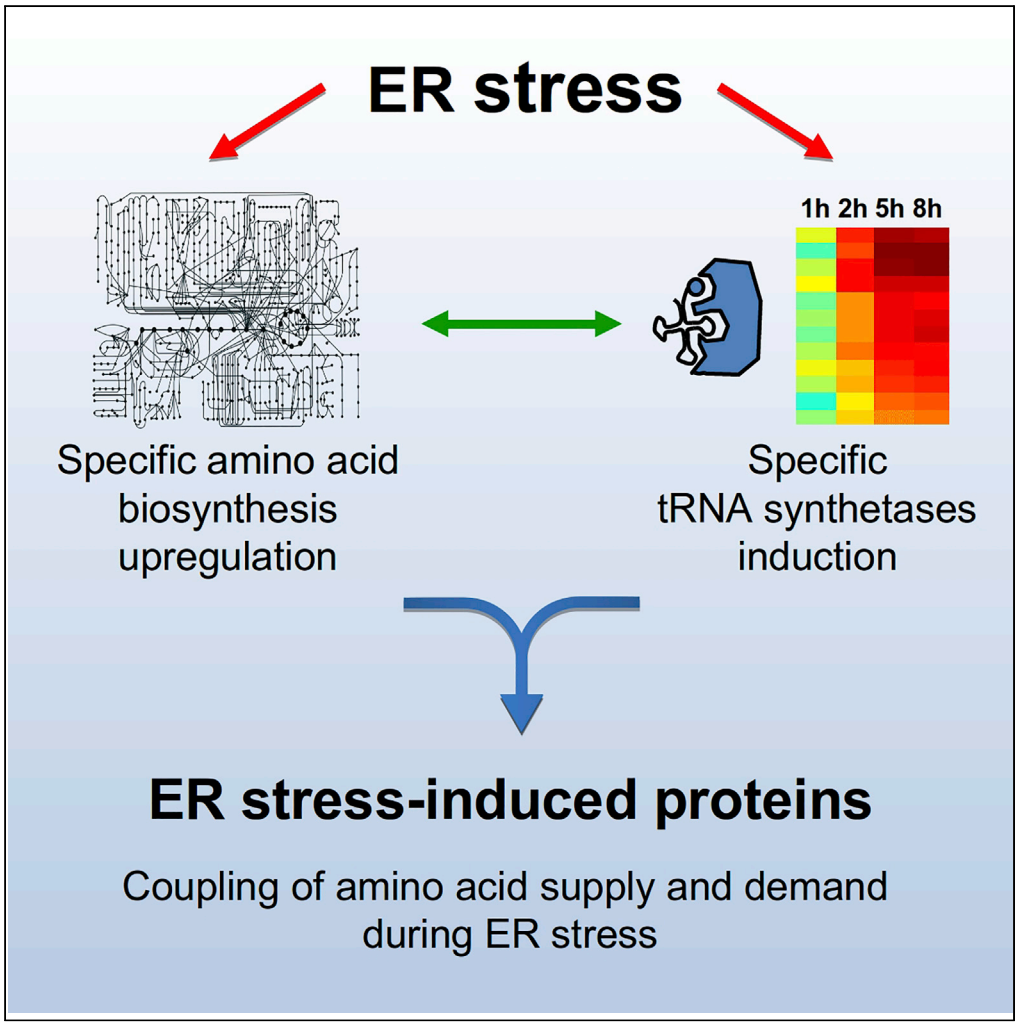
iScience

CellPress

Article

Amino Acid Biosynthesis Regulation during Endoplasmic Reticulum Stress Is Coupled to Protein Expression Demands



Nir Gonen, Anatoly Meller, Niv Sabath, Reut Shalgi

reutshalgi@technion.ac.il

HIGHLIGHTS

Coordination of amino acid supply and protein synthesis demand during ER stress

Specific amino acid biosynthesis and cognate tRNA synthetases induction by the UPR

UPR-induced amino acids support amino acid demand of UPR-upregulated proteins

UPR-induced amino acids are highly enriched within secreted proteins

DATA AND CODE AVAILABILITY

GSE118660
GSE90070

Gonen et al., iScience *19*, 204–213
September 27, 2019
https://doi.org/10.1016/j.isci.2019.07.022

iScience

CellPress

Article

Amino Acid Biosynthesis Regulation during Endoplasmic Reticulum Stress Is Coupled to Protein Expression Demands

Nir Gonen,[1,2] Anatoly Meller,[1,2] Niv Sabath,[1] and Reut Shalgi[1,3,*]

SUMMARY

The endoplasmic reticulum (ER) stress response, also known as the unfolded protein response (UPR), is a complex cellular response to ER protein misfolding that involves transcriptional regulatory branches and a PERK-mediated translational regulatory branch.

Here we revealed that amino acid biosynthesis regulation is coupled to protein synthesis demands during ER stress. Specifically, we demonstrated that the UPR leads to PERK-dependent induction in the biosynthesis of specific amino acids, and to upregulation of their corresponding tRNA synthetases. Furthermore, we found that sequences of UPR-upregulated proteins are significantly enriched with these UPR-induced amino acids. Interestingly, whereas the UPR leads to repression of ER target proteins, we showed that secreted proteins tended to escape this repression and were highly enriched for the UPR-induced amino acids.

Our results unravel coordination between amino acid supply, namely, biosynthesis and tRNA loading, and demand from UPR-induced proteins under ER stress, thus revealing an additional regulatory layer of protein synthesis.

INTRODUCTION

Endoplasmic reticulum (ER) stress elicits a complex cellular program, also termed the *unfolded protein response* (UPR). This response enables cells to cope with dynamic changes in protein folding and processing demands in the ER (Ron and Walter, 2007). The UPR in mammalian cells consists of three major branches: the PERK branch, which leads to global inhibition of translation initiation through eIF2α phosphorylation, and two transcriptional branches, namely, the IRE1-XBP1 and the ATF6 branches (Pavitt and Ron, 2012). Importantly, one of the major secondary effectors downstream of PERK is the transcription factor ATF4, the translation of which is highly induced in an eIF2α phosphorylation-dependent manner during ER stress (Vattem and Wek, 2004) and is known to induce the transcription of genes essential for ER stress adaptation (Rutkowski and Kaufman, 2007).

We and others have characterized another hallmark of the UPR: enhanced repression of ER-targeted proteins (Gonen et al., 2019; Guan et al., 2017; Reid et al., 2014). We found that this repression is mainly at the level of translation and occurs in many cell types (Gonen et al., 2019). Furthermore, using ribosome footprint profiling, we identified three major ER stress gene expression programs in mouse embryonic fibroblasts (MEFs): early induction, late induction, and repression, all of which are PERK dependent (Gonen et al., 2019). Notably, late induction genes are enriched for amino acid biosynthesis pathways (Gonen et al., 2019).

Amino acid biosynthesis genes were previously shown to be induced in response to ER stress. Harding et al. demonstrated that amino acid metabolism is upregulated in response to ER stress, in an ATF4-dependent manner (Harding et al., 2003). Furthermore, ATF4−/− MEFs were found to have an amino acid deficiency, and thus require supplemental non-essential amino acids for their growth (Harding et al., 2003). ATF4−/− cells have also been shown to experience oxidative stress, and supplementation of cysteine, glutathione (GSH), or other reducing agents rescues their growth defects (Harding et al., 2003). In chronic ER stress, ATF4 binds to the promoters of various amino acid biosynthesis genes, and also to amino acid transporters (Han et al., 2013). Moreover, serine metabolism induction was highlighted in response to ER stress (Rendleman et al., 2018). In addition, ATF4 target genes were found to be enriched with tRNA synthetases (Han et al., 2013). In pancreatic β-cells, increased protein synthesis during chronic ER stress involved the transcriptional induction of an amino acid transporter network, and of tRNA synthetases, in an

[1]Department of Biochemistry, Rappaport Faculty of Medicine, Technion–Israel Institute of Technology, Haifa 31096, Israel

[2]These authors contributed equally

[3]Lead Contact

*Correspondence: reutshalgi@technion.ac.il

https://doi.org/10.1016/j.isci.2019.07.022

ATF4-dependent manner (Krokowski et al., 2013). The ATF4-induced network of amino acid transporters has also been shown to be involved in alleviating the inhibition of protein synthesis during prolonged ER stress, by promoting mTOR activation (Guan et al., 2014). In turn, mTOR can contribute to ATF4 expression in an eIF2α phosphorylation-independent manner, by mediating ATF4 mRNA stabilization (Ben-Sahra et al., 2016; Park et al., 2017).

Here, we set out to further characterize the late response to ER stress. Using ribosome footprint profiling in ER-stressed wild-type (WT) and PERK−/− MEFs (treated with thapsigargin [Tg]) (Gonen et al., 2019), we found that ER stress, particularly a late-specific ER stress induction gene expression program, is highly enriched in two categories: amino acid biosynthesis and tRNA synthetases, in line with previous studies. Herein, we observed induction in the biosynthesis of a specific subset of amino acids, with highly significant congruent upregulation of their corresponding tRNA synthetases. This observation led us to hypothesize that this coordinated induction during ER stress might be coupled to differential demands of amino acids from the amino acid sequences of the UPR-induced proteins. Indeed, our analyses showed that sequences of induced proteins are significantly enriched with these amino acids. Finally, we showed that this UPR amino acid signature is preferentially enriched within subclasses of ER targets that tended to escape the enhanced repression during ER stress, namely, secreted and extracellular matrix proteins.

RESULTS

Secreted Proteins Escape the Repression of ER Targets during ER Stress

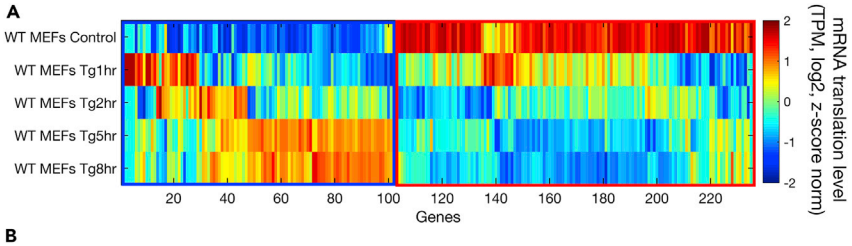
Previously, we and others (Gonen et al., 2019; Guan et al., 2017; Reid et al., 2014) described widespread enhanced repression of ER-targeted proteins during ER stress. Furthermore, we showed that this repression occurs mainly at the level of translation and starts as early as 1 h into the stress, thus affecting the expression of hundreds of mRNAs encoding ER targets (Gonen et al., 2019). These include membrane proteins, glycoproteins, and disulfide bond-containing proteins (Gonen et al., 2019).

We therefore wished to perform a more fine-tuned analysis of expressed ER targets. One of the major mechanisms of ER localization is that of the signal recognition particle, which is responsible for the localization and co-translational import of many ER processed proteins that contain a signal peptide in their N terminus (Keenan et al., 2001). We therefore utilized our ribosome footprint profiling dataset of MEFs that were subjected to short and long ER stress induced by the SERCA inhibitor Tg (Gonen et al., 2019) and aimed to examine all signal peptide-encoding mRNAs (see Methods). We performed hierarchical clustering of the expression levels of the subset of signal peptide-encoding mRNAs that changed at least 2-fold during a time course of short and long ER stress treatments. Our analysis revealed that whereas about 57% of the signal peptide-containing proteins were significantly repressed, the remaining proteins were actually induced (Figure 1A). We then sought to identify the functional differences between the repressed and the enhanced signal peptide-encoding mRNAs. We observed that over 60% of the repressed signal peptide-encoding mRNAs were membrane proteins, consistent with our previous results (Gonen et al., 2019). On the other hand, signal peptide-encoding mRNAs with induced expression during ER stress were selectively enriched for extracellular region and secreted proteins (Figure 1B). Therefore, it seems that different subsets of ER targets are regulated differently, with membrane proteins being selectively repressed, whereas many secreted proteins escaping this repression and being actually enhanced.

Late-Specific ER Stress-Induced Genes Include Amino Acid Biosynthesis Genes and tRNA Synthetases

Next, we set out to further characterize the candidate genes important for adaptation to ER stress, and whose regulation is specific to the late stages of the response. To this end, we performed a differential expression analysis to identify genes that vary in their expression between early (1 and 2 h) and late (5 and 8 h) ER stress treatments (see Methods). The analysis resulted in 120 significantly changing mRNAs, of which 60% were induced at the late time points (Figure 2A). Interestingly, both specific late-induced and late-repressed gene sets were PERK dependent, as their induction or repression in PERK−/− cells was highly impaired (Figures 2A and S1A–S1D). Reassuringly, functional enrichment analysis revealed that the induced gene set was enriched for genes involved in the ER stress response (Figure 2B). Furthermore, the late-specific induced gene set was also highly enriched with two additional functional categories: amino acid biosynthesis and tRNA synthetases (Figure 2B). This result is in agreement with our previously identified late ER stress gene induction program, which was enriched for response to ER stress and amino acid biosynthesis (Gonen et al., 2019), as well as with earlier reports of changes in amino acid biosynthesis

A



B

Category	Term	Count	FDR
GOTERM_CC_DIRECT	GO:0005576~extracellular region	59	0.001251
UP_KEYWORDS	Secreted	56	0.001374

Figure 1. Analysis of Signal Peptide Encoding mRNAs Reveals that Secreted Proteins Tend to Escape Repression upon ER Stress

(A) Hierarchical clustering analysis was performed using 236 signal peptide-encoding mRNAs, with translation change of at least 2-fold compared with control in any of the time points examined following Tg treatment (1, 2, 5, and 8 h) in WT MEFs. The heatmap depicts hierarchal clustering of mRNA translation levels (transcripts per million [TPM] of ribosome footprint profiling in the respective samples, log2), according to Spearman's correlation. *Z* score normalization was further performed for visualization purposes. The mRNAs were clustered into two distinct groups: 102 induced mRNAs (blue square) and 134 repressed mRNAs (red square) following Tg treatment.

(B) Functional enrichment of induced signal peptide encoding genes was analyzed using DAVID (Huang da et al., 2009). The enriched terms were extracellular region and secreted proteins. False discovery rate (FDR)-corrected p values are indicated.

using mRNA expression profiling (Harding et al., 2003). In addition, specific amino acids, such as cysteine and leucine, were shown to be essential for cell survival under prolonged ER stress (Harding et al., 2000, 2003). Interestingly, we found that 25% of the late-specific induced mRNAs were ATF4 targets ($p = 4.4 \times 10^{-49}$, see Methods). Indeed, ATF4 transcriptional output is known to be a component of the adaptive UPR, which occurs at the late time points.

Biosynthesis of a Specific Subset of Amino Acids Is Induced in the Late ER Stress Response

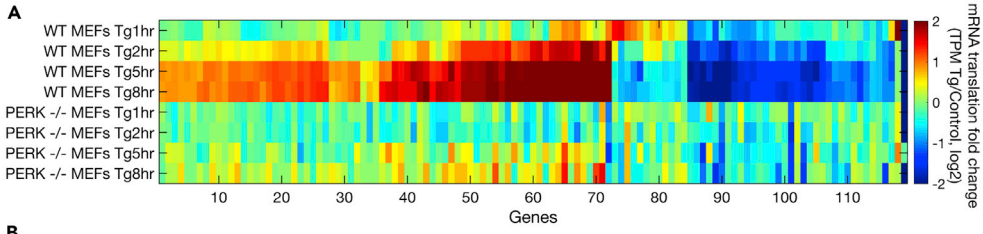
After observing that amino acid biosynthesis pathways are enriched within late-specific ER stress-induced genes, we analyzed the expression patterns of all mRNAs related to amino acid biosynthetic pathways (see Methods). Although overall, amino acid biosynthesis pathway mRNAs showed significant, PERK-dependent, induction in ER stress (Figures S2A and S2B), hierarchical clustering analysis showed that only about half of these mRNAs were induced, whereas the rest were largely unchanged (Figure S2E). To distinguish the pathways that were induced, we examined the amino acid biosynthesis metabolic map (see Methods) and overlaid on it the data of the late-specific induced mRNA set that was characterized above (Figure 2A). Using this map, we identified a subset of eight non- or partly essential amino acids that showed induction of their related biosynthetic pathways: Ser, Cys, Gly, Ala, Asn, Asp, Glu, and Pro (Figure 2C).

Significant Overlap between Amino Acids with ER Stress-Induced Biosynthesis and Upregulated tRNA Synthetases Defines a UPR Amino Acid Signature

Next, having observed that aminoacyl-tRNA synthetases were also enriched within late-specific ER stress-induced mRNAs, we examined the expression pattern of all tRNA synthetases. The overall induction of 37 tRNA synthetases was highly significant and PERK dependent, as tRNA synthetases were not induced in PERK−/− MEFs at any of the ER stress time points (Figures S2C, S2D, and 2D). Nevertheless, clustering analysis revealed that here too, not all tRNA synthetases were equally induced in WT MEFs (Figure 2D). Rather, a subset of 12 cytosolic tRNA synthetases was highly upregulated during ER stress (Figure 2D). Notably, these specific tRNA synthetases were not induced in PERK−/− MEFs (Figure 2D). Other tRNA synthetases were either slightly induced (a subset of mitochondrial tRNA synthetases) or remained unchanged. Importantly, none of the tRNA synthetases was selectively repressed during ER stress.

Having identified a specific subset of eight amino acids whose biosynthesis was induced during the late stages of the ER stress response on one hand, and a subset of 12 highly upregulated tRNA synthetases

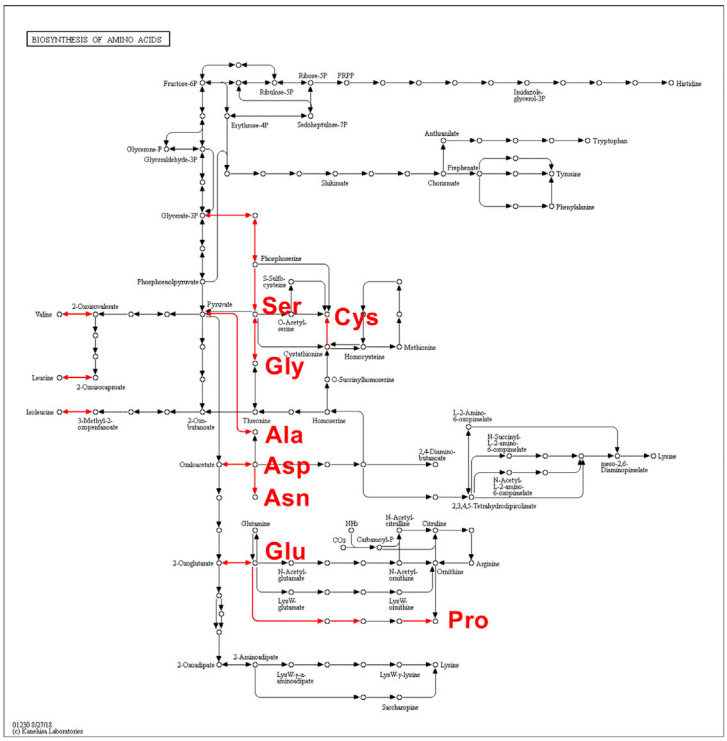
A



B

Category	Term	Count	FDR
UP_KEYWORDS	Amino-acid biosynthesis	10	1.3E-11
GOTERM_BP_DIRECT	GO:0008652~cellular amino acid biosynthetic process	10	4.2E-11
KEGG_PATHWAY	mmu01230:Biosynthesis of amino acids	10	5.4E-06
KEGG_PATHWAY	mmu00260:Glycine, serine and threonine metabolism	6	1.4E-03
GOTERM_BP_DIRECT	GO:0006564~L-serine biosynthetic process	4	2.2E-03
UP_KEYWORDS	Protein biosynthesis	8	2.7E-02
UP_KEYWORDS	Aminoacyl-tRNA synthetase	7	1.9E-04
GOTERM_BP_DIRECT	GO:0006418~tRNA aminoacylation for protein translation	7	2.6E-04
GOTERM_MF_DIRECT	GO:0004812~aminoacyl-tRNA ligase activity	7	4.1E-04
KEGG_PATHWAY	mmu00970:Aminoacyl-tRNA biosynthesis	7	9.3E-03
GOTERM_BP_DIRECT	GO:0034976~response to endoplasmic reticulum stress	8	7.0E-04

C



D

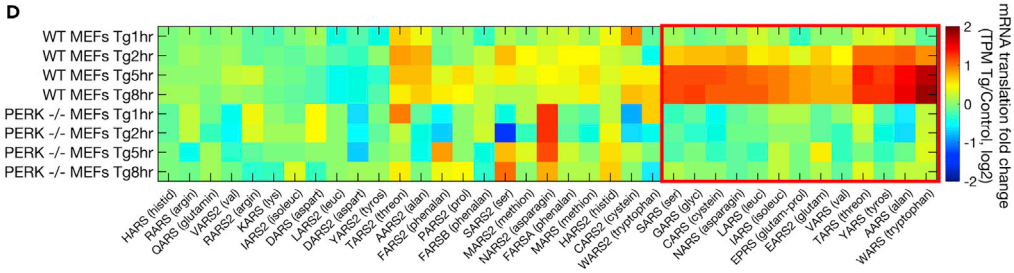


Figure 2. Late-Specific ER Stress-Induced mRNAs Include Specific Amino Acid Biosynthesis Pathways and Their Cognate tRNA Synthetases

(A) Hierarchical clustering analysis was performed using 120 mRNAs that were differentially translated (identified using DESeq2, FDR-corrected p value < 0.1) between the group of 1- and 2-h ER-stressed WT MEFs and a second group of

Figure 2. *Continued*

5- and 8-h treated WT MEFs. The heatmap depicts hierarchically clustered log2 fold change in translation levels (transcripts per million [TPM] of ribosome footprint profiling in the respective samples, Tg/control, log2), according to Spearman's correlation. Clustering was performed according to WT MEF samples. As shown, PERK−/− cells display very little change in these mRNAs.

(B) Functional enrichment of the late-specific induced genes was analyzed using DAVID (Huang da et al., 2009). False discovery rate (FDR)-corrected p values are indicated.

(C) An amino acid biosynthesis metabolic map that was generated using the KEGG mapper (Kanehisa and Goto, 2000) of late-specific ER stress-induced genes shows induction of biosynthesis pathways (induced reaction arrows are colored in red) for eight non- or partly essential amino acids: Ser, Cys, Gly, Ala, Asp, Asn, Glu, and Pro.

(D) Hierarchical clustering analysis of translation log2 fold changes (as described in A) for the group of all tRNA synthetases, identified a cluster of ER stress-induced, PERK-dependent, tRNA synthetases (marked by a red square).

on the other hand, we were prompted to examine the overlap between these two subsets. We found that seven of the eight amino acids whose biosynthesis was induced, showed upregulation of their cognate tRNA synthetase: Ser, Cys, Gly, Ala, Asn, Glu, and Pro (Table S1, p value for the overlap = 0.0012, using the hypergeometric test). We thus termed this subset of seven amino acids with both ER stress-induced biosynthesis and upregulated tRNA synthetases the UPR amino acid signature.

The other induced tRNA synthetases were related to essential amino acids. Indeed, we observed that several amino acid transporters were highly induced during the late stages of ER stress (Figure S3), in line with previous reports (Krokowski et al., 2013). However, as various transporters can import a broad range of amino acids, we decided not to include the essential amino acids in this signature.

The UPR Amino Acid Signature Is Coupled to Differential Amino Acid Demand from ER Stress-Induced Proteins

The identification of a UPR amino acid signature led us to hypothesize that the coordinated induction of biosynthesis of a specific subset of amino acids and their cognate tRNA synthetases might be coupled to changes in the ER stress cellular translatome. More specifically, we hypothesized that induction in the biosynthesis of the identified UPR amino acid signature may correspond with the amino acid composition of proteins that comprise the ER stress induction program. To test this hypothesis, we defined a metric termed *amino acid demand*, which is simply the fraction of an amino acid, or a group of amino acids, within the sequence of each protein. We calculated the UPR amino acid demand as the demand of the set of the seven amino acids comprising the UPR amino acid signature, in each protein sequence in the mouse genome. We then examined the amino acid demand within the three ER stress gene expression programs that we previously identified (Gonen et al., 2019), namely, early UPR induction, late UPR induction, and UPR repression, and compared these with the background UPR amino acid demand of all expressed proteins (Figure 3A). We observed that whereas proteins belonging to the ER stress repression program were significantly depleted for the UPR amino acid signature ($p = 3.8 \times 10^{-5}$), both early and late induction program proteins were highly enriched for this amino acid set (Figure 3A, $p = 3 \times 10^{-4}$ and $p = 1.2 \times 10^{-4}$ for the early and late induction gene expression programs, respectively). Importantly, examination of the proteins that were induced or repressed by ER stress in the PERK−/− MEFs did not reveal enrichment or depletion of the UPR amino acid signature (Figure 3B).

To investigate whether this finding can be recapitulated in another dataset, we examined an ER stress dataset in which MEFs were stressed for 16 h using Tg, and translatome was mapped using polysome sequencing analysis (Guan et al., 2017). Here too, we identified a UPR amino acid signature relevant for a 16-h ER stress, which was similar to the signature that we identified for the 5- and 8-h ER stress in our dataset (Table S1). Almost all cytosolic tRNA synthetases were induced by 16 h of ER stress (Figure S4A), whereas the biosynthesis of fewer amino acids was upregulated at this time point (Figure S4B). Thus, the 16-h UPR amino acid signature consisted of the amino acids Ser, Gly, Cys, Asp, and Glu (Table S1), a similar subset to the late (5–8 h) ER stress signature above. Nevertheless, when considering the amino acid demand originating from the induced proteins at the 16-h ER stress time point of the Guan et al. translatome data, we again observed significant enrichment of the 16-h UPR amino acid signature within induced proteins (Figure S4C, $p = 1.5 \times 10^{-10}$).

Taken together, regulation at the level of amino acid biosynthesis and tRNA synthetases during ER stress is coupled to changes in the demand of amino acids originating from induced proteins.

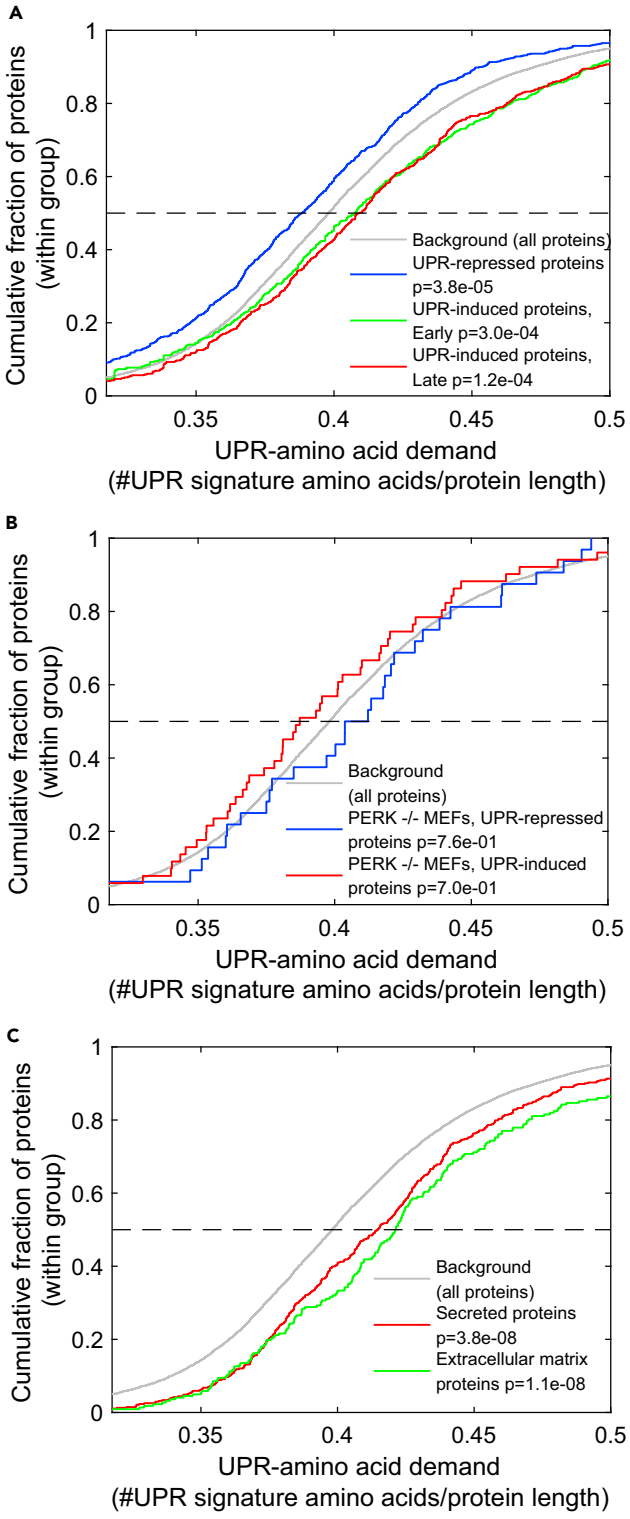


Figure 3. The UPR Amino Acid Signature Is Enriched in UPR-Induced Proteins, and in Secreted Proteins

(A) The UPR amino acid demand, namely, the fraction of the seven amino acids comprising the UPR amino acid signature, in protein sequences within the UPR-repressed (blue), early UPR-induced (green), and late UPR-induced (red) gene expression programs from Gonen et al. (2019) compared with the background of all expressed proteins (gray) using a cumulative distribution function (CDF) plot. The plot shows the cumulative distribution of the fraction of proteins within

Figure 3. *Continued*

each of the groups (y axis), with different UPR amino acid demand scores (x axis). Proteins belonging to the UPR repression program were significantly depleted for the UPR amino acid signature, as indicated by a shift in distribution to more negative values compared with the background distribution. In contrast, both early and late UPR induction program proteins were highly enriched for this amino acid set, as indicated by a significant shift in distribution to more positive values compared with the background distribution. Kolmogorov-Smirnov (KS) test p values are indicated.

(B) Amino acid demand CDF plot (as in A) for all ER stress-induced (red) and ER stress-repressed (blue) mRNAs in the PERK−/− MEFs (at least 2-fold at any of the time points compared with control). None of the groups showed enrichment or depletion for the UPR amino acid signature; KS test p values are indicated.

(C) The UPR amino acid demand of secreted proteins (red) and extracellular matrix proteins (green) compared with the background of all expressed proteins (gray) is shown using a CDF plot (as in A). Both secreted and extracellular matrix proteins were highly enriched for the UPR amino acid signature, KS test p values are indicated.

UPR-Mediated Regulation of Amino Acid Metabolism Coupled with Amino Acid Demand of Secreted Proteins

Next, we sought to examine whether this mode of regulation applies to specific pathways or protein classes. Our observation that secreted proteins tended to escape the ER stress-mediated repression of ER-targeted proteins (Figure 1) prompted us to examine the amino acid composition of this group of proteins. Interestingly, secreted protein sequences demonstrated significant enrichment in the UPR amino acid signature (Figure 3C). Notably, although overall, secreted proteins showed elevated amino acid demand for this signature, secreted proteins that were induced were more significantly enriched for the signature than were those that were repressed (Figure S5D). Even greater enrichment of the UPR amino acid signature was observed in the sequences of extracellular matrix proteins (Figure 3C). This result was not due to the particular amino acid composition of the signal peptide, as exclusion of the signal peptide region from the analyzed protein sequences did not affect any of the trends (Figures S5A and S5B). In addition, this trend was not confounded by overlap with the ER stress induction protein expression programs (from Figure 3A), as it remained significant when we removed these proteins (Figure S5C).

The UPR Amino Acid Signature Shows Higher Codon Optimality in Both UPR-Induced and Secreted Proteins

We next decided to analyze an additional level of protein synthesis regulation, namely, the level of codon optimality. To this end, we calculated for each protein, the tRNA adaptation index (tAI, see Methods), a widely used metric for codon optimality that takes into account the corresponding tRNA copy numbers of each codon (dos Reis et al., 2003). We compared the tAI of secreted proteins to those of all expressed proteins, and observed a highly similar distribution (Figure 4A, $p = 4.7 \times 10^{-1}$). Surprisingly, when the tAI calculation was restricted to the UPR amino acid signature, the codon optimality for secreted proteins was significantly increased ($p = 8.4 \times 10^{-6}$, using Kolmogorov-Smirnov [KS] test, Figures 4B and 4C). Importantly, this finding was not due to the signal peptide region, as it remained after exclusion of this region from the analysis (Figures S6A–S6C). Further examination of the set of late ER stress-induced proteins revealed a similar trend, whereby codon optimality (measured using tAI) for late UPR-induced proteins was elevated for the UPR amino acid signature (Figures S6D–S6F). Thus, the UPR amino acid signature is coupled to UPR-induced protein expression demands, as well as to the amino acid demand of secreted proteins and shows increased codon optimality for these specific groups.

In summary, our analyses unraveled an additional mode of protein synthesis regulation as part of the UPR, which involves rewiring of amino acid metabolism and tRNA synthetase induction, and the coupling of these processes to the amino acid demand of proteins whose synthesis is induced during ER stress.

DISCUSSION

Here we analyzed ribosome footprint profiling data of a time course of ER stress treatments and identified an additional mode of protein synthesis regulation that occurs in the late adaptive stage of the response. Specifically, we found coordinated induction in certain amino acid biosynthetic pathways, together with upregulation of the expression of their cognate tRNA synthetases. In agreement with our hypothesis that this coordination may be related to protein synthesis demands, we found indeed that these signature amino acids are enriched within UPR-induced proteins.

Codon bias was shown to vary according to categories of gene expression. Upon stress, yeast cells were shown to induce the expression of tRNAs whose codons are enriched within stress-induced mRNAs (Torrent et al., 2018). Furthermore, codon usage was found to differ between proliferation-related genes,

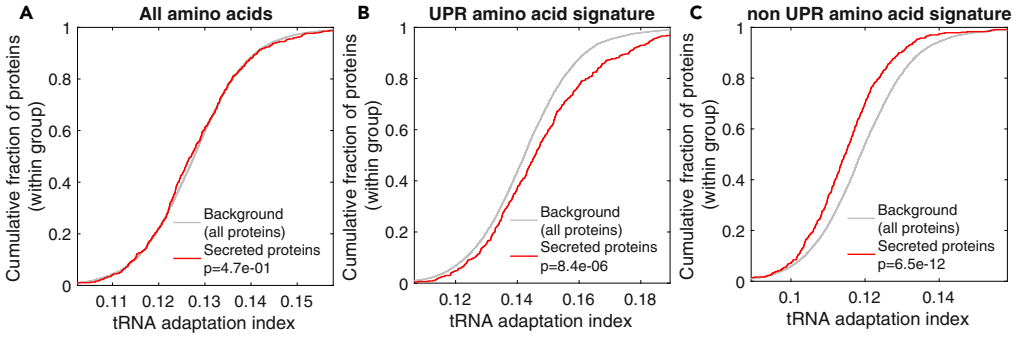


Figure 4. The UPR Amino Acid Signature Shows Higher Codon Optimality for UPR-Induced Proteins and for Secreted Proteins

Codon optimality was calculated for all expressed proteins using the tRNA adaptation index (see Methods), using all amino acids (A), only the amino acids comprising the UPR amino acid signature (B), or only the amino acids that are not part of the UPR amino acid signature (C). tAI distributions were plotted for the group of secreted proteins (red) versus the background distribution of all expressed proteins (gray), using cumulative distribution function (CDF) plots (as in Figure 3). p values for the differences between these two groups, calculated using the Kolmogorov-Smirnov test, showed no difference in the codon optimality of all amino acids (A); however, the codon optimality of the UPR amino acid signature is significantly higher for secreted proteins (B).

whose expression is altered in cancer, and differentiation-related genes, and this was accompanied by shifts in the expression of respective pools of tRNAs (Gingold et al., 2014). In addition, secreted proteins were shown to have particular codon bias properties in several organisms, from yeast to human (Mahlab and Linial, 2014). Therefore, organisms seem to have evolved means of coupling the expression of their tRNA pools to match demands of varying protein expression programs (Quax et al., 2015). Here we describe a phenomenon in which amino acid biosynthesis and tRNA synthetase expression are coupled to the amino acid content of induced proteins during the course of the ER stress response. In codon bias, shifts in tRNA pools involve synonymous codons, whereas the signs of a metabolic shift reported here were related to the amino acid content. Nonetheless, it is interesting that the UPR amino acid signature we identified also shows higher codon optimality for secreted proteins and for late UPR-induced proteins, thus manifesting a combined-level regulation.

Amino acid biosynthesis induction was previously discussed in the context of ER stress. Some studies suggested that this upregulation is adaptive (Harding et al., 2000, 2003), whereas others claimed that the ATF4-mediated induction of amino acid biosynthesis and tRNA synthetases enhances protein synthesis in a manner that could become harmful upon prolonged stress (Krokowski et al., 2013). Here we suggest that the induced biosynthesis of specific amino acids is coupled to protein synthesis demands, which supports the notion that this is indeed an adaptive mechanism. We propose that amino acid composition and shifts in amino acid metabolism have co-evolved such as to yield an improved regulatory program that fine-tunes supply and demand for proteins that are preferentially synthesized upon ER stress.

An alternative possibility is that feedback regulation senses depletion in some amino acids, and thereby acts to enhance biosynthesis, which then meets the increased demand. However, we note that the cells were grown throughout the experiment in a media that was supplemented with non-essential amino acids (Gonen et al., 2019). Therefore, it is highly unlikely that the cells were depleted for these amino acids, but rather that an intrinsic program led to the shift in amino acid metabolism.

We wondered whether the UPR amino acid signature we identified is the optimal one, given the set of induced proteins. Our analysis found that the UPR amino acid signature is only in the top 37th percentile in terms of optimal compatibility to the set of induced proteins and in the top 30th percentile of optimality for secreted proteins (see Figure S7). In other words, other combinations of seven amino acids would more optimally meet the specific demands of amino acids for induced protein expression programs. Importantly, other roles have been ascribed to various amino acid biosynthesis pathways upon ER stress (Harding et al., 2000, 2003). Thus, it is highly likely that amino acid demand is only partly responsible for the observed UPR amino acid signature.

For example, the serine-glycine biosynthesis pathway contains four enzymes, all of which were highly induced in our data in a PERK-dependent manner (Figures 2C and S8A). This pathway was also found to

be elevated, in an ATF4-dependent manner, and to increase glutathione production in non-small-cell lung cancer cells, thus contributing to their aggressiveness (DeNicola et al., 2015). Indeed, glutathione has an important role in protection from oxidative stress in general, and in ATF4−/− cells in particular (Harding et al., 2003). Furthermore, mitochondrial tetrahydrofolate, which is another important derivative of serine-glycine metabolism, has been shown to play important roles in protecting from mitochondrial stress in response to glucose starvation, and in maintaining proper mitochondrial translation upon stress (Minton et al., 2018). Indeed, our data showed that mitochondrial translation is largely maintained during the course of ER stress, as evident by the increased ratio of mitochondrially encoded gene translation (Figure S8C), whereas cytosolic translation was overall repressed (Gonen et al., 2019). Therefore, it is possible that the induction of serine biosynthesis actually leads to increased conversion of serine into glycine and tetrahydrofolate, which is adaptive for cell fate and helps maintain healthy mitochondria during ER stress.

Proline, which is generated by a series of reactions from glutamate, has also been shown to have protective roles in stress. Proline starvation results in unresolved ER stress in cancer cells (Sahu et al., 2016). Furthermore, proline was shown to be limiting in clear cell renal cell carcinoma tumors, in which the PYCR1 enzyme responsible for the final step of proline production from glutamate was induced, whereas proline catabolism enzymes were downregulated (Loayza-Puch et al., 2016). Our data show high induction of PYCR1, along with two additional enzymes in the glutamate to proline biosynthesis pathway, PYCR2 and ALDH18A1, together with a nearly 2-fold reduction in the expression of the proline catabolic enzyme ALDH4A1 (Figure S8B). Therefore, the observed induction in glutamate may be only an intermediate step that promotes the biosynthesis of proline.

Following the above, we investigated whether removal of serine and glutamate would generate a more optimal UPR amino acid signature. Indeed, when we removed one or both of these from the UPR amino acid signature, the signature became more optimal (4.9th and 2nd top percentiles for late-induced proteins and secreted proteins, respectively, Figure S9). Therefore, it is possible that some induced amino acids serve as precursors for certain derivatives, which are themselves involved in protection from various effects of ER stress, or from secondary related oxidative stress. Our observation of a favorable, yet suboptimal set of induced amino acids corroborates with the possibility of other underlying causes for the induction of some of the induced amino acid biosynthesis pathways. This supports the notion that fine-tuned changes in amino acid biosynthesis coincide with amino acid demands that originate particularly from induced protein expression programs. The extent to which these metabolic shifts in the biosynthesis of amino acids or their metabolic derivatives is congruent with specific amino acid-related functions, or with protein synthesis demands, remains to be resolved.

Limitations of the Study

Our study describes an additional regulatory layer that occurs during ER stress, involving coordination between supply and demand of amino acids to fine-tune UPR protein synthesis demands. Furthermore, the same amino acid signature shows improved codon optimality, further supporting the adaptiveness of this mode of regulation. Nevertheless, amino acid biosynthesis generates multiple derivatives, some of which have been shown to play specific roles in protection from prolonged ER stress and other secondary stresses, as discussed above. Future additional experiments, such as tRNA-charging assays and metabolomics, would be required to fully disentangle the effects of amino acid supply-demand regulation from those of secondary derivatives and fully quantify their specific contributions.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

DATA AND CODE AVAILABILITY

The datasets analyzed in this study are available on GEO. Gonen et al. (Gonen et al., 2019) data are available as GEO: GSE118660. Guan et al. polysome sequencing data (Guan et al., 2017) are available as GEO: GSE90070.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.07.022.

ACKNOWLEDGMENTS

We would like to thank Oded Lewinson for useful discussions of the results and critical reading of the manuscript. We thank Ruth Hershberg for critical reading of the manuscript. This project received

funding from the European Research Council under the European Union's Horizon 2020 program Grant 677776.

AUTHOR CONTRIBUTIONS

R.S. conceived and supervised the study. N.G., A.M., and N.S. performed all data analyses. R.S. wrote the manuscript with input from N.S. and A.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 15, 2019
Revised: June 11, 2019
Accepted: July 15, 2019
Published: September 27, 2019

REFERENCES

Ben-Sahra, I., Hoxhaj, G., Ricoult, S.J.H., Asara, J.M., and Manning, B.D. (2016). mTORC1 induces purine synthesis through control of the mitochondrial tetrahydrofolate cycle. Science *351*, 728–733.

DeNicola, G.M., Chen, P.H., Mullarky, E., Sudderth, J.A., Hu, Z., Wu, D., Tang, H., Xie, Y., Asara, J.M., Huffman, K.E., et al. (2015). NRF2 regulates serine biosynthesis in non-small cell lung cancer. Nat. Genet. *47*, 1475–1481.

dos Reis, M., Wernisch, L., and Savva, R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Res. *31*, 6976–6985.

Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christophersen, N.S., Christensen, L.L., Borre, M., Sorensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. Cell *158*, 1281–1292.

Gonen, N., Sabath, N., Burge, C.B., and Shalgi, R. (2019). Widespread PERK-dependent repression of ER targets in response to ER stress. Sci. Rep. *9*, 4330.

Guan, B.J., Krokowski, D., Majumder, M., Schmotzer, C.L., Kimball, S.R., Merrick, W.C., Koromilas, A.E., and Hatzoglou, M. (2014). Translational control during endoplasmic reticulum stress beyond phosphorylation of the translation initiation factor eIF2alpha. J. Biol. Chem. *289*, 12593–12611.

Guan, B.J., van Hoef, V., Jobava, R., Elroy-Stein, O., Valasek, L.S., Cargnello, M., Gao, X.H., Krokowski, D., Merrick, W.C., Kimball, S.R., et al. (2017). A unique ISR program determines cellular responses to chronic stress. Mol. Cell *68*, 885–900.e6.

Han, J., Back, S.H., Hur, J., Lin, Y.H., Gildersleeve, R., Shan, J., Yuan, C.L., Krokowski, D., Wang, S., Hatzoglou, M., et al. (2013). ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. Nat. Cell Biol. *15*, 481–490.

Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., and Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol. Cell *6*, 1099–1108.

Harding, H.P., Zhang, Y., Zeng, H., Novoa, I., Lu, P.D., Calfon, M., Sadri, N., Yun, C., Popko, B., Paules, R., et al. (2003). An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. Mol. Cell *11*, 619–633.

Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The signal recognition particle. Annu. Rev. Biochem. *70*, 755–775.

Krokowski, D., Han, J., Saikia, M., Majumder, M., Yuan, C.L., Guan, B.J., Bevilacqua, E., Bussolati, O., Broer, S., Arvan, P., et al. (2013). A self-defeating anabolic program leads to beta-cell apoptosis in endoplasmic reticulum stress-induced diabetes via regulation of amino acid flux. J. Biol. Chem. *288*, 17202–17213.

Loayza-Puch, F., Rooijers, K., Buil, L.C., Zijlstra, J., Oude Vrielink, J.F., Lopes, R., Ugalde, A.P., van Breugel, P., Hofland, I., Wesseling, J., et al. (2016). Tumour-specific proline vulnerability uncovered by differential ribosome codon reading. Nature *530*, 490–494.

Mahlab, S., and Linial, M. (2014). Speed controls in translating secretory proteins in eukaryotes–an evolutionary perspective. PLoS Comput. Biol. *10*, e1003294.

Minton, D.R., Nam, M., McLaughlin, D.J., Shin, J., Bayraktar, E.C., Alvarez, S.W., Sviderskiy, V.O., Papagiannakopoulos, T., Sabatini, D.M., Birsoy, K., et al. (2018). Serine catabolism by SHMT2 is required for proper mitochondrial translation initiation and maintenance of formylmethionyl-tRNAs. Mol. Cell *69*, 610–621.e5.

Park, Y., Reyna-Neyra, A., Philippe, L., and Thoreen, C.C. (2017). mTORC1 balances cellular amino acid supply with demand for protein synthesis through post-transcriptional control of ATF4. Cell Rep. *19*, 1083–1090.

Pavitt, G.D., and Ron, D. (2012). New insights into translational regulation in the endoplasmic reticulum unfolded protein response. Cold Spring Harb. Perspect. Biol. *4*, a012278.

Quax, T.E., Claassens, N.J., Soll, D., and van der Oost, J. (2015). Codon bias as a means to fine-tune gene expression. Mol. Cell *59*, 149–161.

Reid, D.W., Chen, Q., Tay, A.S., Shenolikar, S., and Nicchitta, C.V. (2014). The unfolded protein response triggers selective mRNA release from the endoplasmic reticulum. Cell *158*, 1362–1374.

Rendleman, J., Cheng, Z., Maity, S., Kastelic, N., Munschauer, M., Allgoewer, K., Teo, G., Zhang, Y.B.M., Lei, A., Parker, B., et al. (2018). New insights into the cellular temporal response to proteostatic stress. Elife *7*, e39054.

Ron, D., and Walter, P. (2007). Signal integration in the endoplasmic reticulum unfolded protein response. Nat. Rev. Mol. Cell Biol. *8*, 519–529.

Rutkowski, D.T., and Kaufman, R.J. (2007). That which does not kill me makes me stronger: adapting to chronic ER stress. Trends Biochem. Sci. *32*, 469–476.

Sahu, N., Dela Cruz, D., Gao, M., Sandoval, W., Haverty, P.M., Liu, J., Stephan, J.P., Haley, B., Classon, M., Hatzivassiliou, G., et al. (2016). Proline starvation induces unresolved ER stress and hinders mTORC1-dependent tumorigenesis. Cell Metab. *24*, 753–761.

Torrent, M., Chalancon, G., de Groot, N.S., Wuster, A., and Madan Babu, M. (2018). Cells alter their tRNA abundance to selectively regulate protein synthesis during stress conditions. Sci. Signal. *11*, eaat6409.

Vattem, K.M., and Wek, R.C. (2004). Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. Proc. Natl. Acad. Sci. U S A *101*, 11269–11274.

ISCI, Volume *19*

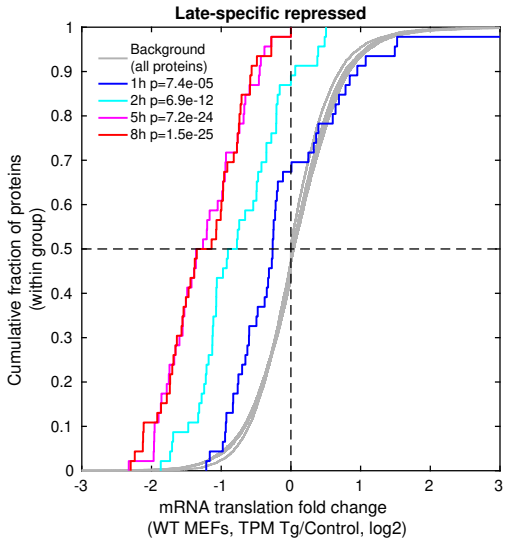
Supplemental Information

Amino Acid Biosynthesis Regulation during Endoplasmic Reticulum Stress Is Coupled to Protein Expression Demands

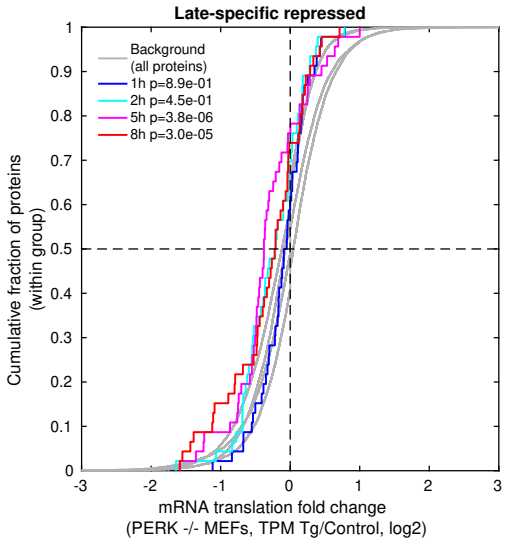
Nir Gonen, Anatoly Meller, Niv Sabath, and Reut Shalgi

Supplemental Figures

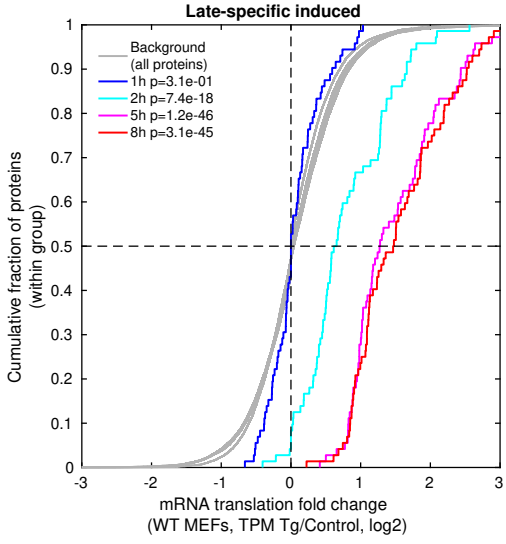
A



B



C



D

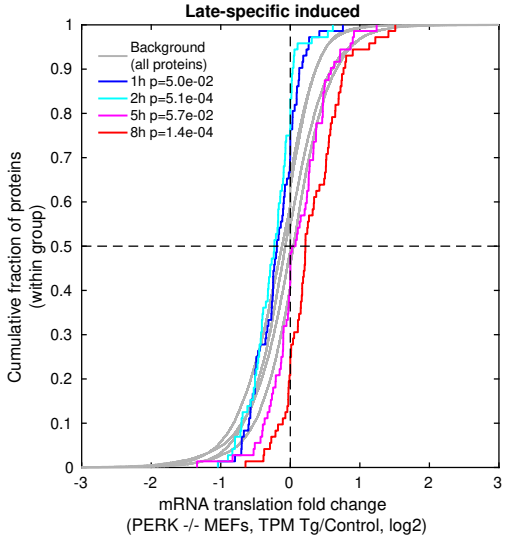
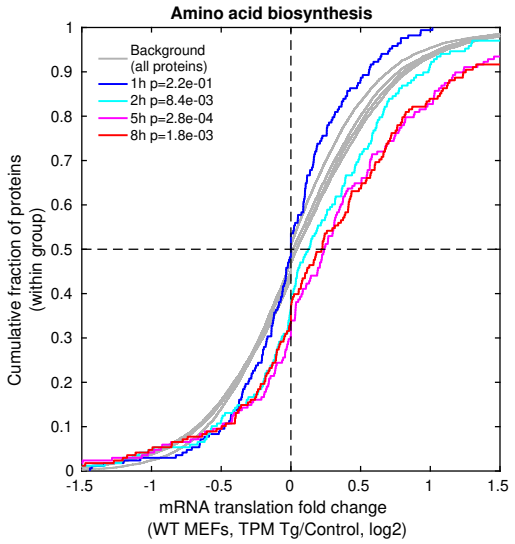


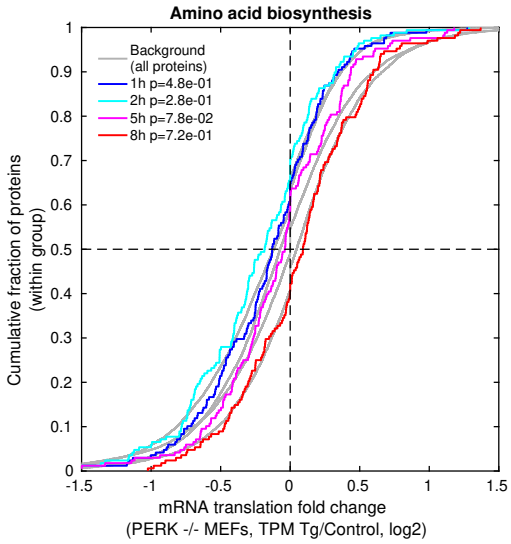
Figure S1. PERK-dependence of late-specific repressed and induced gene sets. Related to Figure 2.

CDF (Cumulative Distribution Function) plots of translation log2 fold change of late-specific repressed (A,B) and induced (C,D) gene sets in WT (A,C) and PERK -/- (B,D) MEFs, show that both repression and induction is largely PERK dependent. The y-axis shows the cumulative fraction of genes corresponding to the log2 fold changes (plotted on the x-axis) in the translation levels of the subset of mRNAs in each timepoint (colored plots) compared to the background of all expressed mRNAs (in grey).

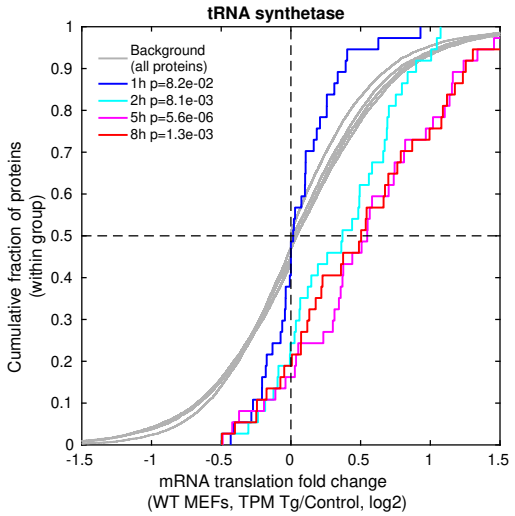
A


Amino acid biosynthesis
Background
(all proteins)
1h p=2.2e-01
2h p=8.4e-03
5h p=2.8e-04
8h p=1.8e-03
Cumulative fraction of proteins
(within group)
mRNA translation fold change
(WT MEFs, TPM Tg/Control, log2)

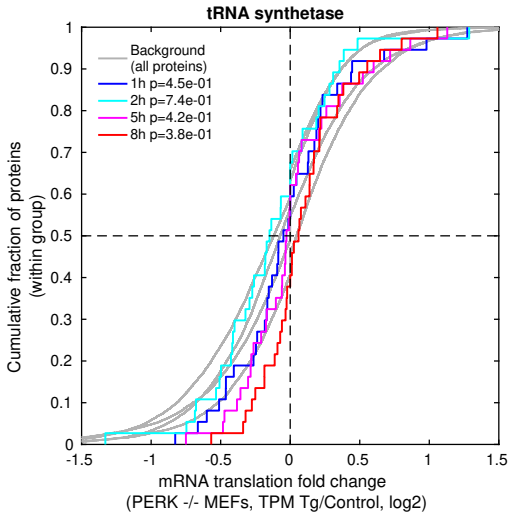
B


Amino acid biosynthesis
Background
(all proteins)
1h p=4.8e-01
2h p=2.8e-01
5h p=7.8e-02
8h p=7.2e-01
Cumulative fraction of proteins
(within group)
mRNA translation fold change
(PERK -/- MEFs, TPM Tg/Control, log2)

C


tRNA synthetase
Background
(all proteins)
1h p=8.2e-02
2h p=8.1e-03
5h p=5.6e-06
8h p=1.3e-03
Cumulative fraction of proteins
(within group)
mRNA translation fold change
(WT MEFs, TPM Tg/Control, log2)

D


tRNA synthetase
Background
(all proteins)
1h p=4.5e-01
2h p=7.4e-01
5h p=4.2e-01
8h p=3.8e-01
Cumulative fraction of proteins
(within group)
mRNA translation fold change
(PERK -/- MEFs, TPM Tg/Control, log2)

E

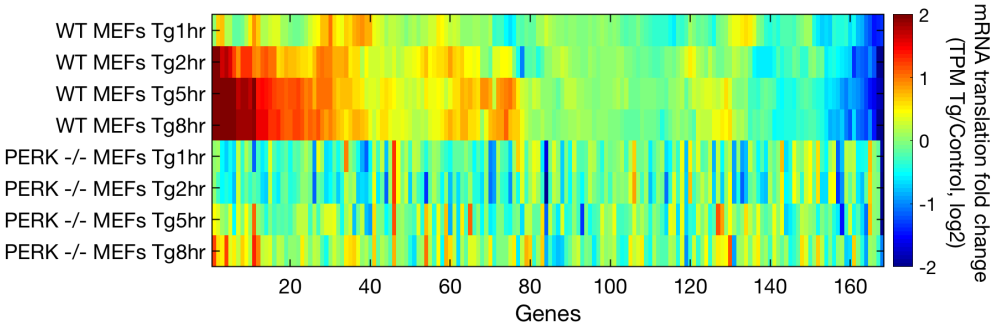

WT MEFs Tg1hr
WT MEFs Tg2hr
WT MEFs Tg5hr
WT MEFs Tg8hr
PERK -/- MEFs Tg1hr
PERK -/- MEFs Tg2hr
PERK -/- MEFs Tg5hr
PERK -/- MEFs Tg8hr
Genes
mRNA translation fold change
(TPM Tg/Control, log2)

Figure S2. Amino acid biosynthesis pathway and tRNA synthetase mRNAs show significant, PERK-dependent, induction in ER stress. Related to Figure 2.

CDF plots of log2 fold changes in the translation levels of amino acid biosynthesis pathway mRNAs (A,B) and tRNA synthetase mRNAs (C,D), in WT (A,C) and PERK -/- (B,D) MEFs, show that induction (in 2h, 5h, and 8h) is PERK dependent. (E) Hierarchical clustering analysis of amino acid biosynthesis pathway genes, showed that ~80 of them were induced while the rest were largely unchanged. The heatmap depicts hierarchically clustered log2 fold change of the TPM of mRNAs, according to their spearman correlation.

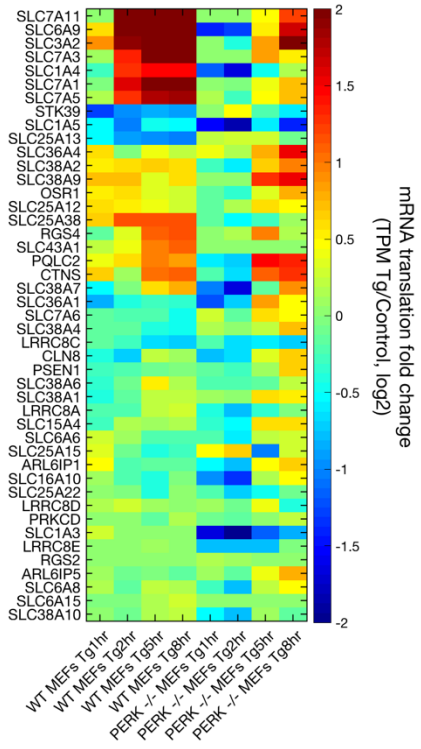
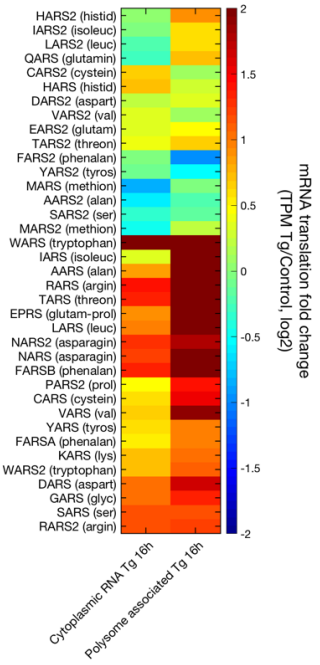


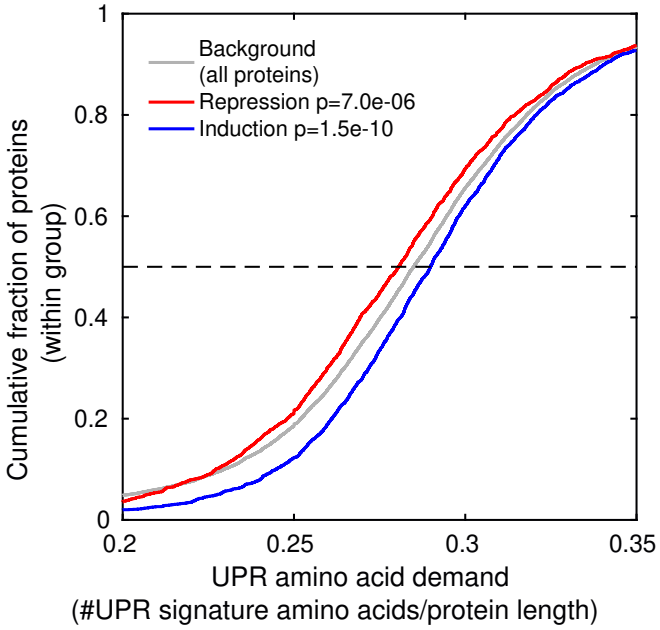
Figure S3. Several amino acid transport mRNAs are induced in ER stress. Related to Figure 2.

Hierarchical clustering analysis was performed using 45 genes related to amino acid transport. Heatmap depicting hierarchically clustered mRNA translation log2 fold changes, according to their spearman correlation.

A


HARS2 (histid)
IARS2 (isoleuc)
LARS2 (leuc)
QARS (glutamin)
CARS2 (cystein)
HARS (histid)
DARS2 (aspart)
VARS2 (val)
EARS2 (glutam)
TARS2 (threon)
FARS2 (phenalan)
YARS2 (tyros)
MARS (methion)
AARS2 (alan)
SARS2 (ser)
MARS2 (methion)
WARS (tryptophan)
IARS (isoleuc)
AARS (alan)
RARS (argin)
TARS (threon)
EPRS (glutam-prol)
LARS (leuc)
NARS2 (asparagin)
NARS (asparagin)
FARSB (phenalan)
PARS2 (prol)
CARS (cystein)
VARS (val)
YARS (tyros)
FARSA (phenalan)
KARS (lys)
WARS2 (tryptophan)
DARS (aspart)
GARS (glyc)
SARS (ser)
RARS2 (argin)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
mRNA translation fold change
(TPM Tg/Control, log2)
Cytoplasmic RNA Tg 16h
Polysome associated Tg 16h

C


Background
(all proteins)
Repression p=7.0e-06
Induction p=1.5e-10
Cumulative fraction of proteins
(within group)
1
0.8
0.6
0.4
0.2
0
0.2
0.25
0.3
0.35
UPR amino acid demand
(#UPR signature amino acids/protein length)

B

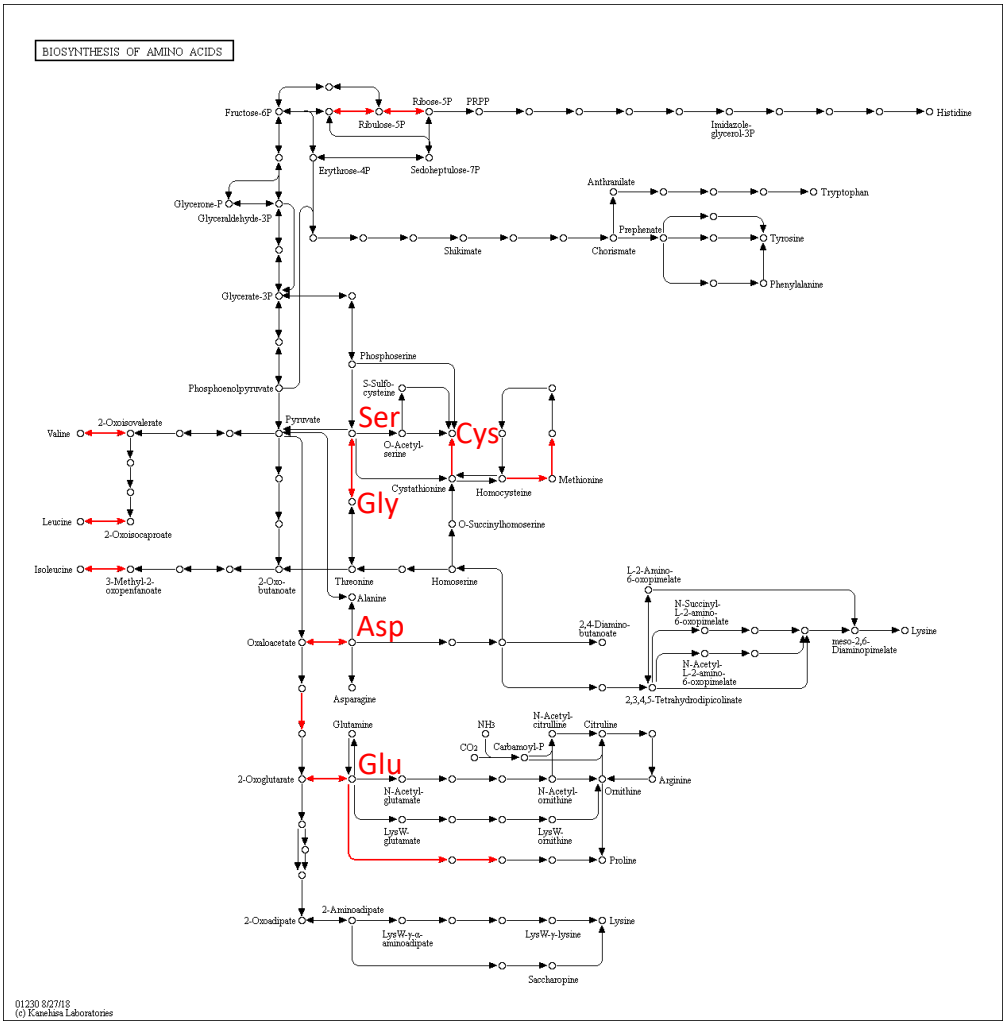

BIOSYNTHESIS OF AMINO ACIDS
Fructose-6P
Ribose-5P
PRPP
Ribulose-5P
Histidine
Imidazole-glycerol-3P
Erythrose-4P
Sedoheptulose-7P
Glycerone-P
Glyceraldehyde-3P
Anthranilate
Tryptophan
Shikimate
Prephenate
Chorismate
Tyrosine
Phenylalanine
Glycerate-3P
Phosphoserine
Phosphoenolpyruvate
S-Sulfo-cysteine
Ser
Cys
Pyruvate
O-Acetyl-serine
Valine
2-Oxoisovalerate
Cystathionine
Homocysteine
Methionine
Gly
Leucine
2-Oxoisocaproate
O-Succinylhomoserine
Isoleucine
3-Methyl-2-oxopentanoate
2-Oxo-butanoate
Threonine
Homoserine
Alanine
Asp
L-2-Amino-6-oxopimelate
N-Succinyl-L-2-amino-6-oxopimelate
Lysine
meso-2,6-Diaminopimelate
2,4-Diamino-butanoate
N-Acetyl-L-2-amino-6-oxopimelate
Oxaloacetate
2,3,4,5-Tetrahydrodipicolinate
Asparagine
Glutamine
NH3
CO2
N-Acetyl-citrulline
Citrulline
Carbamoyl-P
Glu
2-Oxoglutarate
N-Acetyl-glutamate
N-Acetyl-ornithine
Ornithine
Arginine
LysW-glutamate
LysW-ornithine
Proline
2-Aminoadipate
2-Oxoadipate
LysW-γ-α-aminoadipate
LysW-γ-lysine
Lysine
Saccharopine
01230 8/27/18
(c) Kanehisa Laboratories

Figure S4. Analyses of amino acid demand in the Guan *et al.* dataset. Related to Figure 3.

(A) Hierarchical clustering analysis of tRNA synthetase genes in cytoplasmic RNA and polysome-associated RNA from Guan *et al.* (Guan et al., 2017). Heatmap depicting hierarchically clustered mRNA expression (Cytoplasmic RNA) / translation (Polysome-associated) log2 fold changes, according to their spearman correlation. The majority of cytosolic tRNA synthetases were induced at 16h of ER stress. The induction is higher in polysome-associated RNA. (B) Amino acid biosynthesis metabolic map of induced mRNAs within polysome-associated RNA from Guan *et al.* (Guan et al., 2017), shows induction of biosynthesis pathways for five non- or partly-essential amino acids: Ser, Cys, Gly, Asp, and Glu. (C) The fraction of these five amino acids (UPR amino acid demand) in repressed (red) and induced (blue) proteins according to the polysome-associated RNA from Guan *et al.* (Guan et al., 2017), depicted using a CDF plot. Induced proteins are significantly enriched for this 16h UPR amino acid signature, while repressed are depleted. KS-test p-values vs. background distribution (all expressed genes, in grey) are indicated.

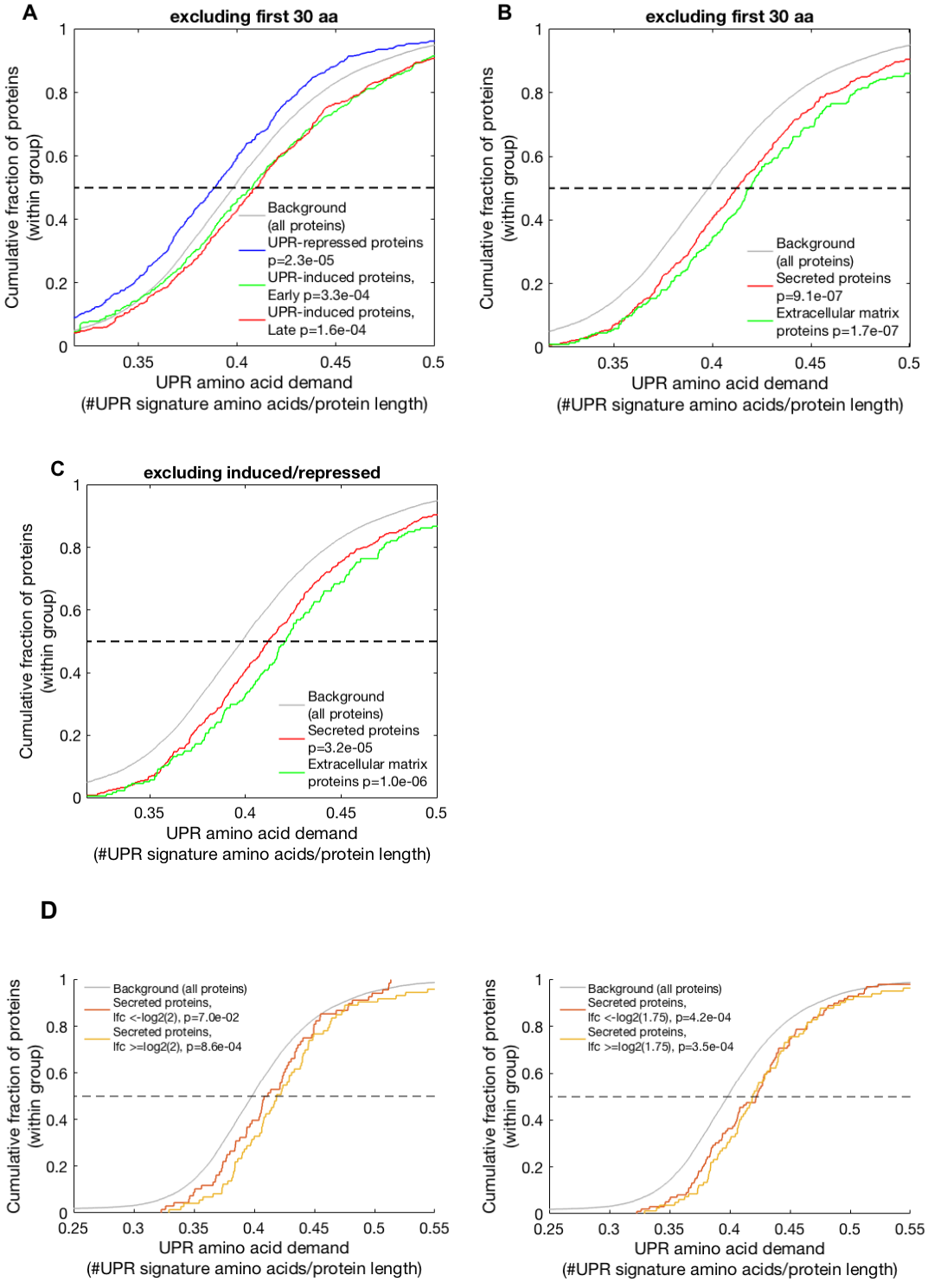

A
excluding first 30 aa
Cumulative fraction of proteins (within group)
UPR amino acid demand (#UPR signature amino acids/protein length)
Background (all proteins)
UPR-repressed proteins p=2.3e-05
UPR-induced proteins, Early p=3.3e-04
UPR-induced proteins, Late p=1.6e-04
B
excluding first 30 aa
Cumulative fraction of proteins (within group)
UPR amino acid demand (#UPR signature amino acids/protein length)
Background (all proteins)
Secreted proteins p=9.1e-07
Extracellular matrix proteins p=1.7e-07
C
excluding induced/repressed
Cumulative fraction of proteins (within group)
UPR amino acid demand (#UPR signature amino acids/protein length)
Background (all proteins)
Secreted proteins p=3.2e-05
Extracellular matrix proteins p=1.0e-06
D
Cumulative fraction of proteins (within group)
UPR amino acid demand (#UPR signature amino acids/protein length)
Background (all proteins)
Secreted proteins, lfc <-log2(2), p=7.0e-02
Secreted proteins, lfc >=log2(2), p=8.6e-04
Cumulative fraction of proteins (within group)
UPR amino acid demand (#UPR signature amino acids/protein length)
Background (all proteins)
Secreted proteins, lfc <-log2(1.75), p=4.2e-04
Secreted proteins, lfc >=log2(1.75), p=3.5e-04

Figure S5. UPR amino acid demand for secreted proteins - additional analyses. Related to Figure 3.

(A-B) UPR amino acid demand CDF plots (as in Fig. 3A,C) were repeated after excluding the first 30 amino acids for all proteins, the region encoding the signal peptide, demonstrate that the same enrichment/depletion trends still hold. (C) Exclusion of mRNAs from the UPR-repression or induction gene expression programs (Gonen et al., 2019) did not affect the trend of increased amino acid demand of the UPR amino acid signature for secreted and extracellular matrix proteins. (D) UPR-induced secreted proteins show more significant amino acid demand profile for the UPR amino acid signature than UPR-repressed secreted proteins. The set of induced and repressed secreted proteins was calculated using two cutoffs (indicated in the insets) and the trend holds for both of them.

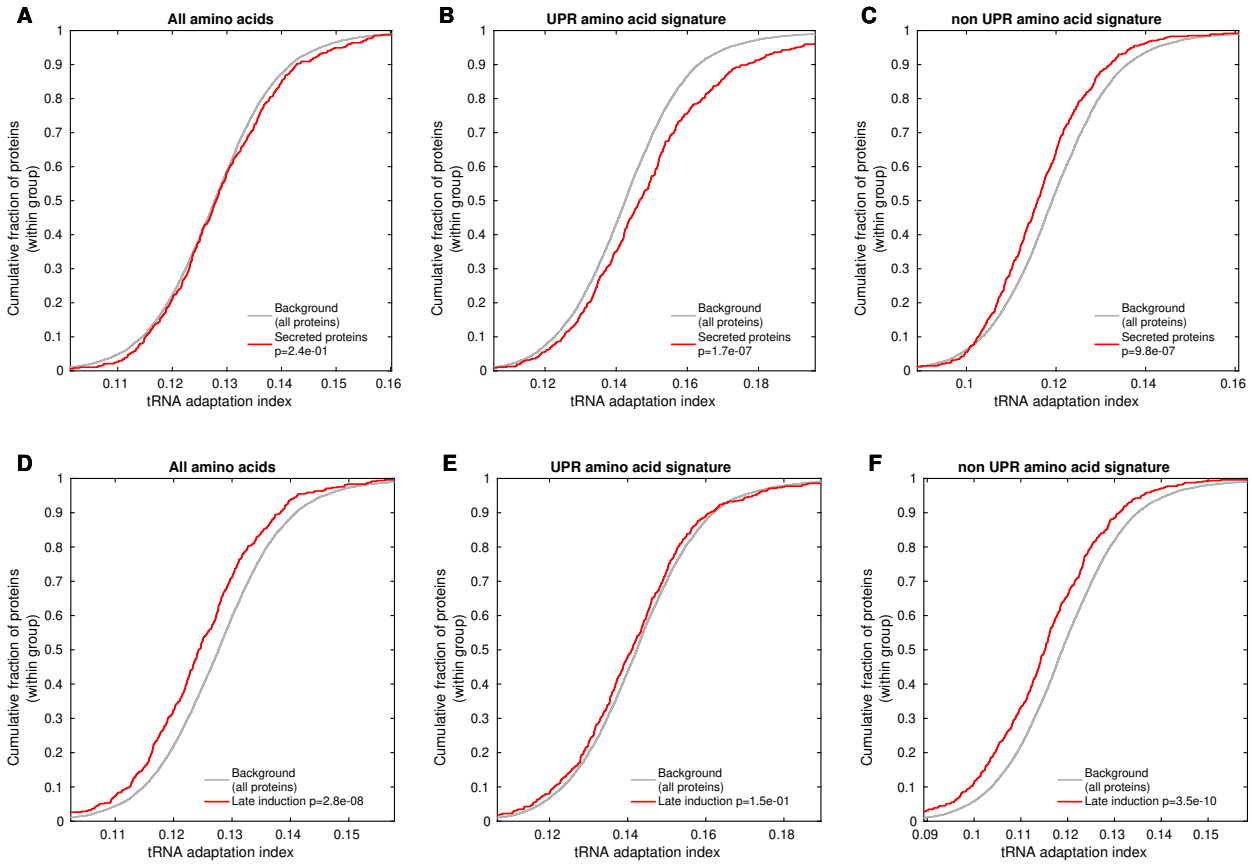


Figure S6: The UPR amino acid signature shows higher codon optimality in UPR-induced as well as secreted proteins. Related to Figure 4.

(A-C) As in Fig. 4, tAI distributions for secreted vs. all proteins calculated for all amino acids (A), only the amino acids comprising the UPR amino acid signature (B), or only amino acids not part of the UPR amino acid signature (C). Protein sequences analyzed were excluded for the first 30 amino acids, a region corresponding to the signal peptide. (D-F) As in Fig. 4, but tAI distributions are compared between the group of late ER stress-induced proteins and all proteins. While the overall tAI distribution of late ER stress-induced proteins is less optimal than the background distribution (all proteins, grey, panel D), considering only the UPR amino acid signature abrogates this difference (E), and shows a relative enhancement of the codon optimality for this group of proteins. p-values for the differenced in the tAI distributions between the two groups and the background distributions were calculated using the KS-test, and indicated.

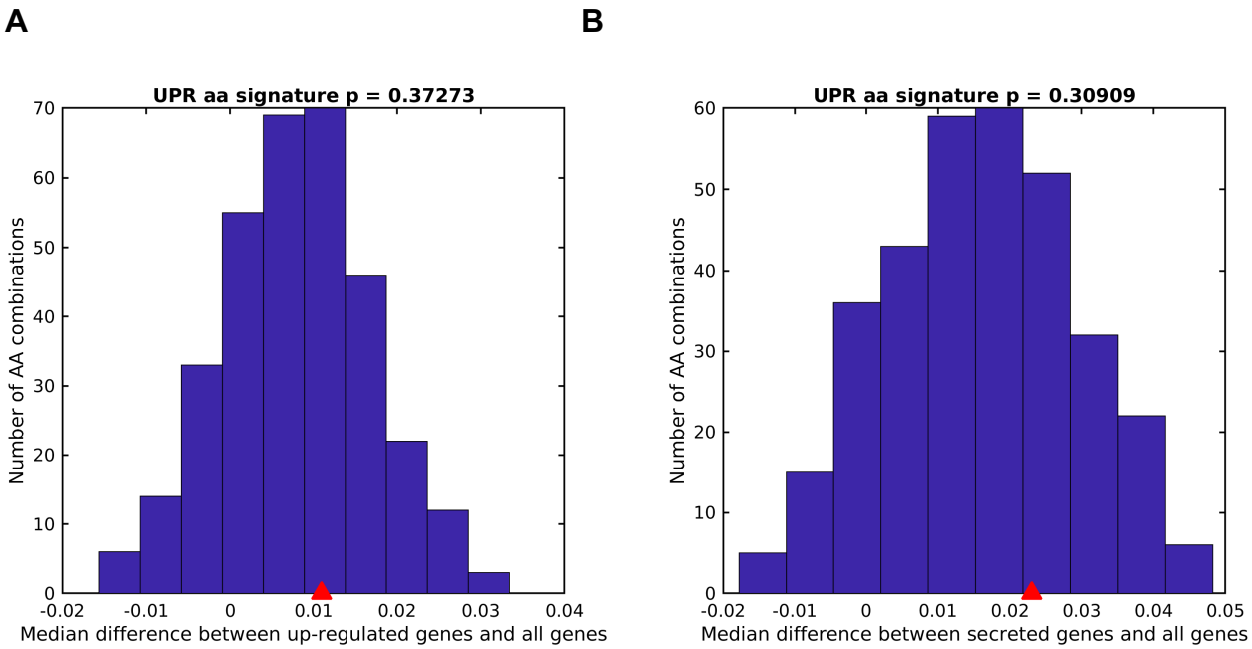
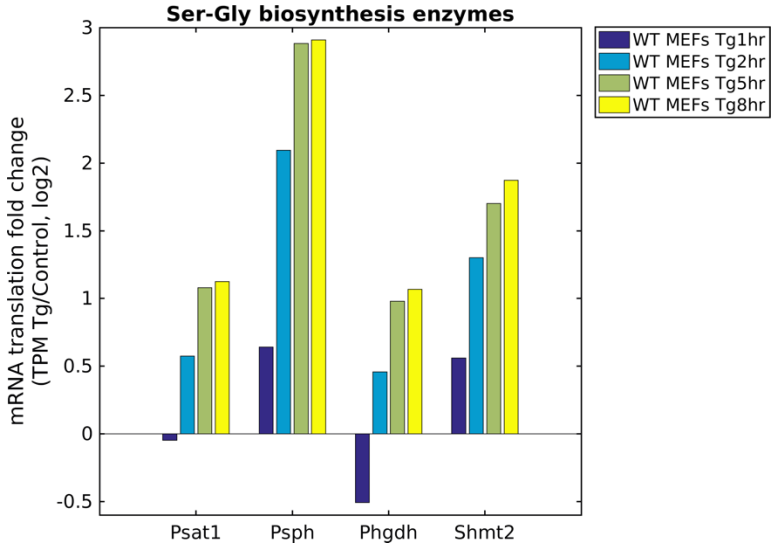


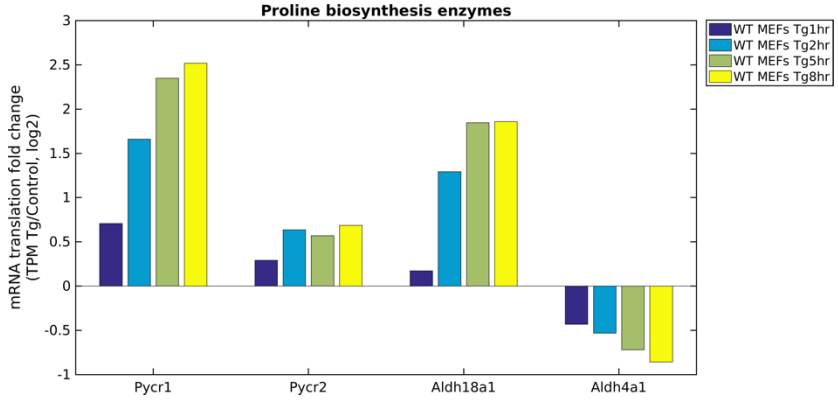
Figure S7. UPR amino acid signature is not optimal for the set of induced proteins. Related to Figure 3.

To test the optimality of the UPR amino acid signature set, we compared the difference in its median amino acid demand between induced genes and all expressed genes to the distribution of differences from all other possible sets of seven non- or partly-essential amino acids. The p represents the number of combinations with differences larger than that of the UPR amino acid signature. The UPR amino acid signature difference for induced mRNAs (A) was at the 63th percentile, and for secreted mRNAs (B) was at the 70th percentile, suggesting that the UPR amino acid signature is not optimal.

A


Ser-Gly biosynthesis enzymes
mRNA translation fold change
(TPM Tg/Control, log2)
WT MEFs Tg1hr
WT MEFs Tg2hr
WT MEFs Tg5hr
WT MEFs Tg8hr
Psat1
Psph
Phgdh
Shmt2

B


Proline biosynthesis enzymes
mRNA translation fold change
(TPM Tg/Control, log2)
WT MEFs Tg1hr
WT MEFs Tg2hr
WT MEFs Tg5hr
WT MEFs Tg8hr
Pycr1
Pycr2
Aldh18a1
Aldh4a1

C

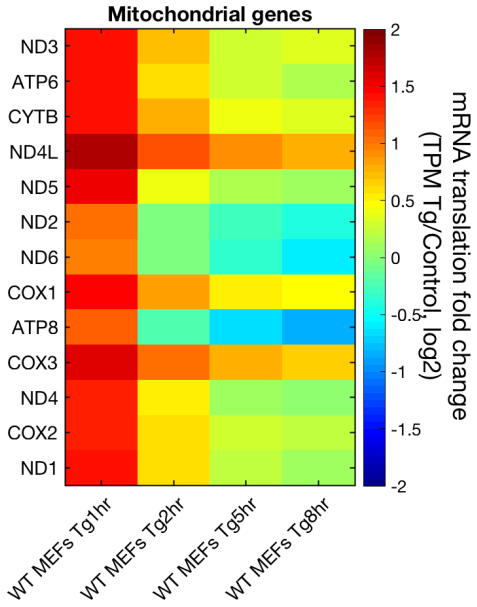
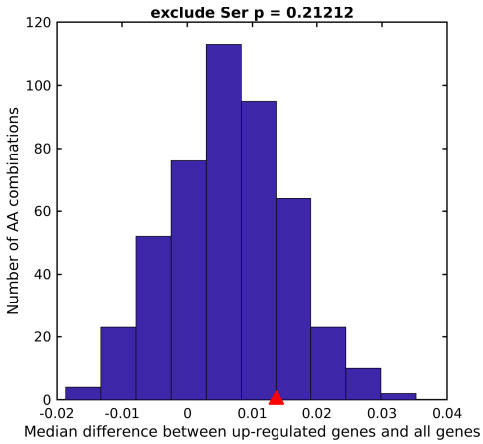
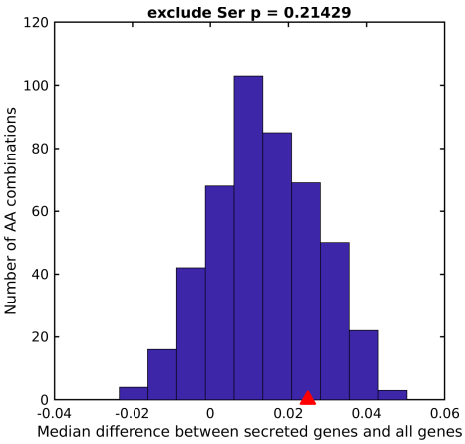

Mitochondrial genes
ND3
ATP6
CYTB
ND4L
ND5
ND2
ND6
COX1
ATP8
COX3
ND4
COX2
ND1
mRNA translation fold change
(TPM Tg/Control, log2)
WT MEFs Tg1hr
WT MEFs Tg2hr
WT MEFs Tg5hr
WT MEFs Tg8hr

Figure S8. Expression changes in the Ser-Gly and Glu-Pro pathways, and in mitochondrially encoded genes. Related to Figure 2.

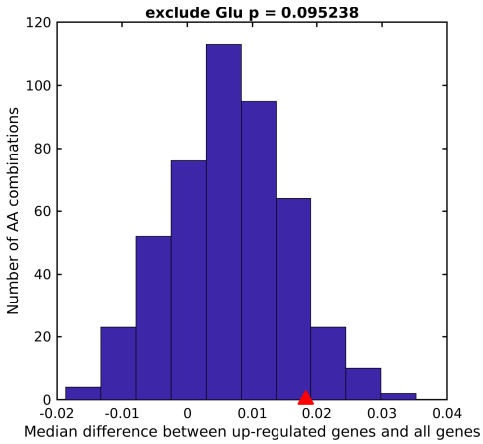
(A) Log2 fold changes (TPM, Tg/Control) in translation levels of four enzymes that are involved in the Serine-Glycine biosynthesis pathway show that all are induced in late ER stress. (B) Log2 fold changes (TPM, Tg/Control) in translation levels of three enzymes that are involved in the Glutamate to Proline biosynthesis pathway, show that all three are induced in late ER stress, whereas the expression of the Proline catabolic enzyme ALDH4A1 is reduced. (C) Hierarchical clustering analysis of the ER stress fold changes (log2) for the 13 mitochondrially encoded mRNAs. Heatmap depicting hierarchically clustered log2 fold changes (TPM, Tg/Control) in translation levels, according to their spearman correlation. The elevation actually reflects steady translation levels of mitochondrially-encoded mRNAs when considering the overall repression of cytosolic translation (see Methods).

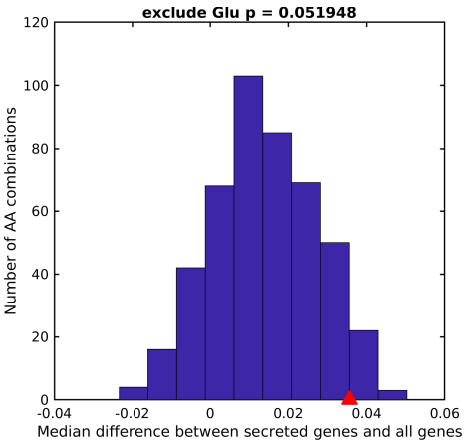
A


exclude Ser p = 0.21212
Number of AA combinations
Median difference between up-regulated genes and all genes

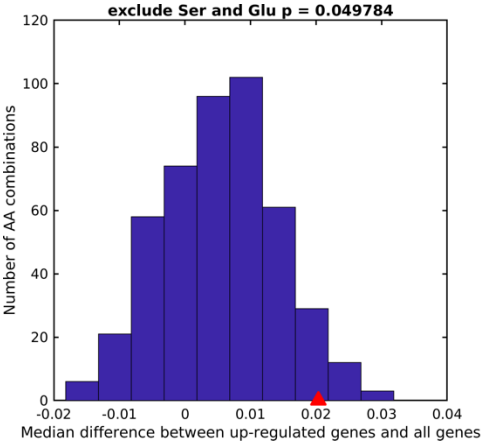

exclude Ser p = 0.21429
Number of AA combinations
Median difference between secreted genes and all genes

B


exclude Glu p = 0.095238
Number of AA combinations
Median difference between up-regulated genes and all genes


exclude Glu p = 0.051948
Number of AA combinations
Median difference between secreted genes and all genes

C


exclude Ser and Glu p = 0.049784
Number of AA combinations
Median difference between up-regulated genes and all genes

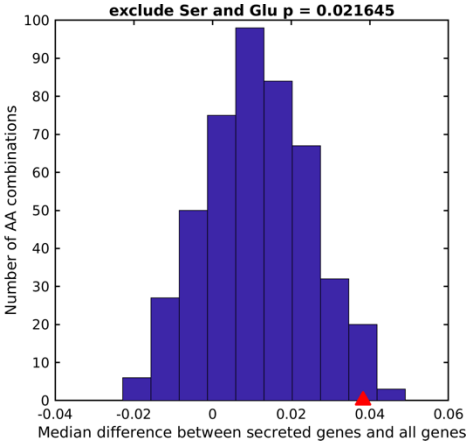

exclude Ser and Glu p = 0.021645
Number of AA combinations
Median difference between secreted genes and all genes

Figure S9. The UPR amino acid signature of UPR-induced proteins and secreted proteins is more optimal after exclusion of Serine and/or Glutamate. Related to Figure 3.

The optimality of the UPR amino acid set (as described in Figure S7), was calculated without Serine (A), Glutamate (B), or both Serine and Glutamate (C), with respect to UPR-induced proteins and secreted proteins. In all cases, the median difference of amino acid demand from the background control was in a higher percentile than that of the complete set (Figure S7). The p represents the number of combinations with difference larger than that of the modified UPR amino acid signature.

Table S1, Related to Figures 2,3: Amino acids whose biosynthesis (blue) and/or tRNA synthetases (green) were induced in Gonen *et al.* (Gonen et al., 2019) and Guan *et al.* (Guan et al., 2017) datasets. The UPR amino acid signature was defined as the set of amino acids with both induced biosynthesis and induced tRNA synthetases during ER stress in each dataset.

Amino acid	Biosynthesis up Gonen et al.	tRNA up Gonen et al.	Biosynthesis up Guan et al.	tRNA up Guan et al.
Ser (S)	✓	✓	✓	✓
Cys (C)	✓	✓	✓	✓
Gly (G)	✓	✓	✓	✓
Ala (A)	✓	✓	✗	✓
Pro (P)	✓	✓	✗	✓
Glu (E)	✓	✓	✓	✓
Asp (D)	✓	✗	✓	✓
Asn (N)	✓	✓	✗	✓
Tyr (Y)	✗	✓	✗	✓
Arg (R)	✗	✗	✗	✓

Transparent Methods

Data analysis

Ribosome footprint profiling data from Gonen *et al.* (Gonen et al., 2019) (GSE118660) were mapped to the mm10 version of the genome using RefSeq CDSs. Transcripts shorter than 100 nucleotides were filtered out and 30 nucleotides were clipped from the start and end of each CDS, similar to Ingolia *et al.* (Ingolia et al., 2012). Ribosome footprint sequences were trimmed to a maximal size of 34, and polyA sequences from the polyA tailing step were removed, such that footprints of lengths 22-34 were considered. FASTQ files were then filtered for rRNAs and tRNAs using STAR (Dobin et al., 2013). Expression levels were quantified using RSEM (Li and Dewey, 2011) after mapping to clipped CDSs using Bowtie2 (Langmead and Salzberg, 2012), to produce transcript and gene level TPM (Transcript Per Million) values.

Lowly expressed genes with a TPM below 2 across all samples were filtered out and the expression of all other genes was thresholded to a TPM of 4. For each experimental group (WT or PERK -/- MEFs), fold changes were calculated as a ratio between each condition and its respective control. Regulated mRNAs were designated as mRNAs that changed at least two-fold relative to their designated controls. mRNAs were clustered hierarchically based on Spearman's correlation using the clustergram MATLAB function. Cumulative Distribution Function (CDF) plots were used to compare the cumulative distributions of log2 fold changes in the expression of mRNAs/protein groups of interest to their respective controls, and to determine significant shifts in these distributions relative to the overall distribution of log2 fold changes of the entire transcriptome (expressed genes, as defined above). The statistical significance of the comparisons for different gene groups was calculated using the Kolmogorov-Smirnov (KS) statistical test.

Guan *et al.* polysome-seq data (Guan et al., 2017) was downloaded from GEO (GSE90070). RNA-Seq reads from polysome-seq were filtered for rRNAs using STAR (Dobin et al., 2013). The remaining reads were mapped to mm10 versions of the mouse genome, using STAR (Dobin et al., 2013). Expression levels were quantified using RSEM (Li and Dewey, 2011). TPM values were averaged between sample replicates.

Differentially expressed genes

We applied the R package DSeq2 (Love et al., 2014) to the read-count tables resulting from RSEM, to identify genes that were differentially expressed (DEG) between samples (False Discovery Rate (FDR)-corrected p-value < 0.1). Specifically, we examined genes that were shown to vary in their responses

between early (1h and 2h) and late (5h and 8h) ER stress treatments, by grouping the 1h and 2h samples into one group, and the 5h and 8h samples into a second group.

Gene groups

The classification of ER stress gene expression programs: early UPR induction, late UPR induction and UPR repression was taken from Gonen *et al.* (Gonen et al., 2019) (supplementary table S5). Signal peptide encoding mRNAs were defined using the signalP program (Nielsen, 2017). Genes related to amino acid biosynthesis pathways were defined according to GO:0006520 (cellular amino acid metabolic process). Genes related to amino acid transport were defined according to GO:0003333 (amino acid transmembrane transport). Genes annotated as extracellular matrix were defined according to GO:0031012. A list of secreted mouse proteins was derived from UniProt (www.uniprot.org) by using the query: keyword:"Secreted [KW-0964]" AND reviewed:yes AND organism:"Mus musculus (Mouse) [10090]". A list of ATF4 target genes was taken from Han *et al.* (Han et al., 2013), and the significance of the overlap between this target set and the late specific ER stress-induced DEG group was calculated using a hypergeometric p-value.

Pathway enrichment analyses

Pathway enrichment was analysed using the DAVID 6.8 functional annotation tool (Jiao et al., 2012), with all expressed genes as background. Pathways with FDR corrected p-value <0.05 were designated as enriched. Induced and repressed signal peptide-encoding mRNAs were analyzed for pathway enrichment against a background of all expressed signal peptide-encoding mRNAs.

Amino acid biosynthesis metabolic map

We used KEGG Mapper – Search&Color Pathway (https://www.genome.jp/kegg/tool/map_pathway2.html) (Kanehisa and Goto, 2000) to examine the metabolic map of amino acid biosynthesis. We used the *Homo sapience* (hsa) database, as it is better annotated than the mouse database, with all other default settings. We loaded the list of upregulated late specific ER stress DEGs (described above), and these were overlaid on the map in red by the KEGG website.

Amino acid demand

We defined a metric called "amino acid demand" for each protein, which is the fraction of a group of amino acids within the sequence of that protein. Protein sequences were taken from mm10 Mus musculus NCBI RefSeq gene annotation database, with each gene represented by a single RefSeq protein (NP) and isoform (NM) sequences for the amino acid demand analyses and codon optimality analysis (see below). We calculated the amino acid demand for the set of amino acids comprising the UPR amino acid signature, in each protein sequence in the mouse genome. We further plotted the CDF of this metric in different protein groups compared to the background set of all expressed proteins.

tRNA Adaptation Index and codon optimality analysis

Codon optimality was calculated using the tRNA Adaptation Index (tAI) metric (dos Reis et al., 2003), which takes into account the tRNA copy number and the wobble interaction for each codon. We calculated tAI as in (dos Reis et al., 2003). Briefly, for each codon *i*, the number of tRNA genes in the mouse genome of each of the codon's anticodons (tGCNij) was extracted from the GtRNAdb database (Chan and Lowe, 2016). First, a *Wi* coefficient was calculated for each codon i as appears below. Every anticodon was assigned a coefficient; $s_{ij} = 0$ for perfectly matching codon-anticodon pairs, and $s_{ij} = 0.5$ for the codons with mismatch of the anticodon at the wobble position, according to the wobble rules.

$$W_i = \sum_{j=1}^{n} (1 - s_{ij}) tGCN_{ij}$$

Then, *Wi* values were normalized by dividing all *Wi* values by the maximal *Wi* value to generate w_i values. Codons with a w_i of zero were assigned the minimal w_i score. The tRNA Adaptation Index of each coding sequence was then calculated as a geometric mean of all tAI values of all the codons in the sequence (of length *L*), excluding the start and stop codons.

$$tAI = \left(\prod_{k=1}^{L} w_k\right)^{1/L}$$

To calculate the geometric mean for the tRNA Adaptation Index for a subset of amino acids, such as the UPR amino acid signature, only codons of the amino acids belonging to the subset were taken into account, both for protein groups of interest and for the background set of proteins.

Analysis of the optimality of the amino acid signature

To investigate the optimality of the UPR amino acid signature, comprising K amino acids, we compared the difference in median amino acid demand between UPR-induced proteins or secreted proteins, and all expressed mRNAs, to the distribution of these median differences according to all other possible sets of K non- or partly-essential amino acids (Fig. S7 and S9).

Mitochondrial translation

All 13 mitochondrially-encoded mRNAs showed highly increased levels following 1h of ER stress, and relative relaxation at subsequent timepoints (Fig. S8C). However, this seeming induction actually indicates the maintenance of non-repressed mitochondrial translation, upon overall cytoplasmic translation repression. Since overall cytosolic translation is inhibited during ER stress, with the highest degree of inhibition occurring at 1h (Gonen et al., 2019), the number of overall translating cytosolic ribosomes is reduced. Thus, the proportion of non-repressed mitochondrially-translated mRNAs would be higher at the earlier timepoints. A similar trend was demonstrated previously in other translation inhibition conditions (Iwasaki et al., 2016).

References

Chan, P.P., and Lowe, T.M. (2016). GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res *44*, D184-189.

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

dos Reis, M., Wernisch, L., and Savva, R. (2003). Unexpected correlations between gene expression and codon usage bias from microarray data for the whole Escherichia coli K-12 genome. Nucleic Acids Res *31*, 6976-6985.

Gonen, N., Sabath, N., Burge, C.B., and Shalgi, R. (2019). Widespread PERK-dependent repression of ER targets in response to ER stress. Sci Rep *9*, 4330.

Guan, B.J., van Hoef, V., Jobava, R., Elroy-Stein, O., Valasek, L.S., Cargnello, M., Gao, X.H., Krokowski, D., Merrick, W.C., Kimball, S.R*., et al.* (2017). A Unique ISR Program Determines Cellular Responses to Chronic Stress. Mol Cell *68*, 885-900 e886.

Han, J., Back, S.H., Hur, J., Lin, Y.H., Gildersleeve, R., Shan, J., Yuan, C.L., Krokowski, D., Wang, S., Hatzoglou, M*., et al.* (2013). ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. Nat Cell Biol *15*, 481-490.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature protocols *7*, 1534-1550.

Iwasaki, S., Floor, S.N., and Ingolia, N.T. (2016). Rocaglates convert DEAD-box protein eIF4A into a sequence-selective translational repressor. Nature *534*, 558-561.

Jiao, X., Sherman, B.T., Huang da, W., Stephens, R., Baseler, M.W., Lane, H.C., and Lempicki, R.A. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. Bioinformatics *28*, 1805-1806.

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res *28*, 27-30.

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

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

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol *15*, 550.

Nielsen, H. (2017). Predicting Secretory Proteins with SignalP. Methods Mol Biol *1611*, 59-73.