Post-transcriptional mechanisms modulate the consequences of adaptive copy number variation

Pieter Spealman[1], Carolina de Santana[2], Titir De[1], and David Gresham[1,3]
[1]Center for Genomics and Systems Biology, Department of Biology, New York University
[2]Laboratório de Microbiologia Ambiental e Saúde Pública - Universidade Estadual de Feira de Santana (UEFS), Bahia
[3]Correspondence: dgresham@nyu.edu

Abstract

Copy-number variants (CNVs) are large-scale amplifications or deletions of DNA that can drive rapid adaptive evolution and result in large-scale changes in gene expression. Whereas alterations in the copy number of one or more genes within a CNV can confer a selective advantage, other genes within a CNV can decrease fitness when their dosage is changed. Dosage compensation - in which the gene expression output from multiple gene copies is less than expected - is one means by which an organism can mitigate the fitness costs of deleterious gene amplification. Previous research has shown evidence for dosage compensation at both the transcriptional level and at the level of protein expression; however, the extent of compensation differs substantially between genes, strains, and studies. Here, we investigated sources of dosage compensation at multiple levels of gene expression regulation by defining the transcriptome, translatome and proteome of experimentally evolved yeast (*Saccharomyces cerevisiae)* strains containing adaptive CNVs. We quantified the gene expression output at each step and found evidence of widespread dosage compensation at the protein abundance (~47%) level. By contrast we find only limited evidence for dosage compensation at the transcriptional (~8%) and translational (~3%) level. We also find substantial divergence in the expression of unamplified genes in evolved strains that could be due to either the presence of a CNV or adaptation to the environment. Detailed analysis of 82 amplified and 411 unamplified genes with significantly discrepant relationships between RNA and protein abundances identified enrichment for upstream open reading frames (uORFs). These uORFs are enriched for binding site motifs for SSD1, an RNA binding protein that has previously been associated with tolerance of aneuploidy. Our findings suggest that, in the presence of CNVs, SSD1 may act to alter the expression of specific genes by potentiating uORF mediated translational regulation.

Introduction

Copy-number variants (CNVs) are amplifications or deletions of DNA that can span dozens of nucleotides to whole chromosomes. CNVs are frequently observed over both short and long evolutionary time spans, although their selective advantage may be different between the two (Kuzmin et al. 2022). In the short term CNVs can result in large changes in gene expression and protein abundance, which provides a selective advantage further driving rapid adaptive evolution (Kondrashov 2012; Myhre et al. 2013; Hull et al. 2017; Robinson et al. 2023). Gene amplification has been shown to mediate rapid adaptation to a variety of selective pressures from nutrient limitation to antibiotics in both natural and experimental populations of microbes (Lauer et al. 2018; Selmecki et al. 2009; Todd and Selmecki 2020; Gresham et al. 2008; Hong and Gresham 2014; Dhami et al. 2016; Nair et al. 2008; Pränting and Andersson 2011; Paulander et al. 2010; Hull et al. 2017). CNVs are also common in cancers, where they can promote tumorigenesis by oncogene amplification (Ben-David and Amon 2020) and drive genome wide changes in gene expression (Shao et al. 2019).

A simple model of the short-term fitness effects of CNVs is a variant of the “driver-hitchhiker” model (Smith and Haigh 1974; Golzio and Katsanis 2013) wherein the fitness benefit of a CNV is derived from the amplification of a single gene or “driver”, while fitness costs arise from the amplification of all “hitchhikers”. Since CNVs include up to hundreds of genes the potential fitness costs of hitchhikers can be high (Makanae et al. 2013; Rice and McLysaght 2017; Avecilla and Spealman et al. 2023). These fitness costs can be categorized as “dosage burden” wherein the fitness cost arises from the burden of the additional DNA replication and gene expression (Dekel and Alon 2005; Wagner 2005; Makanae et al. 2013; Bonney et al. 2015). Conversely, fitness costs may arise from the stoichiometric imbalance of specific “dosage sensitive” genes. The archetypal example of dosage sensitivity is the imbalance of proteins involved in a heteromeric protein complex, leading to the complex having altered function (Wagner 2005; Veitia et al. 2008; Rice and McLysaght 2017). Importantly, the fitness effect of any CNV is highly dependent on the genetic background and environmental context (Sunshine et al. 2015; Robinson et al. 2023).

Dosage compensation (DC) is one mechanism by which an organism could mitigate the fitness costs of hitchhiker gene amplification. While classically associated with sex chromosome inactivation through chromatin silencing (Straub and Becker 2007; Jordan et al. 2019), DC can also be gene specific (Straub and Becker 2007). Furthermore, while DC is often conceptualized as 'complete', such that additional gene copies result in no additional expression, DC can also result in range of increased gene expression levels that are significantly less than what would be expected based on the copy-number (Itoh et al. 2007; Brockdorff and Turner 2015). This 'incomplete' dosage compensation is sometimes referred to as attenuation (Ascencio et al. 2021). Here, we use the term dosage compensation (DC) to refer to both 'complete' (e.g., the gene expression with two copies of the allele is identical to the gene expression from one copy) and 'incomplete' (e.g., the gene expression with two copies of the allele is greater than the gene expression from one copy, but less than two times the expression of one copy).

Recent work on the transcriptional regulation of CNV amplified genes has found that, depending on genetic background and environment 10-60% of amplified genes feature transcriptional dosage compensation (Springer et al. 2010; Hose et al. 2015; Gasch et al. 2016; Avecilla and Spealman et al. 2023). The mechanism underlying this DC is unclear but changes in transcription factor abundances and their concentrations is one possible mechanism (Prestel et al. 2010).

Translational DC has recently become a focus of interest with the advent of ribosome profiling (Ingolia et al. 2009; Zhang and Presgraves 2017) but has been studied using other methods (Gebauer et al. 1999; Uebbing et al. 2015; Krasovec et al. 2023). At the translational level some global mechanisms have been proposed including spliceosomal or ribosomal components acting to limit the global expression of the organism (Veitia et al. 2013), and co-translational mechanisms, such as the unfolded protein response (Kane et al. 2021), although the universality of these mechanisms remains unresolved (Larrimore et al. 2020). Gene specific mechanisms have also been proposed and identified, such as RNA-binding proteins (RBPs) enabling the identification of specific genes and altering their post-transcriptional and translational regulation (Mayr 2017; Harvey et al. 2018; Schultz et al. 2020). The archetypal case of which is the sex-chromosome dosage compensation in Drosophila, enacted by the RNA-binding protein SXL altering the efficacy of alternative translation initiation at an upstream open reading frame (uORF) (Medenbach et al. 2011).

Upstream open reading frames (uORFs) are relatively short (6-300 nucleotide long) ORFs that are positioned upstream, in the transcript leader (TL, or 5' UTR) of the main ORF (mORF) of a gene. uORFs have either canonical start codons (e.g. AUG) or near-cognate codon (NCC, Zitomer et al. 1984) translation initiation sites (TIS) that allow for alternative translation initiation separate from the mORF, and as such can operate as translational regulatory elements, either preventing or enabling the translation of the downstream mORF (Hinnebusch 2005). While uORFs have been suggested to be a broad class of stress responsive regulatory elements (Lawless et al. 2009) their role in the CNVs and aneuploids has not previously been evaluated. Similarly, start codons that are upstream, in-frame, and lack a stop codon separating them from the mORF are capable of adding additional peptides to downstream protein. These N-terminal extensions (NTEs) have been shown to alter the function of the protein product (Chang and Wang 2004; Tang et al. 2004; Suomi et al. 2014), with one recent study identifying an enrichment of mitochondrial targeting signals encoded by NTEs in yeast (Monteuuis et al. 2019).

One interesting candidate for gene specific translational regulation in yeast is SSD1, which encodes an RBP found to preferential bind to cis-elements in TLs (Bayne et al. 2022) and has been shown to alter translation (Hogan et al. 2008; Jansen et al. 2009; Ohyama et al. 2010). SSD1 is believed to play an important role in the response of yeast to aneuploid stress (Hose et al. 2015; Hose 2020). While initial research suggested that yeast exhibited a particular sensitivity to aneuploidy, subsequent research has shown that this sensitivity is

strain dependent, and due to the loss of *SSD1* in the lab strain W303 (Hose et al. 2015, Hose et al. 2020). Whether SSD1 has a role in accommodating CNVs in the genome is unknown.

Quantitative mass spectrometry has enabled the identification of DC at the level of protein abundance (Stingele et al. 2012; Dephoure et al. 2014; Ishikawa et al. 2017; Taggart et al. 2020; Schukken and Sheltzer 2022; Senger et al. 2022). At the protein level, work in yeast has suggested 10-20% of gene amplifications have significant DC (Dephoure et al. 2014; Ishikawa et al. 2017), although genetic differences are a known source of variance (Larrimore et al. 2020; Hose et al. 2020). One proposed mechanism for DC at the protein level is that excess subunits of proteins associated with heteromeric complexes are selectively targeted for protein aggregation (Samant et al. 2019) and degradation by ubiquitination (Ishikawa et al. 2020; Senger et al. 2022). A recent study using hundreds of natural isolates reports that nearly all aneuploid yeast strains observed had consistent protein level DC with an average decrease of 25% from the expected abundance (Muenzner et al. 2022).

Here, we undertook an analysis of multi-level gene expression regulation in long-term experimentally evolved strains of *Saccharomyces cerevisiae* that acquired adaptive CNVs and their ancestor. These strains evolved over the course of hundreds of generations to glutamine-limited growth media and contain distinct adaptive CNVs containing the same driver gene *GAP1* (Lauer et al. 2018) but distinct sets of additional hitchhiker genes within the CNV. We quantified expression at the transcriptional, translational, and protein levels in the glutamine-limited selective conditions in which the CNVs were selected. We find evidence of significant dosage compensation at the transcriptional (~8%), translational (~3%), and protein abundance (~47%) levels. We find that many genes with significant differences in expression are enriched in potential RBP binding sites for SSD1 as well as potential uORFs. Our results suggest that SSD1 and uORFs may act synergistically to limit the protein production of specific genes thereby modulating the impact of CNVs on gene expression.

Results

Topological structure of copy number variant alleles

We isolated four lineages that contained CNVs at the *GAP1* locus following long term selection in glutamine-limited chemostats. CNV-containing lineages were identified through the use of a CNV reporter system (Lauer et al. 2018; Spealman et al. 2023) and isolated using FACS from heterogeneous populations following more than 150 generations of selection. Defining the gene expression consequences of altered copy number requires precise definition of the copy number of each gene. Therefore, we performed hybrid *de novo* genome assembly, using a combination of long-read (ONT) and short-read (Illumina) sequencing to determine the topological structures for the CNVs in each strain (**Figure 1B**). On the basis of the CNV structures we also inferred the likely mechanisms of formation, which include transposon mediated tandem amplification and origin of replication dependent amplification (ODIRA, Brewer 2015) and named the CNV alleles to reflect these mechanisms. Because of the difficulty in ONT sequencing of inverted repeat sequences (Spealman et al. 2020) all ODIRA events were fully resolved using custom CNV identification (**methods**).

CNV structures include “ODIRA_A”, an ODIRA event triplication spanning 77 genes (*ALY1-SIR1*) and “ODIRA_B”, an ODIRA event triplication spanning 91 genes (*VPS1-ESL2*). “Trans”, a transposon mediated tandem duplication (*TOF2-GAP1*) nested within a translocation and telomere replacement on ChrVIII. Finally, a complex quadruplex or “ComQuad" resulting from a ODIRA driven triplication followed by a singular amplification of *GAP1* by a transposon event that spans the common core 17 genes from (*SET3-TRK2*). A complete list of genes and their copy number is available (**Table S1**)**.** Collectively, the four CNV strains include between 17 - 91 genes present at 2, 3, or 4 additional copies. A set of 17 amplified genes centered on the *GAP1* gene is common to all strains.

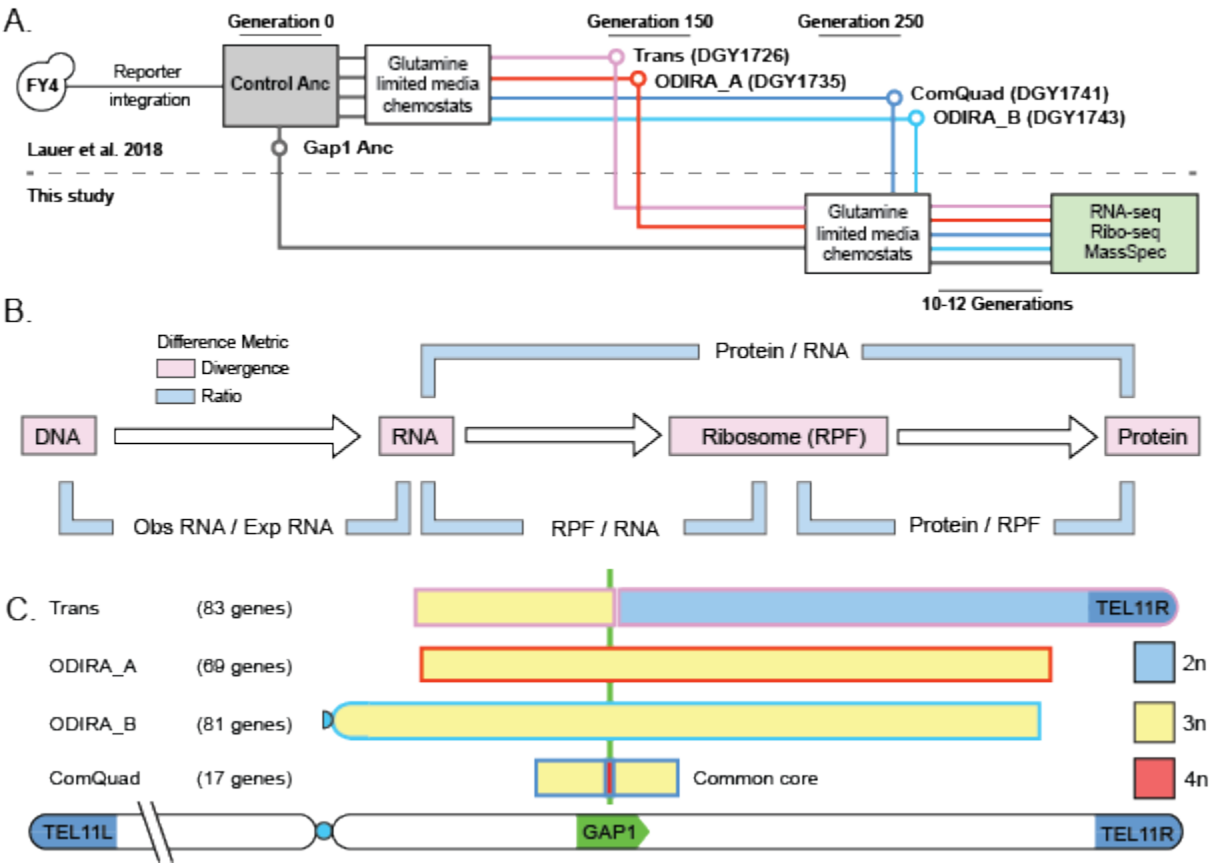

Figure 1. Study design and characterization of CNV strains. (**A**) Strain provenance (Lauer et al. 2018, above the dotted line) and experimental design to assess gene expression effects using RNA-seq, ribosome profiling (Ribo-seq), and TMT-labeled mass spectrometry (this study, beneath dotted line) in glutamine-limited media conditions. (**B**) Definition of terminology used in this study. "Divergence" (**pink**) is used to describe the difference between an evolved CNV strain and the ancestor without copy-number correction or accounting for preceding levels of expression. "Ratio metrics" (**blue**) quantifies the difference in observed and expected abundances using 1) an evolved CNV strain and the ancestor using a relative ratio to correct for gene copy number increase, or 2) the abundance of the preceding step of gene expression. (**C**) A schematic showing the *GAP1*-locus amplified genes in each strain and their copy-number. A common core of 17 genes is amplified in every evolved strain.

Increased levels of gene expression are associated with CNV amplified genes

To assess the gene expression consequences of CNVs we analyzed the four evolved CNV strains and the ancestral wildtype in glutamine-limited chemostats. Strains were grown until reaching a steady-state condition in which culture density remained constant (~10-12 generations). Cycloheximide (100µg/L) was added to cells for 2 minutes before harvesting by vacuum filtration. All RNA-seq and ribosome profiling were performed in tandem using two biological replicates for each strain. Mass spectrometry was conducted using 5 biological replicates.

To define a dataset for quantitative analysis of gene expression we set a threshold for each gene of at least 100 RNAseq reads in at least one sample, at least 10 RPF reads in at least one sample, and have at least one unique peptide detected by mass spectrometry in at least one sample. Following filtering, 4236 genes remained in our dataset, of which 91 were amplified within a CNV in at least one evolved strain. These 91 amplified genes include the common core of 17 genes which are amplified in every evolved strain.

We first sought to compare gene expression differences for CNV amplified genes between each evolved strain and the ancestor (**Figure 1B, pink**). We define gene expression differences between an evolved CNV strain and the ancestor as the gene expression divergence, which we determine for all genes regardless of copy number. To visualize divergence in gene expression within the CNV amplified region we generated heatmaps of robust scaled expression data (**methods**). In all four CNV strains we find increased levels of gene expression associated with CNV amplified genes. A median of 96% of genes within a CNV have significantly higher mRNA expression than the ancestor (DESeq2, adj.pvalue < 0.05), which is substantially higher than the median proportion of genes (9%) that are increased in expression for unamplified genes in the same strains (**Figure 2**). For RPF abundances we find that 94% of genes show significant increases for CNV amplified genes compared with a median of 8% for unamplified genes in the same strains. By contrast, the proportion of amplified genes with higher protein abundance is only 77% (Welch's t-test, FDR adj.pvalue < 0.05), compared with a median of 10% for unamplified genes. This discrepancy between an increase in transcript and RPF abundance for virtually all CNV genes whereas almost one-quarter of the corresponding proteins are not increased in abundance suggests that regulatory mechanisms act on successive levels of expression to modulate protein expression of CNV amplified genes.

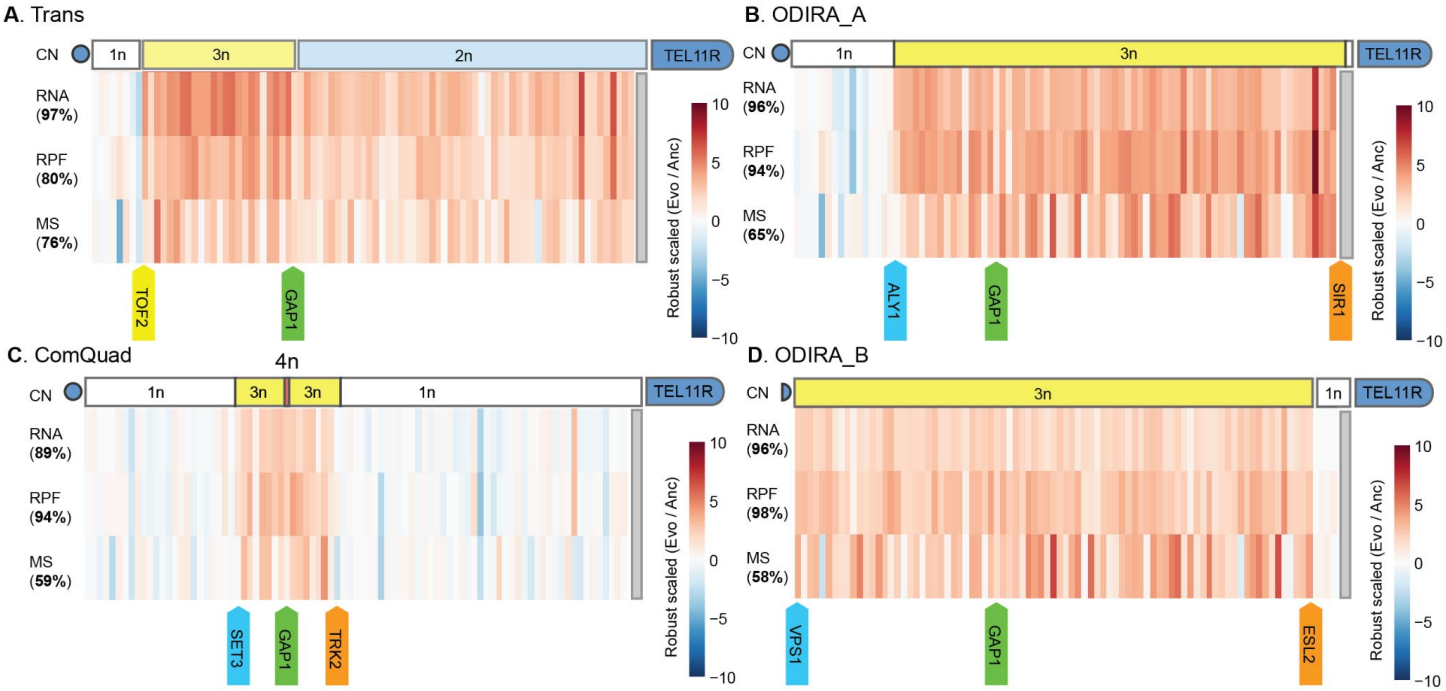

Figure 2. Gene amplification results in increased gene expression at multiple levels. For each strain we show a simplified schematic of the CNV topology of the right arm of ChrXI and the copy-number (CN) gene within the CNV. *GAP1* is highlighted (Green) as well as the genes most proximal to the right (Blue) and left (Orange) CNV breakpoints. The telomere adjacent genes *SIR1*, *FLO10*, *NFT1*, *VBA5* and *GEX2* were excluded due to insufficient data (gray bar). The percentage of amplified genes within the CNV that have significantly higher expression at the level of transcripts (RNA), ribosome footprints (RPF), and proteins (MS) is indicated.

Global patterns of divergence in gene expression

In addition to adaptive CNVs causing changes in gene expression for genes that are located within the amplified region, genes in the rest of the genome that are unaltered in copy number may differ in expression as result of the CNV or other adaptive variation. Therefore, we quantified the global frequency of divergence in gene expression for each evolved strain at each level of gene expression.

Divergence in RNA abundance

In total, 85 of the 91 CNV amplified genes were significantly diverged in RNA abundance (BH adj.p-value < 0.05). All genes that were significantly diverged had increased RNA abundance compared with the ancestor. A small number of genes in the common CNV core including, *DAL80, UTH1*, *PLN1*, *TRK2,* and an additional gene not in the common core, *ECM4*, exhibited increases that were not statistically significant in at least one CNV strain despite being amplified. We also find substantial numbers of unamplified genes throughout the rest of the genome with significant divergence in mRNA abundance relative to the ancestor (median per strain of 1020 genes, (**Table S2, Figure S1**). Of these genes, only 206 were significantly diverged in every background, with 40 being consistently increased and 156 being consistently decreased in abundance. These 206 genes are significantly enriched in gene ontologies associated with metabolism, stress, transport, and ER / protein-folding (**Figure S1**).

Divergence in Ribosome (RPF) abundance

Of the 91 CNV amplified genes, 77 were significantly diverged in RPF abundance (BH adj.p-value < 0.05) and exhibited increased RPF abundance compared with the ancestor. A total of 13 genes were not significantly diverged in RPF abundance in at least one strain despite being amplified including *TIF1, RPS21A,* and genes within the common core including *DAL80*, *UTH1*, and *GAP1*. Genome-wide, we find a substantial number of unamplified genes are diverged in RPF abundance (median per strain of 724 genes, (**Table S2, Figure S2**). Of these genes, only 95 are significantly divergent in every background, with 22 being consistently increased and 69 being consistently decreased in RPF abundance. These 95 genes are significantly enriched in gene ontologies associated with metabolism (**Figure S2**).

Divergence in Protein abundance

We found that 44 of 91 CNV amplified genes were significantly diverged in protein abundance (Welch's t-test, BH adj.pval < 0.05) and had higher protein abundance than the ancestor (**Table S3**). However, 27 genes exhibit less than significant divergence in protein abundance in at least one strain they are amplified in, including *TOF2, LAS1, ALY1,* and members of the common core including *DAL80*, *UTH1*, and *GAP1*. We evaluated changes in protein abundance of unamplified genes. We find the median number of genes in each strain with significantly divergent protein abundances (Welch's t-test, BH adj.pval < 0.05) from the ancestor is ~573, with 85 genes being significantly divergent in every strain, of which 43 are consistently higher and 41 are consistently lower. These 85 genes are enriched with gene ontologies associated with metabolism as well as mitochondria and ATP synthesis (**Figure S3**).

The discrepancy between the patterns of divergence at the RNA, RPF, and protein level are noteworthy as our experimental design is likely to result in increased statistical power for mass spectrometry data, which were analyzed in five biological replicates compared with RNAseq and RPF analysis. Despite this, we identified a smaller proportion of CNV amplified genes that result in increased protein expression than the proportion that result in increased RNA and RPF abundance suggesting that DC mechanisms act primarily at the level of protein abundance.

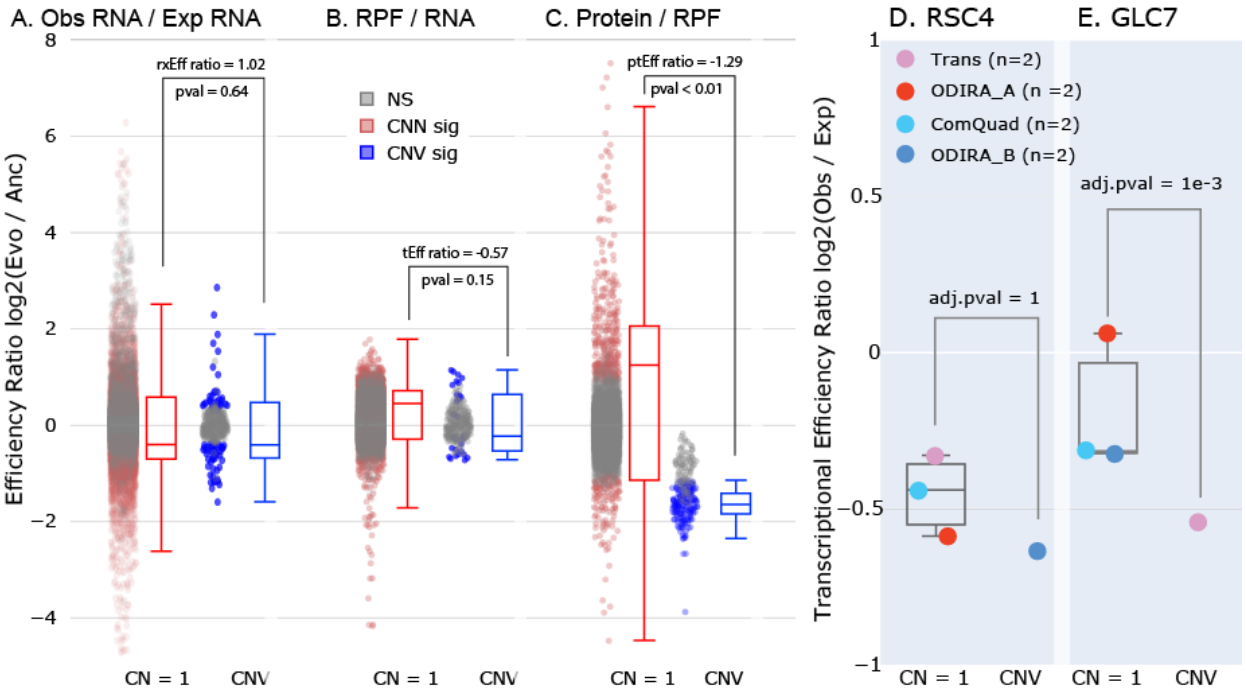

Figure 3. Gene expression effects at multiple levels of regulation. We evaluated changes in gene expression by estimating the expected abundance of each level of gene expression and comparison with the observed values to generate a ratio metric. The resulting log2 fold-change ratio metric between the evolved and ancestral strain is shown on the vertical axis. Genes are separated into CNV amplified genes (**blue**) and unamplified (**red**) genes. Box plots show the distribution for genes with significantly different ratios. (**A**) The Observed to Expected RNA ratio of CNV amplified and unamplified genes is nearly identical (Kruskal-Wallis (KW), ratio = 1.0, pvalue = 0.64). (**B**) The RPF to RNA ratio of CNV amplified and unamplified genes does not differ significantly (KW, ratio = -0.6, pvalue = 0.15). (**C**) The Protein to RPF ratio of CNV amplified genes is significantly reduced compared with unamplified genes (KW, ratio = -1.3, pvalue < 0.01). (**D**) Although CNV amplified genes can have significantly lower ratios than the ancestor, this is not necessarily evidence of dosage compensation. For example, *RSC4* does not have significantly lower Observed RNA / Expected RNA ratios in strains when amplified versus when unamplified. (**E**) Conversely, *GLC7* shows evidence of dosage compensation as has significantly lower Observed RNA to Expected RNA ratios in the CNV strain (Trans) than the other evolved CNV strains in which it is unamplified.

Gene expression ratios as measures of gene regulation

Observed RNA to Expected RNA ratio

Typically, in differential expression analysis the copy-number of the underlying genes are assumed to be uniform throughout the genome and between genotypes. As that is not the case in our study, we first sought to correct for changes in copy-number (**Figure 1B, blue**). Here, we define a null model in which the expected RNA abundance for genes within a CNV is determined by the abundance in the ancestor multiplied by the copy-number in the evolved strain. For example, a twofold increase in gene copy-number is expected to result in a two-fold increase in mRNA. We determined the copy number corrected gene expression for all genes and then compared the observed RNA abundance to this expected value (**methods**, **Table S4**), (**Figure 3A**).

In total, more than half (54 of 91) of the CNV amplified genes were found to have significantly different Observed to Expected RNA ratios. Of these 54 CNV amplified genes 27 had significantly lower ratios in at least one strain, whereas 28 had significantly higher, in at least one strain. One gene, *PTR2*, exhibited both increased and decreased ratios in different strains. For the 17 common core genes we find 2 genes (*KAE1*, *SHB17*) have significantly higher ratios in more than one strain whereas 5 genes have significantly lower ratios in more than one (*DAL80*, *UTH1*, *PLN1*, *TRK2*, and, notably, *GAP1*) (**Table S2**). Interestingly, we find *GAP1*, the hypothesized driver of CNV formation, to have lower ratios in all evolved strains, although this reduction is

only significant in two of the CNV strains (Trans and ODIRA_A). *DAL80*, a transcription factor that is the negative regulator of *GAP1* (Daugherty et al. 1993), has significantly lower ratios in all evolved strains.

We found that 2,934 unamplified genes had significantly different Observed to Expected RNA ratios (**Figure 3A**, BH adj.p-value < 0.05) in at least one evolved strain (**Figure S5**). 1450 of these were significantly higher in at least one strain and 1635 were lower. Of these 366, (12.4%) were significantly different in all evolved strains and are enriched for gene functions associated with metabolism, stress response, chaperone cofactor-dependent protein refolding, and ER organization.

RPF to RNA ratios

To investigate the consequences of CNVs at the level of translation regulation we normalized the ribosome abundance, as measured by RPFs aligned to mORFs only, by transcript abundance. This translation efficiency ratio is an estimate of the amount of translating ribosomes per transcript (Ingolia et al. 2009). Changes in translation efficiency are associated with altered translation initiation rates at the start codon as well as differences in elongation rates throughout the CDS. In addition, differential usage of uORFs can result in changes in ribosome association with transcripts.

In total, 16 CNV amplified genes were found to have significantly different RPF to RNA ratios. Notably, only one of these, *PAM17*, had increased RPF to RNA ratios whereas the other 15 CNV amplified genes exhibited decreased ratios (**Table S5**). Of these, *GAP1* is the only one of the 18 genes in the common core of CNV amplified genes with lower ratios in every evolved strain (median ratio of 0.82 or 1.22-fold less than the ancestor).

For the unamplified genes, we found that 802 genes had significantly different RPF to RNA ratios (**Figure 3B**, FET, p-value < 0.05) relative to the ancestor, 288 (36%) of these had significantly lower ratios than the ancestor, while the majority (549, 68%) had significantly higher ratios than the ancestor (**Table S5**). Of the 802 genes that had significantly different ratios, only 65 were significantly different from the ancestor in all evolved strains. These genes were enriched in functions associated with ribosome assembly and translational fidelity in addition to metabolism (**Figure S6**).

Protein to RPF ratios

Differences in the ratio of protein to RPF abundance within the mORF can be evidence of post-translational regulation, likely the result of changes in protein stability (Ishikawa et al. 2020; Senger et al. 2022). We find that most common CNV amplified genes (16 of 17) had significantly lower Protein to RPF ratios in at least one strain (**Table S6**). This includes *GAP1*, which had significantly lower ratios in ODIRA_B and ComQuad (median 2.6-fold less than the ancestor) but not Trans and ODIRA_A (median 1.4-fold less). Similar to RPF to RNA ratios, we find a significantly lower Protein to RPF ratios in 19 of 91 CNV amplified genes, compared to 98 of 4209 unamplified genes (FET, 11-fold higher, pval < 1e-10).

We find 275 unamplified genes (**Figure 3C**) have significantly different Protein to RPF ratios between the evolved and the ancestor strains (FET, pval <0.05, **Table S6**). Of the 275 genes with significantly different ratios only 21 were significantly different from the ancestor in all evolved strains. These genes were enriched in ontologies associated with metabolism and catabolism (**Figure S7**).

Dosage Compensation

We sought to identify evidence of dosage compensation of amplified genes. Because all evolved strains have adapted to both glutamine-limited growth media as well as the CNVs themselves, a given gene may have deviations in gene expression regardless of the copy-number. One example of this is RSC4 (**Figure 3E**) which is only amplified in ODIRA_B but has lower Observed to Expected RNA ratios in every evolved strain and

significantly so in both ODIRA_A and ComQuad (DESeq2, adj.pval <= 0.05). Conceptually, dosage compensation should require that the expression efficiency of a gene, when amplified, is significantly lower than when unamplified. An example of this is GLC7 (**Figure 3E**), which is only amplified in ODIRA_B but has significantly lower Observed to Expected RNA ratios when amplified compared with the unamplified copy in the other evolved strains.

To identify genes with potential transcription dosage compensation we calculated the median Observed to Expected RNA ratios of all genes when amplified versus when unamplified in evolved strains. Seven genes out of 91 (~7.7%), *DAL80*, *UTH1*, *TIF1*, *GLC7*, *PLN1*, *RPL40B* and *GAP1* had significantly lower ratios (FET, BH adj.pval<0.05) when amplified, consistent with dosage compensation at the transcriptional level. Only one gene, *PCK1*, had significantly higher ratios when amplified (**Table S7**).

To identify genes with potential translational dosage compensation we calculated the median RPF to RNA ratios of each gene when amplified and when unamplified in evolved strains. Three genes out of 91 (~3.3%), *RPS26B*, *RPS21A*, and *GAP1* had significantly lower ratios (FET, BH adj.pval<0.05) when amplified, consistent with dosage compensation at the translational level. Notably, two genes, *PLN1* and *RPL40B* had significantly higher ratios when amplified, potentially counteracting the decreased transcriptional efficiency (**Table S8**).

To identify genes with potential post-translational dosage compensation we calculated the median Protein to RPF ratio of each gene when amplified and when unamplified (FET, BH adj-pval < 0.05). We found 43 genes out of 91 (~47.3%) had significantly lower ratios when amplified. Notably, *GAP1*, the putative target of selection, does have lower ratios, although it is not significant. We did not find any amplified genes with significantly higher ratios (**Table S9**).

Differential positional analysis of ribosome footprint profiles identifies altered translational landscape

In addition to quantifying ribosome abundances, ribosome profiling also captures ribosome positions. Identifying alterations in ribosome positions can be used to identify translational regulatory features, such as the presence of upstream open reading frames (uORFs), N-terminal extensions (NTEs) to existing ORFs, and potential *de novo* genes in the form of novel ORFs (**Figure 3**).

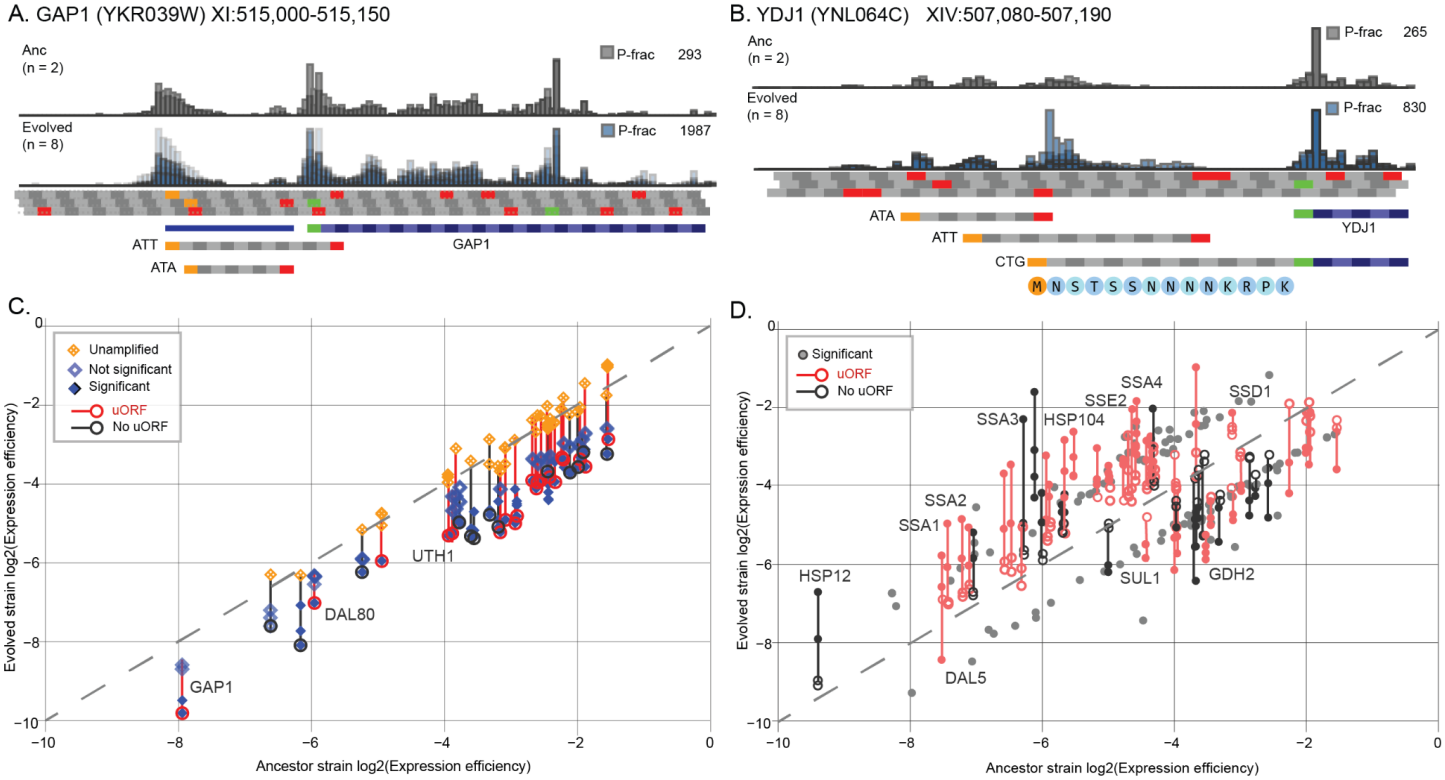

Figure 4 - Ribosome profiling and alternative translational regulation. (**A**) Upstream open reading frames (uORFs) are cis-regulatory elements that can alter the translational efficiency of genes. Using ribosome p-site fractions (vertical axis) that corresponded with AUG (green box), and near-cognate codon (NCC, orange box), translation initiation sites. Here, we show two potential NCC-uORFs upstream of *GAP1*. We find the ratio of ribosomes (median TPM normalized RPF per strain) within the potential uORF region and the main ORF (mORF) to be significantly lower for all evolved strains compared with the ancestor (1.1-fold lower, BH adj.p-value = 0.03). (**B**) Translation initiation sites that are upstream and in-frame to the mORF, and lack an intervening stop codon, can act as N-terminal extensions (NTEs) adding peptides to the N-terminus of the protein. Here, we show a potential NTE upstream of *YDJ1* and the corresponding uniquely matching peptides identified by mass spectrometry. To evaluate the potential role of uORF regulation on gene expression we compared the protein-to-RNA between the ancestor (horizontal axis) and evolved strains (vertical axis) for (**C**) CNV amplified genes and (**D**) unamplified genes. Figure shows select significant genes.

uORFs

Previously published research has identified potential uORFs upstream of several genes that are amplified by CNVs in one or more strains, including *GAP1*, *DAL80*, and *UTH1* (Zhang and Dietrich 2005; Cvijović et al. 2007; Spealman et al. 2018; May et al. 2023). Using ribosome profiling data from each strain we manually identified potential uORFs in 49 of the 109 CNV amplified genes. This frequency is not significantly different from the observed background rate of 84 of 212 genes on the unamplified arm of ChrXI (FET, 1.1-fold higher, p-value = 0.63) (**Table S10**), (**FIGURE 4A**). This suggests that genes within CNVs are not significantly enriched for uORFs and their presence is independent of the CNV itself.

NTEs

Using manual analysis of ribosome profiling data we identified only one N-terminal extension (NTE) associated with a CNV amplified gene (*YKR045C*) , and three NTEs for genes in the unamplified arm of ChrXI (*APE1*, *MRPL31*, and *ZRT3*). Although our mass spectrometry library was not designed to be enriched for NTEs we did identify 4 genes (*PRC1*, *YDJ1*, *ADE12*, *NOB1*) with peptides that uniquely map to non-AUG NTEs (see example **FIGURE 4B**). We also identified an NTE in *YBP2*, although this is most likely a mis-annotated AUG start codon (Eisenberg et al. 2020). Of these eight genes, only *YDJ1* has been described previously as having an NTE (Eisenberg et al. 2020) albeit under different growth conditions. *YDJ1* is a *HSP40* co-chaperone that

enables unfolded protein binding and is involved in the high-molecular-weight complex (HMC) with the stress associated *SSA1* and *SSA2* proteins (Hon et al. 2001). *YDJ1*, *SSA1*, and *SSA2* all exhibit significantly higher expression efficiencies.

Differential expression efficiency

One caveat with the use of ribosome abundance metrics is that not all ribosomes are actively translating or translating at the same speed. Indeed, in addition to ribosome queuing, ribosome collisions, and translation pause sites, elongation rates are part of the regulatory dynamic that responds to cellular and environmental signals (Guzikowski et al. 2022). To best encapsulate these changes, we calculated the ratio of protein to RNA for each gene with the goal of inferring regulatory modulation at both the translational and post-translational levels.

We find that 84 of 91 amplified genes have significantly different Protein to RNA ratios compared to the ancestor and the vast majority (82 or 90%) of these have lower ratios than the ancestor (**Figure 4C**). For the unamplified genes we found 411 genes with significantly different ratios (**Figure 3D**, p-value < 0.05) relative to the ancestor (**Figure 4D**). Of these, 178 (44%) of these had significantly lower ratios than the ancestor, whereas 228 (56%) had significantly higher ratios than the ancestor (**Table S11**).

Enrichment of uORFs in genes with significantly different expression efficiency

Translation efficiency can be altered by uORF mediated translational regulation (Ingolia et al. 2009; Chew et al. 2016; Moro et al. 2021; Spealman et al. 2018). We found that genes with significant RPF to RNA ratios were also significantly enriched in uORFs (Hypergeometric test (HGM), 1.3-fold enriched, pval = 1e-4), suggesting that changes in RPF to RNA ratios could be a product of uORF mediated translational regulation.

This motivated us to test if there was similar enrichment of uORFs in genes with altered overall expression of Protein to RNA. Indeed, we found similar enrichment of predicted uORFs in genes with significantly different Protein to RNA ratios (HGM, 1.3-fold enriched, pval = 1e-4).

While verifying that the predicted uORFs were occupied by ribosomes, we observed many previously unreported uORFs. To better understand the effect uORFs had on expression we manually evaluated all 91 CNV amplified genes (**Figure 4C**) and all 411 genes with significantly different Protein to RNA ratios (**Figure 4D**) for the presence of uORFs. We found that CNV amplified genes with significantly different ratios are enriched for uORFs (FET, 1.9-fold, pval=9.7e-3). Similarly, we found enrichment of uORFs in the 411 genes with significantly different ratios (FET, 2.0-fold, pval=1.2e-4). This is consistent with uORF mediated regulation acting to alter the protein output for a given transcript level.

Notably, CNV amplified genes with uORFs are not significantly enriched in any gene ontology terms, but non-CNV amplified genes with significantly different expression efficiencies and uORFs are significantly enriched in genes associated with metabolism and nitrogen metabolism but also stress response pathways and, intriguingly, protein folding. The stress response and protein folding pathways are particularly interesting as they include *HSP26*, *HSP31*, *HSP32*, *HSP42*, *HSP78*, *HSP82*, and *HSP104* as well as *SSA1*, *SSA2*, *SSA4*, *SSC1*, *SSE1*, *SSE2* and *YDJ1*. This suggests that uORFs mediated translational regulation may be an important factor adaptation to stress, including both nutrient limitation stress and gene amplification stress.

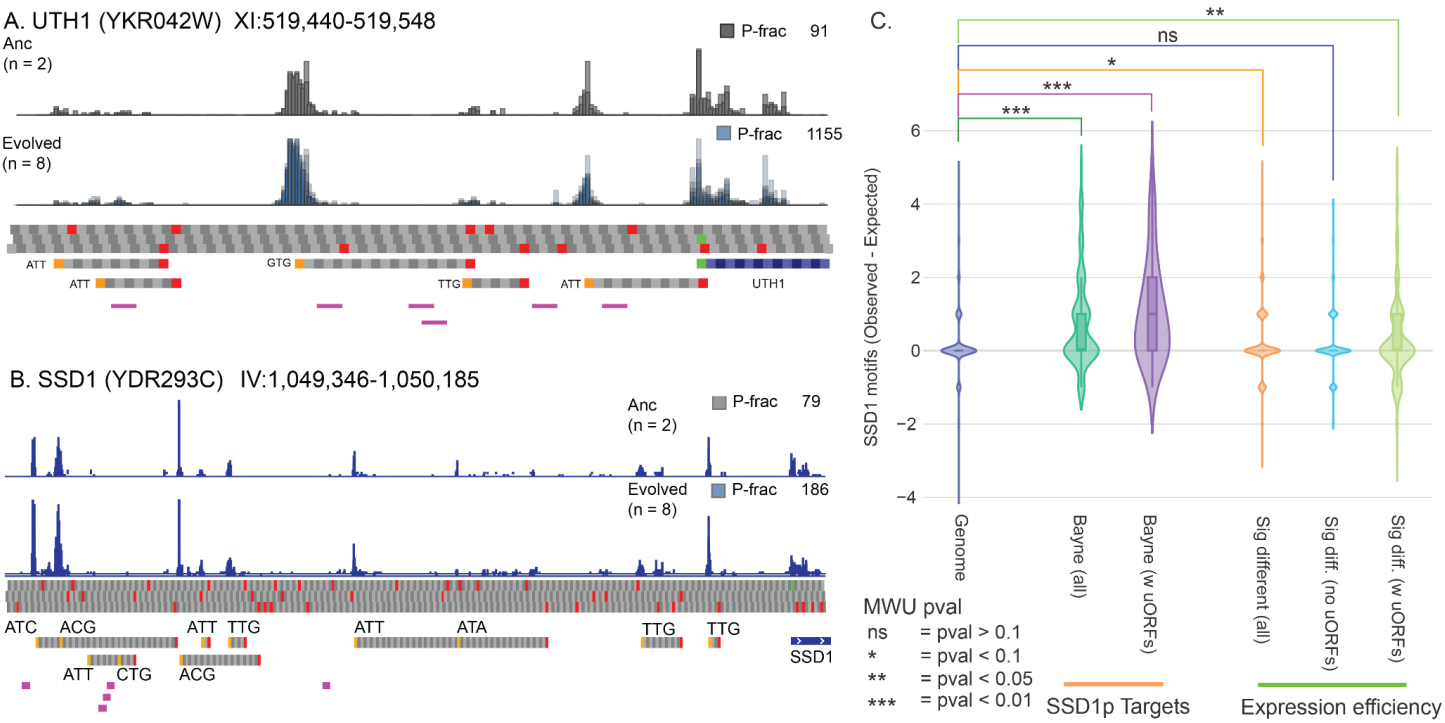

Figure 5. Co-occurrence of SSD1p binding motifs and uORFs. Enrichment of the SSD1p binding motifs upstream of (**A**) *UTH1* and (**B**) *SSD1*. The vertical axis shows the abundance of p-site fractions, which are shown across the length of the transcript leader (horizontal axis). Potential uORFs are indicated beneath the genome track with the start codon indicated for each uORF. SSD1p motifs are indicated by purple bars. Both genes have significant differences in Protein to RNA ratios, relative to the ancestor and have ribosome occupied uORFs. (**C**) In comparing the difference between observed and expected number of SSD1p motifs (Bayne et al.) genes with significantly different Protein to RNA ratios or uORFs are significantly enriched in SSD1p motifs (MWU, pval < 0.05).

RNA binding protein *SSD1* potentially regulates translation by associating with uORFs

Recently, a conserved SSD1p RNA binding motif (CNYUCNYU) was identified using an analysis of high confidence binding sites in optimal and heat-shocked conditions (Bayne et al. 2022). To better understand the role of *SSD1* may play in gene regulation in evolved CNV strains we evaluated genes significantly enriched in the CNYUCNYU motif.

One well known target of SSD1p is *UTH1*, a mitochondrial protein involved in regulating both mitochondrial biogenesis and degradation (Camougrand et al. 2004), the overexpression of *UTH1* has been shown to lead to cell death in yeast strains lacking SSD1 (Camougrand et al. 2003) and is associated with the role of *SSD1* in cellular response to aneuploid stress (Hose 2020). Indeed, we find that *UTH1* is amplified in every evolved strain and has lower Protein to RNA ratios in every evolved strain. A manual analysis of the ribosome profiling data found a large abundance of uORFs and numerous CNYUCNYU motifs in *UTH1* (FET, 3-fold enriched, pval = 0.01) (**Figure 5A)**.

Similar co-occurrences of uORFs and CNYCNYU motifs were observed for several well described *SSD1* targets including *CLN2* (Ohyama et al. 2010) *SUN4*, *CTS1*, *DSE2*, *NSA3*, *SRL1*, *SCW4*, *SCW10*, and *CCW12* (Bayne et al. 2022). This co-occurrence was also observed in genes identified by a SSD1 KO study conducted in an aneuploid background, such as *SCW4*, *SCW10*, *CCW12*, and *TOS1* (Hose 2020). Notably, we also found that *SSD1* itself appears to have extensive uORFs and *SSD1* motifs, suggesting it is subject to self-regulation (**Figure 5B**). We also find that it exhibits higher Protein to RNA ratios in every evolved strain, suggesting that post-transcriptional regulation may be involved. Indeed, SSD1 is transcriptionally repressed in all evolved

strains, but translationally upregulated in all evolved strains resulting in a final protein abundance slightly higher than the ancestor, consistent with translational regulation.

Using the CNYUCNYU motif, we scanned the annotated TLs of the genome to identify genes enriched in the SSD1 motif identifying 185 significantly enriched genes (FET, pval < 0.05) of which 130 were present in our data set. To test if these SSD1 motifs, in co-occurrence with uORFs, could be associated with changes in Protein to RNA ratios we compared the difference in the number observed expected SSD1p sites based on length (**Figure 5C**). SSD1p targets (Bayne et al.) are significantly enriched (MWU, pval = 5.7e-12), the subset of those with uORFs even more so (MWU, pval = 3.5e-15). Similarly, genes with significantly different Protein to RNA ratios are weakly enriched (MWU, pval = 0.096), the subset of those genes without uORFs are not significantly different from the background (MWU, pval = 0.82), while the subset with uORFs is significantly enriched (MWU, pval = 0.034). Taken together this suggests that SSD1 may be driving altered expression efficiency through uORF mediated translational regulation, affecting both CNV amplified and unamplified genes.

Discussion

In this paper we set out to characterize how gene expression changes at the transcriptional, translational, and protein level in response to CNV associated adaptations to selection. To do this we quantified different aspects of gene expression in5 strains of yeast, 4 of which had undergone adaptation to glutamine-limited growth conditions for hundreds of generations and had gained copy-number variants (CNVs), as well as an ancestral strain unadapted to glutamine-limited growth and lacking any CNVs (Figure 1).

In a comparison of gene expression using differential expression analysis (DESeq2 and Perseus), we observed that the majority of genes amplified by CNVs had significantly higher levels of expression than their ancestral counterparts (**Figure 2**) which agrees with previous reports of on the broad effects of CNVs (Springer et al. 2010; Hose et al. 2015; Gasch et al. 2016; Avecilla and Spealman et al. 2023). We also observed that the number of significantly divergent genes (including both amplified and unamplified) decreases at each successive level of expression, from a median of ~1500 per strain for RNA, to ~1100 for ribosome abundance, to ~600 per strain for protein abundance. Again, this agrees with previous reports of buffering of divergent gene expression at the translational (McManus et al. 2014; Wang and Chen 2019; Kusnadi et al. 2022) and post-translational levels (Wang et al. 2018; Cheng et al. 2022; Jiang et al. 2023).

We also sought to identify regulatory modulation of gene expression using ratios of abundance across the levels of gene expression between the evolved strain and ancestor (**Figure 3**). Using these ratios of abundance, we found that numerous genes had significant differences in their RPF to RNA ratios, Protein to RPF ratios, or Protein to RNA, relative to the ancestor. By correcting for changes in gene expression un-associated with CNV amplification, we were able to show that dosage compensation occurs at each level of gene expression, potentially most strongly at the level of protein abundance where it is wide spread and not limited to genes involved in protein complexes, in agreement with recent reports of DC in natural isolates of yeast with adaptive aneuploidy (Muenzner et al. 2022) and humans cancers (Schukken and Sheltzer 2022). It is worth noting however that the current method of identifying dosage compensation may be particularly underpowered for evaluating changes in transcription and translation as only two replicates exist of each, conversely measures utilizing the MS data has 5 replicates and appears to be of much greater sensitivity.

An analysis of genes with significantly different Protein to RNA ratios found them to be significantly enriched with upstream open reading frames (uORFs, **Figure 4A,B**) suggesting that these translational regulatory elements may be playing a role in the post-transcriptional regulation of gene expression (Ingolia et al. 2009; Chew et al. 2016; Moro et al. 2021). A manual analysis of uORFs was performed for all CNV amplified genes

and all genes with significantly different Protein to RNA ratios (**Figure 4C,D**). We found that CNVs were not specifically enriched in uORFs; but all genes with significantly different Protein to RNA ratios were significantly enriched, regardless of being CNV amplified or not. This suggests that global changes in gene expression in response to adaptation to nutrient-limitation and CNV amplification may be translationally mediated by uORFs.

Finally, in the course of our uORF analysis of genes with significantly different Protein to RNA ratios we identified several well-known targets of the RNA-binding protein SSD1 (**Figure 5B**). *SSD1*, has previously been shown to be a stress responsive RBP under glucose starvation (Bresson et al. 2020), heat shock (Bayne et al. 2022) and quiescence (Miles et al. 2019). It is capable of altering translation of bound RNA (Hogan et al. 2008; Jansen et al. 2009; Ohyama et al. 2010), potentially by preferentially binding to TLs and interfering with ribosome scanning (Bayne et al. 2022). To determine if *SSD1* may play a role in altering expression efficiencies, we first identified 130 genes in our data set that had significant enrichment of the conserved SSD1p binding motif. We found the greatest enrichment of SSD1p binding motifs in the subset of SSD1p binding targets with uORFs and the subset of genes with significantly different Protein to RNA ratios with uORFs (**Figure 5C**). These findings suggest that SSD1p may act as a stress responsive trans-acting factor driving uORF mediated translational regulation. Furthermore, these findings help contextualize the role of SSD1 in CNV and aneuploid yeast strains. Previous studies have shown a much higher level of gene specific dosage compensation in aneuploid yeast strains with a functional copy of *SSD1* (such as the S288c strain, as was used in this study) compared to yeast strains lacking a functioning *SSD1* (such as W303). This suggests that the importance of *SSD1* in mitigating the fitness defects of CNVs may be a consequence of the broader role *SSD1* plays in the translational regulation of gene expression under stress, and not specific to CNVs *per se*. Indeed, this may explain some of the observed but varying and incomplete overlap seen in SSD1 targets, genes involved in the hypo-osmotic stress pathway (Bayne et al. 2022), CAGE (Tsai et al. 2019), and analyses that sought to identify the CNV and aneuploid stress response at the transcriptional level (Avecilla and Spealman et al. 2023).

Our study is not without its limitations. Notably, the number of strains evaluated, and the number of replicates limit the scope of our inferences and the generalizability of our results. Beyond that, technical limitations of TMT-labeling (eg. ratio compression) makes some analysis difficult and prevents a straightforward fold-change analysis of protein abundance. Furthermore, the reliance on a manual, heuristic approach for uORF identification limits the pool of genes evaluated to only those of the highest relevance (eg. CNV amplified genes, genes with significantly different Protein to RNA ratios, and all genes on the left arm of chromosome XII as a background control).

That said, this work presents several avenues for ongoing research, such as the potential for RBPs to interact with uORFs, and thereby alter downstream gene expression. While this is not unprecedented (Medenbach et al. 2011; Spealman et al. 2018) but it is largely unexplored. We also observe a shift in expression from an early “protein divergent” phase to a late “transcription / translation divergent” phase. If we consider the number of generations diverged from the ancestor, the generation 150 strains (Trans and ODIRA_A) to have we 2.4-fold more significantly diverged genes at the level of protein abundance relative to the generation 250 strains (ComQuad, ODIRA_B) (average of 1078 genes versus 444). Conversely, at the level of transcription this is reversed with the generation 250 strains having more significantly diverged genes relative to the generation 150 (1740 versus 546) with similar divergence seen at the translational level (1381 versus 361). While it is difficult to meaningfully extrapolate from such a small number of observations it supports previous observations of rapidly evolving regulatory networks acting to mitigate protein imbalances (King and Wilson 1975; Kafri et al. 2016; Aubé et al. 2023; Jiang et al. 2023).

Methods

Strains and growth media

Briefly, the strains evaluated in this paper were originally published in Lauer et al. 2018, they include both long-term experimentally evolved strains as well as their ancestor. The naming convention in this paper is as follows: Ancestor (DGY1657); Trans (DGY1726); ODIRA_A (DGY1735); ComQuad (DGY1741); and ODIRA_B (DGY1743). Trans and ODIRA_A are from population samples after ~150 generations, ComQuad and ODIRA_B are from populations samples after ~250 generations. All strains are derived from FY4 (BY1747) and modified by the inclusion of a GFP and KanMX reporter cassette. All evolved strains were evolved in miniature chemostats under minimal glutamine growth media for either 150 or 250 generations, as described previously. For this study, strains were struck out on YEPD plates from -80°C freezer stocks and grown for 2 days at 30°C. Single colonies were selected based on the presence of fluorescence produced from the previously described mCitrine reporter (Lauer et al. 2018).

Illumina genomic DNA sequencing

For short-read Illumina sequencing of strains we relied on previously generated data (Lauer et al. 2018, NCBI SRA accession SRP142330). Briefly, we performed genomic DNA extraction (Hoffman and Winston 1987), followed by quantification using SYBR Green I and standardized to 2.5 ng/µL. Libraries were constructed using Illumina Nextera tagmentation (Baym et al. 2015). Final concentrations were measured using SYBR Green I, fragment sizes were measured using Agilent TapeStation 2200 before being balanced and pooled. Illumina DNA libraries were sequenced on an Illumina NextSeq 500 using 2x75 paired-end protocol.

Illumina genomic DNA alignment and genotyping

All sequences were aligned against the Ensembl_R64.1.50 reference genome. We aligned reads using bwa mem (v0.7.17, (Li and Durbin 2009)) and generated BAM files using samtools (v1.14, (Danecek et al. 2021)). FASTQ files for all sequencing are available from the SRA (accession SRP142330).

Potential SNV and indel identification of each strain was performed using GATK's HaplotypeCaller (ver 4.1.9.0, (Van der Auwera et al. 2013)) in single-sample mode and annotated using Ensembl VEP (release 107, (McLaren et al. 2016)). Potential structural and copy-number variants using Illumina paired end data were identified using our custom analysis tool CVish (ver 1.x, (Lauer et al. 2018)).

Nanopore genomic DNA sequencing, alignment, and genotyping

All yeast strains were grown to greater than 1 × 10^7 cells/mL in 300 mL Glutamine minimal media. Genomic DNA from each strain was extracted using Qiagen 20/G Genomic tips from ∼1.5 × 10^9 cells using the manufacturer's protocol.

All genomic DNA was barcoded using Oxford-Nanopore's native barcoding genomic DNA kit (EXP-NBD104), adapters were added using the ligation sequencing kit (SQK-LSK109). The manufacturer's protocol (versions NBE_9065_v109_revB_23May2018 and NBE_9065_V109_revP_14Aug2019) was followed with the following exceptions: incubation times for enzymatic repair step were increased to 15 min. All Agencourt AMPure XP beads were incubated for 30 min at 37°C before elution. Adapter ligation time was increased to 10 min. Multiplexed libraries were loaded on MinION flowcells (FLO-MIN106D R9) and run on a MinION sequencer (MIN-101B) (Quinlan and Hall 2010).

All sequences were aligned against the Ensembl_R64.1.50 reference genome (Cunningham et al. 2022). ONT long-read sequences were aligned using minimap2 (2.22, (Li 2018)), with variant detection using sniffles2 (2.0.6, (Li 2018; Sedlazeck et al. 2018)), potential ODIRA CNV breakpoints were evaluated with mugio (v1.7, (Spealman et al. 2020)). DNA depth across genomes was calculated per nucleotide using bedtools 'genomecov' (ver 2.29.2, (Quinlan and Hall 2010)).

DNA reads and depth estimation:

We identified ~15.4% of genes in the ancestral strain had long-read DNA read-depth not near one copy. Some of these are known deletions (MATALPHA) or are regions with known properties of known copy number expansion (CUP1 region, ENA region, and ASP region) or transposons. (**Table S12**) To prevent the over or underestimation of expected

expression in evolved strains, we manually analyzed the locus and flanking regions for signs of CNV breakpoints using both short and long-reads. If no evidence was found for a breakpoint the locus was assigned a copy-number of 1 (**Table S13)**. (Supplement Materials "**Gene copy number estimation using long and short reads")**

Growth conditions and cell harvesting:

Colonies were grown overnight on YEPD then used to inoculate 500 mL of minimal glutamine growth media in chemostats. These were then grown to saturation (approximately 48 hours ~1e7 cells / mL, measured using Coulter Counter) at 30°C in aerobic conditions. After saturation was achieved chemostats were switched to continuous mode with an inflow rate of fresh minimal glutamine growth media at a rate of 0.12 hour / L (corresponding to a population doubling time of ~5.8 hours). This was maintained for 24 hours before the cells were harvested. This was performed in duplicate for each strain for the simultaneous extraction of RNA and RPFs. This was performed for 5 replicates for the generation of material for mass spectrometry.

Cells were harvested following previously described methods (Spealman et al. 2016). Briefly, cycloheximide was added to the media to a final concentration of 100µg/L and incubated for 2 min, before being separated from the media using rapid vacuum filtration using Millipore 0.22µm filters. Cells were resuspended from filters using polysome lysis buffer with cycloheximide and then flash frozen immediately in liquid nitrogen before being transferred and stored at -80°C. RNAseq and ribosome profiling was performed by TB-SEQ, Inc. (Palo Alto, CA). TMT-labeling and LC-MS was performed by Proteomics Laboratory, NYU Langone Medical Center (New York, NY).

Mass spectrometry and analysis

We used 16Plex TMT labeling to enable pooling of peptides from 16 samples (Ch1-16) for offline fractionation and LC-MS/MS detection and quantification. Five replicates were generated for each strain and the total run was performed in two batches will be 2 batches ("A","B"). Ch1 was used as the common reference standard, an equimolar mix of all the samples, labeled with TMT Ch1. All samples were lysed using a TFA approach (Doellinger et al. 2020). Both batches were fractionated by offline HPLC chromatography using reverse phase C18 stationary phase at high pH. Peptides from collected fractions were separated by online HPLC on C18 at low pH coupled to the MS instrument.

Spectra were analyzed using MaxQuant (v2.1.0.0, (Tyanova et al. 2016a)). Settings and parameters are described in the supplemental materials and the mqpar.xml file is provided (**File S1**). Differential abundance analysis is performed using Perseus (v1.6.15.0, (Tyanova et al. 2016b)). Data preprocessing and normalization procedures were performed as described (Yu et al. 2020) are detailed in the supplemental materials. The statistical test for significance in differential abundance is Welch's t-test of the difference between the normalized protein group abundances in each strain relative to the ancestor. Significance here is defined as q-value < 0.05 using BH multiple hypothesis test correction (ie. BH adj.pval < 0.05), (**Table S3**). Volcano plots indicating the protein abundances of CNV in each strain are in the Supplemental section.

Expression profiling and analysis:

RNA and RPFs were aligned using STAR (ver 2.7.6a, (Dobin et al. 2013)) filtering for known non-polyadenylated ncRNA (snRNA, rRNA, and tRNA) using Ensembl_R64.1.1 ncRNA fasta (**File S2**). Counts per gene were calculated using bedtools (v2.29.2, (Quinlan and Hall 2010)) coverage -counts -s -b against Saccharomyces_cerevisiae.R64-1-1.50.gff3. Raw read count tables are available for both RNA (**Table S14**) and RPF (**Table S15**).

For quality control purposes we set an arbitrary expression threshold. Genes were required to not be transposon element associated genes, have at least 100 aligned RNA reads, 10 RPFs, and be detected by MS. This reduced our set to 4236 genes in total.

Expected RNA abundances were calculated using a custom script that multiplies the observed abundance in ancestor by the copy-number of that gene in the evolved strain, these are available as supplemental files (**Table S16**) Differential abundance analysis was performed on both unnormalized observed and unnormalized expected RNAseq reads using DESeq2 (v1.6.3, (Love et al. 2014)). Results of DESeq2 are available as supplemental files (**File S3**). Significance here is defined as BH adj.pval < 0.05.

Gene ontology term (GO term) enrichment was calculated using Yeastmine (Balakrishnan et al. 2012) with significance cut off at BH adj-pvalue < 0.05. Redundant terms were then reduced using REVIGO (Supek et al. 2011)

using the Tiny option, UNIPROT *Saccharomyces cerevisiae* S288c database, and SimRel semantic similarity measure.

All Venn diagrams were originally visualized using Venny (v2.1.0). All expression track data was visualized using IGV browser (v2.9.4, (Robinson et al. 2011)).

Robust Scaling

For the purposes of visualization, sequencing depth differences for CDS aligned reads from RNAseq and ribosome profiling were normalized using TPM. Log2 transformed TPM ratios were calculated for each strain relative to the ancestor. To compare between RNA, RPF, and the run-relative normalized intensities of detected peptides we scaled the data using RobustScaler (SciKit Learn, ver1.2.2).

Differential expression ratio analysis

Conceptually, an increase in ribosome protected fragment (RPF) abundances over a region could be caused by an increase in the amount of ribosomes per transcript or by an increase in the transcript abundance itself. Translation efficiency (RPF to RNA ratios) attempts to deconvolute these options by normalizing the RPF abundances by the RNA abundances (Ingolia et al. 2009). It is important to note that increased translation efficiency does not equate to increased protein abundance. While both the relative proportion of RNA (median *rho* = 0.29) and RPF (median *rho* = 0.33) in the evolved strain relative to the ancestor correlate with the relative difference in protein abundance, the same is not true for the translational efficiency (median *rho* = 0.03) which exhibits a large degree of difference from the final protein abundance (**Figure S8**).

Protein to RPF ratios are an extension of this methodology that normalizes the protein intensity by the RPF abundance. Similarly, the ratio of protein abundance to RNA abundance, effectively an empirically derived inversion of the RNA-to-Protein coefficient (Edfors et al. 2016). For the purposes of comparing between replicates and strains, in all cases, ratios are calculated using depth normalized values, either TPM (RNA, RPF) or the MaxQuant reporter intensity corrected value (Tyanova et al. 2016a).

To determine if the ratio of abundances of any given gene was different between strains, we calculated the ratios in each strain as a proportional ratio of medians. For example, (RPF-evolved / RNA-evolved) / (RPF-ancestor / RNA-ancestor), these were then tested using Fischer Exact Test (FET), Benjamani-Hochberg (BH) was used for multiple hypothesis correction of all p-values.

Calculation of Dosage Compensation

When allowing for ‘incomplete’ DC or dosage attenuation it is important to consider that gene expression can change regardless of the copy number of the gene. This is especially critical in organisms that have had time to adapt to multiple selective pressures, one of which may be the CNV. As such, it behooves us to consider dosage compensation to be a recurrent, but identifiable trend. That is, that the expression of an amplified gene shows a recurring, measurable, difference when compared to the expression of an unamplified paralog.

To calculate this, we first binned each gene based on whether it was a CNV amplified or not. We then generated a proportional ratio of medians for the two bins: eg. (RPF-CNV / RNA-CNV) / (RPF-no_CNV / RNA-no_CNV), these were then tested using Fischer Exact Test (FET), Benjamani-Hochberg (BH) was used for multiple hypothesis correction of all p-values. For determination of transcriptional DC, we performed the same analysis using the observed RNA abundance compared to the expected RNA abundance: eg. (RNA-CNV_observed / RNA-CNV_expected) / (RPF-no_CNV_observed / RNA-no_CNV_expected).

Selection of uORFs from previous studies

Because uORFs have been known to play a role in translational regulation we sought to determine if previously identified uORF may be present in our data. We calculated the mean TL length of genes containing either ‘predicted’ uORFs from (Spealman et al. 2018) or ‘validated’ uORFs from (May et al. 2023). Because of the validation methodology used in (May et al. 2023), uORFs greater than 180bp distance from the main ORF TIS were not tested. We calculated the mean length of SGD annotated TLs for each set (**Table S17**). We found the predicted uORFs were significantly longer than the background TL length (MWU, 1.16-fold larger, pval < 1e-10) while the validated uORFs were shorter, (MWU,

1.35-fold shorter, pval = 0.71). Comparing genes with significantly different RPF to RNA ratios (McManus et al. 2014) to these two sets of uORFs we found that the genes were significantly enriched in 'predicted uORFs' in the two sets produced different results; being significantly enriched (HGM, 1.3-fold higher, pval = 0.0) in the predicted uORFs. Conversely, they were significantly de-enriched (HGM, 1.4 lower pval = 0.0) in the validated uORFs. This is in agreement with the observation that longer TLs can contain a greater number of regulatory elements, potentially necessary for fine-grained regulation (Leppek et al. 2018).

For the purposes of this study we restrict ourselves to the 'predicted' uORFs (Spealman et al. 2018). Unlike our manual analysis using this uORF set we determine uORF regulation genome wide.

Differential positional analysis for the identification of uORFs and NTEs

Because translation of uORFs is known to be condition sensitive (Hinnebusch 2005 ; Moro et al. 2021) and no previous study has been conducted using glutamine-limited growth media or CNVs, we sought to identify them using our ribosome profiling data. For each gene in each strain that exhibited a significant difference in Protein to RNA ratio relative to the ancestor we looked for the presence of ribosome loading with the transcript leader (TL or 5' UTR). Candidate uORFs were required to have ribosomes inframe to potential alternative translation initiation sites (aTISs) be they canonical AUG start codons or NCC start codons. We used p-site fractionation (Spealman et al. 2018) to determine ribosomal reading frame, followed by start codon identification. As some genes (for example, UTH1, SSD1, SSA4, etc.) had potential uORFs in complex, overlapping, or nested configurations we summed across the TL from the most distal TIS to one codon upstream of the CDS.

We evaluated all CNV amplified genes (e.g., ChrXI right arm) and all 411 genes with significantly different Protein to RNA ratios. As a background control we also evaluated all genes located on ChrXI left arm. We then sought to compare our results to the previously reported (e. "predicted") uORFs. As both the CNV genes and the 411 genes were selected for evaluation they were withheld from the comparison to prevent bias. Comparing the identification on ChrXI left arm we found a 34% agreement (30 out of 88) with 3 uORFs (3%) being only previously identified and 55 (63%) being unique to this study.

SSD1 motif analysis

To identify genes enriched in *SSD1* motifs in their TLs we first generated a list of TLs for all genes using the Yeastmine (Balakrishnan et al. 2012) annotated 'Five Prime UTR' (Pelechano et al. 2013) or from (Spealman et al. 2018), taking the longest of either recorded instance. Combined these encompassed 6119 genes, 6080 of which exceeded the length of the motif. We converted these to fasta using bedtools 'getfasta' (ver 2.29.2, (Quinlan and Hall 2010)). These were then scanned for perfect matches to the CNYUCNYU motif (Bayne et al. 2022). The background rate genome wide was calculated per nucleotide as 1.19e-3, the background rate using only TLs was 1.48e-3 per nucleotide. Significance was calculated using FET, using the proportional ratio of (observed hits of TL / length of TL) versus (Total hits / Total length). We found that of the 186 genes that were significantly enriched in SSD1 motifs only 130 were present in our data set at or above our QC threshold (1e2).

Acknowledgements

We thank members of the Gresham lab for their insight and feedback. This work was made possible by grants from the National Institutes of Health (National Institute of General Medical Sciences) to P.S. (F32-GM131573) and D.G. (R01-GM134066).

Works cited.

Ascencio D, Diss G, Gagnon-Arsenault I, Dubé AK, DeLuna A, Landry CR. 2021. Expression attenuation as a mechanism of robustness against gene duplication. *Proc Natl Acad Sci U S A* **118**. http://dx.doi.org/10.1073/pnas.2014345118.

Aubé S, Nielly-Thibault L, Landry CR. 2023. Evolutionary trade-off and mutational bias could favor transcriptional over translational divergence within paralog pairs. *PLoS Genet* **19**: e1010756. http://dx.doi.org/10.1371/journal.pgen.1010756.

Avecilla G and Spealman P, Matthews J, Caudal E, Schacherer J, Gresham D. 2023. Copy number variation alters local and global mutational tolerance. *Genome Res*. http://dx.doi.org/10.1101/gr.277625.122.

Balakrishnan R, Park J, Karra K, Hitz BC, Binkley G, Hong EL, Sullivan J, Micklem G, Cherry JM. 2012. YeastMine--an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. *Database* **2012**: bar062. http://dx.doi.org/10.1093/database/bar062.

Baym M, Kryazhimskiy S, Lieberman TD, Chung H, Desai MM, Kishony R. 2015. Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* **10**: e0128036. http://dx.doi.org/10.1371/journal.pone.0128036.

Bayne RA, Jayachandran U, Kasprowicz A, Bresson S, Tollervey D, Wallace EWJ, Cook AG. 2022. Yeast Ssd1 is a non-enzymatic member of the RNase II family with an alternative RNA recognition site. *Nucleic Acids Res* **50**: 2923–2937. http://dx.doi.org/10.1093/nar/gkab615.

Ben-David U, Amon A. 2020. Context is everything: aneuploidy in cancer. *Nat Rev Genet* **21**: 44–62. http://dx.doi.org/10.1038/s41576-019-0171-x.

Bonney ME, Moriya H, Amon A. 2015. Aneuploid proliferation defects in yeast are not driven by copy number changes of a few dosage-sensitive genes. *Genes Dev* **29**: 898–903. http://dx.doi.org/10.1101/gad.261743.115.

Bresson S, Shchepachev V, Spanos C, Turowski TW, Rappsilber J, Tollervey D. 2020. Stress-Induced Translation Inhibition through Rapid Displacement of Scanning Initiation Factors. *Mol Cell* **80**: 470–484.e8. http://dx.doi.org/10.1016/j.molcel.2020.09.021.

Brockdorff N, Turner BM. 2015. Dosage Compensation in Mammals. *Cold Spring Harb Perspect Biol* **7**: a019406. http://cshperspectives.cshlp.org/content/7/3/a019406.abstract (Accessed September 6, 2023).

Camougrand N, Grelaud-Coq A, Marza E, Priault M, Bessoule J-J, Manon S. 2003. The product of the UTH1 gene, required for Bax-induced cell death in yeast, is involved in the response to rapamycin. *Mol Microbiol* **47**: 495–506. http://dx.doi.org/10.1046/j.1365-2958.2003.03311.x.

Camougrand N, Kiššová I, Velours G, Manon S. 2004. Uth1p: a yeast mitochondrial protein at the crossroads of stress, degradation and cell death. *FEMS Yeast Res* **5**: 133–140. https://academic.oup.com/femsyr/article-pdf/5/2/133/17937835/5-2-133.pdf (Accessed February 22, 2022).

Chang K-J, Wang C-C. 2004. Translation initiation from a naturally occurring non-AUG codon in Saccharomyces cerevisiae. *J Biol Chem* **279**: 13778–13785. http://dx.doi.org/10.1074/jbc.M311269200.

Cheng P, Zhao X, Katsnelson L, Camacho-Hernandez EM, Mermerian A, Mays JC, Lippman SM, Rosales-Alvarez RE, Moya R, Shwetar J, et al. 2022. Proteogenomic analysis of cancer aneuploidy and normal tissues reveals divergent modes of gene regulation across cellular pathways. *Elife* **11**. http://dx.doi.org/10.7554/eLife.75227.

Chew G-L, Pauli A, Schier AF. 2016. Conservation of uORF repressiveness and sequence features in mouse, human and zebrafish. *Nat Commun* **7**: 11663. http://dx.doi.org/10.1038/ncomms11663.

Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R, et al. 2022. Ensembl 2022. *Nucleic Acids Res* **50**: D988–D995. http://dx.doi.org/10.1093/nar/gkab1049.

Cvijović M, Dalevi D, Bilsland E, Kemp GJL, Sunnerhagen P. 2007. Identification of putative regulatory upstream ORFs in the yeast genome using heuristics and evolutionary conservation. *BMC Bioinformatics* **8**: 295. http://dx.doi.org/10.1186/1471-2105-8-295.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* **10**. http://dx.doi.org/10.1093/gigascience/giab008.

Daugherty JR, Rai R, el Berry HM, Cooper TG. 1993. Regulatory circuit for responses of nitrogen catabolic gene expression to the GLN3 and DAL80 proteins and nitrogen catabolite repression in Saccharomyces cerevisiae. *J Bacteriol* **175**: 64–73. http://dx.doi.org/10.1128/jb.175.1.64-73.1993.

Dekel E, Alon U. 2005. Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**: 588–592. http://dx.doi.org/10.1038/nature03842.

Dephoure N, Hwang S, O'Sullivan C, Dodgson SE, Gygi SP, Amon A, Torres EM. 2014. Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. https://elifesciences.org/articles/03023 (Accessed August 28, 2023).

Dhami MK, Hartwig T, Fukami T. 2016. Genetic basis of priority effects: insights from nectar yeast. *Proc Biol Sci* **283**. http://dx.doi.org/10.1098/rspb.2016.1455.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. http://dx.doi.org/10.1093/bioinformatics/bts635.

Doellinger J, Schneider A, Hoeller M, Lasch P. 2020. Sample Preparation by Easy Extraction and Digestion (SPEED) - A Universal, Rapid, and Detergent-free Protocol for Proteomics Based on Acid Extraction. *Mol Cell Proteomics* **19**: 209–222. http://dx.doi.org/10.1074/mcp.TIR119.001616.

Edfors F, Danielsson F, Hallström BM, Käll L, Lundberg E, Pontén F, Forsström B, Uhlén M. 2016. Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol Syst Biol* **12**: 883. http://dx.doi.org/10.15252/msb.20167144.

Eisenberg AR, Higdon AL, Hollerer I, Fields AP, Jungreis I, Diamond PD, Kellis M, Jovanovic M, Brar GA. 2020. Translation Initiation Site Profiling Reveals Widespread Synthesis of Non-AUG-Initiated Protein Isoforms in Yeast. *Cell Syst* **11**: 145–160.e5. http://dx.doi.org/10.1016/j.cels.2020.06.011.

Gasch AP, Hose J, Newton MA, Sardi M, Yong M, Wang Z. 2016. Further support for aneuploidy tolerance in wild yeast and effects of dosage compensation on gene copy-number evolution. *Elife* **5**: e14409. http://dx.doi.org/10.7554/eLife.14409.

Gebauer F, Corona DFV, Preiss T, Becker PB, Hentze MW. 1999. Translational control of dosage compensation in Drosophila by Sex-lethal: cooperative silencing via the 5′ and 3′ UTRs of msl-2 mRNA is independent of the poly(A) tail. *EMBO J* **18**: 6146–6154. https://onlinelibrary.wiley.com/doi/abs/10.1093/emboj/18.21.6146 (Accessed August 28, 2023).

Golzio C, Katsanis N. 2013. Genetic architecture of reciprocal CNVs. *Curr Opin Genet Dev* **23**: 240–248. http://dx.doi.org/10.1016/j.gde.2013.04.013.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet* **4**: e1000303. http://dx.doi.org/10.1371/journal.pgen.1000303.

Guzikowski AR, Harvey AT, Zhang J, Zhu S, Begovich K, Cohn MH, Wilhelm JE, Zid BM. 2022. Differential translation elongation directs protein synthesis in response to acute glucose deprivation in yeast. *RNA Biol* **19**: 636–649. http://dx.doi.org/10.1080/15476286.2022.2065784.

Harvey RF, Smith TS, Mulroney T, Queiroz RML, Pizzinga M, Dezi V, Villenueva E, Ramakrishna M, Lilley KS, Willis AE. 2018. Trans-acting translational regulatory RNA binding proteins. *Wiley Interdiscip Rev RNA* **9**: e1465. https://onlinelibrary.wiley.com/doi/abs/10.1002/wrna.1465 (Accessed August 28, 2023).

Hinnebusch AG. 2005. Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* **59**: 407–450. http://dx.doi.org/10.1146/annurev.micro.59.031805.133833.

Hoffman CS, Winston F. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli. *Gene* **57**: 267–272. http://dx.doi.org/10.1016/0378-1119(87)90131-4.

Hogan DJ, Riordan DP, Gerber AP, Herschlag D, Brown PO. 2008. Diverse RNA-binding proteins interact with functionally related sets of RNAs, suggesting an extensive regulatory system. *PLoS Biol* **6**: e255. http://dx.doi.org/10.1371/journal.pbio.0060255.

Hong J, Gresham D. 2014. Molecular specificity, convergence and constraint shape adaptive evolution in nutrient-poor environments. *PLoS Genet* **10**: e1004041. http://dx.doi.org/10.1371/journal.pgen.1004041.

Hon T, Lee HC, Hach A, Johnson JL, Craig EA, Erdjument-Bromage H, Tempst P, Zhang L. 2001. The Hsp70-Ydj1 molecular chaperone represses the activity of the heme activator protein Hap1 in the absence of heme. *Mol Cell Biol* **21**: 7923–7932. http://dx.doi.org/10.1128/MCB.21.23.7923-7932.2001.

Hose J, Escalante LE, Clowers KJ, Dutcher HA, Robinson D, Bouriakov V, Coon JJ, Shishkova E, Gasch AP. 2020. The genetic basis of aneuploidy tolerance in wild yeast. *Elife* **9**. http://dx.doi.org/10.7554/eLife.52063.

Hose J, Yong CM, Sardi M, Wang Z, Newton MA, Gasch AP. 2015. Dosage compensation can buffer copy-number variation in wild yeast. https://elifesciences.org/articles/05462 (Accessed September 26, 2022).

Hull RM, Cruz C, Jack CV, Houseley J. 2017. Environmental change drives accelerated adaptation through stimulated copy number variation. *PLoS Biol* **15**: e2001333. http://dx.doi.org/10.1371/journal.pbio.2001333.

Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223. http://dx.doi.org/10.1126/science.1168978.

Ishikawa K, Ishihara A, Moriya H. 2020. Exploring the Complexity of Protein-Level Dosage Compensation that Fine-Tunes Stoichiometry of Multiprotein Complexes. *PLoS Genet* **16**: e1009091. http://dx.doi.org/10.1371/journal.pgen.1009091.

Ishikawa K, Makanae K, Iwasaki S, Ingolia NT, Moriya H. 2017. Post-Translational Dosage Compensation Buffers Genetic Perturbations to Stoichiometry of Protein Complexes. *PLoS Genet* **13**: e1006554. http://dx.doi.org/10.1371/journal.pgen.1006554.

Itoh Y, Melamed E, Yang X, Kampf K, Wang S, Yehya N, Van Nas A, Replogle K, Band MR, Clayton DF, et al. 2007. Dosage compensation is less effective in birds than in mammals. *J Biol* **6**: 1–15. https://jbiol.biomedcentral.com/articles/10.1186/jbiol53 (Accessed September 6, 2023).

Jansen JM, Wanless AG, Seidel CW, Weiss EL. 2009. Cbk1 regulation of the RNA-binding protein Ssd1 integrates cell fate with translational control. *Curr Biol* **19**: 2114–2120. http://dx.doi.org/10.1016/j.cub.2009.10.071.

Jiang D, Cope AL, Zhang J, Pennell M. 2023. On the Decoupling of Evolutionary Changes in mRNA and

Protein Levels. *Mol Biol Evol* **40**. http://dx.doi.org/10.1093/molbev/msad169.

Jordan W 3rd, Rieder LE, Larschan E. 2019. Diverse Genome Topologies Characterize Dosage Compensation across Species. *Trends Genet* **35**: 308–315. http://dx.doi.org/10.1016/j.tig.2019.02.001.

Kafri M, Metzl-Raz E, Jona G, Barkai N. 2016. The Cost of Protein Production. *Cell Rep* **14**: 22–31. http://dx.doi.org/10.1016/j.celrep.2015.12.015.

Kane AJ, Brennan CM, Xu AE, Solís EJ, Terhorst A, Denic V, Amon A. 2021. Cell adaptation to aneuploidy by the environmental stress response dampens induction of the cytosolic unfolded-protein response. *Mol Biol Cell* **32**: 1557–1564. http://dx.doi.org/10.1091/mbc.E21-03-0104.

King MC, Wilson AC. 1975. Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116. http://dx.doi.org/10.1126/science.1090005.

Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. *Proc Biol Sci* **279**: 5048–5057. http://dx.doi.org/10.1098/rspb.2012.1108.

Krasovec M, Merret R, Sanchez F, Sanchez-Brosseau S, Piganeau G. 2023. A High Frequency of Chromosomal Duplications in Unicellular Algae Is Compensated by Translational Regulation. *Genome Biol Evol* **15**: evad086. https://academic.oup.com/gbe/article-pdf/15/6/evad086/50518048/evad086.pdf (Accessed August 28, 2023).

Kusnadi EP, Timpone C, Topisirovic I, Larsson O, Furic L. 2022. Regulation of gene expression via translational buffering. *Biochim Biophys Acta Mol Cell Res* **1869**: 119140. http://dx.doi.org/10.1016/j.bbamcr.2021.119140.

Kuzmin E, Taylor JS, Boone C. 2022. Retention of duplicated genes in evolution. *Trends Genet* **38**: 59–72. http://dx.doi.org/10.1016/j.tig.2021.06.016.

Larrimore KE, Barattin-Voynova NS, Reid DW, Ng DTW. 2020. Aneuploidy-induced proteotoxic stress can be effectively tolerated without dosage compensation, genetic mutations, or stress responses. *BMC Biol* **18**: 1–19. https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-020-00852-x (Accessed August 28, 2023).

Lauer S, Avecilla G, Spealman P, Sethia G, Brandt N, Levy SF, Gresham D. 2018. Single-cell copy number variant detection reveals the dynamics and diversity of adaptation. *PLoS Biol* **16**: e3000069. http://dx.doi.org/10.1371/journal.pbio.3000069.

Lawless C, Pearson RD, Selley JN, Smirnova JB, Grant CM, Ashe MP, Pavitt GD, Hubbard SJ. 2009. Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics* **10**: 7. http://dx.doi.org/10.1186/1471-2164-10-7.

Leppek K, Das R, Barna M. 2018. Functional 5' UTR mRNA structures in eukaryotic translation regulation and how to find them. *Nat Rev Mol Cell Biol* **19**: 158–174. http://dx.doi.org/10.1038/nrm.2017.103.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. http://dx.doi.org/10.1093/bioinformatics/bty191.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. http://dx.doi.org/10.1186/s13059-014-0550-8.

Makanae K, Kintaka R, Makino T, Kitano H, Moriya H. 2013. Identification of dosage-sensitive genes in Saccharomyces cerevisiae using the genetic tug-of-war method. *Genome Res* **23**: 300–311. http://dx.doi.org/10.1101/gr.146662.112.

May GE, Akirtava C, Agar-Johnson M, Micic J, Woolford J, McManus J. 2023. Unraveling the influences of sequence and position on yeast uORF activity using massively parallel reporter systems and machine learning. *Elife* **12**. http://dx.doi.org/10.7554/eLife.69611.

Mayr C. 2017. Regulation by 3'-Untranslated Regions. *Annu Rev Genet* **51**: 171–194. http://dx.doi.org/10.1146/annurev-genet-120116-024704.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**: 122. http://dx.doi.org/10.1186/s13059-016-0974-4.

McManus CJ, May GE, Spealman P, Shteyman A. 2014. Ribosome profiling reveals post-transcriptional buffering of divergent gene expression in yeast. *Genome Res* **24**: 422–430. http://dx.doi.org/10.1101/gr.164996.113.

Medenbach J, Seiler M, Hentze MW. 2011. Translational control via protein-regulated upstream open reading frames. *Cell* **145**: 902–913. http://dx.doi.org/10.1016/j.cell.2011.05.005.

Miles S, Li LH, Melville Z, Breeden LL. 2019. Ssd1 and the cell wall integrity pathway promote entry, maintenance, and recovery from quiescence in budding yeast. *Mol Biol Cell* **30**: 2205–2217. http://dx.doi.org/10.1091/mbc.E19-04-0190.

Monteuuis G, Miścicka A, Świrski M, Zenad L, Niemitalo O, Wrobel L, Alam J, Chacinska A, Kastaniotis AJ, Kufel J. 2019. Non-canonical translation initiation in yeast generates a cryptic pool of mitochondrial proteins. *Nucleic Acids Res* **47**: 5777–5791. http://dx.doi.org/10.1093/nar/gkz301.

Moro SG, Hermans C, Ruiz-Orera J, Albà MM. 2021. Impact of uORFs in mediating regulation of translation in stress conditions. *BMC Mol Cell Biol* **22**: 29. http://dx.doi.org/10.1186/s12860-021-00363-9.

Muenzner J, Trébulle P, Agostini F, Messner CB, Steger M, Lehmann A, Caudal E, Egger A-S, Amari F, Barthel N, et al. 2022. The natural diversity of the yeast proteome reveals chromosome-wide dosage compensation in aneuploids. *bioRxiv*. http://biorxiv.org/lookup/doi/10.1101/2022.04.06.487392.

Myhre S, Lingjærde O-C, Hennessy BT, Aure MR, Carey MS, Alsner J, Tramm T, Overgaard J, Mills GB, Børresen-Dale A-L, et al. 2013. Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol Oncol* **7**: 704–718. http://dx.doi.org/10.1016/j.molonc.2013.02.018.

Nair S, Miller B, Barends M, Jaidee A, Patel J, Mayxay M, Newton P, Nosten F, Ferdig MT, Anderson TJC. 2008. Adaptive copy number evolution in malaria parasites. *PLoS Genet* **4**: e1000243. http://dx.doi.org/10.1371/journal.pgen.1000243.

Ohyama Y, Kasahara K, Kokubo T. 2010. Saccharomyces cerevisiae Ssd1p promotes CLN2 expression by binding to the 5'-untranslated region of CLN2 mRNA. *Genes Cells* **15**: 1169–1188. http://dx.doi.org/10.1111/j.1365-2443.2010.01452.x.

Paulander W, Andersson DI, Maisnier-Patin S. 2010. Amplification of the Gene for Isoleucyl–tRNA Synthetase Facilitates Adaptation to the Fitness Cost of Mupirocin Resistance in *Salmonella enterica*. *Genetics* **185**: 305–312. http://dx.doi.org/10.1534/genetics.109.113514.

Pelechano V, Wei W, Steinmetz LM. 2013. Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**: 127–131. http://dx.doi.org/10.1038/nature12121.

Pränting M, Andersson DI. 2011. Escape from growth restriction in small colony variants of Salmonella typhimurium by gene amplification and mutation. *Mol Microbiol* **79**: 305–315. http://dx.doi.org/10.1111/j.1365-2958.2010.07458.x.

Prestel M, Feller C, Becker PB. 2010. Dosage compensation and the global re-balancing of aneuploid genomes. *Genome Biol* **11**: 216. http://dx.doi.org/10.1186/gb-2010-11-8-216.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features.

Bioinformatics **26**: 841–842. http://dx.doi.org/10.1093/bioinformatics/btq033.

Rice AM, McLysaght A. 2017. Dosage-sensitive genes in evolution and disease. *BMC Biol* **15**: 78. http://dx.doi.org/10.1186/s12915-017-0418-y.

Robinson D, Vanacloig-Pedros E, Cai R, Place M, Hose J, Gasch AP. 2023. Gene-by-environment interactions influence the fitness cost of gene copy-number variation in yeast. *G3* . http://dx.doi.org/10.1093/g3journal/jkad159.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* **29**: 24–26. http://dx.doi.org/10.1038/nbt.1754.

Samant RS, Masto VB, Frydman J. 2019. Dosage compensation plans: protein aggregation provides additional insurance against aneuploidy. *Genes Dev* **33**: 1027–1030. http://dx.doi.org/10.1101/gad.329383.119.

Schukken KM, Sheltzer JM. 2022. Extensive protein dosage compensation in aneuploid human cancers. *Genome Res* **32**. https://pubmed.ncbi.nlm.nih.gov/35701073/ (Accessed August 28, 2023).

Schultz CW, Preet R, Dhir T, Dixon DA, Brody JR. 2020. Understanding and targeting the disease-related RNA binding protein human antigen R (HuR). *Wiley Interdiscip Rev RNA* **11**: e1581. http://dx.doi.org/10.1002/wrna.1581.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**: 461–468. http://dx.doi.org/10.1038/s41592-018-0001-7.

Selmecki AM, Dulmage K, Cowen LE, Anderson JB, Berman J. 2009. Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genet* **5**: e1000705. http://dx.doi.org/10.1371/journal.pgen.1000705.

Senger G, Santaguida S, Schaefer MH. 2022. Regulation of protein complex partners as a compensatory mechanism in aneuploid tumors. https://elifesciences.org/articles/75526 (Accessed August 28, 2023).

Shao X, Lv N, Liao J, Long J, Xue R, Ai N, Xu D, Fan X. 2019. Copy number variation is highly correlated with differential gene expression: a pan-cancer study. *BMC Med Genet* **20**: 175. http://dx.doi.org/10.1186/s12881-019-0909-5.

Smith JM, Haigh J. 1974. The hitch-hiking effect of a favourable gene. *Genet Res* **23**: 23–35. https://www.ncbi.nlm.nih.gov/pubmed/4407212.

Spealman P, Burrell J, Gresham D. 2020. Inverted duplicate DNA sequences increase translocation rates through sequencing nanopores resulting in reduced base calling accuracy. *Nucleic Acids Res*. http://dx.doi.org/10.1093/nar/gkaa206.

Spealman P, De T, Chuong JN, Gresham D. 2023. Best Practices in Microbial Experimental Evolution: Using Reporters and Long-Read Sequencing to Identify Copy Number Variation in Experimental Evolution. *J Mol Evol* **91**: 356–368. http://dx.doi.org/10.1007/s00239-023-10102-7.

Spealman P, Naik AW, May GE, Kuersten S, Freeberg L, Murphy RF, McManus J. 2018. Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res* **28**: 214–222. http://dx.doi.org/10.1101/gr.221507.117.

Spealman P, Wang H, May G, Kingsford C, McManus CJ. 2016. Exploring Ribosome Positioning on Translating Transcripts with Ribosome Profiling. *Methods Mol Biol* **1358**: 71–97. http://dx.doi.org/10.1007/978-1-4939-3067-8_5.

Springer M, Weissman JS, Kirschner MW. 2010. A general lack of compensation for gene dosage in yeast. *Mol Syst Biol* **6**: 368. http://dx.doi.org/10.1038/msb.2010.19.

Stingele S, Stoehr G, Peplowska K, Cox J, Mann M, Storchova Z. 2012. Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol Syst Biol* **8**: 608. https://onlinelibrary.wiley.com/doi/abs/10.1038/msb.2012.40 (Accessed August 28, 2023).

Straub T, Becker PB. 2007. Dosage compensation: the beginning and end of generalization. *Nat Rev Genet* **8**: 47–57. http://dx.doi.org/10.1038/nrg2013.

Sunshine AB, Payen C, Ong GT, Liachko I, Tan KM, Dunham MJ. 2015. The fitness consequences of aneuploidy are driven by condition-dependent gene effects. *PLoS Biol* **13**: e1002155. http://dx.doi.org/10.1371/journal.pbio.1002155.

Suomi F, Menger KE, Monteuuis G, Naumann U, Kursu VAS, Shvetsova A, Kastaniotis AJ. 2014. Expression and evolution of the non-canonically translated yeast mitochondrial acetyl-CoA carboxylase Hfa1p. *PLoS One* **9**: e114738. http://dx.doi.org/10.1371/journal.pone.0114738.

Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* **6**: e21800. http://dx.doi.org/10.1371/journal.pone.0021800.

Taggart JC, Zauber H, Selbach M, Li G-W, McShane E. 2020. Keeping the Proportions of Protein Complex Components in Check. *Cell Syst* **10**: 125–132. http://dx.doi.org/10.1016/j.cels.2020.01.004.

Tang H-L, Yeh L-S, Chen N-K, Ripmaster T, Schimmel P, Wang C-C. 2004. Translation of a yeast mitochondrial tRNA synthetase initiated at redundant non-AUG codons. *J Biol Chem* **279**: 49656–49663. http://dx.doi.org/10.1074/jbc.M408081200.

Todd RT, Selmecki A. 2020. Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *Elife* **9**. http://dx.doi.org/10.7554/eLife.58349.

Tsai H-J, Nelliat AR, Choudhury MI, Kucharavy A, Bradford WD, Cook ME, Kim J, Mair DB, Sun SX, Schatz MC, et al. 2019. Hypo-osmotic-like stress underlies general cellular defects of aneuploidy. *Nature* **570**: 117–121. http://dx.doi.org/10.1038/s41586-019-1187-2.

Tyanova S, Temu T, Cox J. 2016a. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **11**: 2301–2319. http://dx.doi.org/10.1038/nprot.2016.136.

Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M, Cox J. 2016b. The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**: 731–740. http://dx.doi.org/10.1038/nmeth.3901.

Uebbing S, Konzer A, Xu L, Backström N, Brunström B, Bergquist J, Ellegren H. 2015. Quantitative Mass Spectrometry Reveals Partial Translational Regulation for Dosage Compensation in Chicken. *Mol Biol Evol* **32**: 2716–2725. http://dx.doi.org/10.1093/molbev/msv147.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* **43**: 11.10.1–11.10.33. http://dx.doi.org/10.1002/0471250953.bi1110s43.

Veitia RA, Bottani S, Birchler JA. 2008. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. *Trends Genet* **24**: 390–397. http://dx.doi.org/10.1016/j.tig.2008.05.005.

Veitia RA, Bottani S, Birchler JA. 2013. Gene dosage effects: nonlinearities, genetic interactions, and dosage compensation. *Trends Genet* **29**: 385–393. http://dx.doi.org/10.1016/j.tig.2013.04.004.

Wagner A. 2005. Energy constraints on the evolution of gene expression. *Mol Biol Evol* **22**: 1365–1374. http://dx.doi.org/10.1093/molbev/msi126.

Wang S, Chen Y. 2019. Fine-Tuning the Expression of Duplicate Genes by Translational Regulation in and Maize. *Front Plant Sci* **10**: 534. http://dx.doi.org/10.3389/fpls.2019.00534.

Wang SH, Hsiao CJ, Khan Z, Pritchard JK. 2018. Post-translational buffering leads to convergent protein expression levels between primates. *Genome Biol* **19**: 83. http://dx.doi.org/10.1186/s13059-018-1451-z.

Yu S-H, Kyriakidou P, Cox J. 2020. Isobaric Matching between Runs and Novel PSM-Level Normalization in MaxQuant Strongly Improve Reporter Ion-Based Quantification. *J Proteome Res* **19**: 3945–3954. http://dx.doi.org/10.1021/acs.jproteome.0c00209.

Zhang Z, Dietrich FS. 2005. Identification and characterization of upstream open reading frames (uORF) in the 5' untranslated regions (UTR) of genes in Saccharomyces cerevisiae. *Curr Genet* **48**: 77–87. http://dx.doi.org/10.1007/s00294-005-0001-x.

Zhang Z, Presgraves DC. 2017. Translational compensation of gene copy number alterations by aneuploidy in Drosophila melanogaster. *Nucleic Acids Res* **45**: 2986–2993. http://dx.doi.org/10.1093/nar/gkx106.

Zitomer RS, Walthall DA, Rymond BC, Hollenberg CP. 1984. Saccharomyces cerevisiae ribosomes recognize non-AUG initiation codons. *Mol Cell Biol* **4**: 1191–1197. http://dx.doi.org/10.1128/mcb.4.7.1191-1197.1984.