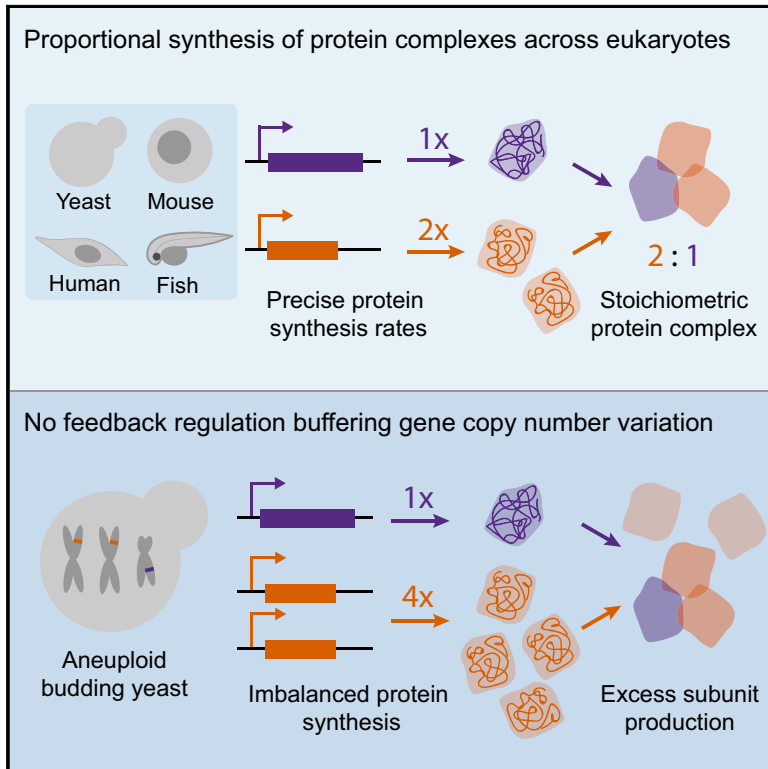


# Cell Systems

## Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes

### Graphical Abstract



### Authors

James C. Taggart, Gene-Wei Li

### Correspondence

gwli@mit.edu

### In Brief

Careful reanalysis of ribosome profiling data revealed proportional synthesis for the vast majority of protein complexes in budding yeast and large complexes in higher eukaryotes. Furthermore, systematic perturbation to chromosome copy number demonstrated that precise rates of protein synthesis are hard coded in the genome rather than actively monitored and maintained through feedback.

### Highlights

- Improved estimate of protein synthesis using authentic ribosome footprint density
- Nearly all protein complexes in yeast are made in proportion to their stoichiometry
- Large protein complexes in higher eukaryotes follow proportional synthesis
- Proportional synthesis is not maintained by feedback regulation for most complexes



# Production of Protein-Complex Components Is Stoichiometric and Lacks General Feedback Regulation in Eukaryotes

James C. Taggart<sup>1</sup> and Gene-Wei Li<sup>1,2,\*</sup>

<sup>1</sup>Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

<sup>2</sup>Lead Contact

\*Correspondence: [gwli@mit.edu](mailto:gwli@mit.edu)

<https://doi.org/10.1016/j.cels.2018.11.003>

## SUMMARY

Constituents of multiprotein complexes are required at well-defined levels relative to each other. However, it remains unknown whether eukaryotic cells typically produce precise amounts of subunits, or instead rely on degradation to mitigate imprecise production. Here, we quantified the production rates of multiprotein complexes in unicellular and multicellular eukaryotes using ribosome profiling. By resolving read-mapping ambiguities, which occur for a large fraction of ribosome footprints and distort quantitation accuracy in eukaryotes, we found that obligate components of multiprotein complexes are produced in proportion to their stoichiometry, indicating that their abundances are already precisely tuned at the synthesis level. By systematically interrogating the impact of gene dosage variations in budding yeast, we found a general lack of negative feedback regulation protecting the normally precise rates of subunit synthesis. These results reveal a core principle of proteome homeostasis and highlight the evolution toward quantitative control at every step in the central dogma.

## INTRODUCTION

How precisely cells tune gene expression is a fundamental question in quantitative biology. A sufficient amount of each protein must be produced to satisfy the demand for its activities in the cell, and excess production can be wasteful. But to synthesize proteins at levels that are “just enough” requires precisely tuned rates of transcription, translation, and mRNA decay, as well as selective pressure against overproduction. Elucidating the general principles underlying the set-points of protein synthesis rates is an important step toward understanding the design constraints of cells and the physiological impacts of gene expression perturbations.

Obligate subunits of multiprotein complexes are an abundant class of proteins whose minimally required levels are precisely defined. At steady state, the synthesis rate of each subunit must exceed or equal the minimum of its binding partners, but

the benefit quickly diminishes when the subunit is made in excess. In bacteria, a systematic study of the rates of protein synthesis using ribosome profiling showed that nearly all obligate subunits of protein complexes are produced at rates proportional to their stoichiometry (Li et al., 2014). This principle, termed proportional synthesis, indicates that bacteria have evolved the ability to synthesize precise amounts of proteins and that degradation of unassembled subunits is not the primary process controlling protein abundance.

In contrast to bacteria, it remains unclear whether precise proportional synthesis is a general strategy of gene expression in eukaryotes. Several lines of evidence suggest that the evolutionary constraints on protein production may be different. First, eukaryotic cells have a much more elaborate capacity for, and regulation of, protein degradation, which could efficiently buffer against imprecise synthesis (Juzskiewicz and Hegde, 2018). Second, whereas bacteria often face limited nutrients and ribosome capacity, it has been suggested that eukaryotes have much lower selective pressure against wasteful production, especially in multicellular organisms (Kafri et al., 2016; Lynch and Marinov, 2015). Third, many bacterial protein complexes are regulated by negative feedback that maintains expression levels (Freedman et al., 1987; Mattheakis and Nomura, 1988; Shen-Orr et al., 2002), but such intricate regulation is less well documented in eukaryotes, even for the abundant ribosomal proteins (Warner and McIntosh, 2009). In general, negative feedback regulation is uncommon in yeast, although it remains to be determined whether subunits of complexes are exceptions to this rule (Ishikawa et al., 2017; Springer et al., 2010). Finally, isolated studies in eukaryotes have reported both complexes that follow proportional synthesis and those that do not (Cleveland et al., 1981; Lehnert and Lodish, 1988; Lodish, 1974). A systematic analysis of obligate protein complexes and their synthesis rates is currently lacking.

An accurate and global assessment of proportional synthesis in eukaryotes remains difficult for several reasons. In principle, ribosome profiling can provide a genome-wide estimate of protein synthesis rates under steady-state growth, when most ribosomes complete full-length proteins and have similar elongation rates on average (Li et al., 2014). However, quantitation of protein production by ribosome profiling requires faithful assignment of short cDNA sequences derived from ribosome-protected mRNA fragments to coding regions, which is challenging to realize for eukaryotic genomes with many paralogous genes and pseudogenes that share regions of identical sequence.



Commonly used analysis pipelines for ribosome profiling data are not equipped to correctly handle ambiguous ribosome footprints generated from these regions. Additionally, proportional synthesis is expected only for stable and obligate complexes with well-defined stoichiometry, but a comprehensive curation of such complexes does not yet exist. Finally, quantitative comparison of expression among subunits should be made in cells with regular or well-controlled gene dosage and ploidy, which are often not satisfied in mammalian cell lines.

In this work, we refined the quantitation of protein synthesis rates by correctly calculating ribosome footprint density for genes that share regions of identical sequence with others. By manually curating a comprehensive list of stable and obligate protein complexes with well-defined stoichiometry in budding yeast, we found that the majority of complex subunits are synthesized at rates proportional to their stoichiometry. Proportional synthesis was also observed for large protein complexes in higher eukaryotic cells derived from primary tissues or whole animals. To investigate whether this precise synthesis is ensured by negative feedback loops, we systematically examined the response to altered gene dosage in yeast strains harboring an extra copy of single chromosomes. A 2-fold increase in synthesis rate was observed for the majority of complex subunits encoded by duplicated genes, including most ribosomal proteins. Together, these results show that eukaryotes evolved precise rates of protein synthesis, which are hard-coded in the genome without widespread feedback regulation.

## RESULTS

### Accurate Quantification of Ribosome Footprint Density for Genes with Regions of Identical Sequence

A prerequisite for estimating rates of protein synthesis from ribosome profiling data is accurate accounting of the number of ribosome footprints per unit length of a gene, herein referred to as ribosome density. Given the small ribosome footprint size ( $\approx 30$  nucleotides), a sequenced fragment may be mapped to multiple locations in the transcriptome, especially for organisms with many gene duplication events. A universal strategy to handle multimapping reads has not yet emerged. Commonly, multimapping reads in ribosome profiling are either discarded (Anders et al., 2015; Baudin-Baillieu et al., 2014; Guo et al., 2010; Xiao et al., 2016) or arbitrarily (randomly or multiply) assigned to their best alignments (Gardin et al., 2014; Ingolia et al., 2009; Trapnell et al., 2009). Neither of these read assignment strategies accurately represent the density of ribosome footprints for genes that share identical sequences (STAR Methods; Figure S1). In the budding yeast *S. cerevisiae* in standard YPED media, 16% of ribosome footprints can be multimapped and therefore incorrectly assigned. To remove any ambiguity, we used a strategy that considers only ribosome footprints in a computationally reduced transcriptome space without identical sequences. Any regions of  $\geq 27$  nt that have identical sequences elsewhere, i.e., regions that can generate multimapping reads, are masked when calculating ribosome density (reads per unit length) by excluding both their read count and length (Figure S1). Importantly, we keep track of the coordinate along open reading frames for the unmasked region, so that a position-dependent correction factor can be applied to account for

differences in ribosome elongation rates or drop-off (STAR Methods; Figures 1A and S2).

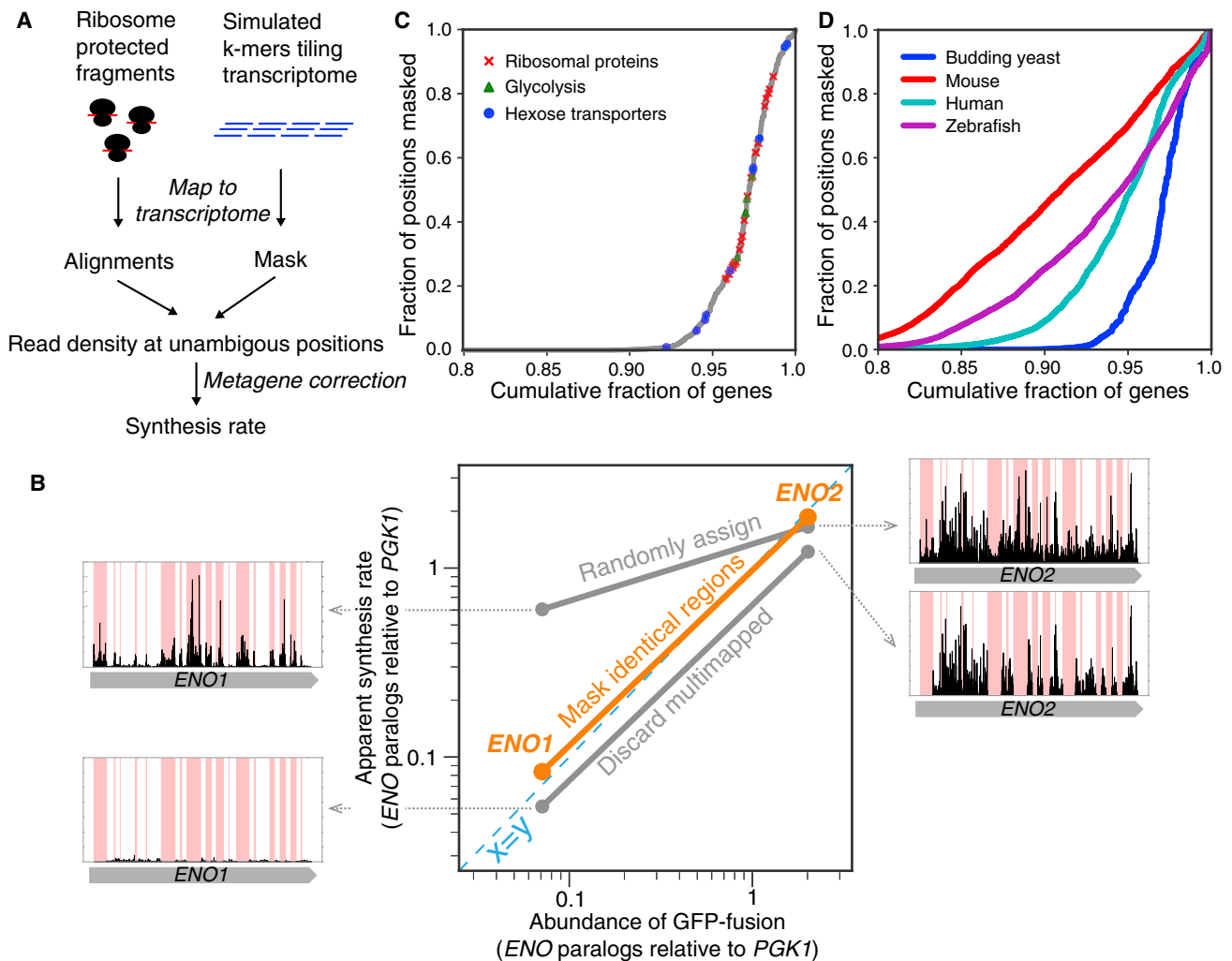
The result of this masking strategy is illustrated for the two paralogous genes for enolase (*ENO1* and *ENO2*) in *S. cerevisiae*, which have substantial portions of their transcripts lacking uniquely mapped reads (Figure 1B). Whereas discarding multimapping reads underestimates footprint density for both genes, and random assignment overestimates footprint density for the lowly expressed *ENO1*, the masking strategy ensures that density is calculated only for regions without ambiguity. The resultant synthesis rates of Eno1 and Eno2, relative to their upstream glycolytic enzyme Pgc1 (3-phosphoglycerate kinase), agree with their respective protein abundances (Figure 1B) (Newman et al., 2006), consistent with the evidence that these proteins share the same half-lives by dilution and therefore have abundances proportional to their synthesis rates (Christiano et al., 2014). This result is also consistent with a recent probabilistic RNA sequencing (RNA-seq) quantification strategy designed for distinguishing overlapping transcripts (Figure S1B) (Bray et al., 2016).

Overall, a substantial fraction of eukaryotic genes requires masking for accurate quantitation. In budding yeast, 5.4% of annotated open reading frames have over 10% of their positions masked (Figure 1C). This group contains many highly expressed proteins, such as glycolytic enzymes, ribosomal proteins, and hexose transporters (Figure 1C). Faithful counting of ribosome footprint density for these proteins is crucial when comparing synthesis rates among components of complexes, as described below (Figures S1C–S1E). In higher eukaryotes such as zebrafish, mice, and humans, an even greater fraction of transcripts has regions that can be erroneously mapped (14%, 17%, and 10% of annotated open reading frames, respectively; Figure 1D). A proper strategy for handling multimapping reads is therefore essential for estimating protein synthesis rates in higher eukaryotes.

### Proportional Synthesis of Complex Subunits in Budding Yeast

With a rational footprint counting strategy, we systematically examined the precision of protein synthesis rates for multi-member complexes in the budding yeast *S. cerevisiae*. We focused on proteins that are stably associated and obligatory to complexes, without other known functions outside the complex. The budding yeast is an ideal eukaryotic model for this global analysis because of its well-characterized protein biochemical properties, its relatively simple transcriptome without extensive alternative mRNA isoforms, and its well-annotated genome.

By manual curation, we first generated a stringent set of 577 proteins in 113 multiprotein complexes that have well-defined and obligate stoichiometry (Table S2). Starting with an existing database of complexes (Meldal et al., 2015), we applied three additional criteria for inclusion. First, each member of a complex must be stably associated as indicated by stoichiometric co-purification, without evidence for regulated association which would cause a large portion of that subunit to be unbound. Second, subunits must lack evidence for promiscuous interactions or association with another complex, which might cause a super-stoichiometric demand on its abundance. Finally, the



**Figure 1. Simple Masking Strategy Ensures Correct Estimate of Protein Synthesis Rates**

(A) Computational workflow for the masking strategy.

(B) Comparison of enolase paralog abundances (relative to the upstream phosphoglycerate kinase, Pgc1) with apparent synthesis rates estimated by three read-mapping strategies. Misassigned ribosome density for enolase paralogs was shown for each of the commonly used read-mapping strategies. Red indicates shared sequences between the paralogs.

(C) Yeast genes ranked by the fraction of their sequences that are not unique. Genes whose products are ribosomal proteins, glycolytic enzymes, and hexose transporters are highlighted.

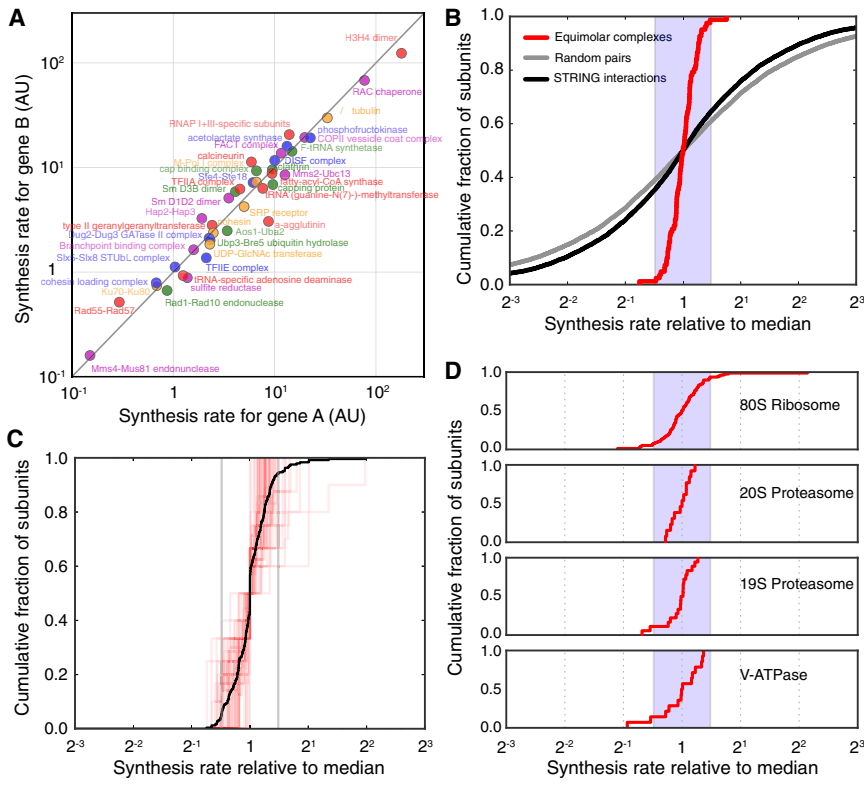
(D) Same as (C), with data for higher eukaryotes.

stoichiometry of each subunit within the complex must be clearly and quantitatively defined, as this sets our expectation for proportional synthesis (STAR Methods). A list of complex subunits excluded in our analysis can be found in Table S3.

Using a deep ribosome profiling dataset for *S. cerevisiae* to ensure sufficient read coverage for most complex subunits (Subtelny et al., 2014) (Table S4; STAR Methods), we found that the majority of stable, obligatory protein complexes in the cytosol and nucleus follow proportional synthesis. Among 41 small complexes made up of two subunits of equal stoichiometry, all but one have closely matched synthesis rates with <2-fold difference (75% within 1.42-fold difference). The absolute expression levels among these complexes span three orders of magnitude, ranging from the lowly expressed Mms4/Mus81 DNA endonuclease to the highly expressed histone H3/4 proteins (Figures

2A and S3A). The concordance between obligate partners is in stark contrast to the synthesis rates of randomly sampled interacting protein pairs (Szklarczyk et al., 2017) (Figure 2B).

For medium-sized complexes, such as the three-component eIF2 and the nine-component core exosome, the subunit synthesis rates are also tightly distributed relative to each other after normalization by their respective stoichiometry (75% within 1.61-fold difference) (Figures 2C and S3B). Similarly, the core subunits of large protein complexes, such as the ribosome and proteasome, have synthesis rates distributed almost completely within a 2-fold range (86%, 100%, 89%, and 86% of subunits for 80S ribosome, 20S proteasome, and 19S proteasome, and V-ATPase, respectively) (Figure 2D). Our masking strategy is particularly important for subunits with paralogs that have regions of identical coding sequences, such as ribosomal proteins



**Figure 2. Proportional Synthesis of Multi-protein Complex Subunits Is Ubiquitous in Budding Yeast**

(A) Synthesis rates for stable complexes composed of two different subunits (encoded by genes “A” and “B”) with equal stoichiometry. 82 subunits and 41 complexes are included.

(B) Cumulative distribution of synthesis rates for equimolar complexes (red), interacting protein pairs as annotated by STRING (Szkarczyk et al., 2017) (black), and random pairs selected from the proteome (gray). Synthesis rates are shown relative to the logarithmically transformed median of each complex. Shaded region indicates a 2-fold spread (max to min).

(C) Cumulative distribution of synthesis rates for individual medium-sized stable protein complexes (3–10 members) (red). Synthesis rates for subunits present in more than one copy per complex were divided by the stoichiometry and shown relative to the logarithmically transformed median of each complex. 244 subunits and 52 complexes are included. Black line shows the aggregate distribution among all medium-sized complexes. Gray vertical lines indicate a 2-fold spread (maximum to minimum).

(D) Cumulative distribution of synthesis rates for core subunits of the 80S ribosome, 20S proteasome, 19S proteasome, and V-ATPase, respectively. Synthesis rates are shown relative to the logarithmically transformed median of each complex. Paralogs of ribosomal proteins were grouped together. Shaded region indicates a 2-fold spread (maximum to minimum).

(Figures S1C–S1E). Consistent with a previous observation that yeast genes have a narrow range of translation efficiencies, the transcript abundances for these subunits, as measured by RNA-seq, demonstrate proportionality similar to their synthesis rates (Figure S5) (Weinberg et al., 2016). Notably, among all obligate complexes, the few atypical subunits that are synthesized at a higher rate than their partners tend to have shorter half-lives (Figure 3) (Christiano et al., 2014). This group is enriched in nuclear-encoded subunits of mitochondrial complexes (17 out of 24 atypical subunits) that appear to be synthesized less proportionally than the cytosolic and nuclear complexes (Figure S4). Subunits with secondary binding partners or moonlighting functions, which we excluded from the list of obligatory complexes, were also produced in excess (Figures S3C and S3D; STAR Methods). In summary, the general concordance suggests that protein synthesis rates have evolved to precisely match cellular demand and produce proportional amounts of protein complex subunits in this organism.

### Large Protein Complexes Are Synthesized Proportionally in Higher Eukaryotes

We next examined whether proportional synthesis is also a rule for cells of higher eukaryotes, such as fish and mammals. Compared to budding yeast, the relative scarcity of biochemical and genetic characterization for these organisms makes it challenging to define a high-confidence set of stable and obligatory complexes with known stoichiometry. We therefore restricted the analysis to large protein complexes such as the ribosome and proteasome. However, even in these cases, we note that the constituents of the core complex are still not fully deter-

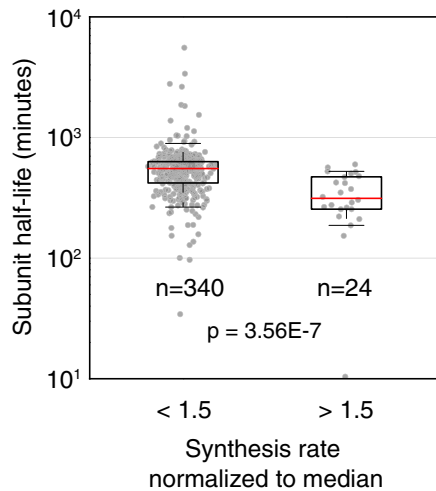
mined, especially given the increasing amount of evidence for specialized ribosomes and proteasome components in higher eukaryotes (Dahlmann, 2016; Shi et al., 2017). It is therefore plausible that some of the subunits analyzed here are not required in stoichiometric amounts.

Nevertheless, we found that the majority of ribosome and proteasome subunits are synthesized at near-stoichiometric ratios. Ribosome profiling data for primary tissue-derived samples showed that 78% of subunits are synthesized with less than 2-fold variation for human foreskin fibroblasts (Tirosh et al., 2015), 79% for mouse embryonic stem cells (Ingolia et al., 2014), and 86% for zebrafish 5 days post-fertilization (Chew et al., 2013) (Figure 4; Table S5). Therefore, despite differences in the prevalence of protein degradation, energetic constraints, and growth rates between the unicellular and multicellular organisms, higher eukaryotes have also evolved precise rates of synthesis for multiprotein complexes.

### Probing Regulation of Precise Synthesis Using Aneuploid Yeast

Precise rates of protein synthesis can be hard-coded through tuned *cis*-regulatory sequences or maintained through feedback loops. The presence of negative feedback regulation is typically detected by targeted perturbation to gene copy number, with a proportional increase in synthesis indicating the absence of feedback regulation. Aneuploid yeast strains represent a convenient system to test these possibilities on a genomic scale. Haploid lines have been constructed, which stably retain an extra copy of one chromosome (referred to as disomic strains), allowing measurements of dosage compensation during





**Figure 3. Complex Subunits Produced in Excess of Stoichiometry Are More Rapidly Turned Over**

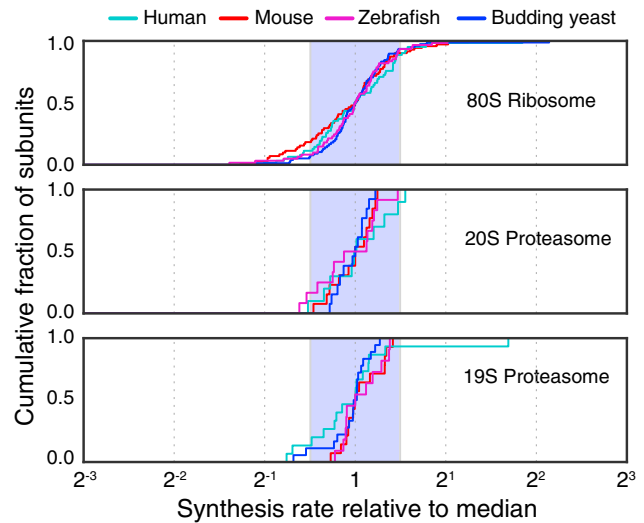
Boxplot showing the half-life of all complex subunits synthesized at levels below or above 1.5-fold times their complex median after normalization for their stoichiometry. Proteins produced in excess appear to have significantly shorter half-lives. Statistics were calculated as Wilcoxon rank-sum test between the bins shown in boxplot.

steady-state growth (Torres et al., 2007). Previous characterization has shown that dosage compensation is largely absent at the level of transcription for most disomic strains (Dephoure et al., 2014). Ribosome profiling for two disomic strains did not detect widespread translational compensation, although the quantitation of paralogs may be influenced by ambiguous assignment of multimapped reads (Dephoure et al., 2014; Thorburn et al., 2013). On the other hand, at the level of protein abundance, a subset of the proteome is attenuated to euploid levels (Dephoure et al., 2014). This subset is enriched for protein complex subunits, and it is possible that translational feedback regulation is a driver of this attenuation.

To determine the mode by which proportional synthesis is maintained and systematically characterize the sources of protein abundance attenuation, we performed ribosome profiling on 11 disomic strains (chromosomes I, II, V, VIII, IX, X, XI, XII, XIII, XV, and XVI) and their corresponding parental strain. We found that disome XV showed a large deviation of protein synthesis from its parental strain, likely reflecting a global transcriptional response, and we therefore excluded it from our analysis (Figure 5) (Sheltzer et al., 2012). A number of other disomic strains (including II, XIII, and XVI) appear to demonstrate such responses, albeit milder than disome XV. For all expressed genes encoded by each duplicated chromosome, their protein synthesis rates were tightly distributed around 2-fold of that measured in the parental (77% between 1.5 and 2.5; Figure 5). This suggests an absence of dosage compensation of protein synthesis for most genes, consistent with the genome-wide heterozygous deletion study (Springer et al., 2010).

### Feedback Control for Proportional Synthesis Is Rare in Budding Yeast

Although most proteins are produced in proportion to their gene copy number, it is possible that obligate subunits of protein com-

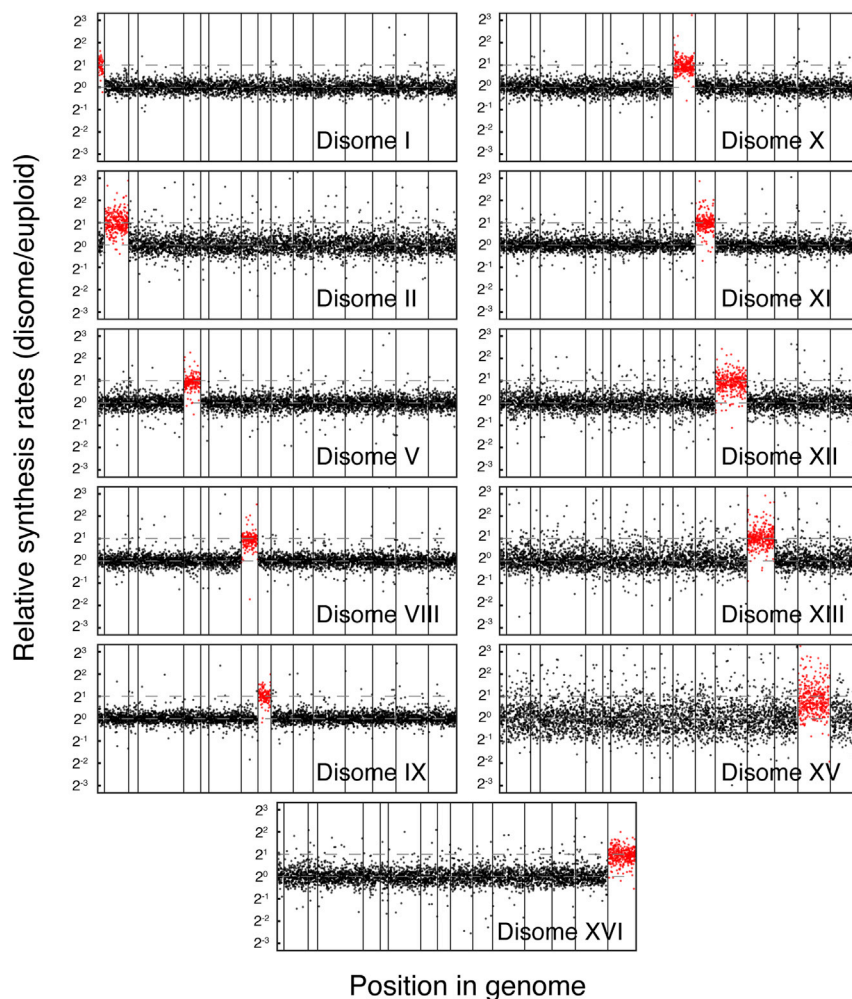


**Figure 4. Proportional Synthesis of Large Protein Complexes in Higher Eukaryotes**

Cumulative distribution of synthesis rates for core subunits of the 80S ribosome, 20S proteasome, and 19S proteasome, respectively. Data for zebrafish 5 days post-fertilization are shown in purple, for mouse embryonic stem cells in red, and for human foreskin fibroblast cells in cyan. For comparison, data for exponentially growing budding yeast are also shown (blue). Shaded region indicates a 2-fold spread (maximum to minimum). Genes that overlap or encode multiple unique protein products were not considered in this analysis (see STAR Methods).

plexes represent a subset that is feedback regulated. Feedback regulation of protein synthesis might be more common for ribosomal proteins as their abundances have been shown to be compensated for gene dosage (Dephoure et al., 2014). To test this possibility, we first examined the synthesis of ribosomal proteins encoded on the duplicated chromosome across disomic strains (excluding disome XV) and compared it with the parental strain.

Mapping ribosome footprint density for ribosomal protein genes has traditionally been challenging due to the widespread presence of highly similar paralogs, which can now be circumvented by the masking strategy presented in this work (Figures S1 and 6A). By quantifying the specific contribution of each gene locus, we found that the majority of 69 paralogs on the duplicated chromosomes are synthesized 2-fold higher compared to euploid levels, whereas their non-duplicated counterparts remain at euploid levels (Figures 6A and 6B). To determine whether the combined synthesis rate from all paralogous loci is subject to dosage compensation, we compared the observed synthesis rates of each ribosomal protein in disomic strains with what would be expected from the effective dosage change without compensation, i.e., the sum of relative synthesis rates weighted by gene dosage (Figure 6C). The majority of ribosomal proteins displayed increased synthesis rates matching the effective dosage change, indicating a general lack of negative feedback control under such perturbations (Figure 6C). Interestingly, a small number of proteins deviate from this trend, and these include the ribosomal proteins whose production were previously shown to be negatively autoregulated



**Figure 5. General Lack of Dosage Compensation at the Synthesis Level in Budding Yeast**

Manhattan plots for protein synthesis rates across the yeast genome for different disomic strains. Dashed lines indicate 2-fold increase from euploid levels. A small number of genes exceed the  $\pm 10$ -fold range on the vertical axes, and these values can be found in [Table S6](#).

## DISCUSSION

Ribosome profiling has revolutionized our ability to quantitate protein synthesis across the genome. Here, we presented a simple solution to the challenge of handling ribosome footprints of ambiguous origin, which can account for 16% of footprints in yeast. This rational approach to measure ribosome density for genes with paralogous partners enabled quantitative comparison of synthesis rates among components of multiprotein complexes, revealing the general principle of proportional synthesis in eukaryotes. The same strategy to process ribosome profiling data should also enable more quantitative analyses for studies of translation efficiency, ribosome pausing, and noncanonical translational events, especially for genomes with histories of regional or global chromosome duplications.

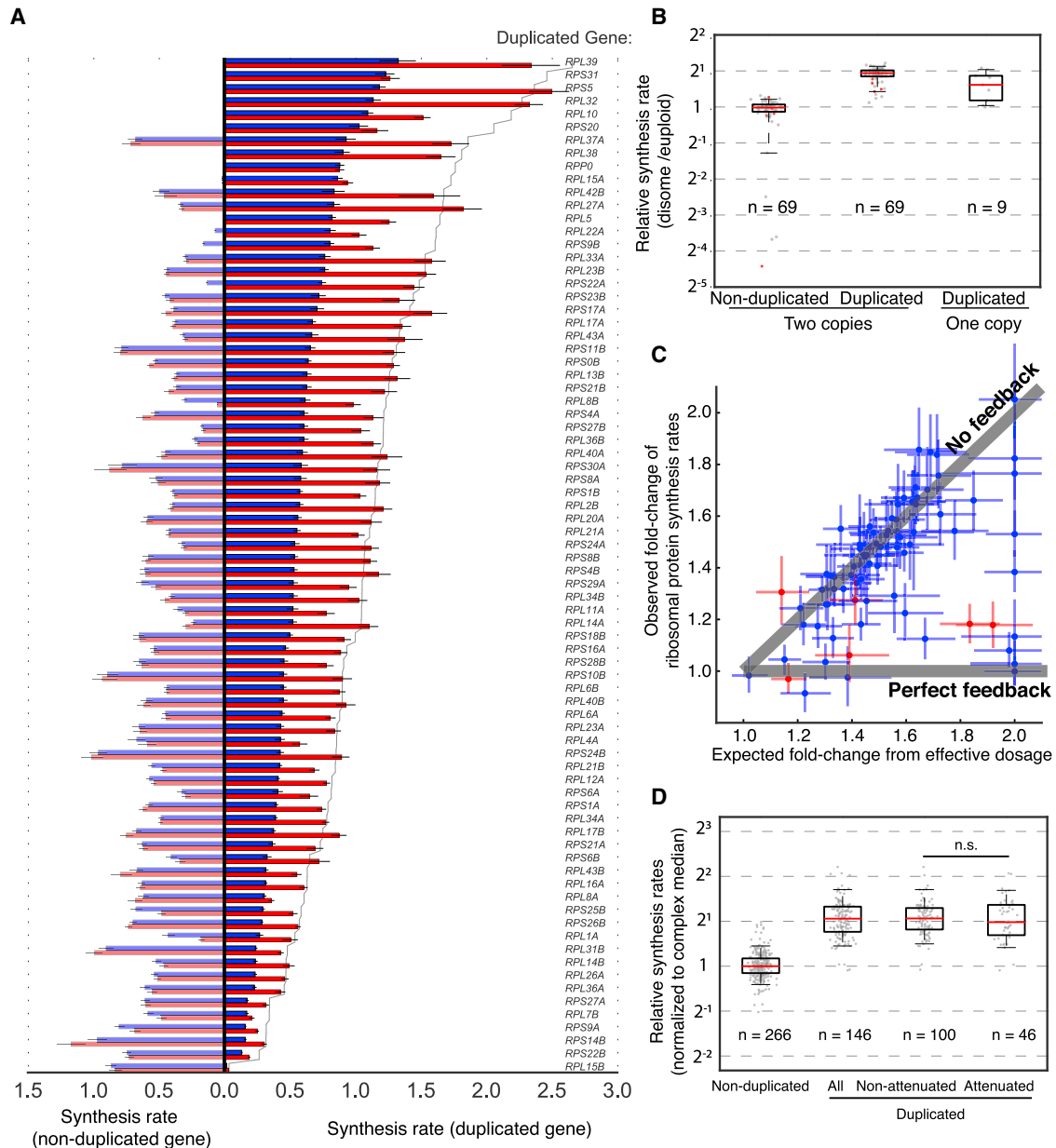
The widespread occurrence of proportional synthesis suggests that eukaryotes have evolved to produce precise amounts

of proteins, despite seemingly small costs of overproduction and an extensive proteostasis network to buffer against mismatches in synthesis. Post-translational quality control, such as degradation of excess subunits, is likely to serve as an important fail-safe mechanism and not the main gatekeeper for stoichiometric abundance under normal conditions. Because the synthesis of most proteins is not subject to negative feedback regulation, precise rates of synthesis imply that there is a low margin of error for every molecular step leading to protein production—transcription, splicing, export, translation, silencing, and degradation—and perturbations to these steps should be associated with selective disadvantage of the organism. Understanding the functional consequence of imbalanced subunit synthesis under the protective proteostasis network in eukaryotes will provide insights into the evolution of gene expression and dysregulation in diseases.

Our observation of disproportionate subunit synthesis in aneuploid yeast cells highlights a crucial scenario in which widespread degradation has to be activated. Without substantial dosage compensation until subunits are synthesized, as demonstrated here, the proteostasis network must encounter a much larger flux of unassembled proteins than it normally does when subunits are synthesized proportionally. The fact that the

by unassembled, excess subunits, such as L4, S9, L22, and L3 ([Gabunilas and Chanfreau, 2016](#); [Pearson et al., 1982](#); [Plocik and Guthrie, 2012](#); [Presutti et al., 1991](#)). Our data suggest that only a few other proteins are regulated this way ([Figures 6A and 6C](#)).

Although previous results demonstrate that subunits of non-ribosomal complexes are enriched for proteins whose abundances are dosage compensated ([Dephoure et al., 2014](#)), our data suggest that the compensation does not generally occur at the level of synthesis. We compared the synthesis of our curated non-ribosomal subunits encoded on the duplicated chromosome across disomic strains (excluding XV) to the median synthesis rate of their non-duplicated partners ([Figure 6D](#)). Whereas the non-duplicated subunits maintain their relative synthesis rates, the majority of duplicated subunits were synthesized 2-fold more than their binding partners (88% above 1.5-fold), indicating a limited number of genes feedback-regulated at the level of synthesis. Even among the subunits whose protein abundance is attenuated in disomes as measured by mass spectrometry ([Dephoure et al., 2014](#)), most are still synthesized in excess (83% above 1.5-fold). Taken together, these data suggest that precise synthesis rates are not actively maintained via feedback regulation for most protein complexes.



**Figure 6. Synthesis of Protein Complex Subunits Scales with Gene Dosage**

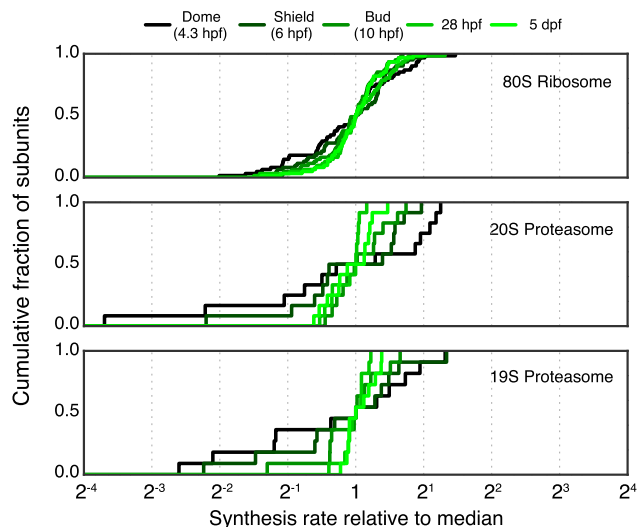
(A) Synthesis rates for all ribosomal proteins encoded on a duplicated chromosome, normalized to that of the median ribosomal protein within each sample. The duplicated paralogs are shown on the right, and the non-duplicated paralogs are shown on the left. Synthesis rates can be found in Table S7. Blue bars show synthesis rates in wild-type, and red bars show synthesis rates in disomic strains. Solid line indicates 2-fold increase from the wild-type level. Error bars represent 25<sup>th</sup> and 75<sup>th</sup> percentiles of bootstrapped distributions.

(B) Boxplot showing synthesis rates of individual paralogs between disomic and wild-type strains for all ribosomal proteins genes encoding a duplicated ribosomal protein. Synthesis rates are normalized to that of the median ribosomal protein within each sample. Whiskers show 5<sup>th</sup> and 95<sup>th</sup> percentile. For ribosomal proteins encoded by two genes, synthesis rates from the duplicated and non-duplicated paralogs are considered separately. Red points indicate proteins for which dosage compensation has been observed at the level of synthesis.

(C) Scatterplot showing synthesis rates of individual ribosomal proteins in the disomic strains. The y axis illustrates the observed fold-change of synthesis for each ribosomal protein (for proteins encoded by multiple paralogs, their synthesis rates are summed). The x axis represents the effective increase in gene dosage for each ribosomal protein in the disomic strain, dictated by the fraction of synthesis that comes from the duplicated gene. Gray lines indicate the expected synthesis rates given no feedback and perfect feedback. Red points indicate proteins for which dosage compensation has previously been observed at the level of synthesis. Error bars represent 25<sup>th</sup> and 75<sup>th</sup> percentiles of bootstrapped distributions.

(D) Boxplots show synthesis rates of complex subunits expressed from non-duplicated or duplicated chromosomes. Complexes with three or more subunits were considered, with synthesis rates divided by subunit stoichiometry. Subunits from the duplicated chromosomes were further separated by the presence of dosage compensation observed at the final protein level (attenuated and non-attenuated). Ribosomal and mitochondrial proteins were excluded, as well as proteins encoded by multiple genes. Statistics were calculated as Wilcoxon rank-sum test between indicated groups, with n.s. indicating p value > 0.05.





**Figure 7. Synthesis of Large Protein Complexes in Higher Eukaryotes during Zebrafish Development**

Cumulative distribution of synthesis rates for core subunits of the 80S ribosome, 20S proteasome, and 19S proteasome at 5 developmental time points. Genes that overlap or encode multiple distinct protein products were not considered in this analysis (see [STAR Methods](#)).

abundances of complex subunits remain largely invariant in aneuploid strains ([Dephousse et al., 2014](#)) indeed suggests that the excess proteins are constantly turned over in this situation. Abnormal futile cycles of protein synthesis and rapid degradation likely restrict the capacity of quality control pathways, leading to the widely observed proteotoxic stress in aneuploid cells ([Oromendia et al., 2012](#); [Torres et al., 2008](#)).

Are there other physiological conditions in which proportional synthesis is not expected to hold? Our analysis focused on populations of cells during steady-state growth, in which imbalanced synthesis would lead to a constant flux of unassembled subunits. The selective pressure to maintain precise synthesis may differ when cells are under transition and the flux of unassembled subunits is only transient or small in magnitude. For example, for *E. coli* following cold shock, obligate complex subunits do not appear to be produced proportionally for the first few hours ([Zhang et al., 2018](#)). However, because overall protein synthesis is drastically reduced during this time period, the accumulation of unassembled proteins may be too insignificant to be selected against. Similarly, translation is low in animal oocytes following fertilization, and we found evidence that protein complexes are also produced less proportionally ([Chew et al., 2013](#)) (Figure 7). In these situations, transiently discordant synthesis is not necessarily accompanied with increased degradation, as the newly synthesized proteins account for a small fraction of the total pool. Technically, measuring the precise rates of synthesis using ribosome profiling is challenging for cells transitioning from or to a slow-growth state, because some of the assumptions that are required, e.g., constant translation elongation rate and little ribosome drop-off, may not hold. A critical evaluation of this and alternative tools will be necessary to quantitatively examine the precision of protein synthesis for cells out of steady-state.

In all domains of life, protein synthesis is an engine for proliferation and a hub for regulation. Taken together with previous studies, our findings suggest that both bacteria and eukaryotes adhere to a general principle of precise synthesis, which provides stoichiometric amounts of components for the cell. Our study also highlights a key difference in the underlying regulation between the two domains, with negative feedback loops largely absent in eukaryotic protein synthesis. More broadly, the improved ability to quantitate protein synthesis in organisms with complex genomes should open new avenues to precisely define and interrogate regulation at the proteomic level.

## 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](#)
  - Strains and Culture Conditions
- [METHOD DETAILS](#)
  - Ribosome Profiling
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
  - Genome Sequence and Annotation
  - Read Trimming and Alignment
  - Rationale for Masking Identical Sequences
  - Mask Generation
  - Synthesis Rate Calculation
  - Curation of Obligate Protein Complexes
  - Bootstrapping Error Estimates
- [DATA AND SOFTWARE AVAILABILITY](#)
  - Sequencing Data

## SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.cels.2018.11.003>.

## ACKNOWLEDGMENTS

We thank A. Amon for providing disomic yeast strains, A. Amon and J.-B. Lalanne for comments on the manuscript, and MIT BioMicro Center for high-throughput DNA sequencing. This research is supported by NIH R35GM124732, Pew Biomedical Scholars Program, a Sloan Research Fellowship, Searle Scholars Program, the Smith Family Award for Excellence in Biomedical Research, a National Science Foundation Graduate Research Fellowship (to J.C.T.), and an NIH Pre-Doctoral Training Grant (T32 GM007287, to J.C.T.).

## AUTHOR CONTRIBUTIONS

J.C.T. and G.-W.L. designed experiments and analysis. J.C.T. collected data and performed analysis. J.C.T. and G.-W.L. wrote the manuscript.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 23, 2018

Revised: October 5, 2018

Accepted: November 18, 2018

Published: December 12, 2018

## REFERENCES

- Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* *31*, 166–169.
- Baudin-Baillieu, A., Legendre, R., Kuchly, C., Hatin, I., Demais, S., Mestdagh, C., Gautheret, D., and Namy, O. (2014). Genome-wide translational changes induced by the prion [PSI<sup>+</sup>]. *Cell Rep.* *8*, 439–448.
- Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal RNA-seq quantification. *Nat. Biotechnol.* *34*, 525–527.
- 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*, D700–D705.
- Chew, G.L., Pauli, A., Rinn, J.L., Regev, A., Schier, A.F., and Valen, E. (2013). Ribosome profiling reveals resemblance between long non-coding RNAs and 5' leaders of coding RNAs. *Development* *140*, 2828–2834.
- Christiano, R., Nagaraj, N., Fröhlich, F., and Walther, T.C. (2014). Global proteome turnover analyses of the yeasts *S.cerevisiae* and *S.pombe*. *Cell Rep.* *9*, 1959–1965.
- Cleveland, D.W., Lopata, M.A., Sherline, P., and Kirschner, M.W. (1981). Unpolymerized tubulin modulates the level of tubulin mRNAs. *Cell* *25*, 537–546.
- Dahlmann, B. (2016). Mammalian proteasome subtypes: their diversity in structure and function. *Arch. Biochem. Biophys.* *591*, 132–140.
- Dephore, N., Hwang, S., O'Sullivan, C., Dodgson, S.E., Gygi, S.P., Amon, A., and Torres, E.M. (2014). Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *Elife* *3*, e03023.
- Dunn, J.G., and Weissman, J.S. (2016). Plastid: nucleotide-resolution analysis of next-generation sequencing and genomics data. *BMC Genomics* *17*, 958.
- Freedman, L.P., Zengel, J.M., Archer, R.H., and Lindahl, L. (1987). Autogenous control of the S10 ribosomal protein operon of *Escherichia coli*: genetic dissection of transcriptional and posttranscriptional regulation. *Proc. Natl. Acad. Sci. USA* *84*, 6516–6520.
- Gabunilas, J., and Chanfreau, G. (2016). Splicing-mediated autoregulation modulates Rpl22p expression in *Saccharomyces cerevisiae*. *PLoS Genet.* *12*, e1005999.
- Gardin, J., Yeasmin, R., Yurovsky, A., Cai, Y., Skiena, S., and Futcher, B. (2014). Measurement of average decoding rates of the 61 sense codons in vivo. *Elife* *3*, 1–20.
- Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature* *466*, 835–840.
- Hashimoto, T., de Hoon, M.J.L., Grimmond, S.M., Daub, C.O., Hayashizaki, Y., and Faulkner, G.J. (2009). Probabilistic resolution of multi-mapping reads in massively parallel sequencing data using MuMRRescueLite. *Bioinformatics* *25*, 2613–2614.
- 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.
- Ingolia, N.T., Brar, G.A., Stern-Ginossar, N., Harris, M.S., Talhouarne, G.J.S., Jackson, S.E., Wills, M.R., and Weissman, J.S. (2014). Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep.* *8*, 1365–1379.
- Ishikawa, K., Makanae, K., Iwasaki, S., Ingolia, N.T., and Moriya, H. (2017). Post-translational dosage compensation buffers genetic perturbations to stoichiometry of protein complexes. *PLoS Genet.* *13*, e1006554.
- Juszkiewicz, S., and Hegde, R.S. (2018). Quality control of orphaned proteins. *Mol. Cell* *71*, 443–457.
- Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. (2016). The cost of protein production. *Cell Rep.* *14*, 22–31.
- Kahles, A., Behr, J., and Rättsch, G. (2016). MMR: A tool for read multi-mapper resolution. *Bioinformatics* *32*, 770–772.
- 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.
- Lalanne, J.B., Taggart, J.C., Guo, M.S., Herzel, L., Schieler, A., and Li, G.W. (2018). Evolutionary convergence of pathway-specific enzyme expression stoichiometry. *Cell* *173*, 749–761.e38.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.
- Lehnert, M.E., and Lodish, H.F. (1988). Unequal synthesis and differential degradation of  $\alpha$  and  $\beta$  spectrin during murine erythroid differentiation. *J. Cell Biol.* *107*, 413–426.
- Li, G.W. (2015). How do bacteria tune translation efficiency? *Curr. Opin. Microbiol.* *24*, 66–71.
- Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* *157*, 624–635.
- Lodish, H.F. (1974). Model for the regulation of mRNA translation applied to haemoglobin synthesis. *Nature* *251*, 385–388.
- Lynch, M., and Marinov, G.K. (2015). The bioenergetic costs of a gene. *Proc. Natl. Acad. Sci. USA* *112*, 15690–15695.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* *17*, 10.
- Mattheakis, L.C., and Nomura, M. (1988). Feedback regulation of the spc operon in *Escherichia coli*: translational coupling and mRNA processing. *J. Bacteriol.* *170*, 4484–4492.
- Meldal, B.H.M., Forner-Martinez, O., Costanzo, M.C., Dana, J., Demeter, J., Dumousseau, M., Dwight, S.S., Gaulton, A., Licata, L., Melidoni, A.N., et al. (2015). The complex portal - an encyclopaedia of macromolecular complexes. *Nucleic Acids Res.* *43*, D479–D484.
- Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* *441*, 840–846.
- Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* *26*, 2696–2708.
- Pearson, N.J., Fried, H.M., and Warner, J.R. (1982). Yeast use translational control to compensate for extra copies of a ribosomal protein gene. *Cell* *29*, 347–355.
- Plocik, A.M., and Guthrie, C. (2012). Diverse forms of RPS9 splicing are part of an evolving autoregulatory circuit. *PLoS Genet.* *8*, e1002620.
- Presutti, C., Ciafré, S.A., and Bozzoni, I. (1991). The ribosomal protein L2 in *S. cerevisiae* controls the level of accumulation of its own mRNA. *EMBO J.* *10*, 2215–2221.
- Rodriguez, J.M., Maietta, P., Ezkurdia, I., Pietrelli, A., Wesselink, J.J., Lopez, G., Valencia, A., and Tress, M.L. (2013). APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* *41*, D110–D117.
- Sheltzer, J.M., Torres, E.M., Dunham, M.J., and Amon, A. (2012). Transcriptional consequences of aneuploidy. *Proc. Natl. Acad. Sci. USA* *109*, 12644–12649.
- Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* *31*, 64–68.
- Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Röst, H.L., Teruel, M.N., and Barna, M. (2017). Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* *67*, 71–83.e7.
- Springer, M., Weissman, J.S., and Kirschner, M.W. (2010). A general lack of compensation for gene dosage in yeast. *Mol. Syst. Biol.* *6*, 1–8.
- Subtelny, A.O., Eichhorn, S.W., Chen, G.R., Sive, H., and Bartel, D.P. (2014). Poly(A)-tail profiling reveals an embryonic switch in translational control. *Nature* *508*, 66–71.
- Szklarczyk, D., Morris, J.H., Cook, H., Kuhn, M., Wyder, S., Simonovic, M., Santos, A., Doncheva, N.T., Roth, A., Bork, P., et al. (2017). The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* *45*, D362–D368.

- Thorburn, R.R., Gonzalez, C., Brar, G.A., Christen, S., Carlile, T.M., Ingolia, N.T., Sauer, U., Weissman, J.S., and Amon, A. (2013). Aneuploid yeast strains exhibit defects in cell growth and passage through START. *Mol. Biol. Cell* **24**, 1274–1289.
- Tirosh, O., Cohen, Y., Shitrit, A., Shani, O., Le-Trilling, V.T.K., Trilling, M., Friedlander, G., Tanenbaum, M., and Stern-Ginossar, N. (2015). The transcription and translation landscapes during human Cytomegalovirus infection reveal novel host-pathogen interactions. *PLoS Pathog.* **11**, e1005288.
- Torres, E.M., Sokolsky, T., Tucker, C.M., Chan, L.Y., Boselli, M., Dunham, M.J., and Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science* **317**, 916–924.
- Torres, E.M., Williams, B.R., and Amon, A. (2008). Aneuploidy: cells losing their balance. *Genetics* **179**, 737–746.
- Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111.
- Warner, J.R., and McIntosh, K.B. (2009). How common are extraribosomal functions of ribosomal proteins? *Mol. Cell* **34**, 3–11.
- Weinberg, D.E., Shah, P., Eichhorn, W., Hussmann, J.A., Plotkin, J.B., Bartel, D.P., Weinberg, D.E., Shah, P., Eichhorn, S.W., Hussmann, J.A., et al. (2016). Improved ribosome-footprint and mRNA measurements provide insights into dynamics and regulation of yeast translation. *Cell Rep.* **14**, 1787–1799.
- Xiao, Z., Zou, Q., Liu, Y., and Yang, X. (2016). Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* **7**, 11194.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Girón, C.G., et al. (2018). Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761.
- Zhang, Y., Burkhardt, D.H., Rouskin, S., Li, G.W., Weissman, J.S., and Gross, C.A. (2018). A stress response that monitors and regulates mRNA structure is central to cold shock adaptation. *Mol. Cell* **70**, 274–286.e7.

## STAR★METHODS

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
G418 Sulfate (Geneticin)	ThermoFisher Scientific	Cat# 11811-023
Dropout media supplement (-His)	Sigma-Aldrich	Cat# Y1751
SupraseIN	ThermoFisher Scientific	Cat# AM2696
cOmplete EDTA-free protease inhibitor	Sigma-Aldrich	Cat# 4693159001
Cycloheximide	Sigma-Aldrich	Cat# C7698
RNase I	Ambion	Cat# AM2295
T4 Polynucleotide Kinase	New England Biolabs	Cat# M0201S
SuperScript III Reverse Transcriptase	ThermoFisher Scientific	Cat# 18080093
CircLigase ssDNA ligase	Epicentre	Not available anymore from same provider
Dynabeads MyOne Streptavidin C1	ThermoFisher Scientific	Cat# 65001
Phusion High-Fidelity DNA Polymerase	New England Biolabs	Cat# M0530S
T4 RNA ligase 2 truncated K277Q	J. Weissman	N/A
Critical Commercial Assays		
SMARTer smRNA-seq kit	Clontech	Cat# 635029
Deposited Data		
Raw sequencing data and pile-up wig files for ribosome profiling	This paper	GEO: GSE118676
Experimental Models: Organisms/Strains		
Disomic yeast strains	A. Amon	<a href="#">Table S1</a>
Oligonucleotides		
rRNA removal oligonucleotides	This paper	<a href="#">STAR Methods</a>
Software and Algorithms		
Bowtie v2.0.0b3	<a href="#">Langmead et al. (2009)</a>	<a href="http://bowtie-bio.sourceforge.net/index.shtml">http://bowtie-bio.sourceforge.net/index.shtml</a>
Tophat v1.4.1	<a href="#">Trapnell et al. (2009)</a>	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Tophat v2.1.1	<a href="#">Kim et al. (2013)</a>	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
Python 2.7.10	Python Software Foundation	<a href="http://www.python.org">http://www.python.org</a>
Plastid v0.4.8	<a href="#">Dunn and Weissman (2016)</a>	<a href="https://plastid.readthedocs.io/en/latest/index.html">https://plastid.readthedocs.io/en/latest/index.html</a>
Mask generation and protein synthesis quantification scripts	This paper	<a href="https://github.com/jamesctaggart/masking_and_rpf_quantification">https://github.com/jamesctaggart/masking_and_rpf_quantification</a>
Other custom Python scripts for data analysis	This paper	Available upon reasonable request
Other		
Supor Membrane Disc Filters, Pall Laboratory (450 nm)	VWR	Cat# 28147-661
Grinding Jar Set, Stainless Steel	QIAGEN	Cat# 69985
TissueLyser II	QIAGEN	Cat# 85300
Novex TBE-Urea gels, 15%	ThermoFisher Scientific	Cat# EC6885BOX
Novex TBE-Urea gels, 10%	ThermoFisher Scientific	Cat# EC6875BOX
Novex TBE gels, 8%	ThermoFisher Scientific	Cat# EC62152BOX
Oligo Clean & Concentrator	Zymo Research	Cat# D4060

## CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and request for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gene-Wei Li ([gwli@mit.edu](mailto:gwli@mit.edu)).

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Strains and Culture Conditions

Disomic yeast strains are derivatives of W303 (details in Table S1). Overnight cultures for all strains except disome V and its corresponding wild-type were grown under selection in synthetic complete (SC) histidine drop-out medium with 1 g/L monosodium glutamate (1.7 g/L yeast nitrogen base without amino acids or ammonium sulfate (Sigma Aldrich), 1.92 g/L drop-out medium supplement (Sigma Aldrich), 2% glucose). Selection was maintained with 200  $\mu$ g/mL G418 (Thermo Fisher Scientific). Disome V and its paired wild-type were grown with 5 g/L ammonium sulfate instead of monosodium glutamate, which could reduce the efficacy of G418 selection. Ribosome profiling indicated, however, that this had minimal impact on retention of the duplicated chromosome (Figure 5). Outgrowth prior to harvesting for ribosome profiling was performed in YPED (20 g/L peptone, 10 g/L yeast extract, 2% glucose) without selection for all strains. All growth in liquid culture was carried out shaking at 30°C.

## METHOD DETAILS

### Ribosome Profiling

Overnight cultures were grown from a single colony in SC medium –His+G418 and back-diluted to  $OD_{590} = 0.002$  in 500 mL YPED. Cultures were grown to  $OD_{590} = 0.5$  and harvested at 30°C by rapid filtration through a filter with a 450 nm pore-size (Supor Membrane Disc Filters, Sigma Aldrich) and flash freezing in liquid nitrogen. Cell pellets were combined with 2.5 mL of frozen lysis buffer (10 mM Tris pH 7.4, 5 mM  $MgCl_2$ , 100 mM KCl, 200  $\mu$ g/mL heparin, 1% sodium deoxycholate, 1% Triton X-100, 2mM DTT, 20U/mL SupersaseIN, 100ug/mL cycloheximide, with 1 tablet cComplete EDTA-free protease inhibitor per 10 mL buffer). The cell and buffer pellet mixture was ground in pre-chilled 25 mL grinding jars (QIAGEN) for 6 cycles of 3 minutes at 15 Hz, cooling in liquid nitrogen between cycles. Ground lysates were thawed and clarified by centrifugation at 20,000 RCF at 4°C for 10 minutes. 20  $A_{260}$  units of lysate were digested with 300 U RNase I (Ambion) at room temperature for 1 hour. Immediately following RNase I digestion, digested lysates were separated on a sucrose gradient and monosomes were collected.

For all samples, RNA was extracted by hot-phenol extraction. Ribosome footprints were purified by denaturing polyacrylamide gel electrophoresis (15% TBE-Urea, Thermo Fisher Scientific), excising fragments bounded by 28 and 31-nucleotide size standards. For disome V and its paired wild-type sample, subsequent library prep and sequencing were carried out nearly as described previously (Lalanne et al., 2018). The only modification to this protocol was an rRNA removal step prior to reverse transcription, rather than following circularization, utilizing selective hybridization to six 5'-biotinylated oligonucleotides (IDT, 5'GATCGGTCGATTGTGCACC, 5'CCGCTTCATTGAATAAGTAAAGAAAC, 5'GACGCCTTATTCGTATCCATCTATA, 5'GGGACCTTGAATGCTAGAACGTGGAAAA, 5'TTAGCCAGAAGGAAAGGCCCGTTGGAA, 5'CATTGATTTTTTATCTAATAAATACATCTCT) followed by MyOne Streptavidin C1 Dynabead purification (Thermo Fisher Scientific).

For all other ribosome profiling samples, footprints were dephosphorylated with 20 units of T4 polynucleotide kinase (NEB) at 37°C for 1 hour, and cDNA libraries were prepared using the SMARTer smRNA sequencing kit (Clontech) as per manufacturer's instructions. Prior to PCR amplification, rRNA sequences were depleted in all libraries through selective hybridization with a mixture of six 5'-biotinylated oligonucleotides (IDT, 5'TTTTCCACGTTCTAGCATTCAAGGTCCC, 5'TTCCAACGGGGCCTTTCTTCTGGCTAA, 5'AGAGATGTATTTATTAGATAAAAAATCAATG, 5'GGTGCACAATCGACCGATC, 5'GTTTCTTTACTTATTCAATGAAGCGG, 5'TATAGATGGATACGAATAAGGCGTC) and MyOne Streptavidin C1 Dynabead purification. PCR reactions were then set up as per kit instructions.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Genome Sequence and Annotation

For *Saccharomyces cerevisiae*, the S288C genome sequence and annotation (release R64-2-1) were downloaded from SGD (Cherry et al., 2012). *Mus musculus* (GRCm38.90), *Homo sapiens* (GRCh38.91), and *Danio rerio* (GRCz10.91) genome sequences and annotations were downloaded from Ensembl (Zerbino et al., 2018). All annotation files were filtered to remove dubious transcript isoforms prior to use. For *S. cerevisiae*, this was achieved by removing all transcripts tagged as dubious in the annotation file. For all other species, any transcript isoform which was not annotated as principal or alternative in the APPRIS database (Rodriguez et al., 2013) was removed from the annotation. Each reference genome was then parsed and any contig which did not encode a transcript in the filtered reference was removed.

### Read Trimming and Alignment

Raw FASTQ files were trimmed of appropriate poly(A) or linker sequences using Cutadapt (Martin, 2011) with parameters  $-u 1 -m 24$ . The first nucleotide was trimmed to reduce mismatches due to non-template addition during reverse transcription common in many library preparation protocols. For all samples prepared with the SMARTer smRNA sequencing kit,  $-u 3$  was used instead of  $-u 1$ . Data obtained from (Subtelny et al., 2014) were trimmed of their adapter and first 8 nucleotides using an in-house Python script rather than Cutadapt. RNA-seq (non-ribosome profiling) data from this study were further trimmed from their 3' end to a uniform length of 27 nt. Trimmed reads were depleted for ribosomal RNA sequences prior to genome alignment by alignment to pre-rRNA transcript sequences using bowtie v2.0.0b3 (Langmead et al., 2009) with parameters  $-v 0 -k 1 -best$ . Unmapped reads



were then aligned to the appropriate filtered reference transcriptome using tophat v1.4.1 (Trapnell et al., 2009) using parameters `-G $annotation_file -no_novel-juncs -transcriptome-only -N 1`. For analyses which require random assignment of multimapping reads, `-g 1` was also specified in tophat. For analysis which required discarding multimapping reads, bowtie v2.0.0b3 was used for alignment to the genome using parameters `-m1 -v1 -k1 -best -strata`. The 5' end of all alignments were mapped with an offset of +13 nucleotides.

For all analyses, aligned ribosome protected fragments were read with custom python scripts written utilizing the plastid library (Dunn and Weissman, 2016). To enrich for high-quality ribosome footprints, only reads between 27 and 33 nucleotides were considered in our analyses. Read length distributions from the disomic libraries demonstrated slightly elevated lengths (with a mode of 31 for the wild-type), likely due to a combination of slight under-digestion with RNase I and additional nucleotides being added to the 5' end during SMARTer smRNA-seq template switching. Due to the polyadenylation and template switching strategy used during library preparation, the 5' and 3' ends of ribosome footprints in our dataset cannot be precisely resolved, limiting codon-resolution analysis. This does not impact our analyses, which focus only on transcript-level read density.

### Rationale for Masking Identical Sequences

To illustrate the potential artifacts in conventional read-counting strategies, consider two genes that share some regions of identity and are expressed at different levels (Figure S1A). While discarding multimapping reads (Common strategy 1) preserves the density ratio between the two genes, it artificially reduces the measured density for both genes (Figure S1A). On the other hand, strategies that retain all reads, such as random assignment (Common strategy 2) or multiply counting, artificially shift read density from the more highly expressed gene to the lowly one (Figure S1A). A third and less commonly used strategy is to assign multimapping reads probabilistically based on prior knowledge of the read density averaged over unique neighboring regions, which could in principle provide more accurate quantitation (Bray et al., 2016; Hashimoto et al., 2009; Kahles et al., 2016) (Figure S1B). However, the probabilistic assignment does not add any additional information to ribosome flux if there is already prior knowledge based on unique regions. The latter can be determined by our masking strategy.

### Mask Generation

Masked positions were stored as BED files, encoding regions of ambiguous sequence within each reference genome. To generate these files, a modified version of the *crossmap* program in the plastid library (Dunn and Weissman, 2016) was used. In this approach, filtered reference transcriptomes were broken into 27-mers and aligned back on themselves using tophat2 v2.1.1 (Kim et al., 2013) with parameters `-G $annotation_file -bowtie1 -transcriptome_only -no_novel-juncs -N 1`. The resulting BAM files were then filtered by MAPQ score, flagging any position which corresponds to the 5' end of a 27-mer that mapped with a score not equal to 50, which corresponds to uniquely mapping reads. These positions were then compiled as a BED file. The positions in this file were offset by +13 nucleotides to match the offset applied to ribosome profiling reads after mapping.

### Synthesis Rate Calculation

Ribosome profiling allows calculation of protein synthesis rates if two assumptions are made: no premature ribosome drop-off and equal average elongation rates across mRNAs (Li, 2015). Many eukaryotic datasets demonstrate significant bias in read coverage towards the 5' of coding regions (Figure S2A), suggesting some degree of deviation from these assumptions. This bias may lead to overestimation of ribosome density for short genes and underestimation of ribosome density for long genes, making it difficult to evaluate proportional synthesis. Our quantification pipeline must therefore explicitly correct for this positional bias in observed ribosome density for accurate estimation of protein synthesis rates, as described below.

Elevated ribosome density towards the 5' of coding regions is likely to be due to either of two events: aborted translation as ribosomes move downstream or slow elongation rates at early stages of translation. In the first case, the actual rates of protein synthesis can be estimated by correcting the observed ribosome occupancy profile by an expected drop-off profile at each position, thereby removing contributions of ribosomes that do not finish translation. In the second case, the actual rates can be estimated by correcting the observed ribosome occupancy profile by an expected elongation rate profile, thereby removing contributions of uneven elongation. For both of these cases, the expected profile used for correction can be estimated using a metagene analysis which generates an averaged position-wise ribosome density across all genes.

To do so, we extracted read count profiles across each protein coding region, excluding the first and last 5 codons (yeast and mouse datasets) or 25 codons (human and zebrafish datasets) of the open reading frame. These additional nucleotides are trimmed from human and zebrafish coding regions due to high local variability at the 5' end of their metagene profiles. Coding regions are excluded if their corresponding gene encodes two or more isoforms with non-identical coding regions, if more than 50% of positions are included in our mask file, or if fewer than 128 reads map. The signal within each codon was averaged and the smoothed profile was then normalized by dividing the signal at each position by the mean signal of the first 150 positions. We then calculated a 90% winsorized mean of the normalized signal at each position relative to the trimmed CDS start across all genes. This metagene profile was smoothed by averaging over a 51-nucleotide sliding window, generating a position-wise correction factor used in downstream synthesis rate calculations.

To convert read density to protein synthesis rate, we extracted the read count profiles for the coding region of each gene. Genes encoding at least two isoforms distinct coding regions were flagged for removal in downstream analysis. Genes with overlapping coding regions were also omitted from our analysis. The ends of the coding region were trimmed as during the metagene calculation,

and the position-wise metagene correction was applied by dividing each position by the corresponding value in the smoothed metagene curve. Because the number of genes used to calculate the metagene at each position decreases with length, this correction factor becomes increasingly noisy at positions far from the CDS start. For this reason, at positions beyond 3000, the median value of the metagene curve between positions 2900 and 3000 was used. Synthesis rates were calculated as the mean read count in this corrected profile, excluding positions which are defined by the mask for the relevant reference transcriptome. Overall, this metagene correction should have the greatest effect on relative quantification between short and long genes, and was found to improve the observed proportionality of synthesis rates between complex subunits of greatly differing lengths (Figure S2B).

### Curation of Obligate Protein Complexes

To generate a database of obligate *S. cerevisiae* protein complex subunits, we used the EMBL Complex Portal (Meldal et al., 2015) as a starting point. This database was last accessed in July of 2018. Complexes were individually sorted, using functional annotation from SGD (Cherry et al., 2012) and review of relevant literature to determine the potential for promiscuous or non-obligate interaction of any subunit. For complexes in which a stable, non-promiscuous core could be identified, this reduced complex was included. In the cases of functionally redundant genes encoded by paralogs, these proteins were considered as a single functional unit and combined in subsequent analysis. Additional complexes and information about subunit stoichiometry not contained in the EMBL database was mined from primary literature. A full accounting of the included complexes can be found in Table S2 and excluded complexes in Table S3.

In total, 577 subunits of 113 complexes were included in our curation. Excluding subunits localized to the mitochondria, 459 subunits of 101 complexes remain. In the wild-type budding yeast dataset studied in Figure 2 (Subtelny et al., 2014), all but one complex (a heterodimer) passed a cutoff of 128 reads per quantified region, a threshold above which counting statistics likely do not contribute to the spread of estimated synthesis rates (Ingolia et al., 2009). This evidenced by Figures S3A and S3B, which show that the spread in synthesis rates among complex subunits does not appear to be strongly determined by absolute expression level.

Considering the subunits which were removed due to moonlighting functions, we found that these subunits are produced in excess of their complex stoichiometry (Figure S3C). Further, we found that moonlighting subunits which participate in a well-defined secondary complex are produced at a rate which is equal to the sum of synthesis rates of each complex individually (Figure S3D). These results highlight the necessity of curation of our complex annotation and show that expectations for precise synthesis based on our curation are met.

### Bootstrapping Error Estimates

To estimate error of our estimates for protein synthesis rates, we generated a bootstrapped distribution of ribosome density. To do this, we randomly re-sampled each gene's corrected and masked read count profile with replacement 1000 times, generating simulated profiles equal in length to the original. We then calculated the mean for each of these profiles, generating a distribution of simulated synthesis rates. When calculating relative synthesis rates between samples, as in Figure 6C, we generated 10,000 individual bootstrapped ratios, each time resampling to generate a synthesis rate for all genes of interest and subsequently calculating the desired quantity.

## DATA AND SOFTWARE AVAILABILITY

### Sequencing Data

The 13 ribosome profiling datasets generated in this paper are available at the Gene Expression Omnibus with accession number GSE118676. Additional ribosome profiling datasets analyzed in this work are available at accession numbers GSE53313 (Subtelny et al., 2014), GSE60095 (Ingolia et al., 2014), GSE69906 (Tirosh et al., 2015), and GSE46512 (Chew et al., 2013).