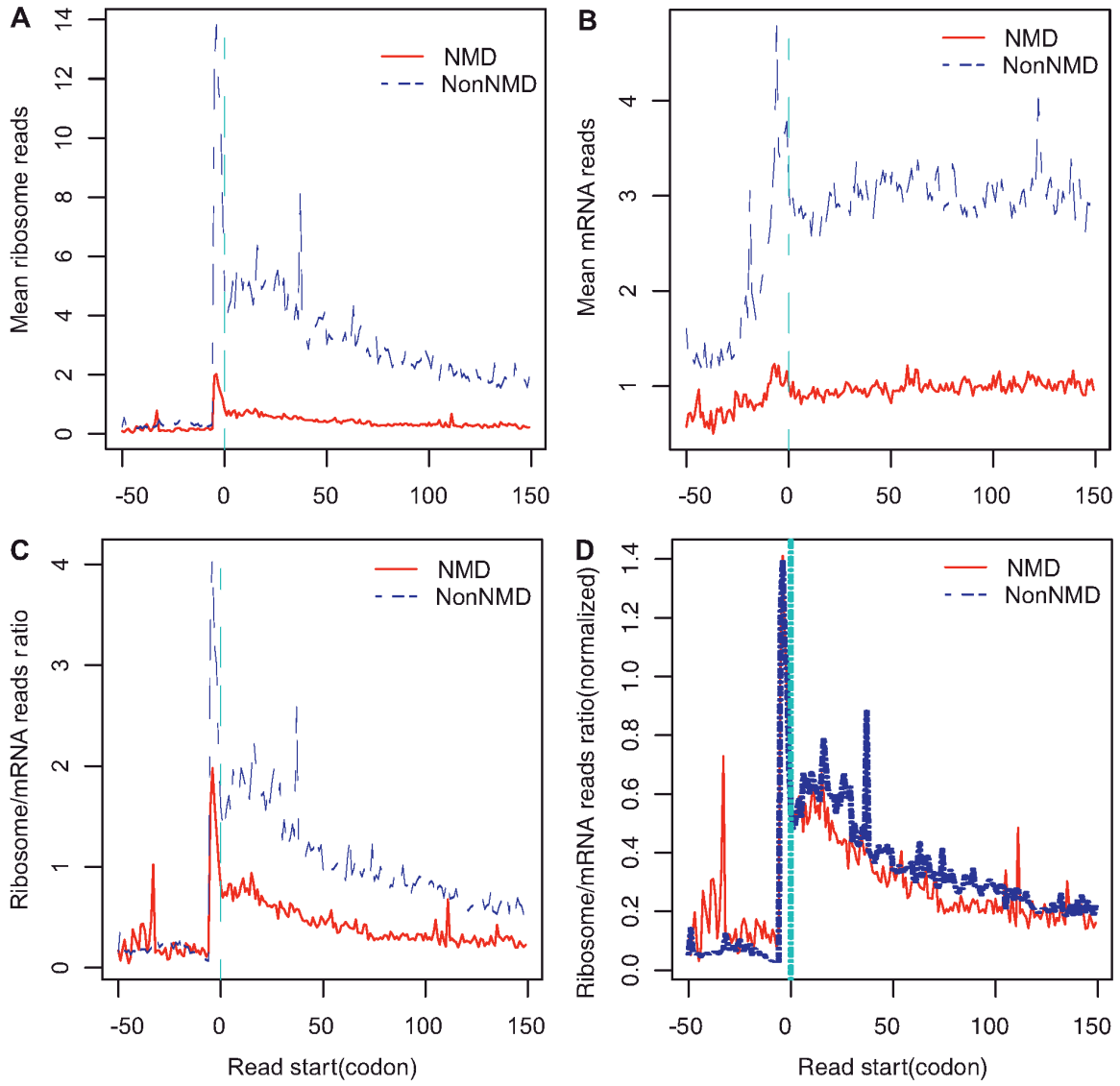

Figure S12. NMD targets have lower ribosome densities near the start codon and along the CDS.

(A) Lower ribosome read density; (B) lower mRNA read density and (C) lower ratio of them in NMD targets. (D) is the same as (C) except that values were normalized by the mean at -15 to +3 for each ORF. The first position of start codon is marked with the vertical dashed line in each plot.

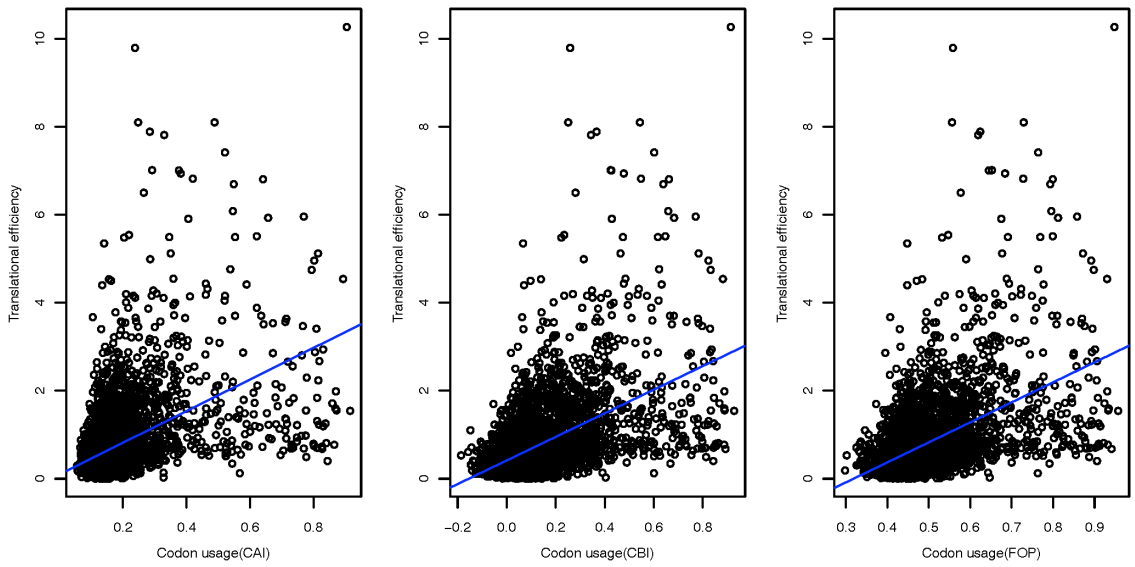

Figure S13. Translational efficiency and codon usage index are highly positively correlated.

Spearman rank correlation, $p < 2.2e\text{-}16$. (A) Codon adaptation index (CAI), (B) codon bias index (CBI), and (C) frequency of optimal codons (FOP).

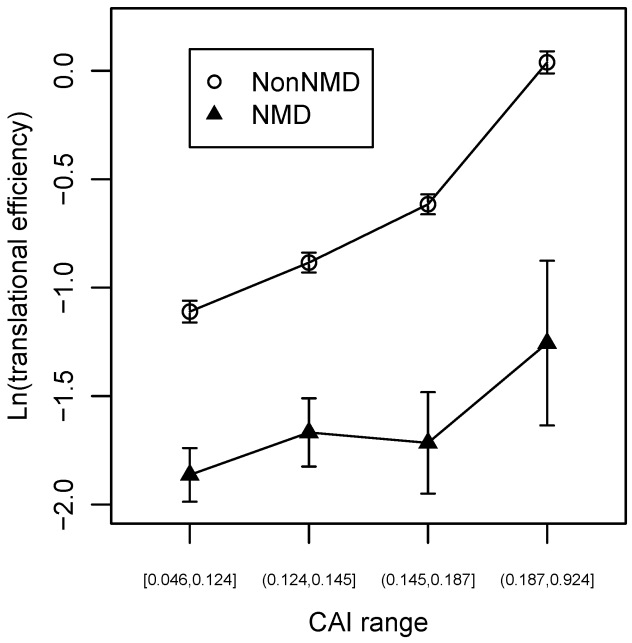

Figure S14. NMD targets have significantly lower translational efficiency in each group of ORFs (P < 2.2e-16).

Here the CAI scores have no significant difference between NMD and non-NMD ORFs.

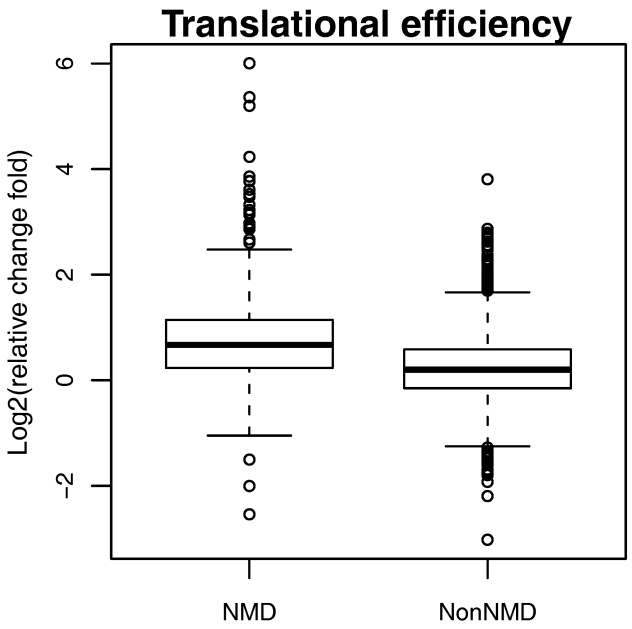

Figure S15. NMD targets show stronger translational up-regulation upon starvation.

The change folds for all the NMD and non-NMD ORFs are compared ($P < 2.2e-16$).

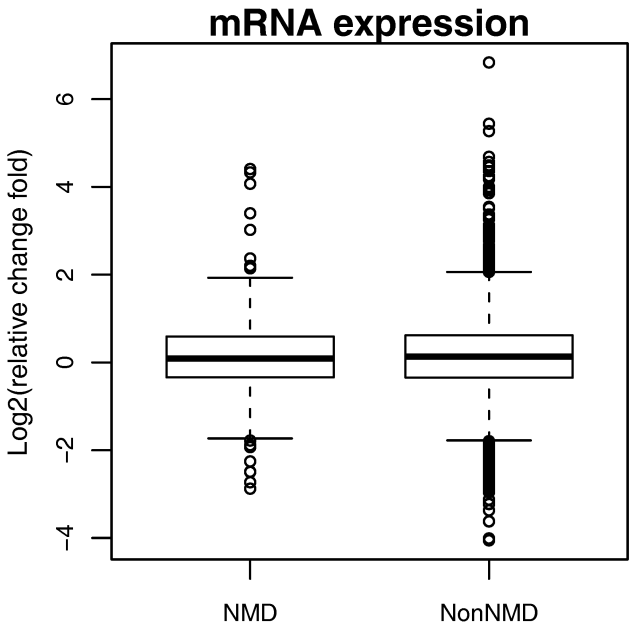

Figure S16. No significant difference of mRNA change folds upon starvation between NMD and non-NMD ORFs

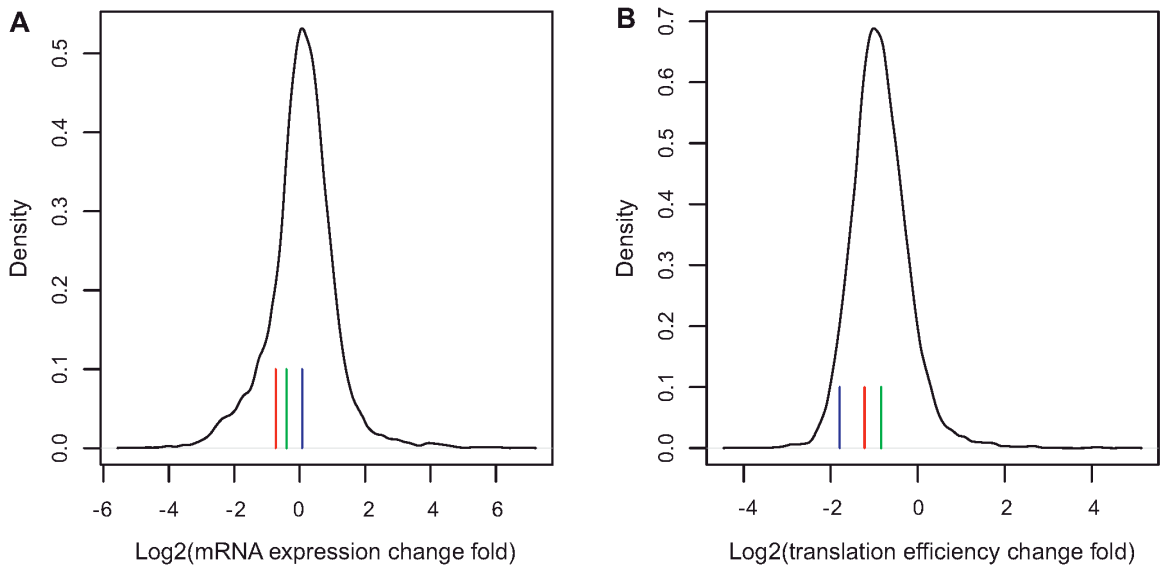

Figure S17. NMD major effectors UPF1 (red), UPF2 (blue) and UPF3 (green) do not change significantly in terms of mRNA expression (A) or TE (B) upon starvation.
In each figure the distribution of change folds for all the ORFs is given for mRNA (A) and TE (B), respectively.

4. Supplementary Tables

Table S1 Comparison of TE between different sets of NMD targets and non-NMD ORFs.

		A[a] Plos genetics	B Molecular cell	C PNAS Upf2	D PNAS Upf1p	E D excluding those in C	Intersecti-on of A,B,C,D	Union of A,B,C,D	Non-NMD
Ribosome-based[b]	P	< 2.2e-16	< 2.2e-16	< 2.2e-16	< 2.2e-16	2.24E-12	< 2.2e-16	< 2.2e-16	
	Count	299	353	270	359	247	73	681	3829
Ratio of protein to mRNA levels[c]	P	< 2.2e-16	3.88E-15	5.89E-16	4.16E-09	1.39E-05	0.000254	< 2.2e-16	
	Count	117	139	103	183	149	19	339	3168

a: Each set of NMD targets (definition is given in Figure S1) is compared to non-NMD ORFs to test if there is a significant lower TE with Wilcoxon rank sum test. Two TE (translational efficiency) estimates are used:
b: the ribosome density normalized by mRNA abundance and
c: ratio of protein to mRNA levels.

Table S2 Correlation between each parameter and mRNA expression and ANCOVA on NMD status with mRNA level as a covariate.

		log2(TE)	Free energy centered at -4	AUG adaptation index	CAI
Correlation with mRNA level	Spearman rho	0.0215	0.135294	0.197891	0.372
	P	0.1683	1.45E-13	< 2.2e-16	< 2.2e-16
ANCOVA for NMD with mRNA level as covariate	P[a]	<2e-16	0.00473	0.312	5.57E-06

a: The P value is for ANCOVA, where each parameter in each column is compared between NMD and non-NMD with mRNA level as a covariate.

Table S3 ANCOVA analysis of codon usage bias for NMD status with mRNA level as a covariate

	CAI		CBI		FOP	
	Coefficient[a]	p	Coefficient[a]	p	Coefficient[a]	p
mRNA abundance	0.028374	<2e-16	0.039567	<2e-16	0.0233	<2e-16
NMD[b]	-0.01524	0.000077	-0.032842	1.3E-08	-0.018803	2.28E-08
intercept	-0.00626	0.218	-0.14672	<2e-16	0.323627	<2e-16

a: Coefficient is the raw values and not comparable among variables.
b: In the model, NMD and nonNMD were given 1 and 0, respectively. Thus a negative coefficient means NMD group has smaller values.

Table S4 Correlation between initiation codon adaptation index and codon usage bias

	Pearson	Spearman
Correlation coefficient	0.298562	0.22885
P	< 2.2e-16	< 2.2e-16

5. Supplementary Dataset (legend)

Dataset S1: the data used in this study, codon usage bias, including TE, TE change upon starvation, folding free energy, and AUG initiation score. Tab-separated. Data in separate file.

6. References

Brumbaugh KM, Otterness DM, Geisen C, Oliveira V, Brognard J, Li X, Lejeune F, Tibbetts RS, Maquat LE, Abraham RT (2004) The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. ***Mol Cell*** **14: 585-598**

Clement SL, Lykke-Andersen J (2006) No mercy for messages that mess with the ribosome. ***Nat Struct Mol Biol*** **13: 299-301**

Coller J, Parker R (2005) General translational repression by activators of mRNA decapping. ***Cell*** **122: 875-886**

Dittmar KA, Sorensen MA, Elf J, Ehrenberg M, Pan T (2005) Selective charging of tRNA isoacceptors induced by amino-acid starvation. ***EMBO Rep*** **6: 151-157**

Doma MK, Parker R (2006) Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. ***Nature*** **440: 561-564**

Elf J, Nilsson D, Tenson T, Ehrenberg M (2003) Selective charging of tRNA isoacceptors explains patterns of codon usage. ***Science*** **300: 1718-1722**

Gardner LB (2008) Hypoxic inhibition of nonsense-mediated RNA decay regulates gene expression and the integrated stress response. ***Mol Cell Biol*** **28: 3729-3741**

Gatfield D, Izaurralde E (2004) Nonsense-mediated messenger RNA decay is initiated by endonucleolytic cleavage in Drosophila. ***Nature*** **429: 575-578**

Gu W, Zhou T, Wilke CO (2010) A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. ***PLoS Comput Biol*** **6: e1000664**

Guan Q, Zheng W, Tang S, Liu X, Zinkel RA, Tsui KW, Yandell BS, Culbertson MR (2006) Impact of nonsense-mediated mRNA decay on the global expression profile of budding yeast. ***PLoS Genet*** **2: e203**

He F, Li X, Spatrick P, Casillo R, Dong S, Jacobson A (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. ***Mol Cell*** **12: 1439-1452**

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**

Isken O, Kim YK, Hosoda N, Mayeur GL, Hershey JW, Maquat LE (2008) Upf1 phosphorylation triggers translational repression during nonsense-mediated mRNA decay. ***Cell*** **133: 314-327**

Isken O, Maquat LE (2008) The multiple lives of NMD factors: balancing roles in gene and genome

regulation. ***Nat Rev Genet*** **9: 699-712**

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 U S A*** **104: 20872-20877**

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**

Kudla G, Murray AW, Tollervey D, Plotkin JB (2009) Coding-sequence determinants of gene expression in Escherichia coli. ***Science*** **324: 255-258**

Mendell JT, ap Rhys CM, Dietz HC (2002) Separable roles for rent1/hUpf1 in altered splicing and decay of nonsense transcripts. ***Science*** **298: 419-422**

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**

Mitchell P, Tollervey D (2003) An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. ***Mol Cell*** **11: 1405-1413**

Miyasaka H (1999) The positive relationship between codon usage bias and translation initiation AUG context in Saccharomyces cerevisiae. ***Yeast*** **15: 633-637**

Parker R, Sheth U (2007) P bodies and the control of mRNA translation and degradation. ***Mol Cell*** **25: 635-646**

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N (2008) Widespread changes in protein synthesis induced by microRNAs. ***Nature*** **455: 58-63**

Sheth U, Parker R (2006) Targeting of aberrant mRNAs to cytoplasmic processing bodies. ***Cell*** **125: 1095-1109**

Stoletzki N, Eyre-Walker A (2007) Synonymous codon usage in Escherichia coli: selection for translational accuracy. ***Mol Biol Evol*** **24: 374-381**

Takahashi S, Araki Y, Sakuno T, Katada T (2003) Interaction between Ski7p and Upf1p is required for nonsense-mediated 3'-to-5' mRNA decay in yeast. ***Embo J*** **22: 3951-3959**

Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, Zaborske J, Pan T, Dahan O, Furman I, Pilpel Y (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. ***Cell*** **141: 344-354**

Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO (2002) Precision and functional specificity in mRNA decay. ***Proc Natl Acad Sci U S A*** **99: 5860-5865**

Zhang Z, Xin D, Wang P, Zhou L, Hu L, Kong X, Hurst LD (2009) Noisy splicing, more than expression regulation, explains why some exons are subject to nonsense-mediated mRNA decay. ***BMC Biol*** **7: 23**