RESEARCH ARTICLE

Trm9-Catalyzed tRNA Modifications Regulate Global Protein Expression by Codon-Biased Translation

Wenjun Deng[1¤a], I. Ramesh Babu[1], Dan Su[1¤b], Shanye Yin[2], Thomas J. Begley[3,4], Peter C. Dedon[1,5]*

1 Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 2 Department of Cell Biology, Harvard Medical School, Boston, Massachusetts, United States of America, 3 SUNY College of Nanoscale Science and Engineering, Albany, New York, United States of America, 4 RNA Institute and Cancer Research Center, University at Albany, State University of New York, Albany, New York, United States of America, 5 Singapore-MIT Alliance for Research and Technology, Singapore

¤a Current address: Clinical Proteomics Research Center, Department of Neurology, WACC 7–739, Massachusetts General Hospital, Boston, Massachusetts, United States of America
¤b Current address: LifeTech, Guilford, Connecticut, United States of America
* pcdedon@mit.edu

Abstract

Post-transcriptional modifications of transfer RNAs (tRNAs) have long been recognized to play crucial roles in regulating the rate and fidelity of translation. However, the extent to which they determine global protein production remains poorly understood. Here we use quantitative proteomics to show a direct link between wobble uridine 5-methoxycarbonyl-methyl (mcm^5) and 5-methoxy-carbonyl-methyl-2-thio (mcm^5s^2) modifications catalyzed by tRNA methyltransferase 9 (Trm9) in $tRNA^{Arg(UCU)}$ and $tRNA^{Glu(UUC)}$ and selective translation of proteins from genes enriched with their cognate codons. Controlling for bias in protein expression and alternations in mRNA expression, we find that loss of Trm9 selectively impairs expression of proteins from genes enriched with AGA and GAA codons under both normal and stress conditions. Moreover, we show that AGA and GAA codons occur with high frequency in clusters along the transcripts, which may play a role in modulating translation. Consistent with these results, proteins subject to enhanced ribosome pausing in yeast lacking mcm^5U and mcm^5s^2U are more likely to be down-regulated and contain a larger number of AGA/GAA clusters. Together, these results suggest that Trm9-catalyzed tRNA modifications play a significant role in regulating protein expression within the cell.

Author Summary

Here we present evidence for a more complicated role for transfer RNAs (tRNAs) than as mere adapters that link the genetic code in messenger RNA (mRNA) to the amino acid sequence of a protein during translation. tRNAs have long been known to be modified with dozens of different chemical structures other than the 4 canonical ribonucleosides,

though the role of these modifications in controlling translation is poorly understood. By quantifying the expression of thousands of proteins in the yeast *S. cerevisiae*, we identified a mechanistic link between modified ribonucleosides located at the wobble position of two tRNAs, tRNA$^{Arg(UCU)}$ and tRNA$^{Glu(UUC)}$, and the translation of proteins derived from genes enriched with codons read by these tRNAs: AGA and GAA. In cells lacking the enzyme that inserts these modifications, tRNA methyltransferase 9 (Trm9), we found a significant reduction in proteins from genes enriched with AGA and GAA codons and with runs of these codons. Also, mRNAs enriched with runs of AGA and GAA codons are subject to stalled translation on ribosomes in yeast lacking mcm^{5}U and mcm^{5}s^{2}U. Together, these results reveal a distinct role for Trm9-catalyzed tRNA modifications in selectively regulating the expression of proteins enriched with AGA and GAA codons.

Introduction

A striking feature of tRNA molecules is the large number of post-transcriptional modifications, representing up to 10% of the ribonucleoside content [1,2]. Ranging from simple base methylation to complex modifications involving multiple enzymatic steps, modified ribonucleosides are phylogenetically widespread and have long been recognized to play crucial roles in tRNA functions [1,3–5]. Modifications in or around the anticodon loop of tRNA affect translation rate and fidelity through stabilization of codon-anticodon pairing, while other modifications remote from the anticodon loop have specific roles in regulating tRNA stability and folding [1,3,4,6–10]. These observations fuel the hypothesis that tRNA modifications play a broader role in regulating global protein expression, with a focus here on wobble uridine modifications catalyzed by tRNA methyltransferase 9 (Trm9) in budding yeast.

Modification of the wobble uridine in tRNA$^{Arg(UCU)}$, tRNA$^{Gly(UCC)}$, tRNA$^{Lys(UUU)}$, tRNA$^{Gln(UUG)}$ and tRNA$^{Glu(UUC)}$ requires a number of key activities (Fig 1). The Elongator complex (ELP1-ELP6) uses uridine as a substrate and catalyzes the formation of 5-carboxy-methyluridine (cm^{5}U). In association with Trm112, Trm9 will use cm^{5}U as a substrate and catalyze the formation of 5-methoxycarbonyl-methyluridine (mcm^{5}U) at the wobble position of tRNA$^{Arg(UCU)}$, tRNA$^{Gly(UCC)}$, tRNA$^{Lys(UUU)}$, tRNA$^{Gln(UUG)}$ and tRNA$^{Glu(UUC)}$ (Fig 1) [11–14]. The wobble position of tRNA$^{Lys(UUU)}$, tRNA$^{Gln(UUG)}$ and tRNA$^{Glu(UUC)}$ is further thiolated by an enzyme cascade involving Urm1, Uba4, Ctu1, Ncs2 and Ncs6 to yield mcm^{5}s^{2}U (Fig 1) [11,12,14].

We and others have performed studies in budding yeast, which suggest that stress-induced reprogramming of wobble modifications in tRNA leads to enhanced translation of codon-biased mRNAs for critical stress response genes. [15–19]. For example, we showed that deficiencies in Trm9 and another anticodon loop tRNA methyltransferase, Trm4, cause sensitivity to DNA alkylating agents (methylmethane sulfonate, MMS) and reactive oxygen species (H$_2$O$_2$), respectively [16,20–23]. The roles of the tRNA modifications catalyzed by Trm4 and Trm9 in the cell response were found to involve stress-induced increases in wobble m^{5}C in tRNA$^{Leu(CAA)}$ and wobble mcm^{5}U in tRNA$^{Arg(UCU)}$ and tRNA$^{Gln(UUG)}$, respectively. With Trm9, these changes were directly linked, by Western blots, to enhanced translation of several stress response proteins enriched with AGA and GAA codons [20]. These results not only support the idea that dynamic changes in tRNA wobble modifications facilitate translation of critical stress response proteins [19,20], but they also raise the possibility of an alternative or accessory genetic code involving selective use of degenerate codons as an adaptation for translational regulation by tRNA modifications.
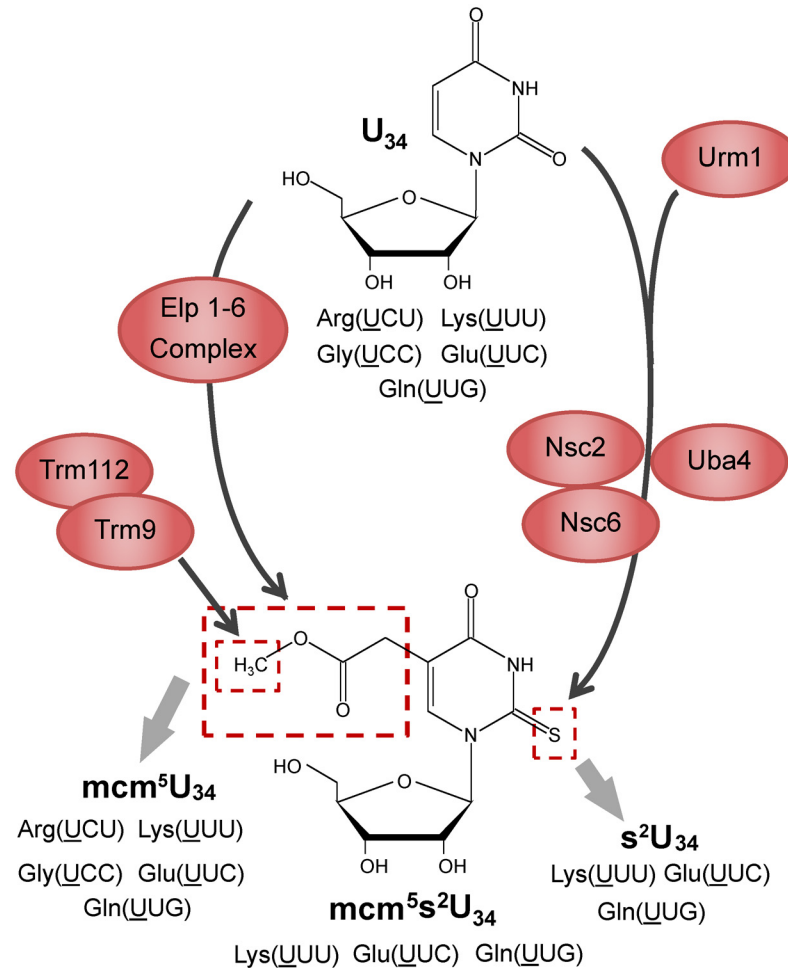
Fig 1. Pathways for the synthesis of modified wobble uridines in tRNA substrates for Trm9 and other enzymes. Synthesis of 2- and 4-thiouridine can precede C5-modification in mitochrondrial tRNAs.

Although there are several examples of individual genes supporting this notion [8,10,17,20,24], what is lacking in this model for translational control of stress response is the larger view of the role of tRNA modifications in regulating global protein expression. Regulation of protein expression occurs at a variety of different levels [25–27], for example, by regulating transcription activity, splicing efficiency, and mRNA export and stability [25,28–33]. Moreover, different stages of protein synthesis are also subject to regulation to ensure efficiency and to preserve fidelity [25,26,34]. Among the variety of factors regulating gene expression, very little is known about the role of tRNA modifications as determinants of global protein translation. Recent studies have shown that loss of Ctu1 or ELP3 result in a moderate reduction in the global protein expression [10], while *urm1* and *elp3* knockout impairs translation of proteins with high usage of AAA, CAA, and GAA codons [9]. Using ribosomal footprinting, two recent genome-based studies measured average ribosome occupancy on each codon type in several yeast U34 modification mutants [24,35]. While there were striking discrepancies between the two similar studies, the consistent results suggested that loss of wobble uridine modifications in *ncs2Δ*, *ncs6Δ*, *elp3Δ* and *elp6Δ* mutants leads to changes in ribosome occupancy for some codons associated with tRNA$^{Arg(UCU)}$, tRNA$^{Gly(UCC)}$, tRNA$^{Lys(UUU)}$, tRNA$^{Gln(UUG)}$ and tRNA$^{Glu(UUC)}$. Of particular note, however, is that the effects caused by loss

of U34 modifying enzymes on overall ribosome occupancy on each codon type were averaged over the genome and thus did not reflect codon usage patterns in individual genes, which greatly limits their regulatory conclusions. To address these problems and identify gene-specific regulatory rules, we performed an integrated analysis of proteome, transcriptome and gene-specific ribosome footprinting to investigate the role of Trm9-catalyzed tRNA modifications, mcm⁵U and mcm⁵s²U, in regulating global protein expression.

Results

Loss of Trm9 affects levels of mcm⁵U and mcm⁵s²U but not Trm9-dependent tRNA species

With the overall goal of assessing the effects of loss of Trm9 and its products mcm⁵U and mcm⁵s²U on global protein translation, we first verified that modified ribonucleosides mcm⁵U and mcm⁵s²U were absent in the *trm9Δ* cells (S1A and S1B Fig) while the abundance of the hypomodified tRNA species were not significantly affected (S1C Fig) under both normal and stress conditions. These results corroborate previous studies [4,11,12,16,20,36] and establish the *trm9Δ* cells as a well-controlled model system for analyzing the influence of tRNA modification on global protein expression.

SILAC proteomics identifies differentially expressed proteins in the absence of Trm9-catalyzed tRNA modifications

We then used a SILAC-based quantitative proteomic analysis to assess global protein expression in unexposed and MMS-exposed wild-type (WT) and *trm9Δ* yeast [37]. Proteins derived from *lys1Δ* yeast grown with $[^{13}C_6,^{15}N_2]$-L-lysine were used as an internal standard that was added to protein extracts in each sample, with quantitation of proteins relative to this standard accomplished by LC-MS/MS analysis of protein digests [37]. Protein coverage was maximized by extensive peptide fractionation using an off-gel isoelectric focusing system [38]. Using this approach, we achieved high-confidence identification of 2,408 proteins with a false-discovery rate of 0.46% (S1 Table) and good reproducibility for the three biological replicates analyzed for each cell type and treatment condition (S1D and S1E Fig). Interestingly, protein expression patterns showed greater similarity between the same yeast strains under different treatment conditions than between different yeast strains under the same conditions, which indicates that loss of Trm9 has a stronger influence on global protein expression than MMS treatment. Altogether, we identified 231 proteins that were significantly down-regulated and 95 up-regulated proteins in *trm9Δ* cells compared with WT cells during normal growth (p <0.05, Student's t-test and fold-change >1.2) (S2 Table). We also identified 195 significantly down-regulated proteins and 137 significantly up-regulated proteins in *trm9Δ* cells in response to MMS treatment (S2 Table).

Proteins enriched with AGA and GAA codons are preferentially down-regulated in *trm9Δ* cells

We first examined whether changes in protein expression are highly selective given the assumption that, after controlling for protein length, proteins with enhanced usage of codons dependent on these wobble modifications are more likely to be down-regulated in *trm9Δ* cells. In this case, we would expect to see genes enriched with mcm⁵U-dependent codons AGA and GGA, as well as mcm⁵s²U-dependent codons, CAA, GAA and AAA, to be selectively down-regulated. Moreover, it has been shown that the mcm⁵ side chain facilitates wobble decoding

for G-ending codons [4]. Accordingly, proteins enriched with AGG and GGG codons may likewise be affected by loss of mcm^{5}U.

To this end, we analyzed gene-specific codon usage patterns for all 5886 genes in the yeast genome to determine groups of proteins that are significantly enriched with each codon. A Z-score was calculated to indicate whether a certain codon is over- or under-represented in each individual gene compared to the genome average. Hierarchical clustering analysis of Z-scores of all genes revealed clusters of codons with relatively similar patterns of usage across different genes (Fig 2A; codon usage data for individual genes is presented in S3 Table). The heat map in Fig 2A shows clustering of CAA, AGA and GAA codons, which is distinguished from clustering of GGG, AGG and GGA codons, while the AAA codon was separated from the others.

We then asked whether proteins enriched with these wobble modification-dependent codons were selectively down-regulated in *trm9Δ* cells. To this end, we calculated the number of significantly down-regulated proteins in each group of proteins enriched with one of the 61 codons. Moreover, to control for the size of different groups and the randomness of changes in protein expression, the percentage of down-regulated proteins, as well as the ratio of the number of down-regulated proteins to the number of up-regulated proteins (D/U) in each group were calculated. For example, in our proteomic dataset, 148 proteins overrepresented with AGA codon were identified and quantified, of which 54 (36.5%) were significantly down-regulated, while only 10 were significantly up-regulated in *trm9Δ* cells (D/U = 5.4). Among the 196 proteins with high GAA usage, 45 (23.0%) were significantly down-regulated and 10 were significantly up-regulated in *trm9Δ* cells (D/U = 4.5). In contrast, for all 2408 proteins identified, only 231 (9.6%) were down-regulated while 95 were up-regulated (D/U = 2.4). As shown in Fig 2B and 2C, the percentage of down-regulated proteins and D/U ratios were significantly enhanced in AGA- and GAA-enriched groups as compared with the genome average. In addition, the CAA group showed a high D/U ratio but the percentage of down–regulated proteins showed no difference from the genome average. This suggested that proteins from genes enriched with AGA/GAA codons were preferentially down-regulated in *trm9Δ* cells. In contrast, we observed no evidence that expression of proteins from genes enriched with GGA, GGG, AGG or AAA were selectively inhibited in *trm9Δ* cells. One possible explanation for lack of effect is that these codons are all non-optimal codons with low overall usage in the genome (see Discussion).

However, several codons independent of the modifications were also associated with a high proportion of down-regulated proteins, which could be explained by co-enrichment of these codons with AGA and GAA. To investigate the influence of codon co-enrichment, we removed proteins enriched with both AGA and another codon from the group of proteins enriched with AGA, and *vice versa*. As shown in Fig 2D, this analysis revealed that high usage of the AGA codon, to the exclusion of any other codon, remained the single best predictor for protein down-regulation in *trm9Δ* cells. In contrast, as expected, the percentages of down-regulated proteins were reduced after removing proteins whose reduction could be better explained by co-enrichment of AGA codon. A similar result was observed for the GAA enriched group (Fig 2E).

In response to MMS treatment, changes in global protein expression in *trm9Δ* cells were likewise skewed as a function of high usage of AGA and GAA codons, but not the other codons dependent on the wobble modifications (S2A–S2D Fig). Taken together, these results support a role for mcm^{5}U and mcm^{5}s^{2}U modifications in regulating proteins enriched with AGA and GAA codons under both normal growth and stress conditions, establishing that Trm9-catalyzed tRNA modifications play a significant role in regulating protein expression.
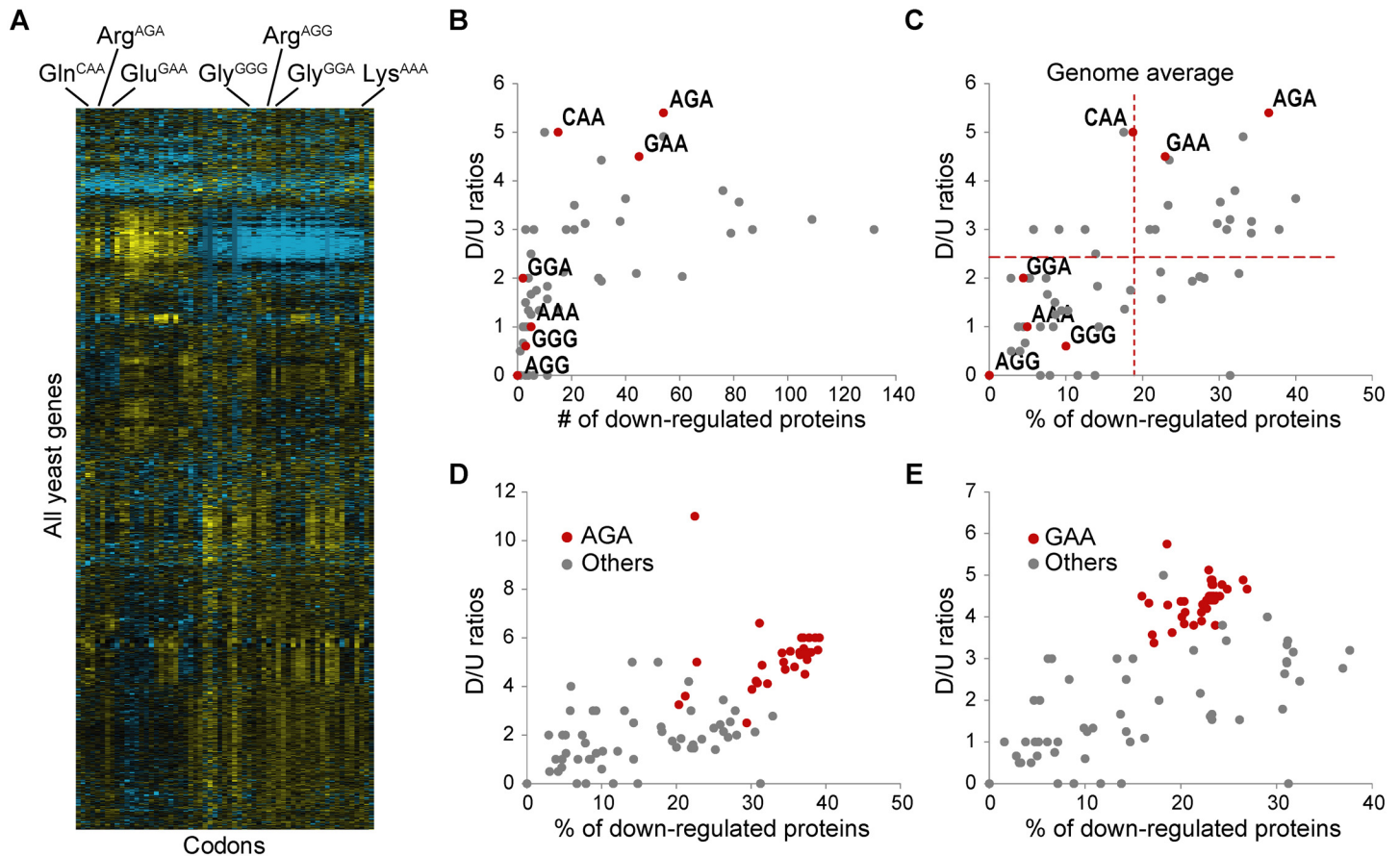
Fig 2. Loss of Trm9 preferentially reduces expression of proteins from genes enriched with AGA and GAA codons, but not other codons. Groups of genes were identified, each with a significantly high use of one of 61 codons. Within each codon group, the number or percentage of significantly altered proteins and the ratio of down- to up-regulated proteins (D/U) were calculated. (**A**) Hierarchical clustering analysis of gene-specific codon usage patterns (Z-scores) for all yeast genes. Z-scores were calculated as the difference between the frequency of each codon used by each transcript and the genome average, divided by the standard deviation, with the scores indicating whether a transcript is over- or under-represented with a specific codon compared to the genome average. Over-represented codons are displayed in yellow and under-represented codons are displayed in blue. Codons showing no difference from the genome average are displayed in black. For clarity dendrograms are not displayed. Data are taken from (**B**) Plot of D/U *versus* the number of significantly down-regulated proteins in each group of proteins enriched with a given codon. The red dots represent data for the groups linked to the codons read by Trm9-modified tRNAs, while grey dots represent data for other codon groups. (**C**) To control for differences in group size, the ratio of down- to up-regulated (D/U) proteins was plotted as a function of the percentage of significantly down-regulated proteins in each group of proteins enriched with a given codon. The red lines indicate the overall percentage of significantly down-regulated proteins and the average D/U ratio in all proteins, respectively. (**D**) To account for codon biases other than AGA, proteins enriched with other codons (except for GAA) were removed from the AGA-enriched group, one codon at a time. Proteins from genes enriched with GAA were likewise removed from groups of proteins enriched in other codons. The values of the remaining proteins enriched with AGA were separately plotted in red dots, while the values of the remaining proteins enriched with other codons were plotted in grey dots. (**E**) Analysis similar to (**D**) repeated for proteins enriched with GAA.

Depletion of mcm⁵U and mcm⁵s²U suppresses expression of proteins from genes enriched with AGA and GAA codons in a highly selective manner

We then examined whether proteins enriched with AGA or GAA codons were more likely to be down-regulated than expected by chance in *trm9Δ* cells and whether the results could be better explained by, for example, changes in mRNA level or biased protein expression. To this end, for proteins enriched with AGA codon (n = 148), we performed 100,000 random samplings of 148 proteins from the proteins that are not enriched with AGA codon, and calculated the percentage of down-regulated proteins and D/U ratio for each sampling. This analysis

demonstrated that groups of proteins from genes enriched with AGA were more likely to possess a higher proportion of down-regulated proteins (Fig 3A) as well as a greater D/U ratio (Fig 3C) than expected by chance. Similarly, proteins enriched with GAA codons (n = 196) were more likely to be down-regulated than genome average in *trm9Δ* cells (Fig 3B and 3D), but not for those enriched with the other codons dependent on the wobble modifications (S3A–S3J Fig). Taken together, these results suggest that depletion of mcm^{5}U and mcm^{5}s^{2}U represses expression of proteins enriched with AGA and GAA codons in a highly selective manner.

However, this analysis could be misleading without controlling for changes in mRNA levels, which may be a major contributor to the changes in protein expression. To this end, we combined the proteomic data with our previous microarray data from *trm9Δ* cells [19]. We found that changes in protein level and changes in mRNA expression were not correlated, with only 4% (13 out of 326) of the significantly changed proteins explained by changes in mRNA level. We then repeated the analysis after removing these proteins from the dataset. As expected, proteins enriched with AGA and GAA codons were still preferentially down-regulated in *trm9Δ* cells (S4A and S4B Fig).

Another feature in question was a bias caused by protein abundance. As shown in S5A and S5B Fig, proteins from genes with high usage of AGA and GAA codons were skewed toward highly expressed proteins. Accordingly, these proteins may be more dramatically down-regulated because they were present at higher levels in WT cells. To control for this, we repeated the analysis by randomly selecting a group of proteins with the same expression level as proteins enriched with AGA or GAA codon, respectively, in each sampling. As shown in S5C and S5D Fig, after controlling for protein abundance, proteins from genes enriched with AGA or GAA were still more likely to be down-regulated in *trm9Δ* cells than expected by chance.

We then examined the possibility that a specific protein is down-regulated in *trm9Δ* cells as a function of increased usage of AGA/GAA codons in all proteins, regardless of whether they are enriched with AGA/GAA codons or not. As shown in Fig 3E and 3F, we found that reduced protein expression in *trm9Δ* cells was significantly correlated with enhanced usage of AGA (r_s = -0.17, p = 2.8E-09) and GAA (r_s = -0.17, p = 7.4E-08), respectively. Furthermore, we binned proteins into seven groups based on their usage of AGA and GAA codons and calculated the D/U ratio in each group. As shown in Fig 3G and 3H, increased usage of AGA and GAA codons additively enhanced the chance of down-regulation in *trm9Δ* cells on a genomic scale. These correlations held when we examined the data for MMS-treated cells (S6 Fig). Together, these results establish that depletion of mcm^{5}U and mcm^{5}s^{2}U selectively repressed expression of proteins with high usage of AGA and GAA codons.

AGA and GAA codons tend to cluster in genes

In addition to overall codon usage, certain features such as codon clustering (i.e., close spacing of codons along a gene sequence) may also regulate the rate of translation along a transcript. We thus scanned each mRNA sequence with a sliding window searching for physical clustering of AGA and GAA codons. As shown in Fig 4A and 4B, for genes enriched with AGA and GAA codons, we observed non-random distributions of these codons along the transcripts. We then tested whether such clustering occurred more frequently than expected by chance. To this end, we counted the number of triplet runs (3-mer) of AGA and GAA combinations in each gene. Maintaining codon composition, we shuffled the codons of each gene and counted the number of triplet runs. After performing this shuffling 10,000 times for each gene, we found that the actual number of codon runs observed was significantly higher than randomization (Fig 4C, p = 2.5E-187, Mann-Whitney U test). The results remained robust when quadruple or quintuple codon combinations were used (S7A and S7B Fig).
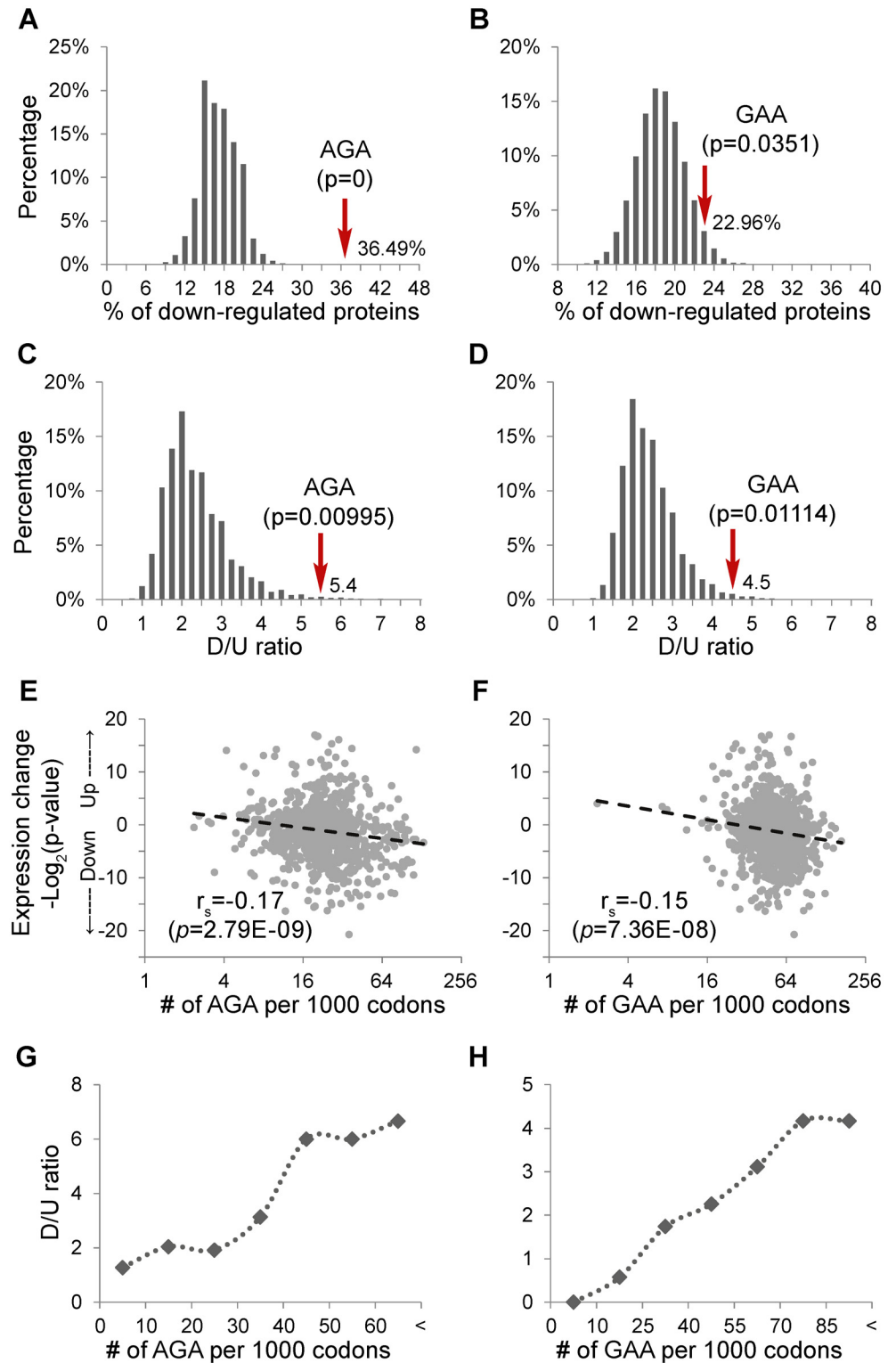
Fig 3. Loss of Trm9 selectively reduces expression of proteins from genes with high usage of AGA and GAA codons. (**A**) Proportion of down-regulated proteins in AGA-enriched proteins *versus* non-enriched proteins. The histogram shows the distribution of proportion of down-regulated proteins of the non-enriched proteins obtained by 100,000 random samplings. The position of AGA-enriched proteins is indicated by an arrow, with the p-values indicating the chance that a random sampling has a ratio no less than that of the

AGA-enriched group. (**B**) Proportion of down-regulated proteins of GAA-enriched proteins *versus* non-enriched proteins, as described above. (**C**) The D/U ratio of AGA-enriched proteins *versus* non-enriched proteins. (**D**) The D/U ratio of GAA-enriched proteins *versus* non-enriched proteins. (**E**) Correlation between AGA or (**F**) GAA codon usage and changes in expression of all proteins in *trm9Δ* cells. Spearman rank correlation coefficient (rs) and p-value are shown. (**G, H**) To further assess codon usage among significantly up- and down-regulated proteins, proteins were divided into seven bins based on AGA (**G**) or GAA (**H**) codon frequency (number per 1000 codons; bins of 0–10, 10–20, *etc*.), and the average codon frequency in each bin was plotted relative to the ratio of down- to up-regulated proteins (D/U ratio for significantly altered proteins) calculated for each binned group.

In the absence of mcm^5U and mcm^5s^2U modifications, these codon runs may generate a local sequence unfavorable for translation by enhancing the chance of ribosomal pausing. In line with this notion, we showed that genes for down-regulated proteins in *trm9Δ* cells contained a significantly higher number of AGA and GAA runs than the other proteins (Fig 4D, p = 7.4E-6, Mann-Whitney U test). However, this could be potentially explained if genes for down-regulated proteins contained more AGA and GAA codons, and as a result, a higher number of codon runs. We controlled for this scenario by limiting our analysis to proteins from genes enriched with AGA and GAA codons. We found no significant difference in the usage of these codons between proteins with codon runs (n = 144) and those without codon runs (n = 154) (mean frequency: with = 11.2% *versus* without = 11.0%; p = 0.84, Mann-Whitney U test). Controlling for codon usage, we found that down-regulated proteins still had a significantly higher number of codon runs than the other proteins (Fig 4E, p = 0.017, Mann-Whitney U test). These results support the notion that clustering of certain codons imposes an additional layer of regulation on translation efficiency and provide independent evidence for selective inhibition of proteins from genes enriched with AGA and GAA codons in *trm9Δ* cells.

Reduced expression of proteins enriched with AGA and GAA codons is linked to enhanced ribosomal pausing in cells lacking mcm^5U and mcm^5s^2U

Ribosome footprinting analysis provides an opportunity to quantify the rate of translation of specific mRNA sequences *in vivo* based on the assumption that the slower a ribosome travels along a specific region of a transcript, the more likely that the ribosomal density in that region will be enhanced. Zinshteyn and Gilbert [24] used ribosome footprinting to assess the effect of mcm^5U and mcm^5s^2U on translation rates in *elp3Δ* yeast cells lacking these modifications and found ribosome accumulations at AAA, CAA, and GAA codons. However, their results were based on genome-average ribosomal occupancy on each codon type, and cannot be used to predict altered expression of individual proteins. We thus integrated this ribosomal footprinting data with our proteomic data to examine whether there is a link between ribosomal pausing and reduced protein expression in cells lacking mcm^5U and mcm^5s^2U.

After controlling for differences in sequencing depth and changes in mRNA expression, we calculated the changes in stringently mapped ribosomal densities that occur within a single transcript between *elp3Δ* cells and WT cells [24]. As shown in Fig 5A, proteins whose transcripts have enhanced ribosomal density (pausing) are preferentially down-regulated compared to those without increase in ribosomal density (*i.e.*, no pausing) (p = 3.56E-05, K-S test). Specifically, as shown in Fig 5B, 75 out of the 292 (26%) transcripts with pausing were found among the significantly down-regulated proteins in *trm9Δ* cells, while only 156 out of the 930 (17%) proteins without pausing were significantly down-regulated in our proteomics dataset (p = 6.92E-4, chi-square test).
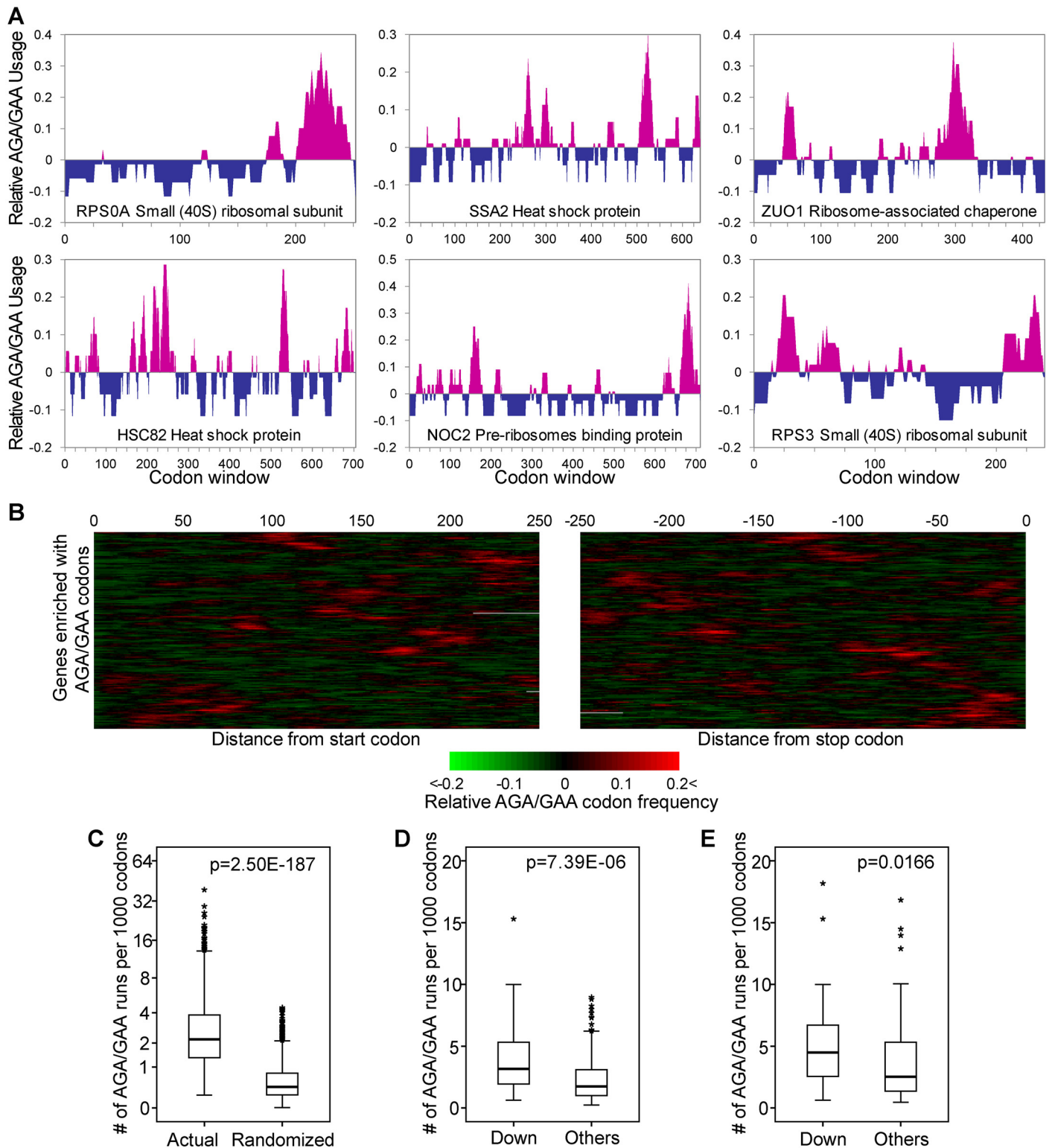
Fig 4. Physical clustering of AGA and GAA codons in genes. (**A**) To illustrate clustering or nonrandom spacing of AGA and GAA codons in several genes, AGA and GAA usage in each gene was calculated in a sliding window size of 15 codons (see Methods) and plotted in the histogram along the length of the gene. Pink peaks correspond to physical clusters of AGA and GAA codons while blue peaks represent regions deficient in these codons. (**B**) The close spacing of AGA and GAA codons is illustrated for all AGA- and GAA-enriched genes in a heat map showing AGA and GAA usage calculated in a sliding window size of 15 codons. For clarity, the map shows regions of genes within 250 codons downstream of the start codon or upstream of the stop codon. (**C**)

Boxplot of the number of 3-mer codon runs identified in each gene (actual) and that of shuffled sequences maintaining the codon composition of each gene (randomized). (**D**) Boxplot of the number of 3-mer codon runs identified in significantly down-regulated proteins and in other proteins. (**E**) Similar to (**D**), but only proteins enriched with AGA/GAA codons in both groups were considered. Asterisks in panels C, D and E represent outliers.

We then asked whether ribosomal pausing is associated with enhanced usage of AGA and GAA codons. We note that all genes with ribosome footprinting information were analyzed, regardless of whether they were identified in our proteomic study. As expected, groups of transcripts with pausing tended to contain a higher proportion of genes with high usage of AGA and GAA codons (Fig 5C, p = 7.16E-11, K-S test). Specifically, a significantly higher rate of genes enriched with AGA and GAA codons was observed in genes with pausing (226/1251, 18.1%) than in genes without pausing (598/4353, 13.7%) (Fig 5D, p = 1.39E-4, chi-square test).

We further asked whether clustering of AGA and GAA codons could enhance ribosomal pausing. As shown in Fig 5E, the genes prone to pausing were skewed toward those with more
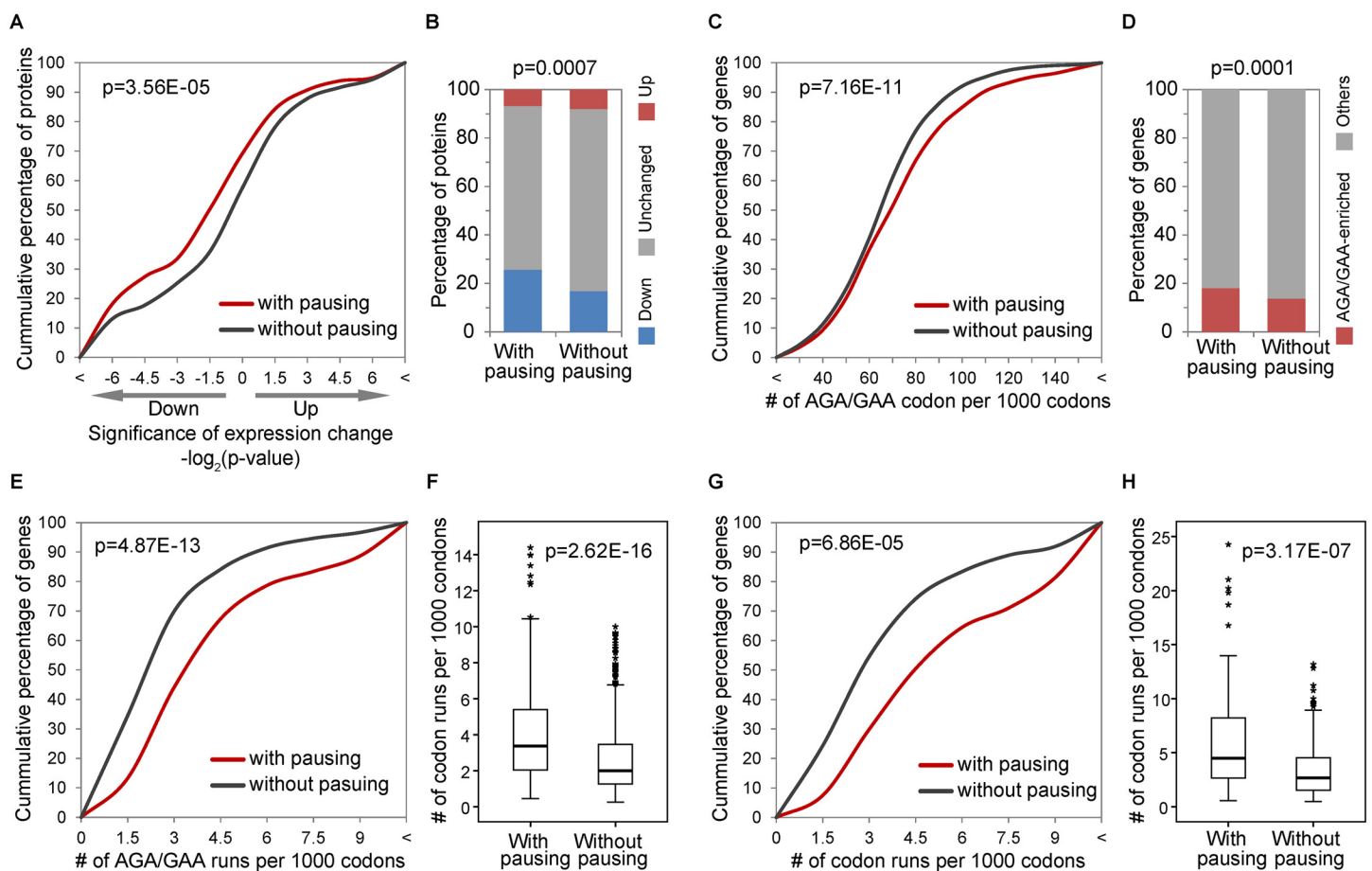

Fig 5. Enhanced ribosomal pausing of proteins enriched with AGA and GAA codons. (**A**) The cumulative percentage of differentially expressed genes representing proteins with or without ribosomal pausing in *trm9Δ* cells. Based on ribosome footprinting data from *elp3Δ* cells lacking mcm⁵U and mcm⁵s²U [24] (**B**) Percentages of significantly down-regulated (blue), up-regulated (red) and unchanged (grey) proteins from transcripts with or without ribosomal pausing. (**C**) The cumulative percentage of genes with a given AGA and GAA codon usage, in groups of proteins with or without ribosomal pausing in *trm9Δ* cells. (**D**) Percentages of proteins enriched with AGA and GAA codons (red) and other proteins (grey) in genes with pausing or without ribosome pausing. (**E**) The cumulative percentage of genes with a given frequency of AGA and GAA codon runs, in groups of proteins with or without ribosomal pausing in *trm9Δ* cells. (**F**) Boxplot of the number of codon runs identified in groups of proteins with and without ribosomal pausing. (**G**) Similar to (**E**), but only proteins enriched with AGA and GAA codons in both groups were considered. (**H**) Similar to (**F**), but only proteins enriched with AGA and GAA codons in both groups were considered.

runs of AGA and GAA codons (p = 4.87E-13, K-S test) and the frequency of AGA and GAA codon runs in stalled genes is significantly higher than those without pausing (Fig 5F, p = 2.62E-16, Mann-Whitney U test). However, this bias toward codon runs could simply result from an association of ribosomal pausing with transcripts possessing high AGA and GAA codon usage. To control for this bias, we limited our analysis to proteins enriched with AGA and GAA codons in both groups. As shown in Fig 5G, genes prone to pausing are still skewed toward higher numbers of codon runs (p = 6.86E-05, K-S test) and the number of codon runs is significantly higher in transcripts on stalled ribosomes (Fig 5H, p = 3.17E-6, Mann-Whitney U test). Taken together, these data provide independent evidence that loss of mcm^5U and mcm^5s^2U selectively reduces translation of genes enriched with AGA and GAA codons by causing ribosomal pausing.

Trm9-dependent proteins are functionally enriched in protein synthesis, cell cycle control and DNA damage response

The fact that loss of Trm9-catalyzed tRNA modifications disrupts expression of proteins from AGA- and GAA-enriched genes led us to explore the Trm9-dependent proteome for a molecular mechanism underlying the associated phenotype of MMS sensitivity. To this end, we analyzed the biological processes associated Trm9-dependent proteins, with comparison of normally growing and MMS-treated *trm9Δ* cells. Using the David program [36], we find that most of the defects in protein expression occurring in response to MMS exposure are readily observed under normal growth conditions in *trm9Δ* cells (S8 Fig; S4 and S5 Tables). Notably, down-regulated proteins with a unique codon usage pattern linked to Trm9 are heavily enriched in translation machinery. For example, 18 out of the 20 components of the 60S ribosomal subunit and all 15 components of the 40S ribosomal subunit are significantly down-regulated under normal and/or stress condition, indicating impaired function of this basic translation machinery (Fig 6A). Paralleling the reduction in ribosomes, we also observed a down-regulation of proteins involved in different steps of translation (Fig 6B), including eIF2, eIF4A, eIF4g and DED1 involved in translation initiation, 15 out of 17 proteins involved in translational elongation, and SSB1, YEF3 and RPL10 involved in translation termination. Moreover, six out of seven proteins involved in protein folding were significantly down-regulated (Fig 6C). Notably, 11 of the 12 aminoacyl-tRNA synthetases were likewise significantly down-regulated. These results revealed an unexpected role for Trm9-catalyzed tRNA modifications in regulating translation, which is consistent with our previous observation that loss of Trm9 impaired expression of proteins involved in translation elongation (YEF3) and DNA damage repair (RNR1 and RNR3), and leaded to translational infidelity, protein errors and activation of protein stress response pathways [19,20]. This is held up to explain the observation that some proteins without high usage of AGA and GAA codons were also down-regulated in *trm9Δ* cells. However, we also suggest that the effect, if any, could not far surpasses that induced by codon usage bias, otherwise we should not observe the selective repression of proteins enriched with AGA and GAA codons. In addition to translation components, we also observed significant down-regulation of proteins involved in DNA damage repair (MPH1, RPL40A and DEF1) and cell cycle control (NBP1, YRB1, CMD1 and MYO1) pathways (Fig 6D). This is consistent with our previous observations that *trm9Δ* cells display delayed transition into S-phase following DNA damage [19].

Discussion

The variety and conservation of modified ribonucleosides in tRNA support the idea that they must play an important role in regulating protein expression [2,3,8,17,19], though the evidence remains largely circumstantial without testing their influence on a global scale *in vivo*. To this
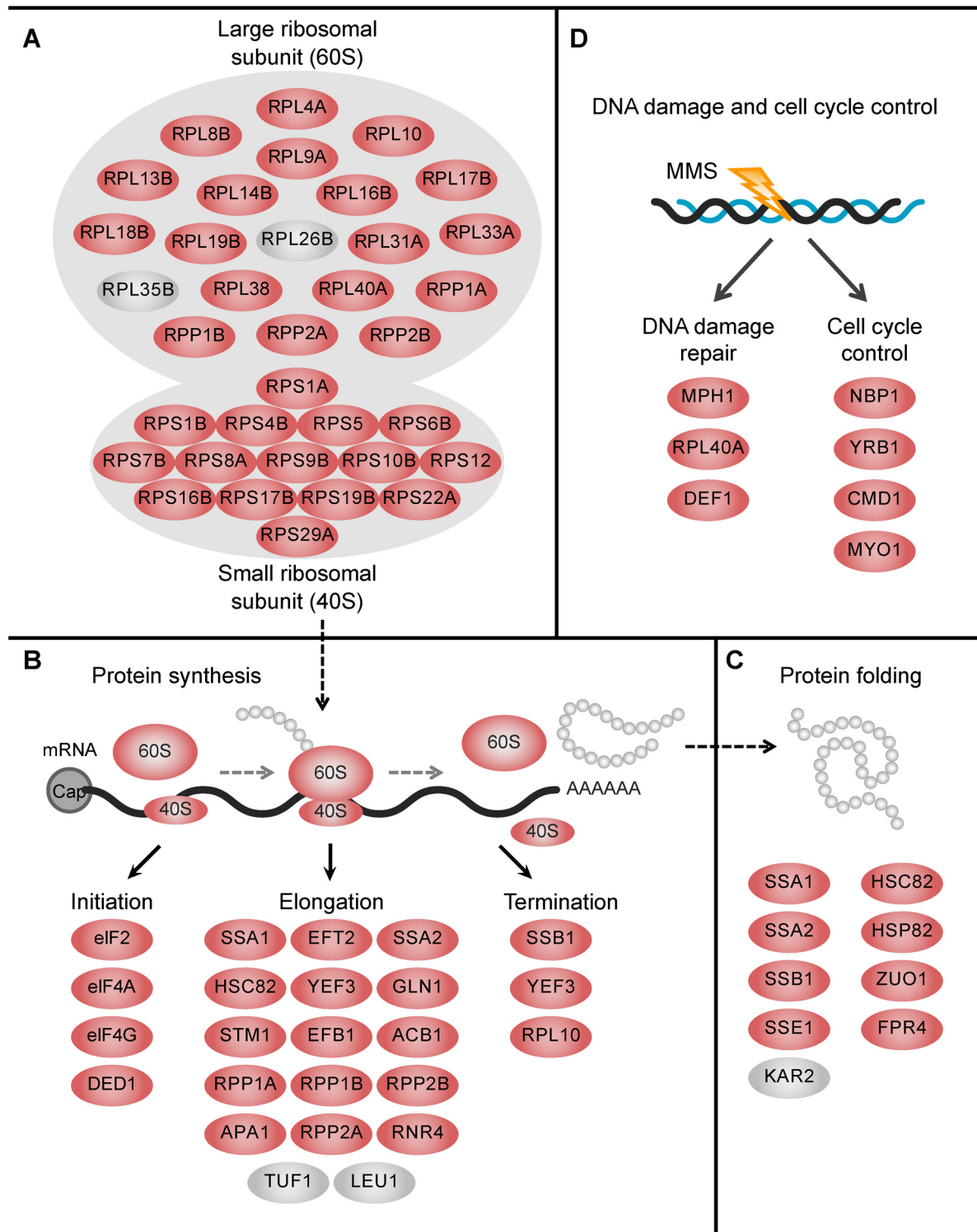
A Large ribosomal subunit (60S)

RPL4A
RPL8B   RPL10
RPL9A
RPL13B   RPL14B   RPL16B   RPL17B
RPL18B   RPL19B   RPL26B   RPL31A   RPL33A
RPL35B   RPL38   RPL40A   RPP1A
RPP1B   RPP2A   RPP2B

RPS1A
RPS1B   RPS4B   RPS5   RPS6B
RPS7B   RPS8A   RPS9B   RPS10B   RPS12
RPS16B   RPS17B   RPS19B   RPS22A
RPS29A

Small ribosomal subunit (40S)

D DNA damage and cell cycle control

MMS

DNA damage repair   Cell cycle control

MPH1   NBP1
RPL40A   YRB1
DEF1   CMD1
                           MYO1

B Protein synthesis

mRNA   60S   60S   60S
Cap   40S   40S   AAAAAA
                                40S

Initiation   Elongation   Termination

eIF2   SSA1   EFT2   SSA2   SSB1
eIF4A   HSC82   YEF3   GLN1   YEF3
eIF4G   STM1   EFB1   ACB1   RPL10
DED1   RPP1A   RPP1B   RPP2B
          APA1   RPP2A   RNR4
              TUF1   LEU1

C Protein folding

SSA1   HSC82
SSA2   HSP82
SSB1   ZUO1
SSE1   FPR4
KAR2

Fig 6. Response to MMS exposure involves Trm9-dependent genes that fall into several tightly connected protein complex and biological processes. (**A**) Ribosomal subunits, (**B**) protein synthesis, including initiation, elongation and termination, (**C**) protein folding and (**D**) DNA damage and cell cycle control. Significantly down-regulated proteins are shown in red and significantly up-regulated proteins are shown in grey. A full list of proteins in different GO contents is provided in S4 and S5 Tables.

doi:10.1371/journal.pgen.1005706.g006

end, we integrated proteome, transcriptome and ribosome footprinting data to elucidate the role of mcm^{5}U and mcm^{5}s^{2}U in regulating global protein expression. After controlling for various confounding factors, such as protein abundance, changes in mRNA levels, and potential influence of other codons, we found robust evidence that expression of proteins enriched with AGA and GAA codons, and to a lesser extent with CAA, are preferentially repressed in cells lacking mcm^{5}U and mcm^{5}s^{2}U under both normal and stress conditions. Consistent with this result, we previously examined expression of several TAP-tagged endogenous proteins, and found that loss of Trm9 only affected expression of proteins overrepresented with AGA and GAA codons [20]. Moreover, we re-engineered a Trm9-depedent gene, ribonucleotide reductase 1 (RNR1), to replace all Trm9-dependent codons with Trm9-independent synonymous codons. In striking contrast to the wild-type gene, the mutant RNR1 gene was largely resistant to *trm9Δ*-induced repression of protein expression [19]. A combined consideration of the proteomic results presented here and previous genetic studies [39] reveal a highly important role of wobble uridine modifications in regulating global gene expression.

Our data clearly revealed that loss of Trm9 and its wobble modifications causes a significant shift to reduced expression of AGA- and GAA-enriched genes. However, it is important to point out that this regulation is not an "all or none" effect—that is, the loss of Trm9 causes a significant shift in translation but not a complete down-regulation of all AGA- and GAA-enriched genes. The observation that not all AGA- and GAA-enriched genes are affected by loss of Trm9 illustrates the fact that gene expression in general and translation in particular are regulated by a complex interplay of different factors that control the efficiency and fidelity of different steps of protein synthesis. It is also important to point out that we compared groups of proteins enriched or not enriched with a single codon as an unbiased test of the hypothesis that the effects of Trm9 loss should be more pronounced for proteins enriched with Trm9-dependent codons. This proved to be the case, but does not imply that genes enriched with AGA and GAA codons are the only ones affected by loss of Trm9, or that all genes enriched with AGA and GAA must be affected by Trm9 loss. However, we did not see any evidence that genes enriched with other Trm9-dependent codons, including GGA, GGG, AGG and AAA, were preferentially down-regulated in cells lacking Trm9. This could be explained by, for example, lower usage of these codons in yeast genes relative to AGA and GAA codons, or that the effect was counterbalanced by poor usage of AGA, GAA and CAA codons, which is supported by the codon usage clustering result in Fig 2A. Nonetheless, the key point is that our results clearly established, as a proof of concept, that Trm9-depedent tRNA modifications play a significant role in regulating protein expression *in vivo*.

An interesting feature of AGA- and GAA-enriched genes is the observation that the codons are more likely to cluster together than expected by chance. Such clustering has been found to affect local translation rate, which has emerged as a mechanism to fine-tune protein expression and minimize protein folding errors, thus providing an additional layer of translational control. For example, biased combinations of codon runs differ in their propensity to cause mistranslation or ribosome pausing [40–42]. Furthermore, large codon clusters could have a greater effect on protein production than an equivalent number of randomly scattered codons, while clustering of rare codons could play an important role in regulating tissue-specific protein expression [40,43,44]. Here we found that Trm9-dependent proteins from genes enriched with AGA and GAA codons showed a significantly increased frequency of AGA and GAA codon runs. Our results provide evidence that 1) proteins with AGA/GAA codon runs, after controlling for codon usage, are more likely to be down-regulated in *trm9Δ* cells than those without codon runs, and that 2) ribosomal pausing in yeast cells lacking mcm^{5}U and mcm^{5}s^{2}U is more likely to occur with transcripts possessing a larger number of AGA/GAA codon runs. These results support the idea that codon clusters add another layer of translational control to protein production.

Since our proteomic analyses provide new insights into the functional complexity of wobble uridine modifications in regulating translation, it is important to place our results in the context of published studies that address U34 modifications from other perspectives and reveal a highly complicated system of regulatory control. Here we compare our results for Trm9-dependent modifications with the genetic studies (tRNA over-expression rescue) of Bjork and coworkers in *ncs6Δ* and *elp3Δ* mutants [39], as well as with ribosome footprinting studies of Zinshteyn and Gilbert [24] and, more recently, Nedialkova and Leidel [35] in *ncs2Δ*, *ncs6Δ*, *elp3Δ* and *elp6Δ* mutants. S6 Table summarizes these studies. Most notably, the footprinting data showed that the loss of *ncs2Δ*, *ncs6Δ* and *elp3Δ* caused ribosomal pausing on codons GAA, CAA, AAA and GAG, but not AGA, and the authors speculated that the effect was too small to influence protein production [24]. In addition to discordant conclusions based on the same yeast mutants [24,35] and misinterpretation of oxidative stress affects on U34 modifications [35], we point to several problems with these footprinting studies in terms of the uncertainty of both RNA-seq technology and data analytics, as detailed in a recent review of ribosome profiling technology [45], problems that obviate stringent comparisons of the ribosome footprinting data sets with the proteomic data. In terms of data analytics, the two studies analyzed the RNA-seq footprinting data in terms of genome-average effects of ribosomal pausing on each codon type and did not specify ribosomal pausing on individual genes at single-codon resolution [24,35]. Ribosomal pausing on the same codon can vary dramatically among genes and even along the same transcript, depending upon the structural and physiochemical properties of the local protein sequence. So the approach to data analysis used in the two ribosome footprinting studies precludes drawing conclusions about ribosomal pausing on individual transcripts or changes in expression of individual genes. To address these problems, we reanalyzed the *elp3Δ* ribosome footprinting data of Zinshteyn and Gilbert [24] in terms of individual genes and found a significant association between enhanced ribosomal pausing and high usage of AGA/GAA codons, as well as the number of AGA/GAA runs along the transcripts. Another feature of the footprinting data, which involves a focus on short mRNA fragments protected by a single ribosome, could explain the apparent absence of AGA codons among the codons associated with ribosome pausing. Close spacing of paused ribosomes has been shown to produce longer protected RNA fragments that are ignored in most ribosomal footprinting methods [45]. Such close spacing could occur by strong pausing at a high density of AGA codons, with a possible contribution from the drag produced when positively charged Arg residues (coded by AGA) interact with the negatively charged ribosomal exit tunnel [46]. This could explain why overall ribosomal occupancy on AGA codons actually decreased in *elp6Δ* cells in the studies of Nedialkova and Leidel [35]. Lastly, all mutants used in the footprinting studies have modification deficiencies that result in the presence of wobble U or s^2U on specific tRNAs, while the *trm9* mutant used here leaves wobble cm^5U or cm^5s^2U and the *nsc6* mutant used in Bjork et al. [39] leaves a wobble mcm^5U. These different wobble modification structures further confound the comparison of the results of the various studies and highlight the complexity of U34 modification effects.

Our analysis of proteins regulated by mcm^5U and mcm^5s^2U modifications provides several insights into the molecular mechanisms underlying the Trm9-dependent alkylation stress response in budding yeast. First, our proteomic analysis revealed 153 of the proteins found to confer sensitivity to MMS in a phenotypic screening study [22], with 30 of these down-regulated and 7 up-regulated by loss of Trm9 (S7 Table). These proteins may thus play a role in the MMS stress response phenotype. An analysis of the functional categories of proteins down-regulated in *trm9Δ* cells provides insights into the molecular mechanisms. One striking feature is that a majority of down-regulated proteins in *trm9Δ* cells are ribosome related or involved in different steps of translation (Fig 6). This behavior has strong parallels with Trm4-dependent

translation of ribosomal protein paralogs from TTG-enriched genes during the response to oxidative stress [17]. As one of the central control points for gene expression, protein synthesis is regulated at multiple levels for translation efficiency and error reduction [19,26], with ribosomal protein variants allowing control of ribosome structure and function for plasticity in the cell response to environmental changes and stress [47]. So it is not surprising that loss of Trm9 causes in large-scale changes in protein expression as a result of perturbations of the translational machinery, even for proteins not enriched with AGA and GAA codons. The additional changes in expression of elongation factors could alter local translation rates, leading to misfolding and impaired protein function [41], which would be compounded by the down-regulation of chaperone proteins observed in trm9Δ cells. Consistent with the idea that cells lacking Trm9 suffer from translational inadequacy, loss of Trm9 was found to perturb polysome profiles of Trm9-dependent transcripts [20], to cause a mis-folded protein stress due to multiple translational errors including mis-incorporations and frame shifts [19], and to disrupt translation of AGA, GAA, CAA, GAG codons [19]. The sum of these dysregulations thus induces accumulation of mis-translated and mis-folded proteins that activates protein stress response pathways in trm9Δ cells, which is consistent with the fact that loss of Trm9 recapitulates the stress response associated with exposure to the protein- and nuclei acid-damaging agent, MMS.

Another feature of the trm9Δ phenotype involves up-regulation of energy production. Since both protein synthesis and rescue of mis-folded proteins by chaperones are energy-dependent processes (S8 Fig; S4 and S5 Tables) [41], it is likely that energy demands are elevated in trm9Δ cells to maintain translation activities and to activate degradation pathways for errors in protein translation and folding. This model is supported by our proteomics data and Gene Ontology analysis. In comparison with WT cells, we find that proteins involved in glucose metabolism and the tricarboxylic acid cycle are coordinately up-regulated in trm9Δ cells under both normal and MMS stress conditions (S8 Fig; S4 and S5 Tables), possibly to keep up with the increased requirement for energy consumption due to translational errors caused by loss of Trm9 [19]. Oxidative stress response proteins were also activated in trm9Δ cells (S8 Fig; S4 and S5 Tables), suggesting that loss of Trm9 confers a state of oxidative stress with elevated reactive oxygen or nitrogen species that are harmful to the cell.

The observed proteomic changes all reflect a phenotype characterized by cell stress, which is consistent with the up-regulation of proteins involved in apoptosis and cell death in trm9Δ cells (S8 Fig). Indeed, loss of Trm9 partially recapitulates the MMS-induced stress response in budding yeast [20]. In the absence of Trm9-catalyzed tRNA modifications, cells experience a dysregulated uncoupling of modified tRNAs from codon-biased translation, which leads to a highly regulated but unfavorable steady-state of altered protein synthesis, energy metabolism and cell death. In this regard, we propose a novel model in which regulation of the spectrum of modified ribonucleoside levels at wobble positions in the system of tRNAs fine-tunes global protein expression by codon-biased translation of mRNAs and reprogramming of translational machinery. Trm9 activity thus illustrates a systems-level mechanism for translational control of cell behavior, with mechanistic parallels in other tRNA modification enzymes.

Materials and Methods

Materials

Urea, sodium chloride, Tris, sodium fluoride, β-glycerophosphate, sodium orthovanadate, sodium pyrophosphate, dithiothreitol, iodoacetamide were purchased from Sigma Chemical Co. (St. Louis, MO). Endoproteinase Lys-C was purchased from Wako (Richmond, VA). All chemicals and reagents were of the highest purity available. All strains of S. cerevisiae BY4741 were purchased from American Type Culture Collections (Manassas, VA).

Yeast strains and treatment

Yeast strain BY4741 was used in this study. WT and *trm9Δ* yeast cells were grown in yeast nitrogen base (YNB) liquid medium containing 30 mg/l normal L-lysine. MMS at a concentration of 0.0125% was used to treat log-phase yeast cells (OD$_{600}$ 0.7) for 1 h, which caused a lethality of <5%. *lys1Δ* yeast cells were grown in YNB medium containing U-[^{13}C$_6$,^{15}N$_2$]-lysine (Sigma) at 30 mg/l for at least 10 generations to log-phase. Cells were harvested by centrifugation for 10 min at 1,500 × g at 4°C and washed twice with cold water.

Quantification of tRNA modifications

Modified ribonucleosides in cytoplasmic tRNA were identified and quantified as reported previously [16,17]. Briefly, WT and *trm9Δ* yeast cells were lysed by lyticase treatment in the presence of deaminase inhibitors (5 μg/ml coformycin, 50 μg/ml tetrahydrouridine) and antioxidants (0.1 mM desferrioxamine, 0.1 mM butylated hydroxytoluene). Total RNA were extracted by the Trizol-chloroform method following manufacturer's instructions, and tRNA-containing small RNA species were enriched using the PureLink miRNA Isolation Kit (Invitrogen). The quantity of tRNA was then determined by UV-vis spectrophotometer (absorbance at 260 nm) and Bioanalyzer analysis (Agilent Bioanalyzer Small RNA Kit). Using identical quantities of tRNA from WT and *trm9Δ* cells, each tRNA sample was mixed with [^{15}N]$_5$-2-deoxyadenosine as an internal standard and the tRNA was enzymatically hydrolyzed with nuclease P1 and RNase A, followed by dephosphorylation by alkaline phosphatase [16,17]. Proteins and other large molecules were removed by ultrafiltration, and the digested ribonucleosides were then resolved by reversed-phase HPLC (Agilent 1100) with a mobile phase of 1~100% acetonitrile in 8 mM ammonium acetate at a flow rate of 300 μl/min for 1 h. Eluted ribonucleosides were analyzed on an Agilent 6410 Triple Quadrupole mass spectrometer. Modified ribonucleosides were identified by HPLC retention time and collision-induced dissociation (CID) fragmentation pattern. The signal intensity for each ribonucleoside was normalized with the signal intensity of [^{15}N]$_5$-dA and abundance of the modified ribonucleosides was compared for WT and *trm9Δ* cells in the presence and absence of MMS.

Northern-blot analysis of tRNA species

DNA probes complementary to 5sRNA (5'-TGGTAGATATGGCCGCAACC-3'), tRNA$^{Arg(UCU)}$ (5'-CACGGCTTAGAAGGCCGTTG-3'), tRNA$^{Glu(UUC)}$ (5'-CTCCGCTACGGGGAGTCGAAC-3'), tRNA$^{Gln(UUG)}$ (5'- GGTCGTACTGGGAATCGAACCCAGG-3'), tRNA$^{Lys(UUU)}$ (5'-CTCCCACTGCGAGATTCGAACTCGC-3') and tRNA$^{Gly(UCC)}$ (5'-GTGTAGTGGTTATCATCCCACCCTTC-3') were end labeled with T4 polynucleotide kinase (NEB) in the presence of [^{32}P]-ATP according to the manufacturer's instructions. Total RNA (10μg) extracted by the Trizol-chloroform method was separated on a 12% polyacrylamide/8M urea/1×TBE gel followed by semi-dry electroblotting onto Hybond N+ nylon membranes (GE Healthcare). Membranes were cross-linked and pre-hybridized for 1 h in hybridization buffer (50% formamide, 0.5% SDS, 5×SSPE, 5×Denhardt's solution, and 20 μg/ml sheared, denatured, salmon sperm DNA). Then the membrane was hybridized in the same solution containing 10 pmol radio labeled probes overnight at 42°C. The membrane was washed 3-times with 4×SSC at ambient temperature for 10 min. The membrane was wrapped in plastic wrap (Saran) and placed in a cassette with Kodak MS film at -80°C overnight.

SILAC-based yeast proteomic analysis

Light SILAC-labeled WT and *trm9Δ* yeast cells as well as heavy-labeled *lys1Δ* yeast cells were disrupted in an alkaline buffer (2 M NaOH, 8% v/v 2-mercaptoethanol), and proteins were

isolated by TCA precipitation. Yeast proteins were pelleted by centrifugation for 15 min at 15,000 × g at 4°C, and resuspended in lysis buffer (8 M urea, 75 mM NaCl, 50 mM Tris, pH 8.2, 50 mM NaF, 50 mM β-glycerophosphate, 1 mM sodium orthovanadate, 10 mM sodium pyrophosphate, 1 mM PMSF). Protein concentration was determined using the Bradford assay. This *lys1Δ* yeast protein extract was used as global internal standard and mixed (1:1, w/w) with WT and *trm9Δ* proteins separately.

The protein mixture was reduced for 2.5 h at 37°C in 1 mM dithiothreitol (DTT), alkylated for 40 min by 5.5 mM iodoacetamide (IAA) at ambient temperature in the dark, and then digested with 50:1 (w/w) endoproteinase lys-C overnight at 37°C. Peptide mixtures were fractionated into 24 fractions according to their isoelectric point using Agilent 3100 OFFGEL Fractionator (Agilent). Each peptide fraction was acidified by adding 0.1% formic acid, and loaded onto a C18 trap column (200Å Magic C18 AQ 5μm, 150 μm × 10 mm) at flow rate of 5 μl/min, with subsequent elution from a coupled analytic column (200Å Magic C18 AQ 5 μm, 75 μm × 150 mm) at 200 nl/min using a 2–98% acetonitrile gradient (180 min) in 0.1% formic acid. Eluted peptides were analyzed on a QSTAR-XL (Applied Biosystems/MDS Sciex) mass spectrometer. Three technical replicates were performed for each sample.

Acquired MS/MS spectra were parsed by Spectrum Mill (Agilent) and searched against the Saccharomyces Genome Database (SGD). SILAC peptide and protein quantitation was performed with DEQ (Differential Expression Quantitation). SILAC protein ratios are determined as the average of all peptide ratios assigned to the protein, and the proteins quantified in at least two replicates of the sample are recruited for further study. Differential protein expression was determined by Student's t-test between different samples.

Gene-specific codon usage

Using protein-coding sequences for 5886 *Saccharomyces cerevisiae* genes downloaded from the SGD database (http://www.yeastgenome.org/), we applied custom scripts to determine the gene-specific codon frequency in terms of the number of a particular codon per thousand codons in the open reading frame. Whether a gene was over- or under-represented with a specific codon relative to the genome average was determined by calculating a Z-score based on a hypergeometric distribution with a cut-off of $p < 0.01$. Gene-specific codon usage data were analyzed by hierarchical clustering using Cluster 3.0, and visualized as a heat map using Treeview [48]. Simulation and ORF shuffling were performed in R project using custom scripts. A spreadsheet containing the gene-specific codon usage data is presented in

AGA and GAA codon cluster

We investigated gene-specific clusters of AGA and GAA by calculating the frequency of AGA/GAA codons over a sliding window which was then subtracted the mean value of that gene. Similar to previous studies [49,50], a window size of 15 nt was used in this study. The data were plotted as a histogram, with positive peaks indicating clusters of AGA and GAA codons in these regions. The number of short codon runs in the form of triplets of AGA, GAA or their combinations in each gene was counted using custom scripts in R project.

Ribosomal footprinting dataset

Ribosomal footprinting data of the wildtype and *elp3Δ* cells as well as the RNA-seq data of the corresponding samples were obtained from a previous study (GSE45366) [24]. The reads per kilobase per million mapped reads (rpkms) of ribosomal footprinting data (FP) and the matched RNA-seq data (T) were used for comparison after normalization for library size. FP/T ratio for each gene was calculated to indicate ribosomal density on each transcript. A cut-off of

>1.2 fold increase in FP/T ratio was used to determine whether ribosomal density on a certain gene was enhanced in cells lacking mcm^{5}U and mcm^{5}s^{2}U.

Supporting Information

S1 Fig. Identification of changes in tRNA modifications and protein expression in *trm9Δ* cells. (**A**) The chromatograms indicate that depletion of Trm9 results in a total loss of the modified ribonucleosides mcm^{5}U and mcm^{5}s^{2}U in yeast cells under normal conditions and (**B**) in response to MMS treatment. Modifications were identified by HPLC retention time and collision-induced dissociation fragmentation patterns. (**C**) Northern-blot analysis of tRNA abundance in wild-type (WT) and *trm9Δ* cells under normal condition and in response to MMS treatment. (**D**) Relative standard deviation (RSD) of SILAC ratios between three replicates of each sample. RSD is the absolute value of the coefficient of variation calculated by dividing the standard deviation by the mean value. The median RSD for SILAC quantification was 0.068. (**E**) Hierarchical clustering of SILAC ratios (which indicate relative protein abundance) of all proteins identified in replicates of WT and *trm9Δ* cells under normal and stress conditions. (JPG)

S2 Fig. Proteins enriched with AGA and GAA codons are preferentially down-regulated in *trm9Δ* cells in response to MMS treatment. (**A**) Plot of D/U ratio versus the number of significantly down-regulated proteins in each group of proteins enriched with one certain codon is shown. The red dots represent codons dependent on the wobble modifications, while grey dots represent data of other codon groups. (**B**) Plot of D/U ratio versus the percentage of significantly down-regulated proteins in each group. The red line indicates the percentage of significantly down-regulated proteins in total proteins. (**C**) To control for biased usage of codons other than AGA (except for GAA), proteins enriched with other codons were removed from the AGA-enriched group, one at a time. Proteins enriched with AGA were likewise removed from other codon groups. The values of the proteins enriched with AGA to the exclusion of other codons were separately plotted in red dots, while the values of the proteins enriched with other codons, after removing AGA-enriched proteins, were plotted in grey dots. (**D**) Analysis similar to (**C**) repeated for proteins enriched with GAA codon. (JPG)

S3 Fig. Depletion of mcm^5 and mcm^{5}s^2 did not selectively repress expression of proteins enriched with other codons. (**A**) The percentage of down-regulated proteins of CAA-enriched proteins *vs* non-enriched proteins. The histogram shows the distribution of the percentages obtained by 100,000 random samplings of the non-enriched proteins. The position of CAA-enriched proteins to that distribution is indicated by arrow, with the p-values indicating the chance that a random sampling has a value no less than that of the CAA-enriched group. Similar results were observed for (**B**) GGA-, (**C**) AAA-, (**D**) AGG- and (**E**) GGG-enriched proteins. The D/U ratio of (**F**) CAA-, (**G**) GGA-, (**H**) AAA-, (**I**) AGG- and (**J**) GGG-enriched proteins. (JPG)

S4 Fig. Controlling for changes in mRNA expression, depletion of mcm^5 and mcm^{5}s^2 still selectively repress expression of proteins enriched with AGA/GAA codons. Proteins whose changes in protein expression could possibly be explained by altered mRNA expression in *trm9Δ* cells were excluded from the analysis. (**A**) The percentage of down-regulated proteins of AGA-enriched proteins *vs* non-enriched proteins. The histogram shows the distribution of the percentages obtained by 100,000 random samplings of the non-enriched proteins. The position of AGA-enriched proteins to that distribution is indicated by arrow, with the p-values indicating the chance that a random sampling has a value no less than that of the AGA-enriched

group. (**B**) Similar analysis performed for GAA-enriched proteins.
(JPG)

S5 Fig. Controlling for protein expression, proteins enriched with AGA and GAA codons are still selectively down-regulated in *trm9Δ* cells. (**A**) A comparison of expression levels for AGA-enriched proteins *vs* non-enriched proteins. Protein abundance was estimated using summed peptide intensity from the mass spectrometric analysis. P-value was calculated using student's t-test. (**B**) A comparison of expression levels for AGA-enriched proteins *vs* non-enriched proteins. P-value was calculated using student's t-test. (**C**) D/U ratios of AGA-enriched proteins *vs* non-enriched proteins controlling for protein expression. Proteins were binned into 10 groups based on expression level, and random sampling of non-enriched population were performed keeping the composition of proteins with different expression levels identical to that of AGA-enriched group. The histogram shows the distribution of D/U ratios of the non-enriched proteins obtained by 100,000 random samplings, the positions of AGA-enriched proteins are indicated by arrow, with the p-values indicating the chance that a random sampling has a ratio no less than that of the AGA/GAA-enriched group. (**D**) D/U ratios of GAA-enriched proteins *vs* non-enriched proteins controlling for protein expression, as described above.
(JPG)

S6 Fig. Proteins enriched with AGA and GAA codons are selectively down-regulated in *trm9Δ* cells in response to MMS treatment. (**A**) Proteins were divided into seven groups based on AGA codon frequency or (**B**) GAA codon frequency, and D/U ratios were calculated in each group, respectively. The plots show a positive correlation between down-regulation and AGA or GAA codon content. (**C**) D/U ratios of AGA-enriched proteins *vs* non-enriched proteins. The histogram shows the distribution of D/U ratios of the non-enriched proteins obtained by 100,000 random samplings, the positions of AGA-enriched proteins are indicated by arrow, with the p-values indicating the chance that a random sampling has a ratio no less than that of the AGA/GAA-enriched group. (**D**) D/U ratios of GAA-enriched proteins *vs* non-enriched proteins, as described above.
(JPG)

S7 Fig. AGA/GAA codon runs occur more frequently than expected by chance. (**A**) Boxplots show the number of 4-mer codon runs identified in each gene (actual) *vs* that of 10,000 shufflsed sequences per gene maintaining the codon composition of each gene (randomized). P-value was calculated by Mann-Whitney U test. (**B**) Similar to (A), but 5-mer codon runs were considered.
(JPG)

S8 Fig. Biological processes affected by Trm9-dependent proteins under normal growth conditions and following MMS treatment. Gene Ontology category enrichment was determined for down- or up-regulated in *trm9Δ* cells compared to wild-type cells under normal growth conditions and following MMS treatment.
(JPG)

S1 Table. Proteins quantified in the SILAC proteomics analysis of wild-type and *trm9Δ* yeast.
(XLSX)

S2 Table. Significantly up- and down-regulated proteins in *trm9Δ* cells under normal growth conditions and following DNA damage stress with MMS.
(XLSX)

S3 Table. Codon usage data for individual genes in *S. cerevisiae*.
(XLSX)

S4 Table. Function categories of differentially expressed proteins enriched with Arg[AGA] **and/or Glu**[GAA] **codons in *trm9Δ* cell in normal and MMS conditions**
(DOCX)

S5 Table. Function categories of differentially expressed proteins in *trm9Δ* cell in normal and MMS conditions.
(DOCX)

S6 Table. Comparison of published studies of wobble U modifications.
(XLSX)

S7 Table. Proteomic analysis in *trm9Δ* cells of genes conferring resistance to MMS exposure.
(XLSX)

Acknowledgments

Author Contributions

Conceived and designed the experiments: WD IRB TJB PCD. Performed the experiments: WD IRB DS SY. Analyzed the data: DS SY TJB PCD. Wrote the paper: WD IRB DS SY TJB PCD.

References

1. Maraia RJ, Blewett NH, Bayfield MA (2008) It's a mod mod tRNA world. Nat Chem Biol 4: 162–164. doi: 10.1038/nchembio0308-162 PMID: 18277975

2. Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L (2012) A role for tRNA modifications in genome structure and codon usage. Cell 149: 202–213. doi: 10.1016/j.cell.2012.01.050 PMID: 22464330

3. Phizicky EM, Hopper AK (2010) tRNA biology charges to the front. Genes Dev 24: 1832–1860. doi: 10.1101/gad.1956510 PMID: 20810645

4. Johansson MJ, Esberg A, Huang B, Bjork GR, Bystrom AS (2008) Eukaryotic wobble uridine modifications promote a functionally redundant decoding system. Mol Cell Biol 28: 3301–3312. doi: 10.1128/MCB.01542-07 PMID: 18332122

5. Chechetkin VR (2006) Genetic code from tRNA point of view. J Theor Biol 242: 922–934. PMID: 16808928

6. Alexandrov A, Chernyakov I, Gu W, Hiley SL, Hughes TR, et al. (2006) Rapid tRNA decay can result from lack of nonessential modifications. Mol Cell 21: 87–96. PMID: 16387656

7. Yarian C, Townsend H, Czestkowski W, Sochacka E, Malkiewicz AJ, et al. (2002) Accurate translation of the genetic code depends on tRNA modified nucleosides. J Biol Chem 277: 16391–16395. PMID: 11861649

8. Leipuviene R, Bjork GR (2005) A reduced level of charged tRNAArgmnm5UCU triggers the wild-type peptidyl-tRNA to frameshift. RNA 11: 796–807. PMID: 15840821

9. Rezgui VA, Tyagi K, Ranjan N, Konevega AL, Mittelstaet J, et al. (2013) tRNA tKUUU, tQUUG, and tEUUC wobble position modifications fine-tune protein translation by promoting ribosome A-site binding. Proc Natl Acad Sci U S A 110: 12289–12294. doi: 10.1073/pnas.1300781110 PMID: 23836657

10. Bauer F, Matsuyama A, Candiracci J, Dieu M, Scheliga J, et al. (2012) Translational control of cell division by Elongator. Cell Rep 1: 424–433. PMID: 22768388

11. Kalhor HR, Clarke S (2003) Novel methyltransferase for modified uridine residues at the wobble position of tRNA. Mol Cell Biol 23: 9283–9292. PMID: 14645538

12. Noma A, Sakaguchi Y, Suzuki T (2009) Mechanistic characterization of the sulfur-relay system for eukaryotic 2-thiouridine biogenesis at tRNA wobble positions. Nucleic Acids Res 37: 1335–1352. doi: 10.1093/nar/gkn1023 PMID: 19151091

13. Mazauric MH, Dirick L, Purushothaman SK, Bjork GR, Lapeyre B (2010) Trm112p is a 15-kDa zinc finger protein essential for the activity of two tRNA and one protein methyltransferases in yeast. J Biol Chem 285: 18505–18515. doi: 10.1074/jbc.M110.113100 PMID: 20400505

14. Huang B, Lu J, Bystrom AS (2008) A genome-wide screen identifies genes required for formation of the wobble nucleoside 5-methoxycarbonylmethyl-2-thiouridine in Saccharomyces cerevisiae. RNA 14: 2183–2194. doi: 10.1261/rna.1184108 PMID: 18755837

15. Begley U, Sosa MS, Avivar-Valderas A, Patil A, Endres L, et al. (2013) A human tRNA methyltransferase 9-like protein prevents tumour growth by regulating LIN9 and HIF1-alpha. EMBO Mol Med. 5: 366–383. doi: 10.1002/emmm.201201161 PMID: 23381944

16. Chan CT, Dyavaiah M, DeMott MS, Taghizadeh K, Dedon PC, et al. (2010) A quantitative systems approach reveals dynamic control of tRNA modifications during cellular stress. PLoS Genet 6: e1001247. doi: 10.1371/journal.pgen.1001247 PMID: 21187895

17. Chan CTY, Pang YLJ, Deng W, Babu IR, Dyavaiah M, et al. (2012) Reprogramming of tRNA modifications controls the oxidative stress response by codon-biased translation of proteins. Nat Commun 3: 937. doi: 10.1038/ncomms1938 PMID: 22760636

18. Fu D, Brophy JA, Chan CT, Atmore KA, Begley U, et al. (2010) Human AlkB homolog ABH8 Is a tRNA methyltransferase required for wobble uridine modification and DNA damage survival. Mol Cell Biol 30: 2449–2459. doi: 10.1128/MCB.01604-09 PMID: 20308323

19. Patil A, Chan CT, Dyavaiah M, Rooney JP, Dedon PC, et al. (2012) Translational infidelity-induced protein stress results from a deficiency in Trm9-catalyzed tRNA modifications. RNA Biol 9: 990–1001. doi: 10.4161/rna.20531 PMID: 22832247

20. Begley U, Dyavaiah M, Patil A, Rooney JP, DiRenzo D, et al. (2007) Trm9-catalyzed tRNA modifications link translation to the DNA damage response. Mol Cell 28: 860–870. PMID: 18082610

21. Begley TJ, Rosenbach AS, Ideker T, Samson LD (2004) Hot spots for modulating toxicity identified by genomic phenotyping and localization mapping. Mol Cell 16: 117–125. PMID: 15469827

22. Begley TJ, Rosenbach AS, Ideker T, Samson LD (2002) Recovery Pathways in S. cerevisiae Revealed by Genomic Phenotyping and Interactome Mapping. Mol Canc Res 1.

23. Chan CT, Deng W, Li F, DeMott MS, Babu IR, et al. (2015) Highly Predictive Reprogramming of tRNA Modifications Is Linked to Selective Expression of Codon-Biased Genes. Chem Res Toxicol 28: 978–988. doi: 10.1021/acs.chemrestox.5b00004 PMID: 25772370

24. Zinshteyn B, Gilbert WV (2013) Loss of a conserved tRNA anticodon modification perturbs cellular signaling. PLoS Genet 9: e1003675. doi: 10.1371/journal.pgen.1003675 PMID: 23935536

25. Romero IG, Ruvinsky I, Gilad Y (2012) Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet 13: 505–516. doi: 10.1038/nrg3229 PMID: 22705669

26. Warnecke T, Hurst LD (2011) Error prevention and mitigation as forces in the evolution of genes and genomes. Nat Rev Genet 12: 875–881. doi: 10.1038/nrg3092 PMID: 22094950

27. Deng WJ, Nie S, Dai J, Wu JR, Zeng R (2010) Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages. Mol Cell Proteomics 9: 100–116. doi: 10.1074/mcp.M900020-MCP200 PMID: 19700791

28. Parmley JL, Urrutia AO, Potrzebowski L, Kaessmann H, Hurst LD (2007) Splicing and the evolution of proteins in mammals. PLoS Biol 5: e14. PMID: 17298171

29. Schwartz S, Ast G (2010) Chromatin density and splicing destiny: on the cross-talk between chromatin structure and splicing. EMBO J 29: 1629–1636. doi: 10.1038/emboj.2010.71 PMID: 20407423

30. Lei H, Zhai B, Yin S, Gygi S, Reed R (2013) Evidence that a consensus element found in naturally intronless mRNAs promotes mRNA export. Nucleic Acids Res 41: 2517–2525. doi: 10.1093/nar/gks1314 PMID: 23275560

31. Wang P, Yin S, Zhang Z, Xin D, Hu L, et al. (2008) Evidence for common short natural trans sense-antisense pairing between transcripts from protein coding genes. Genome Biol 9: R169. doi: 10.1186/gb-2008-9-12-r169 PMID: 19055728

32. Yin S, Deng W, Zheng H, Zhang Z, Hu L, et al. (2009) Evidence that the nonsense-mediated mRNA decay pathway participates in X chromosome dosage compensation in mammals. Biochem Biophys Res Commun 383: 378–382. doi: 10.1016/j.bbrc.2009.04.021 PMID: 19364502

33. Yin S, Wang P, Deng W, Zheng H, Hu L, et al. (2009) Dosage compensation on the active X chromosome minimizes transcriptional noise of X-linked genes in mammals. Genome Biol 10: R74. doi: 10.1186/gb-2009-10-7-r74 PMID: 19594925

34. Warnecke T, Hurst LD (2007) Evidence for a trade-off between translational efficiency and splicing regulation in determining synonymous codon usage in Drosophila melanogaster. Mol Biol Evol 24: 2755–2762. PMID: 17905999

35. Nedialkova DD, Leidel SA (2015) Optimization of Codon Translation Rates via tRNA Modifications Maintains Proteome Integrity. Cell 161: 1606–1618. doi: 10.1016/j.cell.2015.05.022 PMID: 26052047

36. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4: 44–57. doi: 10.1038/nprot.2008.211 PMID: 19131956

37. Ishihama Y, Sato T, Tabata T, Miyamoto N, Sagane K, et al. (2005) Quantitative mouse brain proteomics using culture-derived isotope tags as internal standards. Nat Biotechnol 23: 617–621. PMID: 15834404

38. Kristiansson MH, Bhat VB, Babu IR, Wishnok JS, Tannenbaum SR (2007) Comparative time-dependent analysis of potential inflammation biomarkers in lymphoma-bearing SJL mice. J Proteome Res 6: 1735–1744. PMID: 17388619

39. Bjork GR, Huang B, Persson OP, Bystrom AS (2007) A conserved modified wobble nucleoside (mcm5s2U) in lysyl-tRNA is required for viability in yeast. RNA 13: 1245–1255. PMID: 17592039

40. Parmley JL, Huynen MA (2009) Clustering of codons with rare cognate tRNAs in human genes suggests an extra level of expression regulation. PLoS Genet 5: e1000548. doi: 10.1371/journal.pgen.1000548 PMID: 19578405

41. Warnecke T, Hurst LD (2010) GroEL dependency affects codon usage—support for a critical role of misfolding in gene evolution. Mol Syst Biol 6: 340. doi: 10.1038/msb.2009.94 PMID: 20087338

42. Cannarozzi G, Schraudolph NN, Faty M, von Rohr P, Friberg MT, et al. (2010) A role for codon order in translation dynamics. Cell 141: 355–367. doi: 10.1016/j.cell.2010.02.036 PMID: 20403329

43. Clarke TFt, Clark PL (2008) Rare codons cluster. PLoS One 3: e3412. doi: 10.1371/journal.pone.0003412 PMID: 18923675

44. Chen C, Tuck S, Bystrom AS (2009) Defects in tRNA modification associated with neurological and developmental dysfunctions in Caenorhabditis elegans elongator mutants. PLoS Genet 5: e1000561. doi: 10.1371/journal.pgen.1000561 PMID: 19593383

45. Jackson R, Standart N (2015) The awesome power of ribosome profiling. RNA 21: 652–654. doi: 10.1261/rna.049908.115 PMID: 25780177

46. Charneski CA, Hurst LD (2013) Positively charged residues are the major determinants of ribosomal velocity. PLoS Biol 11: e1001508. doi: 10.1371/journal.pbio.1001508 PMID: 23554576

47. Komili S, Farny NG, Roth FP, Silver PA (2007) Functional specificity among ribosomal proteins regulates gene expression. Cell 131: 557–571. PMID: 17981122

48. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. Bioinformatics 20: 1453–1454. PMID: 14871861

49. Tuller T, Carmi A, Vestsigian K, Navon S, Dorfan Y, et al. (2010) An evolutionarily conserved mechanism for controlling the efficiency of protein translation. Cell 141: 344–354. doi: 10.1016/j.cell.2010.03.031 PMID: 20403328

50. 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. doi: 10.1126/science.1168978 PMID: 19213877