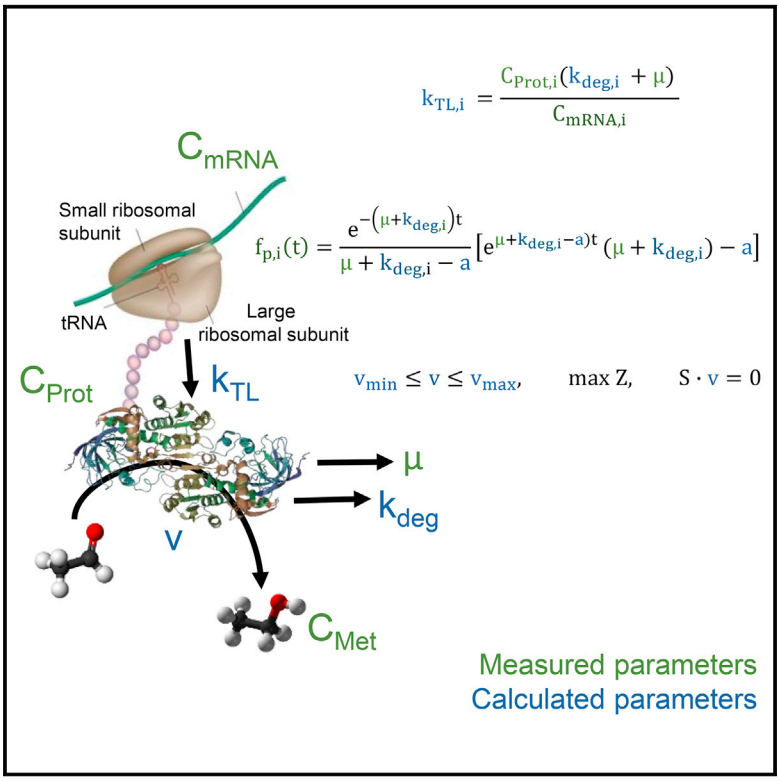
Article

Cell Systems

Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast

Graphical Abstract



Highlights

- mRNA and protein abundances were measured under ten environmental conditions
- Translation efficiency differed >400-fold between individual proteins
- Changes in protein levels were highly correlated with the changes in mRNA levels
- Most of the transcriptionally regulated fluxes were related to mitochondria

Authors

Petri-Jaan Lahtvee, Benjamín J. Sánchez, Agata Smialowska, Sergo Kasvandik, Ibrahim E. Elsemman, Francesco Gatto, Jens Nielsen

Correspondence

nielsenj@chalmers.se

In Brief

Absolute mRNA and protein abundance in combination with protein turnover measurements will give an opportunity to calculate translation efficiencies of individual proteins. In the current study, we cultivated yeast under ten environmental conditions and analyzed the translation efficiencies and studied the levels of regulation of metabolic fluxes.

Lahtvee et al., 2017, Cell Systems *4*, 495–504
May 24, 2017
http://dx.doi.org/10.1016/j.cels.2017.03.003

Cell Systems
Article

Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast

Petri-Jaan Lahtvee,[1,2] Benjamín J. Sánchez,[1,2] Agata Smialowska,[1,3] Sergo Kasvandik,[4] Ibrahim E. Elsemman,[5] Francesco Gatto,[1,6] and Jens Nielsen[1,2,5,7,*]
[1]Department of Biology and Biological Engineering, Systems and Synthetic Biology, Chalmers University of Technology, Kemivägen 10
[2]Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology
412 96 Gothenburg, Sweden
[3]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, 171 65 Solna, Sweden
[4]Institute of Technology, University of Tartu, 50411 Tartu, Estonia
[5]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970 Hørsholm, Denmark
[6]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA
[7]Lead Contact
*Correspondence: nielsenj@chalmers.se
http://dx.doi.org/10.1016/j.cels.2017.03.003

SUMMARY

Protein synthesis is the most energy-consuming process in a proliferating cell, and understanding what controls protein abundances represents a key question in biology and biotechnology. We quantified absolute abundances of 5,354 mRNAs and 2,198 proteins in *Saccharomyces cerevisiae* under ten environmental conditions and protein turnover for 1,384 proteins under a reference condition. The overall correlation between mRNA and protein abundances across all conditions was low (0.46), but for differentially expressed proteins (n = 202), the median mRNA-protein correlation was 0.88. We used these data to model translation efficiencies and found that they vary more than 400-fold between genes. Non-linear regression analysis detected that mRNA abundance and translation elongation were the dominant factors controlling protein synthesis, explaining 61% and 15% of its variance. Metabolic flux balance analysis further showed that only mitochondrial fluxes were positively associated with changes at the transcript level. The present dataset represents a crucial expansion to the current resources for future studies on yeast physiology.

INTRODUCTION

Environmental conditions shape the metabolism of all living cells. Cells respond to these changes by adjusting their metabolic fluxes, which are regulated on various levels. Acquiring a holistic understanding of how cellular metabolism is controlled requires reconstruction of complete regulatory networks (Barabasi et al., 2004; Liu et al., 2011, 2013). Although genome sequencing provides the list of components existing in every system, the whole network of interactions between cellular components has yet to be determined for virtually any organism (Goodwin et al., 2016; Reuter et al., 2015). For a better understanding of the regulatory network and dynamics among metabolic parameters, quantification of multiple levels of metabolic processes for a number of environmental conditions is crucial. As protein synthesis is the most energy-consuming process in proliferating cells, understanding how protein levels are controlled and how they influence metabolic fluxes plays an important role in designing new, more efficient, industrial producer strains or understanding the regulation of a disease. Recently, energetic limitations and proteome allocation has been determined as one of the main contributors controlling important cellular functions, such as overflow metabolism, in bacteria and yeast (Hui et al., 2015; Nilsson and Nielsen, 2016; Peebo et al., 2015; Schmidt et al., 2016). We therefore undertook a study of the budding yeast *Saccharomyces cerevisiae* to quantify at the cellular level mRNA and protein abundances. We aimed to quantify the levels of protein translation and protein degradation, and how protein abundance correlates with metabolic activities, here quantified by flux balance analysis.

Protein abundances can be described as a function of mRNA abundance, translation efficiency, and protein turnover (Figure 1A), where translation efficiency is the rate of mRNA translation into proteins within cells (measured in protein per mRNA per hour; Schwanhäusser et al., 2011), while protein turnover is the balance between protein synthesis and degradation (measured in protein per protein per hour). Relative mRNA levels, representing the expression changes compared with the reference conditions, have been measured for decades using genome-wide techniques such as microarrays and RNA sequencing (RNA-seq). Thus far, to our knowledge there has not yet been generated a quantitative mRNA dataset for budding yeast in which absolute abundances describing the total amount of transcripts in the cells were calibrated against spike-ins of known concentration of the transcript, although absolute quantification has been carried out for a fission yeast (Marguerat et al., 2012). In the yeast *S. cerevisiae*, translation efficiency has been previously estimated based on relative omics data pooled from various experiments (Csárdi et al., 2015) and measured through ribosome

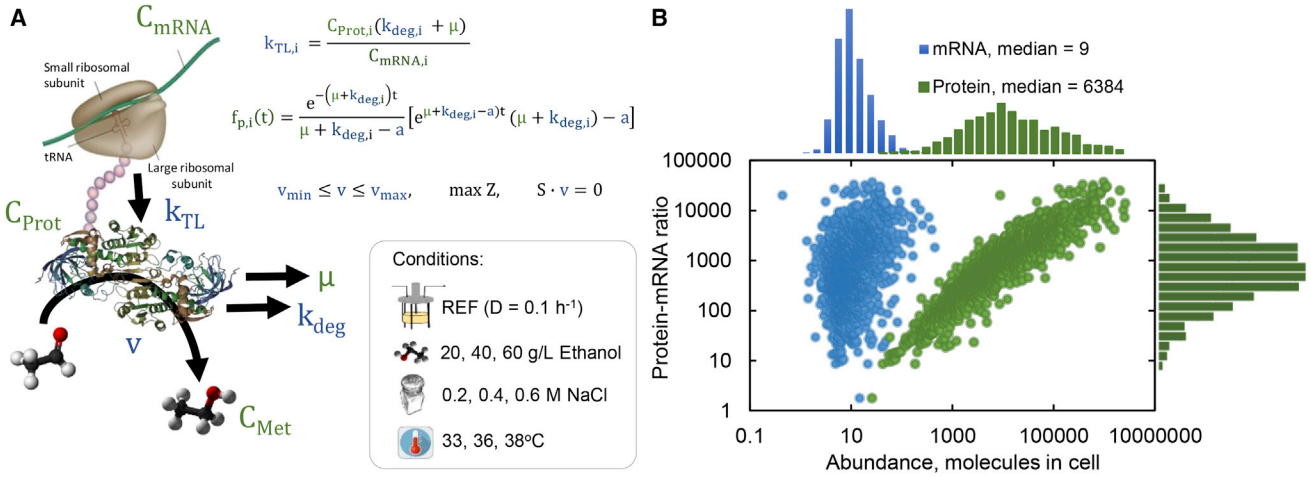


Figure 1. Overview of Collected and Calculated Data and Distribution of Absolute mRNA and Protein Abundances

(A) Representation of the mathematical model adopted to estimate the contribution of transcription, translation, and degradation in the control of protein levels in yeast. Absolute protein and mRNA abundances were quantified together with exometabolome for ten environmental conditions, all carried out in chemostats at constant dilution rate of 0.1 hr^{-1}. Translation efficiencies were estimated from the regression of measured protein versus transcript abundances. Protein turnover rates for individual proteins were estimated for the reference condition from non-linear fitting of measurements on the incorporation rate of labeled amino acid in the intracellular environment versus the proteome. Intracellular metabolic fluxes were estimated by computational modeling using flux balance analysis on experimentally measured constraints on exchange fluxes and biomass growth rate. Fluxes were correlated to mRNA changes to estimate transcriptional control of fluxes. Green and blue variables represent experimental and calculated values, respectively. C_{mRNA}, C_{Prot}, C_{Met}, absolute mRNA, protein, and metabolite concentration, respectively; k_{TL}, translation efficiency; k_{deg}, protein turnover; μ, specific growth rate; D, dilution rate; f_p, unlabeled fraction of the protein in a protein pool; t, time; a, rate at which intracellular unlabeled lysine is replaced by heavy lysine in its free amino acid pool; v, flux; Z, optimization function; S, stoichiometric matrix.

(B) Absolute mRNA and protein abundances were quantified. Absolute numbers for the reference condition are shown (D = 0.1 hr^{-1}, optimal environmental conditions).

profiling, a method whereby ribosome-protected mRNA fragments are sequenced and the translation efficiency is estimated based on the abundance of fragments belonging to the same gene (Brar et al., 2012; Ingolia et al., 2009; McManus et al., 2014; Weinberg et al., 2016). Due to the high coverage and accuracy in RNA sequencing, ribosome profiling is considered the best method for translation efficiency estimations. Nonetheless, measuring ribosome density in mRNAs does not provide a valid prediction of whether translation elongation rate varies from gene to gene or from one environmental condition to another, because this technique measures the density of ribosomes on an mRNA but not how fast they are moving. Lack of genome-scale quantitative data has currently prevented the calculations of translation efficiency based on mRNA abundance, protein abundance, and protein turnover. Protein turnover has been reported to play an important role in fine-tuning protein levels in cells. Previously reported protein turnover measurements have yielded varying results, with reported median protein turnover rates in the range of 0.078–0.98 hr^{-1} (Belle et al., 2006; Helbig et al., 2011; Hong et al., 2012; Khmelinskii et al., 2012; Pratt et al., 2002). There are many reasons for this variability, including differences in strains and experimental conditions as well as complications in the experimental setup and differences in data handling. From these analyses, however, it is clear that protein turnover constitutes a significant part of cellular maintenance and affects cellular energy metabolism (Hong et al., 2012).

Cellular phenotypes are determined largely by metabolic fluxes, and understanding how the cell is controlling metabolism is a central question in biology (Nielsen, 2003, 2011). Metabolic fluxes are determined to some extent by enzyme abundance, but many other factors also play a role, e.g., metabolite levels and enzyme post-transcriptional modifications. No method directly measures metabolic fluxes at the genome-scale level. Hence, computational methods have often been used in yeast to approximate the distribution of intracellular fluxes based on the reconstruction of a genome-scale metabolic network, constraints on the exchange fluxes, and the definition of a biomass composition under the assumption that the steady-state fluxes are mass-balanced and tailored to optimize the biomass growth rate (Kerkhoven et al., 2015; Sánchez and Nielsen, 2015; Simeonidis and Price, 2015). This approach, called flux balance analysis (Orth et al., 2010), enables comparison of metabolic fluxes with transcriptome and proteome data and provides information about the control of fluxes at the genome scale (Bordel et al., 2010).

In this study, we quantified protein synthesis in *S. cerevisiae* at the system level and correlated these data with metabolic fluxes. We obtained a global map of which fluxes are controlled by protein abundance in a eukaryotic cell. To quantify protein synthesis, we measured the absolute levels of individual transcripts and proteins in *S. cerevisiae* grown under ten environmental conditions in triplicate steady-state experiments at a constant specific growth rate. A steady physiological state was chosen to avoid bias caused by delays in protein synthesis occurring under dynamic environmental conditions. Protein turnover was measured and translation efficiency was calculated for 1,117 proteins. To understand the regulation from proteins to metabolic fluxes, we calculated intracellular flux distributions using a genome-scale metabolic model and flux balance analysis.

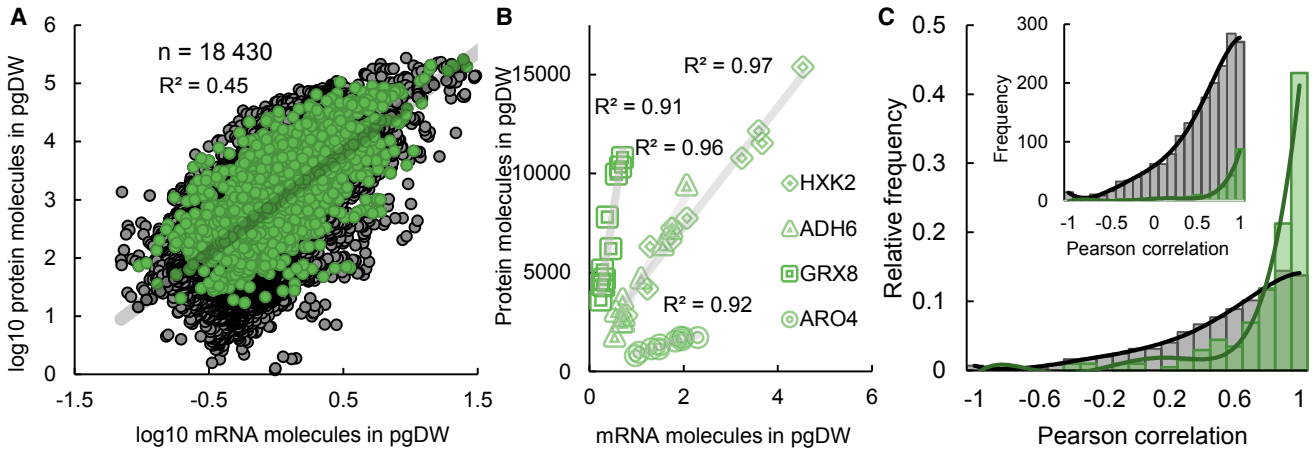


Figure 2. Integration of mRNA and Protein Abundances under Ten Environmental Conditions

(A) Correlation of absolute mRNA and protein abundances under all studied environmental conditions. Green color highlights proteins that show differential expression under some of the environmental conditions compared with the reference condition (adjusted $p < 0.01$).

(B) Correlation of absolute mRNA and protein abundances illustrated for individual proteins across the ten environmental conditions, showing a large variation in their slopes.

(C) Pearson's correlation distribution of absolute mRNA and protein abundances for individual proteins under the studied environmental conditions. The median of 0.88 was detected for the proteins that showed significant change in their transcript level (adjusted $p < 0.01$; green bars), and median of 0.48 for all the detected proteins (gray bars). Both relative and absolute frequency histograms are displayed.

A correlation analysis of these different layers of data allowed us to predict transcriptionally regulated fluxes and identify patterns in translation efficiency.

RESULTS

Quantitative Transcriptome and Proteome under Carbon-Limited Conditions

Absolute levels of the yeast *S. cerevisiae* transcriptome and proteome were quantified under a steady physiological state using chemostat cultures at a fixed dilution rate of 0.1 hr^{-1} (equal to specific growth rate). To harness yeast regulation, we perturbed the environment by applying three different states of stress (ethanol, salt, and increased temperature), each at three gradient steps, resulting in a total of nine perturbed conditions. Cultures were grown at the same dilution rate as for the reference condition (Figure 1A). Genome-wide transcriptome analysis was performed using RNA-seq. RNA-seq read counts for 5,354 transcripts were then calibrated based on the absolute concentrations of 18 mRNAs, which were selected to cover the overall mRNA expression dynamic range (Tables S1 and S2). Heavy (^{15}N,^{13}C) lysine-labeled biomass as an internal standard and Orbitrap mass spectrometry (MS) were used for quantification of 2,198 proteins (Table S3). For the absolute mRNA levels, the average SD of the residuals was calculated as 5.2%, and a bootstrapping method estimated an average abundance error of 1.16-fold with a 95% confidence interval. For the proteome data, the average SD of the residuals was 87.4%, and bootstrapping based on average abundance error with a 95% confidence interval was 2.0-fold. The significance of relative changes between different studied conditions was estimated using the false discovery rate (FDR) corrected p values (STAR Methods).

We counted the cells in the reference condition and measured the dry weight concentration. We then estimated a single cell weight equal to 13 ± 1 pg, similar to values reported in the literature and corresponding to a cell volume of 35 fL (Bryan et al., 2010; Klis et al., 2014; McMurrough and Rose, 1967). We quantified the total amounts of mRNA and proteins per picogram of dry cellular weight (DW) to allow for a cell-size-independent comparison under different environmental conditions. Under the reference condition, we found for each gene a median of 0.7 mRNA molecules and 491 protein molecules per picogram of DW corresponding to 9 and 6,384 molecules per cell under the reference condition, respectively (Figure 1B). To our knowledge, no other group has reported a combined absolute quantitative transcriptome and proteome dataset for *S. cerevisiae*. These numbers were in agreement with the total count calculations for proteins and mRNAs suggested in a prior study (Milo, 2013). As for other organisms, we observed a relatively narrow distribution in the count of mRNA molecules per cell that spans slightly more than two orders of magnitude (Marguerat et al., 2012; Schwanhäusser et al., 2011). Conversely, protein abundances spanned over five orders of magnitude. No gene ontology group-dependent distribution was detected, meaning that the distributions of both transcript and protein abundances are likely unrelated to their functional role in the cell (Figure S1). When mRNA levels were correlated with their corresponding proteins, a Pearson correlation of $R^2 = 0.51$ was detected for the reference condition and $R^2 = 0.45$ when leveraging on all the studied conditions (Figure 2A). Due to the use of steady-state conditions, this value is higher than previously reported for yeast or other organisms (Maier et al., 2009). However, only 46% of protein levels were predicted correctly based on linear correlation while experimental noise was taken into account for the calculations (STAR Methods), where approximately 10% of the detected proteins showed more than a 10-fold difference from the predicted value. Csárdi et al. (2015) suggest, based on pooled data analysis where experimental noise is accounted

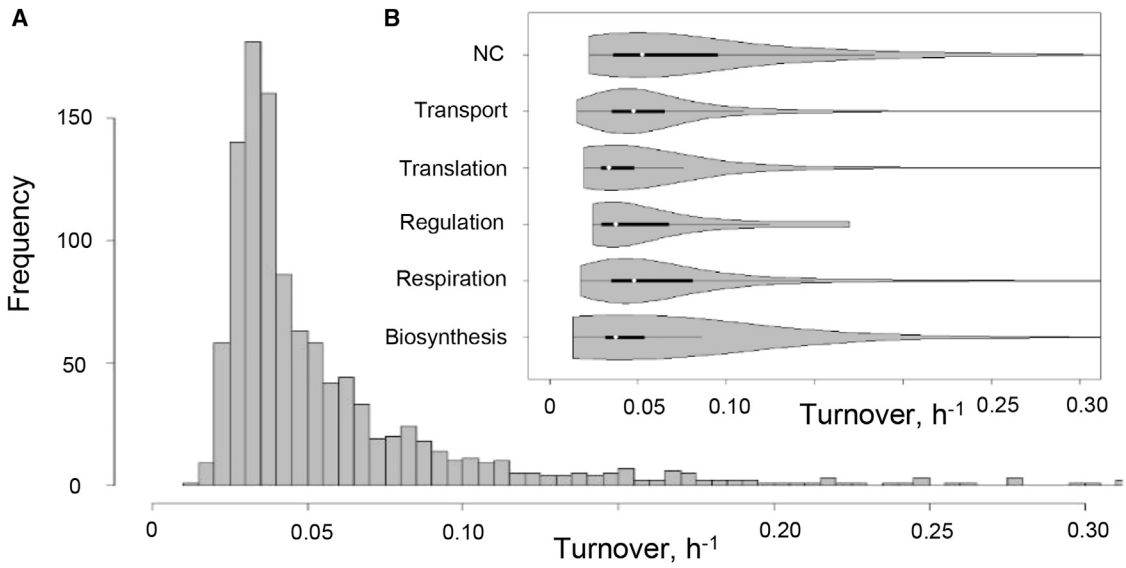


Figure 3. Protein Turnover under Glucose-Limited Chemostat Conditions

(A) Distribution of protein turnover for 1,384 proteins calculated based on 18 time points measuring labeled amino acid incorporation into intracellular amino acid pool and into individual proteins.

(B) No significant changes in protein turnover rates were detected between various functional groups as shown by violin plots (combining information from box plots and kernel density plots). White dot illustrates median value, black bold line 50% of the population, and thinner gray line 75% of the population.

for in exponentially growing yeast cells, that up to 85% of protein abundances can be predicted based on their transcriptome levels. We failed to support this conclusion with our steady-state data.

Transcriptional Control Dominates Protein Expression Changes in Yeast

Although we found relatively low correlation between the total proteome and mRNA concentration (Figure 2A), we took advantage of our data from multiple environmental conditions and examined the mRNA-protein correlation across conditions at the level of single genes. Only 202 proteins (11% of the detected) showed differential expression profile when compared with the reference condition (FDR <0.01). These 202 proteins correlated strongly with the corresponding mRNA level across the ten studied environmental conditions with a median Pearson correlation of 0.88 (representative examples shown in Figure 2B, system-level correlations shown in Figure 2C). This indicates that transcriptional control seems to be the major driver behind changes in protein levels. We speculated that variation in the translation efficiency between different genes could determine the imperfect correlations commonly detected between mRNA and protein levels in previous studies. These correlations were robust to changes in the selected confidence level.

Our integrated transcript-protein data resulted in 1,864 comparable transcript-protein pairs where differential expression was observed in 1,205 transcripts and 202 proteins. Among these pairs, 164 showed differential expression at both the transcript and protein level, which is 81% of all differentially expressed proteins and 15% of the differentially expressed transcripts. Hence, we observed 1,041 genes which changed their expression at the transcriptional level, but no significant change was detected at the protein level. It is possible that many of the 85% of transcriptional changes not reflected at the protein level could be a bias introduced by the lower accuracy in proteome measurements or the fact that the two data types span different orders of magnitude (Figure 1B). Thereby, protein variance may be overestimated and affect the statistical power needed to detect truly differentially expressed transcript versus protein pairs.

Protein Turnover for Individual Proteins

As differences in protein turnover and translation efficiency are expected to explain the observed differences in post-transcriptional control, we quantified individual protein turnover under the reference condition, which in combination with transcript and protein abundances enabled us to estimate translation efficiency. Two parallel chemostat experiments were carried out with a lysine auxotrophic (*Δlys1*) strain, and sampled 18 times during 13 hr after the minimal medium supplemented with light lysine was replaced with an identical medium containing heavy lysine. Based on samples collected at 18 time points in the replicate experiments, the protein turnover was calculated by fitting protein levels exponentially over time while adjusting for intracellular amino acid recycling between the proteome and free amino acid pool (methods similar to Hong et al., 2012; Lahtvee et al., 2014; Figure 1A). We obtained a statistically significant fit for 1,384 proteins (out of 1,591 detected; 87%, FDR <0.01), where the data were biased toward more abundant proteins (Figure S2). The accuracy of each fit was determined by calculating the 95% confidence interval for each estimate of the parameters. On average, we observed a 36% deviation from the median on each estimated confidence interval (Figure S3 and Table S5). The median protein turnover was determined to be 0.043 hr^{-1} (abundance-based weighted average 0.041 hr^{-1}), which corresponds to a median half-life of 17.2 hr (1^{st}–99^{th} quantile range: 1.3–34.0 hr; Figure 3A and Table S5). The median protein half-life

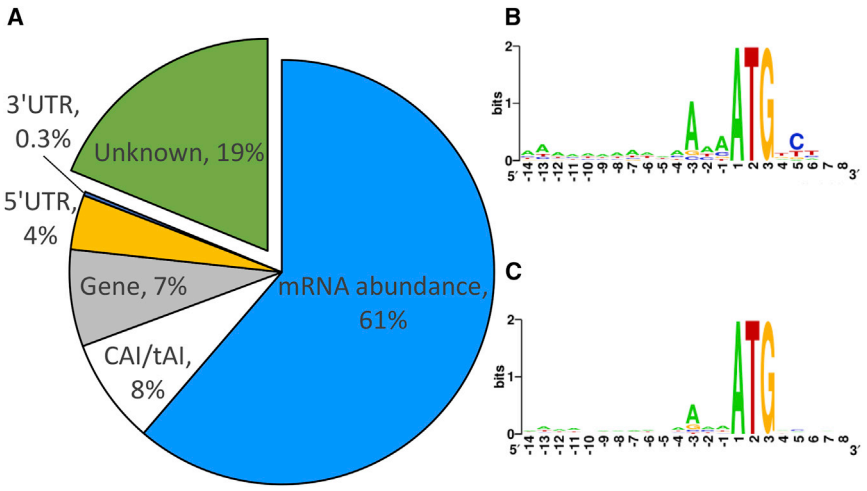


Figure 4. Distribution of Control over Translation Efficiency and Final Protein Abundances

(A) Non-linear regression analysis show the factors that determine the final protein abundance. Sixty-one percent is described by the mRNA abundance; 8% can be described by codon and tRNA adaptation indexes (CAI/tAI); 7% by the nucleotide frequency and the sequence in the coding region of a gene; and 4% and 0.3% by the nucleotide frequency and the sequence in the 5′ and 3′ UTR, respectively. Nineteen percent of the contribution was not described with the currently available data.

(B) Positional conservation analysis in the 5′ UTR and sequencing region of the mRNAs among the highest 10% translation efficiency.

(C) Positional conservation analysis in the 5′ UTR and sequencing region of the mRNAs among the lowest 10% translation efficiencies.

is 2.3-fold longer than the doubling time at the studied conditions. Measured protein turnovers showed a log-normal distribution. No significant differences were detected in the distributions between different functional groups of proteins (Figure 3B). However, a statistically significant enrichment in ribosomes and amino acid metabolic processes was detected among the 100 proteins with the highest detected turnover rate ($p < 0.001$). Only very moderate negative correlation was detected between the protein abundance and turnover ($R^2 = 0.23$, $p = 2.2 \times 10^{-16}$; Figure S4B).

Protein synthesis is an energy intensive process, and protein degradation will therefore contribute significantly to the maintenance ATP requirements. Synthesis of each peptide bond requires a little more than four ATP molecules (Stouthamer, 1973; Stephanopoulos et al., 1998), and degradation of a peptide bond has been estimated to utilize one ATP (Benaroudj et al., 2003). As protein polymerization has been estimated to be the highest cost for biomass synthesis (~23 mmol ATP gDW^{-1} h^{-1}), a median protein turnover of 0.043 hr^{-1} would result in an additional 12 mmol ATP gDW^{-1} h^{-1} for protein synthesis. Previously, using genome-scale models and flux balance analysis it had been estimated that the net production of ATP in the cell forms approximately 59 mmol ATP gDW^{-1} h^{-1} (Forster et al., 2003). Therefore, ~21% of the energy generated for biomass synthesis is used for protein turnover, which makes protein turnover a major client of energy metabolism in yeast.

Translation Efficiency and Its Regulation

Using the quantitative data for proteins, mRNA, and protein turnover, we calculated the translation efficiency for 1,117 proteins (Figure 1A and Table S6). To understand why the translation efficiency can differ between genes by up to 434-fold (1–99 percentile), we applied multivariate adaptive regression spline (MARS) analysis. Non-linear regression analysis MARS was applied in a similar fashion to Vogel et al. (2010) to describe the total protein synthesis. As input data for the regression analysis we used mRNA abundances together with the RNA secondary structure information (combination of singlets, duplexes, and triplexes) and length in the coding region as well as the 5′ and 3′ UTRs. In addition, information related to codon and tRNA adaptation indexes (indicating differences in translation elongation), protein turnover, and co-translational RNA decay (as introduced in Pelechano et al., 2015) were used. In total, the contributions of 259 features were tested individually. In the final round, 27 features were selected to determine the total amount of synthesized proteins, where each factor contributed more than 0.1%. We found that 61% of the final protein abundances can be predicted based on their corresponding mRNA abundance (Figure 4A). This high percentage is consistent with the strong linear correlation between mRNA and protein level found for most of the proteins. An additional 15% is related to the RNA secondary structure in the coding region (information pooled together from several related variables), which is related to the elongation speed of the ribosomes. Translation elongation was contributing mainly through codon and tRNA adaptation indexes, but also due to the high frequency of TT and GT sequences (for more information see Table S9). Four percent of the control was determined by the 5′ UTR of the genes, which indicates a small but significant role of the translation initiation. Only 0.4% of the protein abundances were described by the secondary structure of the 3′ UTR. Together, this analysis described 81% of the control at the final protein levels. An additional 19% of protein abundances were not described by the provided data and could be explained by the regulation of anti-sense RNAs, the effect of folding energy, or other factors.

Although the MARS analysis provided information about the preferred codons and sequence elements in coding and 5′ UTRs, it did not provide information on whether there are preferred positional sequences in the data. Therefore, we also studied the top versus bottom 10% of genes ranked by translation efficiency to detect consensus sequences associated with translation efficiency (in a window 15 bp upstream to 50 bp downstream of the transcription start site). Consistent with a previous study by Robbins-Pianka et al. (2010), which used ribosome profiling data from Ingolia et al. (2009) and Zur and Tuller (2013), we discovered that highly translated proteins showed: (1) high adenine content, especially at positions −1 and −3; and (2) that the highly translated proteins preferred serine encoded by the TCT codon as the first amino acid after the start codon (Figure 4B and Table S10). No significant consensus

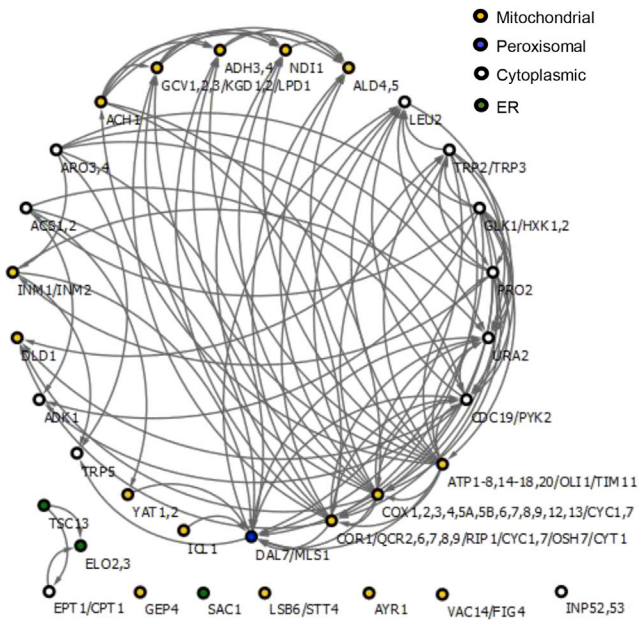


Figure 5. Clustering of Transcriptionally Regulated Fluxes

The majority of transcriptionally regulated fluxes are closely related to mitochondrial functions. Edges on the figure are connecting reactions located in proximity (maximum two reactions apart) according to the yeast consensus model Yeast 7.6. Nodes are color coded based on the protein location under optimal conditions (according to Cherry et al., 2012; Chong et al., 2015): orange, mitochondria; blue, peroxisomal; white, cytoplasmic; green, ER. See Tables S2 and S8 for gene names and according metabolic reactions, respectively. Kiwi analysis was used to determine connecting reactions.

was found among the proteins translated with the lowest studied translation efficiency (Figure 4C).

Is Translation Efficiency Changing under Different Conditions?

We examined for each protein whether its translation efficiency changed across different conditions. For simplicity, we opted to approximate the translation efficiency with the protein-to-mRNA level ratio (this approximation neglects protein turnover; however, differences in translation efficiency values were on average within 30% in the reference condition; Figures S4B and S4C; Table S4). Under the studied stress conditions, the approximate translation efficiency showed high in-between correlation, with an R^2 value of 0.98 between different conditions, while at most 58 proteins (3%) significantly changed translation efficiency. The only exception was high-temperature stress (38°C), where 334 proteins (17%) showed significant differences, equally split in increased versus decreased translation efficiency compared with reference (Figure S5). Among the genes that displayed an increased translation efficiency at 38°C, we detected an enrichment for glycolytic, protein folding, purine-ribonucleotide binding, and methane metabolism genes ($p < 0.001$). Reduced translation efficiency was displayed mainly among mitochondrial proteins involved in oxidative phosphorylation, the tricarboxylic acid cycle, and mitochondrial ribosomes. This outcome indicates that under conditions where respiration is inhibited, translation efficiency of respiratory proteins is also reduced. Even though robustness in translation efficiency could be attributed to the constant specific growth rate across conditions, these results suggest that stresses do not affect translation efficiency, except at high temperature, likely due to inhibition of respiration.

Transcriptionally Regulated Fluxes

Adaptation to environmental perturbations typically entails a reprogramming of cellular metabolism. We set out to infer from our data the contribution of transcriptional regulation to changes in metabolic fluxes. We used a genome-scale metabolic model for yeast, Yeast v7.6 (Aung et al., 2013), and implemented flux balance analysis to calculate intracellular flux distributions across conditions. The model was used to simulate fluxes subject to experimentally measured constraints, such as exchange fluxes and condition-dependent biomass composition (measured lipid composition, total protein, RNA, trehalose, and glycogen content). The simulation assumed that metabolic fluxes were geared to maximize ATP consumption, in other words that the objective function of yeast metabolism is maximum energy yield in the form of ATP. Calculated fluxes were further subjected to random sampling at 90% of the maximal ATP consumption value to estimate a distribution of suboptimal values for each flux (Table S7). The simulation was carried out for each environmental condition, and significantly changed fluxes were compared with changes in the associated protein and mRNA levels. Protein complexes, isozymes, and promiscuous enzymes were taken into account according to the data available in SWISS-PROT (STAR Methods). We integrated 314 flux-protein-transcript combinations, of which 51 (16%) displayed a consistent trend at the transcriptional and protein level (i.e., decrease in flux correlated with decrease in transcript level, or vice versa, $R^2 > 0.5$ and $p < 0.01$) (Figure 5 and Table S8). Enrichment analysis of those genes showed strong enrichment toward oxidative phosphorylation genes ($p < e\text{-}14$) and aromatic amino acid biosynthetic processes ($p < e\text{-}5$). Furthermore, these fluxes are highly connected with each other (Figure 5), which points to a common transcriptional regulation. A possible explanation for transcriptional control of oxidative phosphorylation may be that it is essential to coordinate flux control through all the steps of oxidative phosphorylation in order to avoid the formation of reactive oxidative species and cytochrome *c* leakage, resulting in apoptosis.

DISCUSSION

Here, we provide an absolutely quantified dataset of the transcriptome (5,354 transcripts) and proteome (2,198 proteins) of *S. cerevisiae* grown under ten environmental conditions. Using these data, we estimated the protein turnover rate (for 1,384 proteins) and calculated the translation efficiency for 1,117 proteins. Finally, we simulated the fluxome and ultimately determined the transcriptional control of metabolic regulation.

In contrast to previous studies that reported relative mRNA quantification in yeast, we used spike-in standards to calibrate absolute mRNA abundances in *S. cerevisiae*. Absolute protein levels have been measured before using tandem affinity purification- or GFP-tagged proteins (Chong et al., 2015; Ghaemmaghami et al., 2003) and by a targeted proteomics approach based on selected reaction monitoring (Picotti et al., 2009). However,

these measurements were not collected together with transcriptome analysis under the same experimental conditions nor over a wide range of conditions. In this sense, this study represents a comprehensive catalog of mRNA and protein abundance profiles encompassing respiratory environmental conditions as well as nine environmentally perturbed conditions. The correlation between mRNA and protein levels was found in the same range as reported previously with relative abundance data (Maier et al., 2009). However, because we collected mRNA and protein measurements for a variety of environmental conditions, we could detect high transcriptional control over protein levels when the mRNA and protein levels for individual proteins were correlated over various environmental conditions (Figures 3A and 3B). We found that although changes in protein levels are well correlated with the changes in the transcripts, the translation efficiency of proteins is highly variable between individual proteins. Similar results have been detected for human tissue, whereby data collected from databases were combined and relatively constant protein-mRNA ratios were found for individual proteins (Wilhelm et al., 2014). Recently, another study reported constant tissue-independent protein-mRNA ratios for 55 proteins, for the first time measured from the same sample (Edfors et al., 2016). Our data, however, represent a larger global study for yeast, whereby constant protein-mRNA ratios were quantified for more than 1,800 proteins across the ten environmental conditions studied.

To understand whether differences in protein-mRNA ratios can be caused by significant differences in protein degradation rates, we measured protein turnover rates for 1,384 proteins in the reference condition. The rates have a log-normal distribution with a median value of 0.043 hr^{-1}. Previously, in chemostats at a dilution rate of 0.1 hr^{-1} and under nitrogen limitation conditions, protein turnover was estimated for 641 proteins at a median value of 0.064 hr^{-1} (Helbig et al., 2011). Despite a 33% difference in the median value, histograms illustrating the distribution of individual protein turnover rates showed an almost complete overlap (Figure S6). Under carbon limitation, protein turnover has been measured for the whole proteome using uniformly labeled glucose, whereby measurements showed different values for different amino acids with a median of 0.078 hr^{-1} (Hong et al., 2012). Much more variability has been observed between measurements carried out under batch conditions, often caused by incorrect methods used for protein turnover estimations. In the current study, protein turnover was measured only in the reference condition, and further analysis is required to determine the growth rate and dependence on environmental conditions of protein turnover. By combining protein abundances and measured turnover under the reference conditions, we detected no significant changes for the translation efficiency of individual proteins across the studied conditions. At a median protein turnover of 0.043 hr^{-1} at a specific growth rate of 0.1 hr^{-1}, we estimated that approximately 21% of the total ATP used for biomass synthesis is consumed solely to resynthesize proteins that are degraded.

Translation efficiency calculated in this study varied 434-fold in the 1^{st}–99^{th} percentile range between individual proteins. Similar variation has been reported based on ribosome profiling (Ingolia et al., 2009). However, a more recent study on ribosomal profiling of yeast demonstrated only 15-fold variability among individual proteins in the 1^{st}–99^{th} percentile range (Weinberg et al., 2016). Although the differences are significant, they can be explained by the large effect on translational elongation rate which is not taken into account in ribosomal profiling experiments. To describe this large range of translation efficiencies in our study, we performed position-dependent conservation analysis for different nucleotides. Although environmental conditions chosen for the analysis were significantly different between our data and those from Ingolia et al., we found a similar conservation pattern for the highly translated proteins (Gingold and Pilpel, 2011; Ingolia et al., 2009; Robbins-Pianka et al., 2010), where lysine (encoded by AAA) and serine (encoded by TCT) were the most frequently observed amino acids just before and after the start codon, respectively. Based on the mRNA abundances, content of the sequence information, and codon/tRNA adaptation indexes, non-linear regression analysis was able to describe 81% of the variability in the total protein synthesis. A similar approach has been used previously, whereby 67% of the variability was described for a human cell line (Vogel et al., 2010). Consistently with the observed linear correlation between mRNA and protein levels for many proteins in our study, 61% of the variability was described by the mRNA abundances. In addition, 15% of the control accounted for the features related to translation elongation and only 4% for features related to translation initiation. Previous studies analyzing the translational control for *Escherichia coli* have assumed that translation initiation was the most critical step in regulating translation efficiency, whereas folding energy and codon bias had earlier been claimed to account for the majority of translation control in *S. cerevisiae* (Tuller et al., 2010). Significant differences were found among translation efficiencies between individual proteins, and minimal variability was simultaneously found in translation efficiencies between tested environmental conditions. We showed that mRNA abundances and elongation of translation were the two major factors for determining the amount of total protein synthesis and, hence, describing the translation efficiency.

To add another layer of information, we estimated the distribution of metabolic fluxes in the cell using genome-scale metabolic models, flux balance analysis, and random sampling. This approach allowed us to determine which metabolic fluxes are transcriptionally controlled—the most conventional level of manipulation by metabolic engineers. We detected a highly enriched group of fluxes regulated at the transcriptional level related to mitochondrial enzymes, many of which take part in oxidative phosphorylation. Transcriptional control of mitochondrial biogenesis and function has also been observed for higher eukaryotes (Hock and Kralli, 2009), and is likely to be an evolutionarily conserved feature of eukaryotic cells to ensure balanced flux through oxidative phosphorylation. Lack of flux balance in this pathway will lead to accumulation of reactive oxygen species and possible leakage of cytochrome *c* to the cytoplasm, resulting in induction of apoptosis.

Our findings show that transcriptional regulation of proteins is a common feature. However, metabolic fluxes were only modestly related to transcriptional changes, as observed previously for *Bacillus subtilis* and *E. coli* (Chubukov et al., 2013; Valgepea et al., 2013). This might suggest that fluxes are mainly controlled at a post-translational level. In a study by Sauer and colleagues, post-translational modifications were not found to have a dominant role in the control of metabolism of *B. subtilis* (Chubukov et al., 2013). These results suggest that

most of the control in metabolism is attributable to enzyme properties (perhaps with the exception of mitochondrial proteins, as noted above) and opens a possibility to concentrate on global studies of the behavior of enzyme activities. In addition to the findings presented here, our study compiled an exhaustive set of absolutely quantitative measurements of proteins and mRNA for yeast, having generated over 250,000 quantitative data points under common environmental conditions. We believe that this represents a valuable resource for the systems biology community toward the realization of quantitative models of regulation in eukaryotes.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Sampling from Bioreactor
  - Exometabolome, Biomass Composition Analysis and Cell Counting
  - RNA Sequencing
  - Quantitative Proteome Measurements
  - Nano-LC/MS/MS Sample Preparation
  - Nano-LC/MS/MS Analysis
  - Protein Turnover Measurements
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - RNA Quantification
  - Mass-spectrometric Raw Data Identification and Quantification of Proteome
  - Curve Fitting for Protein Turnover Calculations
  - Noise Adjusted Correlation
  - Flux Balance Analysis (FBA) and Random Sampling
  - Estimation of ATP Expenditure for Protein Turnover
  - Enrichment Analysis
  - Translation Efficiency
  - Multi Array Regression Spline Analysis
  - Positional Analysis
- DATA AND SOFTWARE AVAILABILITY
  - Transcriptomics Data
  - Proteomics Data

SUPPLEMENTAL INFORMATION

Supplemental Information includes 6 figures and 12 tables and can be found with this article online at http://dx.doi.org/10.1016/j.cels.2017.03.003.

AUTHOR CONTRIBUTIONS

Conceptualization, P.-J.L. and J.N.; Investigation, P.-J.L., A.S., and S.K.; Formal Analysis and Methodology, P.-J.L., B.J.S., I.E.E., and F.G.; Writing – Original Draft, P.-J.L.; Writing – Review & Editing, B.J.S., A.S., I.E.E., S.K., F.G., and J.N.; Funding Acquisition, J.N.

ACKNOWLEDGMENTS

Clara Navarrete is acknowledged for her instructions in strain development, and the Center of Food and Fermentation Technologies, Estonia, for the analysis of intracellular lysine labeling patterns. Eduard Kerkhoven and Avlant Nilsson are acknowledged for their critical reading of the manuscript, and Leif Väremo for his consultation. This investigation was financially supported by the Knut and Alice Wallenberg Foundation and the Novo Nordisk Foundation.

Received: September 6, 2016
Revised: January 31, 2017
Accepted: March 1, 2017
Published: March 29, 2017

REFERENCES

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Aung, H.W., Henry, S.A., and Walker, L.P. (2013). Revising the representation of fatty acid, glycerolipid, and glycerophospholipid metabolism in the consensus model of yeast metabolism. Ind. Biotechnol. (New Rochelle N. Y.) *9*, 215–228.

Barabasi, A.-L., Oltvai, Z.N.Z.N., and Barabási, A.-L. (2004). Network biology: understanding the cell's functional organization. Nat. Rev. Genet. *5*, 101–113.

Belle, A., Tanay, A., Bitincka, L., Shamir, R., and O'Shea, E.K. (2006). Quantification of protein half-lives in the budding yeast proteome. Proc. Natl. Acad. Sci. USA *103*, 13004–13009.

Benaroudj, N., Zwickl, P., Seemüller, E., Baumeister, W., and Goldberg, A.L. (2003). ATP hydrolysis by the proteasome regulatory complex PAN serves multiple functions in protein degradation. Mol. Cell *11*, 69–78.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B (Methodol.) *57*, 289–300.

Bordel, S., Agren, R., and Nielsen, J. (2010). Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. PLoS Comput. Biol. *6*, e1000859.

Brar, G.A., Yassour, M., Friedman, N., Regev, A., Ingolia, N.T., and Weissman, J.S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. Science *335*, 552–557.

Bryan, A.K., Goranov, A., Amon, A., and Manalis, S.R. (2010). Measurement of mass, density, and volume during the cell cycle of yeast. Proc. Natl. Acad. Sci. USA *107*, 999–1004.

Charif, D., and Lobry, J.R. (2007). SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In Structural Approaches to Sequence Evolution: Molecules, Networks, Populations, U. Bastolla, M. Porto, E. Roman, and M. Vendruscolo, eds. (Springer), pp. 207–232.

Cherry, J.M., Hong, E.L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E.T., Christie, K.R., Costanzo, M.C., Dwight, S.S., Engel, S.R., et al. (2012). Saccharomyces genome database: the genomics resource of budding yeast. Nucleic Acids Res. *40*, 1–6.

Chong, Y.T., Koh, J.L.Y., Friesen, H., Duffy, K., Cox, M.J., Moses, A., Moffat, J., Boone, C., and Andrews, B.J. (2015). Yeast proteome dynamics from single cell imaging and automated analysis. Cell *161*, 1413–1424.

Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R.J., Jules, M., Link, H., Aymerich, S., Stelling, J., and Sauer, U. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. Mol. Syst. Biol. *9*, 709.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Crooks, G., Hon, G., Chandonia, J., and Brenner, S. (2004). NCBI GenBank FTP Site\nWebLogo: a sequence logo generator. Genome Res. *14*, 1188–1190.

Csárdi, G., Franks, A., Choi, D.S., Airoldi, E.M., and Drummond, D.A. (2015). Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast. PLoS Genet. *11*, e1005206.

Edfors, F., Danielsson, F., Hallström, B.M., Käll, L., Lundberg, E., Pontén, F., and Uhlén, M. (2016). Gene-specific correlation of RNA and protein levels in human cells and tissues. Mol. Syst. Biol. *12*, 883.

Forster, J., Famili, I., Palsson, B.O., and Nielsen, J. (2003). Genome-scale reconstruction of the *Saccharomyces cerevisiae* Metabolic Network. Genome Res. *13*, 244–253.

Ghaemmaghami, S., Huh, W., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. (2003). Global analysis of protein expression in yeast. Nature *108*, 737–741.

Gingold, H., and Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. Mol. Syst. Biol. *7*, 481.

Goodwin, S., McPherson, J.D., and McCombie, W.R. (2016). Coming of age: ten years of next-generation sequencing technologies. Nat. Rev. Genet. *17*, 333–351.

Helbig, A.O., Daran-Lapujade, P., van Maris, A.J., de Hulster, E.A., de Ridder, D., Pronk, J.T., Heck, A.J.R., and Slijper, M. (2011). The diversity of protein turnover and abundance under nitrogen-limited steady-state conditions in *Saccharomyces cerevisiae*. Mol. Biosyst. *7*, 3316–3326.

Hock, M.B., and Kralli, A. (2009). Transcriptional control of mitochondrial biogenesis and function. Annu. Rev. Genet. *71*, 177–203.

Hong, K.-K., Hou, J., Shoaie, S., Nielsen, J., and Bordel, S. (2012). Dynamic (13) C-labeling experiments prove important differences in protein turnover rate between two *Saccharomyces cerevisiae* strains. FEMS Yeast Res. *12*, 741–747.

Hui, S., Silverman, J.M., Chen, S.S., Erickson, D.W., Basan, M., Wang, J., Hwa, T., and Williamson, J.R. (2015). Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Mol. Syst. Biol. *11*, 784.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Kerkhoven, E.J., Lahtvee, P.-J., and Nielsen, J. (2015). Applications of computational modeling in metabolic engineering of yeast. FEMS Yeast Res. *15*, 1–13.

Khmelinskii, A., Keller, P.J., Bartosik, A., Meurer, M., Barry, J.D., Mardin, B.R., Kaufmann, A., Trautmann, S., Wachsmuth, M., Pereira, G., et al. (2012). Tandem fluorescent protein timers for in vivo analysis of protein dynamics. Nat. Biotechnol. *30*, 708–714.

Khoomrung, S., Chumnanpuen, P., Jansa-Ard, S., Ståhlman, M., Nookaew, I., Borén, J., and Nielsen, J. (2013). Rapid quantification of yeast lipid using microwave-assisted total lipid extraction and HPLC-CAD. Anal. Chem. *85*, 4912–4919.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*, R36.

Klis, F.M., de Koster, C.G., and Brul, S. (2014). Cell wall-related bionumbers and bioestimates of *Saccharomyces cerevisiae* and *Candida albicans*. Eukaryot. Cell *13*, 2–9.

Lahtvee, P.-J., Adamberg, K., Arike, L., Nahku, R., Aller, K., and Vilu, R. (2011). Multi-omics approach to study the growth efficiency and amino acid metabolism in *Lactococcus lactis* at various specific growth rates. Microb. Cell Fact *10*, 12.

Lahtvee, P.-J., Seiman, A., Arike, L., Adamberg, K., and Vilu, R. (2014). Protein turnover forms one of the highest maintenance costs in *Lactococcus lactis*. Microbiology *160*, 1501–1512.

Lahtvee, P.J., Kumar, R., Hallström, B.M., and Nielsen, J. (2016). Adaptation to different types of stress converge on mitochondrial metabolism. Mol. Biol. Cell *27*, 2505–2514.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2011). Controllability of complex networks. Nature *473*, 167–173.

Liu, Y.-Y., Slotine, J.-J., and Barabási, A.-L. (2013). Observability of complex systems. Proc. Natl. Acad. Sci. USA *110*, 2460–2465.

Maier, T., Güell, M., and Serrano, L. (2009). Correlation of mRNA and protein in complex biological samples. FEBS Lett. *583*, 3966–3973.

Marguerat, S., Schmidt, A., Codlin, S., Chen, W., Aebersold, R., and Bähler, J. (2012). Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. Cell *151*, 671–683.

McManus, C.J., May, G.E., Spealman, P., and Shteyman, A. (2014). Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. Genome Res. *24*, 422–430.

McMurrough, I., and Rose, A.H. (1967). Effect of growth rate and substrate limitation on the composition and structure of the cell wall of *Saccharomyces cerevisiae*. Biochem. J. *105*, 189–203.

Milborrow, S. (2011). Package 'earth'3.2–1. Multivariate adaptive regression spline models.

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays *35*, 1050–1055.

Nielsen, J. (2003). It is all about metabolic fluxes. J. Bacteriol. *185*, 7031–7035.

Nielsen, J. (2011). Transcriptional control of metabolic fluxes. Mol. Syst. Biol. *7*, 478.

Nilsson, A., and Nielsen, J. (2016). Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. Sci. Rep. *6*, 22264.

Orth, J.D., Thiele, I., and Palsson, B.Ø. (2010). What is flux balance analysis? Nat. Biotechnol. *28*, 245–248.

Peebo, K., Valgepea, K., Maser, A., Nahku, R., Adamberg, K., and Vilu, R. (2015). Proteome reallocation in *Escherichia coli* with increasing specific growth rate. Mol. Biosyst. *11*, 1184–1193.

Pelechano, V., Wei, W., and Steinmetz, L.M. (2015). Widespread co-translational RNA decay reveals ribosome dynamics. Cell *161*, 1400–1412.

Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., and Aebersold, R. (2009). Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Cell *138*, 795–806.

Pratt, J.M., Petty, J., Riba-Garcia, I., Robertson, D.H.L., Gaskell, S.J., Oliver, S.G., and Beynon, R.J. (2002). Dynamics of protein turnover, a missing dimension in proteomics. Mol. Cell. Proteomics *1*, 579–591.

Reimand, J., Arak, T., and Vilo, J. (2011). G: Profiler—a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. *39*, 307–315.

Reuter, J.A., Spacek, D.V., and Snyder, M.P. (2015). High-throughput sequencing technologies. Mol. Cell *58*, 586–597.

Robbins-Pianka, A., Rice, M.D., and Weir, M.P. (2010). The mRNA landscape at yeast translation initiation sites. Bioinformatics *26*, 2651–2655.

Sánchez, B.J., and Nielsen, J. (2015). Genome scale models of yeast: towards standardized evaluation and consistent omic integration. Integr. Biol. *7*, 846–858.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. *6*, 1290–1307.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent *Escherichia coli* proteome. Nat. Biotechnol. *34*, 104–110.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Sharp, P.M., and Cowe, E. (1991). Synonymous codon usage in *Saccharomyces cerevisiae*. Yeast *7*, 657–678.

Simeonidis, E., and Price, N.D. (2015). Genome-scale modeling for metabolic engineering. J. Ind. Microbiol. Biotechnol. *42*, 327–338.

Stephanopoulos, G., Aristidou, A.A., and Nielsen, J.H. (1998). Metabolic Engineering: Principles and Methodologies (Academic Press).

Stouthamer, A.H. (1973). A theoretical study on the amount of ATP required for synthesis of microbial cell material. Antonie Van Leeuwenhoek *39*, 545–565.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tuller, T., Waldman, Y.Y., Kupiec, M., and Ruppin, E. (2010). Translation efficiency is determined by both codon bias and folding energy. TL-107. Proc. Natl. Acad. Sci. USA *107*, 3645–3650.

Valgepea, K., Adamberg, K., Seiman, A., and Vilu, R. (2013). *Escherichia coli* achieves faster growth by increasing catalytic and translation rates of proteins. Mol. Biosyst. *9*, 2344–2358.

Vizcaíno, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The Proteomics Identifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. *41*, D1063–D1069.

Vogel, C., de Sousa Abreu, R., Ko, D., Le, S.-Y., Shapiro, B.A., Burns, S.C., Sandhu, D., Boutz, D.R., Marcotte, E.M., and Penalva, L.O. (2010). Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line. Mol. Syst. Biol. *6*, 400.

Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., Plotkin, J.B., and Bartel, D.P. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. Cell Rep. *14*, 1787–1799.

Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., and Mathieson, T. (2014). Mass-spectrometry-based draft of the human proteome. Nature *509*, 582–587.

Yassour, M., Kaplan, T., Fraser, H.B., Levin, J.Z., Pfiffner, J., Adiconis, X., Schroth, G., Luo, S., Khrebtukova, I., Gnirke, A., et al. (2009). Ab initio construction of a eukaryotic transcriptome by massively parallel mRNA sequencing. Proc. Natl. Acad. Sci. USA *106*, 3264–3269.

Zur, H., and Tuller, T. (2013). New universal rules of eukaryotic translation initiation fidelity. PLoS Comput. Biol. *9*, e1003136.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
L-LYSINE:2HCL (13C6, 99%; 15N2, 99%)	Cambridge Isotope Laboratories	CNLM-291-H-PK
Critical Commercial Assays		
PierceTM BCA Protein Assay Kit	Thermo Scientific	23225
RNA quantification (RNeasy Mini Kit)	Qiagen	74104
Amino acid derivatization kit AccQ•Tag™ Ultra	Waters	186003836
Deposited Data		
RNA sequencing	ArrayExpress	ArrayExpress: E-MTAB-4044
Absolute proteome and protein turnover	ProteomeXchange	PRIDE: PXD005041
Experimental Models: Organisms/Strains		
Saccharomyces cerevisiae CEN.PK113-7D	Euroscarf	889517
Saccharomyces cerevisiae CEN.PK113-7D $_{\Delta}$LYS1	This study	NA
Software and Algorithms		
Multivariate Adaptive Regression Spline analysis in R ("earth" package)	CRAN	https://cran.r-project.org/web/packages/earth/index.html
Proteome turnover calculation algorithm	This study	NA
Yeast v7.6 consensus genome scale model	Sourceforge	yeastnet
COBRA Toolbox	GitHub	opencobra/cobratoolbox
RAVEN Toolbox	GitHub	SysBioChalmers/RAVEN
Matlab 2012b	Mathworks	https://www.mathworks.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact Jens Nielsen (nielsenj@chalmers.se).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The yeast *Saccharomyces cerevisiae* CEN.PK113-7D (MATa, MAL2–8c, SUC2) was used if not stated otherwise. As an internal standard for quantitative proteome analysis and protein turnover measurements *Saccharomyces cerevisiae* CEN.PK113-7D ΔLYS1 strain was used. Cultures were stored in aliquot glycerol stocks at -80°C.

Experiments were carried out under glucose limited chemostat conditions on minimal mineral medium as described in Lahtvee et al. (2016). Briefly, ten different environmental conditions were studied at the constant specific growth rate of 0.1 h^{-1}. In addition to the reference conditions (optimal environmental conditions: temperature 30°C, pH 5.5, $pO_2 > 30\%$, agitation 600 rpm), three stress conditions (ethanol, osmotic pressure and heat) were applied in three gradual steps until the highest level of stress was reached while maintaining the chosen specific growth rate. Selected stress conditions were the following: Ethanol: 20, 40 and 60 g L^{-1}; osmolarity: 0.2, 0.4, 0.6 M NaCl; temperature: 33, 36, 38°C. Experiments were carried out in Dasgip 1 L bioreactors (Jülich, Germany) equipped with off-gas analysis, pH, temperature and dissolved oxygen sensors. Base medium used contained 10 g glucose, 5 g $(NH_4)_2SO_4$, 3 g KH_2PO_4 and 0.5 g $MgSO_4$ per litre, in addition to 1 mL of trace elements solution and 1 mL of vitamin solution. The trace element solution contained, per litre (pH=4): EDTA (sodium salt), 15.0 g; $ZnSO_4 \cdot 7H_2O$, 4.5 g; $MnCl_2 \cdot 2H_2O$, 0.84 g; $CoCl_2 \cdot 6H_2O$, 0.3 g; $CuSO_4 \cdot 5H_2O$, 0.3 g; $Na_2MoO_4 \cdot 2H_2O$, 0.4 g; $CaCl_2 \cdot 2H_2O$, 4.5 g; $FeSO_4 \cdot 7H_2O$, 3.0 g; H_3BO_3, 1.0 g; and KI, 0.10 g.). The vitamin solution contained, per liter (pH=6.5): biotin, 0.05 g; p-amino benzoic acid, 0.2 g; nicotinic acid, 1 g; Ca-pantothenate, 1 g; pyridoxine-HCl, 1 g; thiamine-HCl, 1 g and myo-inositol, 25 g.

METHOD DETAILS

Sampling from Bioreactor

Sampling for exometabolome analysis was carried out by immediate filtration of the culture broth. The supernatant was collected and stored at -20°C until analysis. Sampling for the transcriptome, proteome, and lipid analyses was carried out as follows: the dead

volume in the tubing was collected and discarded; biomass was collected from the reactor with a syringe and immediately injected into Eppendorf tubes already placed in a centrifuge pre-cooled to 4°C; samples were centrifuged for 22 seconds; supernatant was discarded and cell pellets were snap-frozen in liquid nitrogen. The whole procedure from sample collection to the final freezing step took less than 45 seconds. Samples were stored at -80°C until analysis. For biomass concentration measurements ca 40 mL of culture broth was collected from the outflow into falcon tubes placed in ice and measured gravimetrically using pre-weight filter plates.

Exometabolome, Biomass Composition Analysis and Cell Counting

An HPLC (ultimate 3000 HPLC system; Thermo Fisher Scientific, Waltham, MA) equipped with a Bio- Rad (Hercules, CA) HPX-87H column and an IR detector was used for the detection and quantification of the following extracellular metabolites – glucose, ethanol, pyruvate, succinate, acetate. Isocratic elution of 5 mM H_2SO_4 at a flow rate of 0.6 mL min^{-1} and at 45°C was used. Lipids were measured as described in Khoomrung et al. (2013). Briefly, dry biomass was extracted in anaerobic conditions in $CHCl_3$:MeOH (2:1) solution, while microwaved for 10 min at 60°C and 1000 W. After addition of 0.73% NaCl solution, organic layer was transferred to a new tube, concentrated, and analysed using HPLC (Dionex; ultimate 3000 HPLC system, Germany) equipped with CAD detector (Corona; ESA, Chelmsford, Massachusetts, United StatesMA, U.S.A.). Exchange fluxes and biomass composition used for the FBA is described in the Table S12. The RNA concentration was determined using a NanoDrop 2000 (Thermo Fisher Scientific, Wilmington, United States) after the RNA extraction using RNeasy Mini Kit (Qiagen, Hilden, Germany). Total protein was measured using a commercial Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). The cells were disrupted using glass beads and total protein measurements were carried out according to the manufacturer's instructions. Cell counting was carried out in Neubauer 0.100 mm haemocytometer (Assistant, Germany) using a Leica DM E microscope (Leica Microsystems GmbH, Wetzlar, Germany).

RNA Sequencing

RNA from the biomass samples was extracted and purified using Qiagen RNeasy Mini Kit extraction and DNA degradation according to the user's manual (Qiagen, Hilden, Germany). Integrity of the product was verified using 2100 Bioanalyzer instrument according to its user's manual (Agilent Technologies, Santa Clara, California, United States). RNA concentration was determined by a NanoDrop 2000 (Thermo Fisher Scientific).

The Illumina TruSeq sample preparation kit v2 (Illumina, San Diego, California, United States), with poly-A selection, was used to prepare RNA samples for sequencing. Fragments were clustered on cBot and sequenced on two lanes on an Illumina HiSeq 2500 with paired ends (2x100bp), according to the manufacturer's instructions.

The short reads were mapped to the CEN.PK 113-7D reference genome (http://cenpk.tudelft.nl) using TopHat version 2.0.10 (Kim et al., 2013). Each sample had between 8.3 to 16.2 million mappable reads, with an average map rate of 88%. Read counts were determined using the featureCounts software from the subread package, version 1.4.0-p1. (Liao et al., 2014). FPKM-values were calculated using Cufflinks version 2.1.1. (Trapnell et al., 2010).

Read counts were used in the differential expression analysis, with the software DESeq. (Anders and Huber, 2010). P-values were adjusted for multiple testing using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995) as implemented in DESeq. All conditions were compared to the reference samples. Raw data from the experiments were deposited in ArrayExpress and assigned the identifier E-MTAB-4044.

Quantitative Proteome Measurements

The peak intensity-based absolute quantification method iBAQ was chosen for protein quantification(Schwanhäusser et al., 2011). First, a lysine auxotrophic strain was created by deleting the LYS1 gene. Latter strain was cultivated with a labelled heavy ^{15}N, ^{13}C-lysine (Cambridge Isotope Laboratories), and fully labelled biomass was produced (in which 98% of proteogenic lysine was labelled, data not shown) and used as an internal standard in the measurements.

Nano-LC/MS/MS Sample Preparation

Cell pellets were suspended in 4% SDS, 100 mM Tris pH 7.5, 20 mM dithiothreitol (DTT), heated at 95°C 5 min and sonicated with Bioruptor (Diagenonde, Denville, NJ, United States) sonication (15 min, ''High'' setting). After protein concentration measurement with tryptophan, florescence samples spiked at a 1:1 ratio with the heavy lysine labelled standard. For absolute quantification, 6 μg of heavy standard was spiked separately with 1.1 μg of UPS2 protein mix (Sigma Aldrich). Overall, 50 μg of protein was precipitated with a 2:1:3 methanol:chloroform:water extraction. The precipitates were mixed in 7:2 M urea:thiourea and 100 mM ammonium bicarbonate. After a disulfide reduction with 2.5 mM DTT and alkylation with 5 mM iodoacetamide, the proteins were digested with 1:50 LysC overnight at room temperature. The peptides were desalted using C18 material (3M) tips and reconstituted in 0.5% TFA.

Nano-LC/MS/MS Analysis

Injected peptides (2 μg) were separated on an Ultimate 3000 RSLCnano system (Dionex, Sunnyvale, California, United States) using a C18 cartridge trap-column in a backflush configuration and an in-house packed (3 μm C18 particles, Dr Maisch) analytical 50 cm x 75 μm emitter-column (New Objective). The peptides were eluted at 200 nL min^{-1} with an 8%–40% B 240 min gradient (buffer B: 80% acetonitrile + 0.1% formic acid, buffer A: 0.1% formic acid) to a Q Exactive (Thermo Fisher Scientific) tandem mass spectrometer operating with a top-10 strategy and a cycle time of 0.9 seconds. Briefly, one 350-1 400 m/z MS scan at a resolution of R = 70,000 was followed by higher-energy collisional dissociation fragmentation (normalized collision energy of 25) of the

10 most-intense ions (charge states +2 to +6) at R = 17,500. The MS and MS/MS ion target values were 3x 10^6 and 5x 10^4, respectively. Dynamic exclusion was limited to 80 seconds.

Protein Turnover Measurements

Protein turnover measurements were carried out using modified stable isotope labelling with amino acids in cell culture (SILAC) method. Two chemostat experiments were carried out under the previously described reference conditions with two exceptions: the *S. cerevisiae* CEN.PK113-7D lysine auxotrophic strain *Δlys1* was used; and the cultivation medium was supplemented with L-lysine (Sigma-Aldrich, St. Louis, Missouri, United States). Media optimization was carried out to detect conditions for double substrate limitation in chemostats (culture was limited by both glucose and lysine). By using pulse experiments and metabolite measurements, it was determined that the combination of 10 g L^{-1} glucose and 0.22 g L^{-1} of Lysine resulted in double substrate limitation under the studied chemostat conditions. When the culture was in a steady state, media containing unlabelled Lysine was replaced with the media containing heavy ^{13}C, ^{15}N-Lysine (Cambridge Laboratories). Sampling from two parallel bioreactors was carried out sequentially in which samples from the first bioreactor were collected at 0.5, 1.5, 2.5, 4, 6, 8, 10 and 12 hours and at 1, 2, 3, 5, 7, 9, 11 and 13 hours after the light medium was sampled and the medium from the second bioreactor was changed. Sampling was carried out to analyse proteins (as described previously) and the intracellular metabolome. For the intracellular metabolome samples, ~2 mL of the culture was immediately quenched in 40 mL of cold methanol (kept at -50°C in an ethanol bath; the exact amount of sample withdrawn was determined gravimetrically expecting culture density of 1mg mL^{-1}) and centrifuged for 5 min at -14°C at 4500 rpm. The cell pellets were snap-frozen in liquid nitrogen, and metabolites were extracted using 1 mL of hot ethanol (70°C) in three consecutive rounds that were pooled together. The extracts were concentrated under vacuum. The metabolome samples were analysed as described previously (Lahtvee et al., 2011) using a UPLC-MS-TOF instrument (ACQUITY UPLC and LCT Premier, Waters, Milford, MA, United States) and an amino acid detection kit according to the manufacturer's instructions (AccQ•Tag™ Ultra, Waters). Protein samples were analysed as described in the proteome analysis section, except an internal standard was not introduced.

QUANTIFICATION AND STATISTICAL ANALYSIS

RNA Quantification

For mRNA quantification, the QuantiGene assay (Affymetrix, Santa Clara, CA, United States) was used to directly measure absolute concentrations for 18 mRNAs in cell lysates. Ten mg of frozen yeast cell pellets collected under different reference conditions were thawed on ice, resuspended in 100 mM Tris pH 7.5 and lysed mechanically using glass beads. After clearing the lysate by centrifugation, the lysates were diluted to approximately 6x 10^4 cells mL^{-1} in 100 mM Tris pH 7.5 supplemented with 1 U μL^{-1} RiboLock RNase inhibitor (Thermo Fisher Scientific, Waltham, MA, United States). The dilutions used in the experiment were in the range of 6x 10^4–3x 10^2 cells mL^{-1} depending on target abundance. Targets were selected from genes spanning FPKM values of 4x 10^2–3x 10^5. Three dilutions of cell lysate were used for the measurement of each target. The probes for the selected targets were synthesised by the manufacturer on request, and a list of the targets is presented in the Supplementary table (Table S1). RNA transcripts (ArrayControl RNA Spikes, AM1780, Ambion, Thermo Fisher Scientific) were used to construct the calibration curve. The dilutions of standards were approximately 2x 10^{-4}–1x 10^{-6} nM, depending on the target abundance in each plate. Targets and standard dilutions for each plate were matched with respect to signal strength. Each standard was present in 5 concentrations, and two standards were used per plate. The quantification was performed as per the manufacturer's manual: 20 μL of cell lysate was incubated with 80 μL of probe/lysis/blocking mix overnight at 55°C. The following steps were performed as written in the product manual. Intensities were quantified with a FLUOstar Omega multi-well plate reader (BMG Labtech, Germany). Linear correlation with the Pearson R^2 value 0.96 was achieved among 18 absolutely quantified mRNAs and their corresponding FPKM values obtained through RNA sequencing. The same correlation was applied to each mRNA value for quantifying the mRNA levels (Tables S1 and S2).

Mass-spectrometric Raw Data Identification and Quantification of Proteome

Raw data were identified and quantified with MaxQuant 1.4.0.8 (Cox and Mann, 2008). Labelling state (multiplicity) was set to 2, and Lys8 was defined as the heavy label. Methionine oxidation and protein N-terminal acetylation were set as variable modifications, whereas cysteine carbamidomethylation was defined as a fixed modification. A search was performed against the UniProt (www.uniprot.org) *Saccharomyces cerevisiae* reference proteome database (version from Sept 2014) using the LysC/P digestion rule. Protein identifications with a minimum of 1 peptide of 7 amino acids long were accepted, and transfer of peptide identifications between runs was enabled. Protein quantification values were derived by using each protein's median peptide H/L ratio and requiring at least one peptide ratio measurement for reporting quantitative values (i.e. min ratio count set to 1). Signal integration (re-quantification) of missing label channels was used, except for the runs where heavy label incorporation was confirmed for the spike-in standard. Peptide-spectrum match and protein false discovery rate (FDR) was kept below 1% using a target-decoy approach. The heavy spike-in standard was quantified using the iBAQ method as described by Scwanhausser et al (Schwanhäusser et al., 2011). Essentially, protein intensities were divided by the number of theoretically observable peptides, log-transformed and plotted against log-transformed UPS2 mix (48 human proteins) known protein. This regression was then used to derive all other protein absolute quantities using their iBAQ intensities. Other parameters were set at their default values. To account for any mixing errors normalized H/L ratios (by shifting median peptide log H/L ratio to zero) were used for all down-stream quantitative analyses (Cox and Mann, 2008).

Curve Fitting for Protein Turnover Calculations

To calculate protein turnover rates based on protein levels measured with the SILAC approach, both, heavy-light ratio in protein dynamics as well as in the intracellular amino acid pool must be taken into account. As in mass-spectrometry based protein analysis we were measuring peptides containing in average one lysine molecule, we are not able to detect whether the protein where the peptide with a light lysine belongs was coming from the "old" or resynthesized protein. This is because the intracellular amino acid pool, although with a constant intracellular concentration, is not having a constant heavy to light ratio but exponentially approaching the fully labelled form. Hence, the rate of change in the fraction of unlabelled lysine can be described with the differential equation described also previously (Hong et al., 2012; Lahtvee et al., 2014):

$$\frac{df_{p,i}}{dt} = f_a(t)\left(D + k_{deg,i}\right) - f_{p,i}(t)k_{deg,i} - f_{p,i}(t)D \qquad \text{(Equation 1)}$$

Parameters f_a and $f_{p,i}$ are unlabelled fraction of lysine in its pool and in protein *i* at time t. $k_{deg,i}$ is the turnover of protein i. The specific growth rate μ of yeast is equal to the dilution rate D at steady state in continuous culture experiments.

The free lysine pool f_a is defined as the amount of lysine available to yeast for the synthesis of proteins. This free lysine pool is common for all proteins in the cell. The fraction of unlabelled lysine in its pool is a function of time and can be defined as:

$$f_a(t) = e^{-at} \qquad \text{(Equation 2)}$$

where coefficient *a* describes the rate at which intracellular unlabelled lysine is replaced by heavy lysine in its free amino acid pool.

The integration of Equation 1 gives a solution for $f_{p,i}$:

$$f_{p,i}(t) = \frac{e^{-\left(D + k_{deg,i}\right)t}}{D + k_{deg,i} - a}\left[e^{\left(D + k_{deg,i} - a\right)t}\left(D + k_{deg,i}\right) - a\right] \qquad \text{(Equation 3)}$$

Equation 3 contains two unknown parameters *a* and $k_{deg,i}$. Parameter *a* is common for all the proteins and is calculated based on fitting time-dependent intracellular lysine labelling pattern to the Equation 2. Thereafter, *a* is used for calculating individual protein turnover rates ($k_{deg,i}$). Protein turnover data can be found from Table S5.

For every fitted line 95% confidence intervals were calculated (2.5%–97.5% confidence interval) and lower bounds (LB) and upper bounds (UB) were determined with a median value of 36% (Table S5 and Figure S3).

Noise Adjusted Correlation

According to Csárdi et al., 2015, we calculated noise adjusted transcript-protein correlation coefficient for the reference condition

$$\left(\widehat{r}_{true} = \frac{\sqrt[9]{r_{X_1Y_1} * r_{X_1Y_2} * r_{X_1Y_3} * r_{X_2Y_1} * r_{X_2Y_2} * r_{X_2Y_3} * r_{X_3Y_1} * r_{X_3Y_2} * r_{X_3Y_3}}}{\sqrt[6]{r_{X_1X_2} * r_{X_1X_2} * r_{X_2X_3} * r_{Y_1Y_2} * r_{Y_1Y_3} * r_{Y_2Y_3}}}\right) \qquad \text{(Equation 4)}$$

where X_j represents the transcriptome and Y_j proteome data, respectively, and *j=1,2,3*. Input data for the analysis is provided in Table S11.

Flux Balance Analysis (FBA) and Random Sampling

Consensus yeast metabolic model Yeast v7.6 (Aung et al., 2013) was used for all the studied conditions. Condition-dependent biomass composition was introduced into the model and model calculations were constrained by the biomass composition and exchange fluxes (input data can be found in the Table S12). MATLAB 2011B (MathWorks Inc, Natick, MA, United States) with GLPK solver on COBRA Toolbox (Schellenberger et al., 2011) were applied for the calculations. First, ATP drain, representing additionally required maintenance energy, was maximized to calculate the unique pattern of intracellular fluxes. To determine the variability of fluxes, random sampling algorithm in RAVEN was used with 1000 samplings at the 90% previously determined ATP drain value (Bordel et al., 2010). This resulted in an average flux with the standard deviation, representing the flux variability.

Estimation of ATP Expenditure for Protein Turnover

For the estimation of ATP expenditure on protein turnover, we took into account the total ATP spent for the biomass formation according to the Yeast v7.6 genome scale model. As already demonstrated previously, ATP requirement (growth and non-growth related ATP net production) under the reference environmental conditions was 59 mmol-ATP gDW^{-1} h^{-1} (Forster et al., 2003). From the other hand, we calculated ATP expenditure for the biosynthesis of monomers and protein polymerization – the highest polymerization cost in the cell. ATP requirement for monomers was taken from the genome scale FBA analysis, which resulted in 7.5 mmol-ATP gDW^{-1} h^{-1}. ATP requirement for protein polymerization was calculated based on the total number of proteins, average length of a protein and an estimation that 4.306 molecules of ATP are spent for the synthesis of one peptide bond (Stouthamer, 1973; Stephanopoulos et al., 1998), which results in 23 mmol-ATP gDW^{-1} h^{-1}. The remainder of the synthesized ATP could be expected to be used for maintenance. To calculate the ATP requirement for protein turnover, again, 4.306 molecules of ATP were considered for a formation of a peptide bond and 1 molecule of ATP for the degradation of one peptide bond (Benaroudj et al., 2003). Taking into account the measured average protein turnover, it was estimated that 12 mmol-ATP gDW^{-1} h^{-1} was used during protein turnover.

Enrichment Analysis

Enrichment analysis was carried out using g:Profile online software (Reimand et al., 2011).

Translation Efficiency

Translation efficiency (k_{TL}) was calculated accordingly:

$$k_{TL} = \frac{C_{prot,i}\left(k_{deg,i} + \mu\right)}{C_{mRNA,i}} \quad \text{(Equation 11)}$$

Where $C_{prot,i}$ and $C_{mRNA,i}$ refer to the measured absolute protein and mRNA abundance.

Multi Array Regression Spline Analysis

Multi array regression spline (MARS) analysis was run using the earth package on R platform (Milborrow, 2011). The input data for the model comprised from the "Total synthesized protein abundances" as a response variable and 259 features as explanatory variables. Explanatory variables included: (i) mRNA abundances; (ii) Secondary structure information (frequency of nucleotide singlets, duplexes and triplexes, normalized to their total length) in the sequencing region, 5' UTR and 3' UTR; (iii) Length of a sequencing region, 5' UTR and 3' UTR; (iv) Codon and tRNA adaptation indexes of genes; (v) protein turnover; and (vi) information about co-translational RNA decay (as introduced in Pelechano et al., 2015). Genome sequence information was taken from SGD, as of 02.11.2015. Genome coordinates for 5'- and 3'UTRs were adapted from (Yassour et al., 2009). Codon adaptation indexes and tRNA adaptation indexes were adapted from Sharp and Cowe, 1991 and Tuller et al., 2010, respectively. R-script from SeqinR-package was used to calculate CAI and tAI values for individual proteins in yeast (Charif and Lobry, 2007).

Individual contribution was calculated for every feature, ranked according to their individual contribution and followed by the calculation of the combined contribution. Twenty-seven features were selected by the algorithm to contribute to the final protein abundance (impacting the final amount more than 0.1%). It is important to note that combined contribution is smaller than the sum of individual contributions as there will be some overlap in the contributions. Contributions of the features were calculated by adding parameters one by one according to their rank (Table S9).

Model prediction and generality was estimated by applying the model to the perturbed environmental conditions studied. Although we were missing protein turnover under those conditions to calculate total amount of synthesized proteins, the average predictive power to determine the protein abundance in the cell was 63 ± 13% (average predictive power of 9 conditions ± standard deviation).

Positional Analysis

Genes were sorted based on the ratio of protein abundance to mRNA abundance and the top 10% as higly translated and lowest 10% as low translated were selected. The sequence for 15 nts before and 15 nts after the start codon (ATG) was extracted using YeastMine (Saccharomyces Genome Database) and the most frequently observed codons were calculated. Finally, we used the WebLogo service to draw the sequence logos (Crooks et al., 2004).

DATA AND SOFTWARE AVAILABILITY

Transcriptomics Data

The accession number for the RNA-seq data reported in this paper is ArrayExpress: E-MTAB-4044 (Lahtvee et al., 2016). Processed quantitative data are in Table S2.

Proteomics Data

The LC-MS/MS proteomics data have been deposited in the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaíno et al., 2013) and can be retrieved using the dataset identifier PRIDE: PXD005041. Processed quantitative data are in Table S3.

Cell Systems, Volume *4*

Supplemental Information

Absolute Quantification of Protein and mRNA Abundances Demonstrate Variability in Gene-Specific Translation Efficiency in Yeast

Petri-Jaan Lahtvee, Benjamín J. Sánchez, Agata Smialowska, Sergo Kasvandik, Ibrahim E. Elsemman, Francesco Gatto, and Jens Nielsen

Supplementary Figures and Tables

Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast

Petri-Jaan Lahtvee[1,2], Benjamín J. Sánchez[1,2], Agata Smialowska[1,3], Sergo Kasvandik[4], Ibrahim E. Elsemman[5], Francesco Gatto[1,6], Jens Nielsen[1,2,5,7]

[1]Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96, Gothenburg, Sweden

[2]Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, 412 96, Gothenburg, Sweden

[3]National Bioinformatics Infrastructure Sweden (NBIS), Science for Life Laboratory, 17165, Solna, Sweden.

[4]Institute of Technology, University of Tartu, 50411, Tartu, Estonia

[5]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2970, Hørsholm, Denmark

[6]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive La Jolla, CA 92093-0412, United States

[7]Lead Contact author

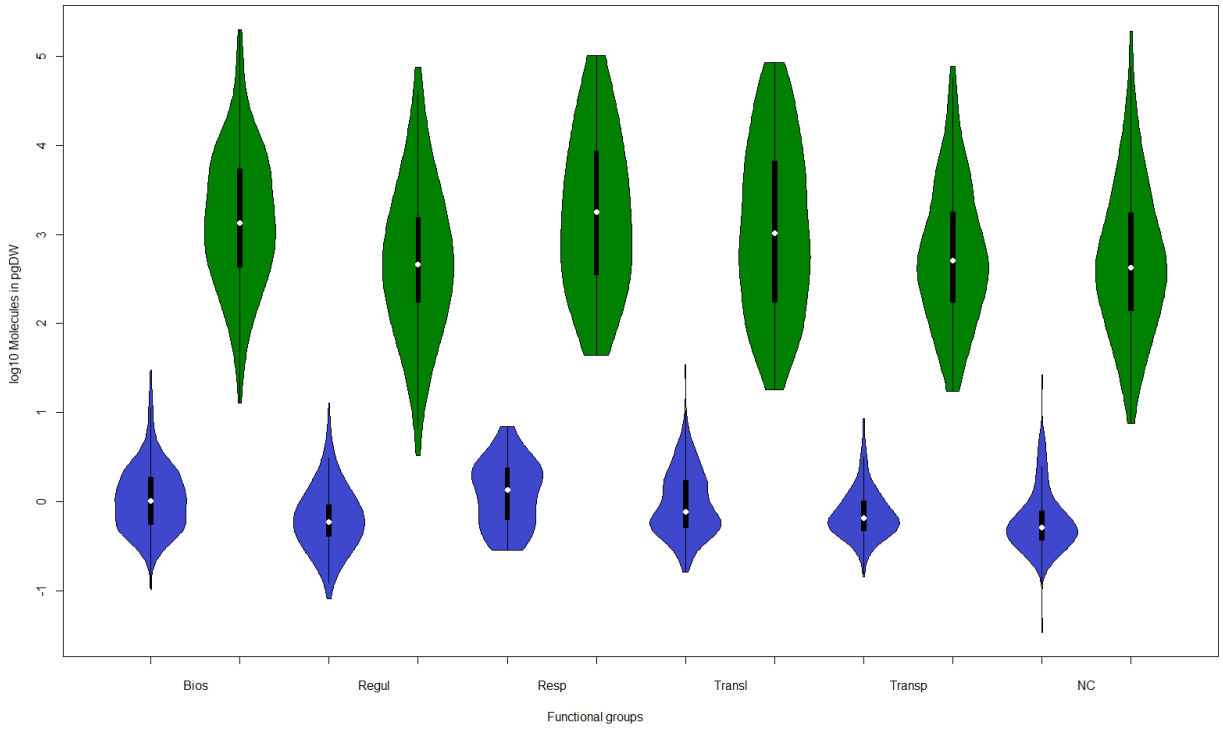


Figure S1. Distribution of mRNA (blue) and protein (green) abundances divided into functional groups. Bios – biosynthesis; Regul – regulation; Resp – respiration; Transl – translation; Transp – transporters; NC – not categorized.

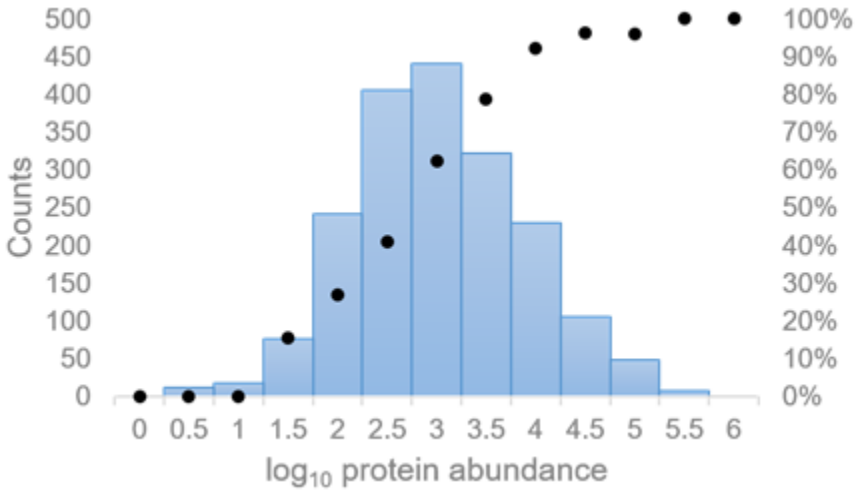


Figure S2. Histogram of a protein abundances (blue bars) together with the percentage of proteins which protein turnover was determined for each abundance bar (black dots).

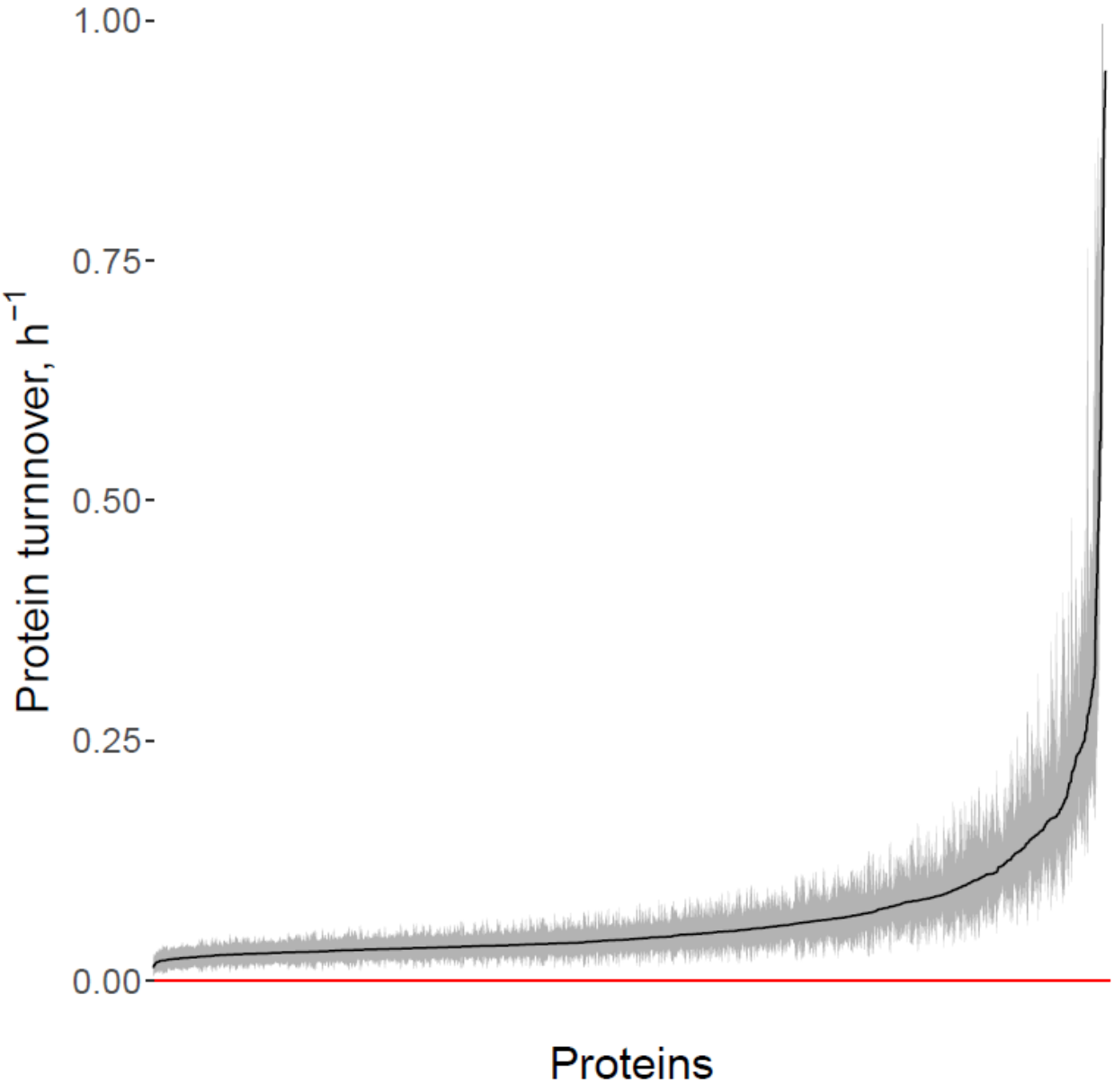


Figure S3. Protein turnover rate estimates (k_{deg}; black dots) sorted increasingly for all measured proteins for which the regression was statistically significant (FDR<0.01). For each estimate, the 95% confidence interval is shown as a grey line.

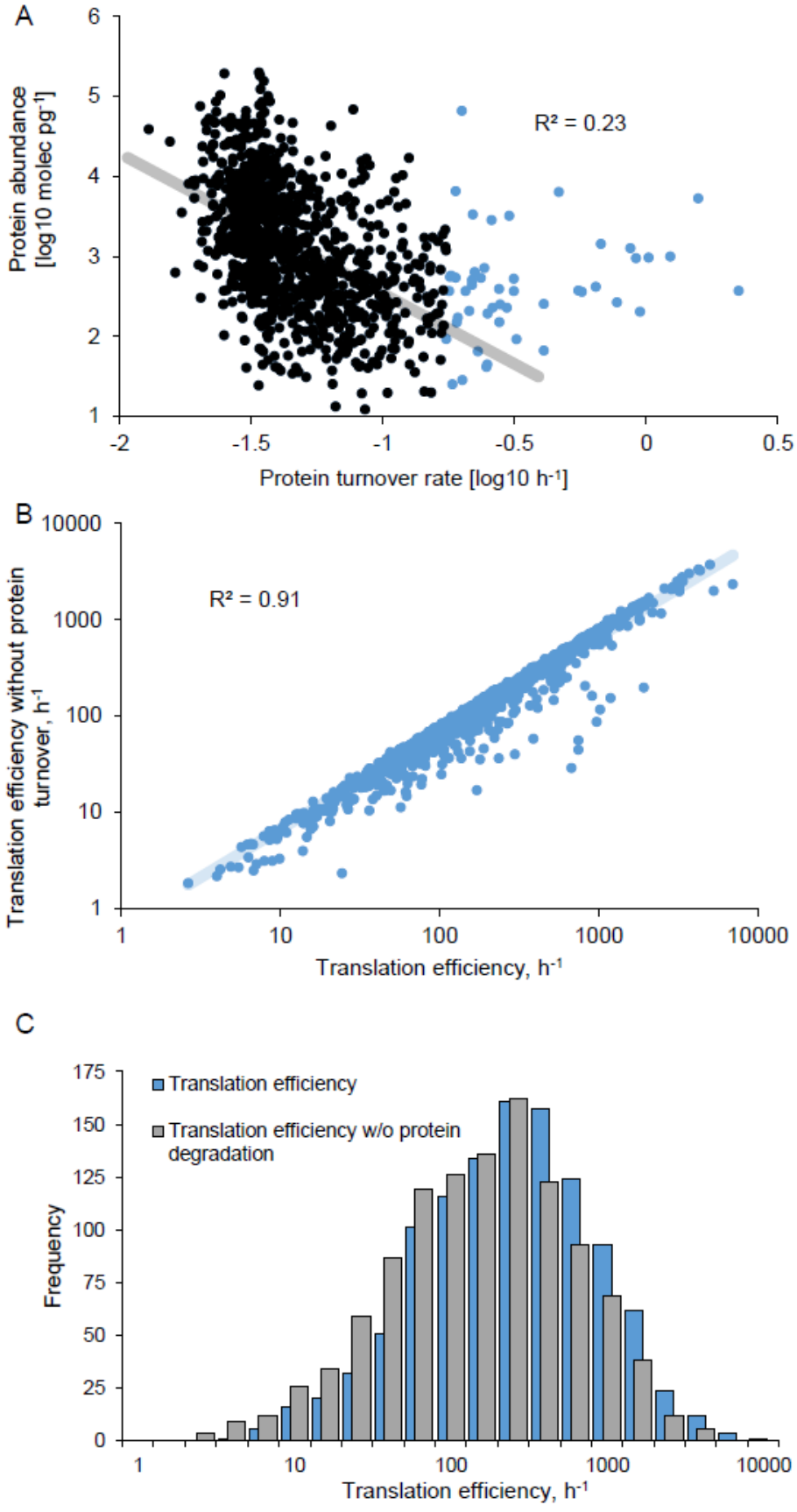

A
Protein abundance [log10 molec pg-1]
R² = 0.23
Protein turnover rate [log10 h-1]
B
Translation efficiency without protein turnover, h-1
R² = 0.91
Translation efficiency, h-1
C
Translation efficiency
Translation efficiency w/o protein degradation
Frequency
Translation efficiency, h-1

Figure S4. The impact of protein turnover for calculations of translation efficiency. (A) Moderate correlation between protein turnover and protein abundance was observed. Black circles exclude 5% the highest protein turnover rates. (B) Correlation between translation efficiency and efficiency calculated without protein turnover taken into account. Less than 30% difference was detected for the majority of individual proteins. (C) Histogram of translation efficiencies when protein turnover is or is not taken into account (represented by blue and grey bars, respectively).

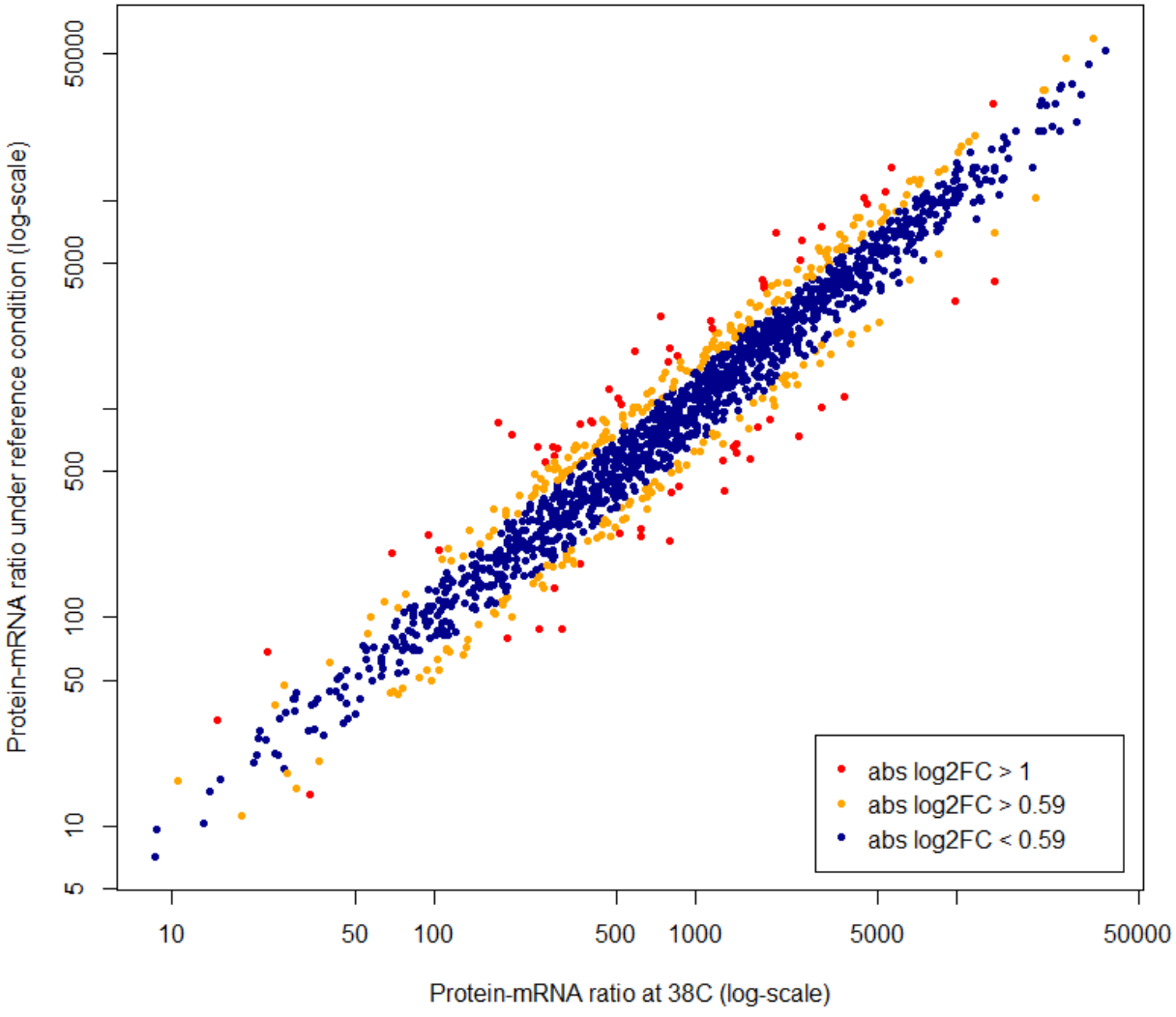


Figure S5. Correlation of protein-mRNA ratios between the reference and the most extreme condition studied (38°C temperature; both studied in chemostats at D = 0.1 h^{-1}). Yellow and red circles indicate protein-mRNA ratio differences more than 1.5- and 2-fold between the conditions, respectively.

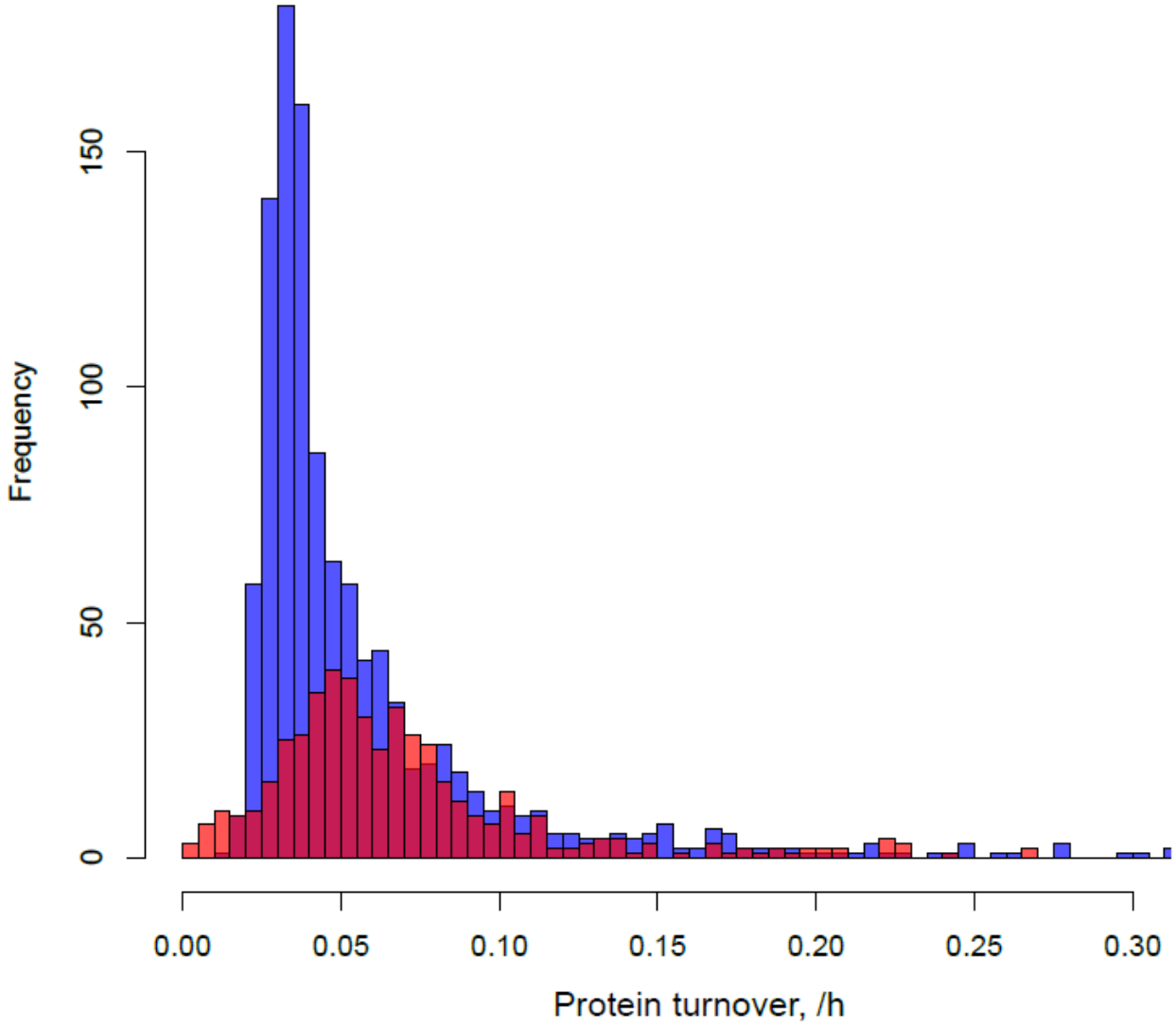


Figure S6. Histogram comparison of the protein turnover data reported in the current study (blue bars) and Helbig et al., 2011 (red bars). Both studies are run under chemostat conditions at D = 0.1 h^{-1}, however, representing aerobic (current study) *versus* anaerobic conditions (Helbig et al. 2011).

Table S1. mRNAs selected for the absolute quantification. Log10 FPKM values from RNAseq and log10 measured concentrations are presented.

Gene ID	Gene symbol	Description	log10 FPKM value	log10 Concentration, fmol/mg
YMR169C	ALD3	Cytoplasmic aldehyde dehydrogenase, involved in beta-alanine synthesis; uses NAD+ as the preferred coenzyme; very similar to Ald2p; expression is induced by stress and repressed by glucose [Source:SGD;Acc:S000004779]	2.635293	-0.78168
YKL166C	TPK3	cAMP-dependent protein kinase catalytic subunit; promotes vegetative growth in response to nutrients via the Ras-cAMP signaling pathway; partially redundant with Tpk1p and Tpk2p; localizes to P-bodies during stationary phase [Source:SGD;Acc:S000001649]	2.995085	-0.35913
YGL248W	PDE1	Low-affinity cyclic AMP phosphodiesterase, controls glucose and intracellular acidification-induced cAMP signaling, target of the cAMP-protein kinase A (PKA) pathway; glucose induces transcription and inhibits translation [Source:SGD;Acc:S000003217]	3.291336	-0.21625
YJL196C	ELO1	Elongase I, medium-chain acyl elongase, catalyzes carboxy-terminal elongation of unsaturated C12-C16 fatty acyl-CoAs to C16-C18 fatty acids [Source:SGD;Acc:S000003732]	3.306934	-0.30395
YMR037C	MSN2	Transcriptional activator related to Msn4p; activated in stress conditions, which results in translocation from the cytoplasm to the nucleus; binds DNA at stress response elements of responsive genes, inducing gene expression [Source:SGD;Acc:S000004640]	3.417923	-0.01738
YPL203W	TPK2	cAMP-dependent protein kinase catalytic subunit; promotes vegetative growth in response to nutrients via the Ras-cAMP signaling pathway; partially redundant with Tpk1p and Tpk3p; localizes to P-bodies during stationary phase [Source:SGD;Acc:S000006124]	3.654339	-0.15492

YCR005C	CIT2	Citrate synthase, catalyzes the condensation of acetyl coenzyme A and oxaloacetate to form citrate, peroxisomal isozyme involved in glyoxylate cycle; expression is controlled by Rtg1p and Rtg2p transcription factors [Source:SGD;Acc:S000000598]	3.728173	-0.00824
YGL205W	POX1	Fatty-acyl coenzyme A oxidase, involved in the fatty acid beta-oxidation pathway; localized to the peroxisomal matrix [Source:SGD;Acc:S000003173]	4.281594	0.252723
YHR183W	GND1	6-phosphogluconate dehydrogenase (decarboxylating), catalyzes an NADPH regenerating reaction in the pentose phosphate pathway; required for growth on D-glucono-delta-lactone and adaptation to oxidative stress [Source:SGD;Acc:S000001226]	4.359378	0.708496
YKL148C	SDH1	Flavoprotein subunit of succinate dehydrogenase (Sdh1p, Sdh2p, Sdh3p, Sdh4p), which couples the oxidation of succinate to the transfer of electrons to ubiquinone as part of the TCA cycle and the mitochondrial respiratory chain [Source:SGD;Acc:S000001631]	4.373571	0.588765
YNR016C	ACC1	Acetyl-CoA carboxylase, biotin containing enzyme that catalyzes the carboxylation of acetyl-CoA to form malonyl-CoA; required for de novo biosynthesis of long-chain fatty acids [Source:SGD;Acc:S000005299]	4.47165	0.612974
YFR053C	HXK1	Hexokinase isoenzyme 1, a cytosolic protein that catalyzes phosphorylation of glucose during glucose metabolism; expression is highest during growth on non-glucose carbon sources; glucose-induced repression involves the hexokinase Hxk2p [Source:SGD;Acc:S000001949]	4.680155	0.956133
YGL055W	OLE1	Delta(9) fatty acid desaturase, required for monounsaturated fatty acid synthesis and for normal distribution of mitochondria [Source:SGD;Acc:S000003023]	4.816657	0.597592

YHR007C	ERG11	Lanosterol 14-alpha-demethylase; catalyzes the C-14 demethylation of lanosterol to form 4,4''-dimethyl cholesta-8,14,24-triene-3-beta-ol in the ergosterol biosynthesis pathway; member of the cytochrome P450 family; associated and coordinately regula /.../th the P450 reductase Ncp1p [Source:SGD;Acc:S000001049]	5.014566	0.807304
YLR044C	PDC1	Major of three pyruvate decarboxylase isozymes, key enzyme in alcoholic fermentation, decarboxylates pyruvate to acetaldehyde; subject to glucose-, ethanol-, and autoregulation; involved in amino acid catabolism [Source:SGD;Acc:S000004034]	5.031084	1.160538
YKL060C	FBA1	Fructose 1,6-bisphosphate aldolase, required for glycolysis and gluconeogenesis; catalyzes conversion of fructose 1,6 bisphosphate to glyceraldehyde-3-P and dihydroxyacetone-P; locates to mitochondrial outer surface upon oxidative stress [Source:SGD;Acc:S000001543]	5.293373	1.078671
YBR072W	HSP26	Small heat shock protein (sHSP) with chaperone activity; forms hollow, sphere-shaped oligomers that suppress unfolded proteins aggregation; oligomer activation requires heat-induced conformational change; also has mRNA binding activity [Source:SGD;Acc:S000000276]	5.4545	1.715457
YGR192C	TDH3	Glyceraldehyde-3-phosphate dehydrogenase, isozyme 3, involved in glycolysis and gluconeogenesis; tetramer that catalyzes the reaction of glyceraldehyde-3-phosphate to 1,3 bis-phosphoglycerate; detected in the cytoplasm and cell wall [Source:SGD;Acc:S000003424]	5.525863	1.766594

Table S8. List of transcriptionally regulated fluxes.

Reaction	Corresponding gene(s)
r_0001	DLD1
r_0002	COR1/QCR2,6,7,8,9/RIP1/CYC1,7/OSH7/CYT1
r_0004	COR1/QCR2,6,7,8,9/RIP1/CYC1,7/OSH7/CYT1
r_0042	ARO3,4
r_0061	LEU2
r_0111	ACH1
r_0112	ACS1,2
r_0114	ACS1,2
r_0148	ADK1
r_0167	ADH3,4
r_0175	ALD4,5
r_0178	ALD4,5
r_0203	TRP2/TRP3
r_0214	URA2
r_0226	ATP1-8,14-18,20/OLI1/TIM11
r_0252	YAT1,2
r_0437	COR1/QCR2,6,7,8,9/RIP1/CYC1,7/OSH7/CYT1
r_0438	COX1,2,3,4,5A,5B,6,7,8,9,12,13/CYC1,7
r_0439	COR1/QCR2,6,7,8,9/RIP1/CYC1,7/OSH7/CYT1
r_0451	FUM1
r_0452	FUM1
r_0473	PRO2
r_0505	GCV1,2,3/KGD1,2/LPD1
r_0533	GLK1/HXK1,2
r_0534	GLK1/HXK1,2
r_0535	GLK1/HXK1,2
r_0566	TRP2/TRP3
r_0659	IDP2
r_0662	ICL1
r_0716	DAL7/MLS1
r_0717	DAL7/MLS1
r_0757	INM1/INM2
r_0773	NDI1
r_0831	GCV1,2,3/KGD1,2/LPD1
r_0832	GCV1,2,3/KGD1,2/LPD1
r_0962	CDC19/PYK2
r_1055	TRP5
r_2157	ELO2,3

r_2177	TSC13
r_2330	AYR1
r_2529	EPT1/CPT1
r_2542	GEP4
r_2543	GEP4
r_2544	GEP4
r_2545	GEP4
r_2546	GEP4
r_2547	GEP4
r_2823	LSB6/STT4
r_3130	SAC1
r_3151	INP52,53
r_3217	VAC14/FIG4

Table S9. Multivariate regression spline (MARS) analysis to determine the control over total amount of proteins synthesized at optimal environmental conditions in chemostat (D = 0.1 /h). One hundred fifty-two variables were used. Variables included: (i) mRNA abundances; (ii) Secondary structure information (frequency of nucleotide singlets, duplexes and triplexes, normalized to their total length) in sequencing region, 5' UTR and 3' UTR; (iii) Length of a sequencing region, 5' UTR and 3' UTR; (iv) Codon and tRNA adaptation indexes of genes; and (v) Information about co-translational RNA decay.

#	Variable	Level of (additional) control	Accumulated control	Group
1	mRNA	61.280%	61.3%	mRNA abundance
2	CAI	5.709%	67.0%	CAI/tAI (translation elongation)
3	tAI	2.340%	69.3%	CAI/tAI (translation elongation)
4	Gene - TT	1.409%	70.7%	Gene (translation elongation)
5	Gene - GT	1.213%	72.0%	Gene (translation elongation)
6	Gene - A	0.869%	72.8%	Gene (translation elongation)
7	5' UTR - CC	0.733%	73.6%	5' UTR (translation initiation)
8	5' UTR - A	0.723%	74.3%	5' UTR (translation initiation)
9	Gene - AT	0.691%	75.0%	Gene (translation elongation)
10	Gene - GC	0.552%	75.5%	Gene (translation elongation)
11	Gene - GG	0.468%	76.0%	Gene (translation elongation)
12	5' UTR - TA	0.465%	76.5%	5' UTR (translation initiation)
13	Length - Gene	0.378%	76.8%	Gene (translation elongation)

14	Gene - C	0.373%	77.2%	Gene (translation elongation)
15	Gene - CT	0.365%	77.6%	Gene (translation elongation)
16	Gene -T	0.360%	77.9%	Gene (translation elongation)
17	5' UTR - CC	0.326%	78.3%	5' UTR (translation initiation)
18	5' UTR - TG	0.314%	78.6%	5' UTR (translation initiation)
19	5' UTR - GC	0.295%	78.9%	5' UTR (translation initiation)
20	5' UTR - AC	0.293%	79.2%	5' UTR (translation initiation)
21	Length - 3' UTR	0.293%	79.5%	3' UTR
22	Gene - CG	0.287%	79.7%	Gene (translation elongation)
23	5' UTR - GG	0.275%	80.0%	5' UTR (translation initiation)
24	5' UTR -T	0.266%	80.3%	5' UTR (translation initiation)
25	5' UTR - AA	0.235%	80.5%	5' UTR (translation initiation)
26	Gene -TC	0.207%	80.7%	Gene (translation elongation)
27	Gene - TA	0.194%	80.9%	Gene (translation elongation)

Table S10. Codon frequencies as the first position after the start codon among 200 the highest protein-mRNA ratios quantified for the yeast Saccharomyces cerevisiae

Unique Codon	**Codon count**
TCT	33
GCT	20
TCC	18
CCA	10
ACT	10
TTG	9
GCC	7
TTT	7
AGC	7
TCA	7
GTT	5
CCT	5
TTA	5

GCA	4
CTA	4
AAG	4
TTC	4
GTC	4
TCG	4
GAG	3
CTT	3
GGT	3
ACC	3
AGT	3
GTA	2
AAA	2
AGA	2
AAT	2
ACA	2
GGA	1
GAC	1
ATT	1
GAT	1
CGT	1
CTG	1
TGT	1
ATC	1