

# Autoregulation of Many Yeast Ribosomal Proteins Discovered by Efficient Search for Feedback Regulation

Basab Roy<sup>1</sup>, David Granas<sup>1</sup>, Fredrick Bragg Jr., Jonathan A. Y. Cher, Michael A. White, Gary D. Stormo

Department of Genetics and Center for Genome Sciences and Systems Biology

Washington University School of Medicine, St Louis, MO 63110

<sup>1</sup> co-first authors

## Abstract

Post-transcriptional autoregulation of gene expression is common in bacterial systems but many fewer examples are known in eukaryotes. We used the yeast collection of genes fused to GFP as a rapid screen for examples of feedback regulation in ribosomal proteins by overexpressing a non-regulatable version of a gene and observing the effects on the expression of the GFP-fused version. We tested 95 ribosomal protein genes and found that 21 of them showed at least a three-fold repression. Two genes, *RPS22B* and *RPL1B*, showed over a 10-fold repression. In both cases the cis-regulatory segment resides in the 5' UTR of the gene as shown by placing that segment of the mRNA upstream of GFP alone and demonstrating it is sufficient to cause repression of GFP when the protein is over-expressed. Further analyses showed that the intron in the 5' UTR of *RPS22B* is required for regulation, presumably because the protein inhibits splicing that is necessary for translation. The 5' UTR of *RPL1B* contains a sequence and structure motif that is conserved in the binding sites of Rpl1 orthologs from bacteria to mammals, and mutations within the motif eliminate repression.

## Introduction

Feedback regulation is common in many biological processes. In metabolic pathways the end product can often inhibit one of the enzymes in the pathway to set the flux through the pathway at the appropriate level<sup>1,2</sup>. Gene expression is also often regulated by feedback mechanisms. This commonly occurs with a transcription factor that regulates its own expression, referred to as autoregulation. One of the first transcription factors to be studied in detail, the lambda repressor, was found to be autoregulated both positively and negatively, allowing it to maintain its *in vivo* concentration in a narrow range<sup>3,4</sup>. Autoregulation has been found to be among the most common network motifs in bacterial transcription<sup>5,6</sup>. Studies on the regulatory network in yeast also identify many examples of autoregulation<sup>7-9</sup>. Mathematical analyses have characterized the properties and advantages of autoregulatory networks<sup>10-12</sup>.

Although less well-studied, autoregulation also occurs for proteins involved in post-transcriptional steps of gene expression. For example, many splicing factors regulate their own expression<sup>13-17</sup>. Recently developed methods for high-throughput analysis of RNA-protein interactions have identified many RNA-binding proteins, some of which are associated with their own mRNAs<sup>18-25</sup>. By itself that doesn't prove they are autoregulatory, but it seems likely to be a consequence of such binding. In most of those cases, both transcription factors and RNA-binding proteins, the normal function of the protein is to bind DNA or RNA and often to regulate gene expression. The fact that they can regulate their own expression is not surprising given the advantages of such feedback processes, it only requires that the gene's own

DNA or RNA be included in the target list for the protein. Binding sites are often short and, because unconstrained nucleic acids can evolve rapidly, sites that offer a selective advantage are likely to be obtained through random mutagenesis processes.

There are also examples of proteins whose primary function is not in gene regulation but have a secondary role in regulating their own expression. Many of these are proteins that bind to RNA, but whose primary functions are not involved in controlling gene expression. For example, most of the ribosomal proteins in *E. coli* are subject to feedback regulation<sup>26,27</sup>. Ribosomal proteins are expressed as part of transcription units (operons) composed of other ribosomal proteins. Autoregulation by one of the proteins in the operon is typically sufficient to control the expression of all the other genes in the operon by translational coupling, where translation of an upstream gene in an operon is required for translation of the downstream genes. The ribosomal proteins are all RNA-binding proteins, having as their primary target the rRNAs of the ribosome. To become autoregulatory, the mRNA simply has to evolve a sequence that is a molecular mimic of the primary target site, but with lower affinity so that binding of the rRNA is saturated before the regulatory site becomes bound by the protein<sup>27-30</sup>. There are also examples of tRNA synthetase genes in bacteria and yeast that have evolved a similar regulatory site, where the mRNA mimics the tRNA that the synthetase gene normally binds to, but with lower affinity<sup>31-35</sup>. Particularly interesting are cases where the protein's normal function does not involve binding to RNA but it is found in screens for RNA-binding proteins<sup>36,37</sup>. In some cases, proteins with alternative primary functions have been shown to be direct regulators of their own translation<sup>38-42</sup>. Such examples highlight the enormous functional capacity of RNA where it can become a sensor of the cellular environment and autonomously regulate its own fate, as exemplified in cases where no protein is required, such as riboswitches<sup>43-47</sup>. When the effector being recognized by the mRNA is its own gene product, the result is autoregulation of gene expression.

The fundamental characteristic of feedback regulation of gene expression is that if the activity of a gene product, which is usually proportional to its concentration, is higher than the set point, or "desired" level for the cell, then its expression is reduced, and conversely, if the activity is too low, the expression is increased. This relationship is true regardless of the mechanism by which the feedback regulation occurs, whether it involves a complex network of interactions or is simply the result of direct autoregulation by the gene product itself. Once examples of feedback regulation of gene expression are obtained, the mechanism can be determined by additional experiments. The collection of yeast strains with genes fused to green fluorescent protein (GFP)<sup>48</sup> provides an excellent resource to screen for examples of feedback regulation. By introducing into those strains an inducible copy of a gene for the same protein, but lacking all potential cis-regulatory elements (cre-less), an observed reduction in the level of GFP after induction indicates some feedback mechanism controlling the expression of the wild-type gene. Further analyses are required to determine the step in the expression process that is being regulated, whether it is transcription initiation, any of the processes leading to the mature mRNA, any step in translation, or even post-translational enhancement of protein degradation. We are particularly interested in finding new examples of post-transcriptional regulation of protein expression so our initial focus is on ribosomal proteins, which are commonly translationally autoregulated in bacteria but<sup>27,28,49,50</sup> for which many fewer examples are known in yeast.

Ribosome synthesis in yeast is subject to feedback regulation in part by alternative functions of ribosomal proteins<sup>51-55</sup>. There are several examples of post-transcriptional autoregulation by yeast ribosomal proteins, most often through inhibition of splicing necessary for protein expression. For

example, *RPL22B*, *RPS14B* and *RPL30* all have introns within the N-terminus of the coding sequence and splicing is inhibited by binding of the encoded protein<sup>56-59</sup>. Remarkably, the ortholog of Rpl30 in the archaeon *Sulfolobus acidocaldarius* can bind to the same mRNA target and inhibit splicing<sup>60</sup>. *RPS9A* and *RPS9B* both have introns within the N-terminus of the coding region and both genes are subject to feedback regulation by inhibition of splicing<sup>61</sup>. The orthologs of Rps9 are involved in autoregulation in several other eukaryotic species and even in bacteria<sup>27,61</sup>. *RPS28B* does not contain an intron but is autoregulated by a different mechanism where binding of the Edc3 decapping enzyme to the 3' UTR is regulated by the Rps28 protein, leading to mRNA degradation<sup>62,63</sup>. These cases are all consistent with examples from bacteria where ribosomal protein synthesis is regulated post-transcriptionally, and it seems likely that a directed search for feedback regulation among yeast ribosomal protein genes could uncover more examples, leading us to utilize the yeast GFP-fusion collection.

## Results and Discussion

Feedback regulation of protein expression requires that when the activity of the protein, usually proportional to its concentration, is higher than the homeostasis point of the cell, its expression is reduced, and when the activity is lower than that point, its expression is increased. This allows the cell to maintain expression in a narrow range around its set point. The collection of yeast genes fused to GFP<sup>48</sup> provides an excellent resource to screen for genes that exhibit feedback regulation. A version of the gene is synthesized that is lacking any potential cis-regulatory elements (a cre-less version of the gene), under the control of an inducible promoter. The cre-less version is synthesized with alternative 5' and 3' UTRs, mCherry is fused to the C-terminus in place of the GFP of the wild-type gene, any introns are removed, and the codons of the gene are shuffled<sup>64</sup> to maintain the wild-type protein sequence while altering the mRNA sequence sufficiently that we expect any cis-regulatory elements that overlap the coding sequence would be eliminated. We use the *GAL1* promoter for induction of the cre-less gene (Figure 1a). If there is feedback regulation, overexpression of the cre-less gene (monitored by mCherry fluorescence) will bind to the regulatory elements of the wild-type gene, leading to a decrease in its expression, which is monitored by GFP fluorescence. Identification of feedback regulated genes does not provide information about the mechanism of action, and further analyses are required to determine the step in the expression process that is regulated. The scheme for gene synthesis and fluorescent detection is summarized in supplementary figure S1.

The pyruvate decarboxylase gene *PDC1* is transcriptionally autoregulated<sup>65</sup>. To test our strategy, we synthesized a cre-less copy of *PDC1*, fused to mCherry and under control of the *GAL1* promoter. Figure 1b shows the change in the fluorescence signal of both mCherry and GFP after 10 hours of induction, when the GFP signal is reduced about 10-fold. Note that there is a subset of cells that do not induce mCherry fluorescence, which results in the shoulder seen on the GFP fluorescence signal. To simplify measurements of the change in expression, in all further examples we use the measurement of GFP after induction only from cells that show induction of the mCherry signal.

### Screen of ribosomal proteins

We are primarily interested in identifying cases of post-transcriptional autoregulation, examples of which are common among ribosomal genes in bacteria<sup>27,29,30,50</sup>. Among the known examples of yeast ribosomal genes that are post-transcriptionally autoregulated, repressing proper splicing of the pre-

mRNA is often the mechanism<sup>56-61</sup>. Many yeast ribosomal proteins have paralogs which are identical, or nearly so, to each other and for those cases it is sufficient to make a cre-less version from only one of the paralogs and to test the effects on expression of both wild-type (with GFP-fusion) paralogs. GFP expression of 60 large subunit ribosomal protein genes, including 25 paralogous pairs, both with and without induction of the cre-less gene, are shown in Figure 2a. Figure 2b shows the same for 35 small subunit ribosomal protein genes, including 16 paralogous pairs, and the fluorescence signal from control cells lacking a GFP-fusion gene. Shown are the median of the log (fluorescence) values of two or more measurements with the standard deviations also shown, which are generally quite small. Figure 2c shows both the induced and uninduced measurements in the same plot for all of the genes. Two genes, *RPL1B* and *RPS22B*, have greater than one log (>10-fold) decrease in expression after induction (marked with \*\* and above the 1-log dotted line), and another 19 genes have greater than ½ log change in expression (marked with \*). One pair of paralogs, *RPL26A* and *RPL26B*, are both decreased by greater than ½ log, and a few others have similar changes of both paralogs, such as *RPL24A* and *RPL24B*, but only one exceeds the ½ log threshold. Cases where both paralogs are decreased may represent examples of increased rates of protein degradation of the overexpressed proteins<sup>52</sup>. But in most cases only one paralog has a significant reduction in expression and is therefore likely to indicate direct repression of one gene's expression. Of the known examples described above, we see a large reduction in expression of *RPL22B*. We did not see a large reduction in expression for *RPS14B*, but it is expressed at very low levels, consistent with previous reports that the ratio of *RSP14A* to *RSP14B* is 10:1<sup>56</sup>. We also did not see a large reduction in expression of *RPS28B*, likely because its regulation requires the 3' UTR which is disrupted in the GFP-fusion genes<sup>62,63</sup>. This indicates one limitation of our approach, that the GFP fusion to the C-terminus of the protein disrupts the normal 3' UTR, and cis-regulatory elements residing in that region will likely be missed. *RPL30* and *RPS9A/B* are missing from our GFP collection and could not be tested. The two genes we observe with the largest effects, *RPS22B* and *RPL1B*, have not, to our knowledge, been previously shown to be autoregulated, nor have most of the regulated examples we observe.

### Autoregulation of *RPS22B*

Upon induction of the cre-less version of *RPS22* (the two paralogs code for identical proteins), the expression of wild-type *RPS22B* (with GFP fusion) is reduced over 10-fold at 10 hours while *RPS22A* has only a modest reduction (Fig. 2b). The *RPS22A* gene has no introns whereas the *RPS22B* gene has two introns, including one in the 5' UTR that contains a conserved, predicted secondary structure<sup>66</sup>. That intron contains seven AUG codons and is a substrate for RNase III-mediated cleavage if not spliced<sup>67</sup>, both of which suggest that splicing is required for translation of the mRNA. In fact, deletion of the 5' UTR intron increases expression of the *RPS22B* gene several-fold<sup>68</sup>. The simplest hypothesis is that the 5' UTR intron of the *RPS22B* gene is the cis-regulatory site required for autoregulation, probably via inhibition of splicing. To test that we integrated the gene for GFP expressed from the constitutive *TEF2* promoter, whose activity is unaffected by galactose induction<sup>69</sup>, into the yeast chromosome in place of a putative gene of unknown function (chromosome II: YBR032W). Two different 5' UTRs were placed upstream of GFP, the complete *RPS22B* UTR and a "post-spliced" *RSP22B* UTR with the intron removed. Both strains were transformed with the plasmid containing the cre-less version of *RPS22B* (Figure 3a). GFP fluorescence was measured for each strain, with and without the plasmid, and with and without induction (Figure 3b and 3c). The strain with the spliced 5' UTR showed nearly identical expression with or without the plasmid and with or without induction (Figure 3b and 3c, blue traces), indicating that

translation of the spliced mRNA is not repressed. In the strain with the complete *RPS22B* 5' UTR, GFP expression is nearly 3-fold lower in cells without the plasmid and with the plasmid but without induction. We expect this is due to repression by the endogenous Rps22 protein in the cells. When the cre-less plasmid is induced to overexpress Rps22, the expression of GFP is further reduced about 6-fold (figure 3b and 3c, green traces), similar to the repression of the wild-type *RPS22B*-GFP strain (Figure 2b). This indicates that, similar to several other autoregulated ribosomal protein genes, repression occurs by inhibiting splicing. Interestingly, the bacterial ortholog of Rps22 is S8<sup>70</sup> and it is also involved in autoregulation<sup>49</sup>. One anomaly is worth noting. The original cre-less version of Rps22 had unintentionally left the stop codon at the end of the gene, so that it was not fused with mCherry, but it showed the autoregulation. When that was corrected to make a cre-less version fused to mCherry, autoregulation was no longer observed, indicating interference with RNA binding by the mCherry fusion. It is possible the same thing happens with other of our mCherry fusion constructs, which is another reason we could be underestimating the true number of autoregulated ribosomal protein genes.

### **Autoregulation of *RPL1B***

The expression of the *RPL1B* gene is reduced over 10-fold after 10 hours of induction of the cre-less version (Figure 2a). The paralog *RPL1A* showed only about a 2-fold reduction in expression (Figure 2a). *RPL1B* has no introns and we surmised that the 5' UTR may be the regulatory region. The gene has a short 5' UTR of 64 bases that is highly conserved within the *senso stricto* yeast species<sup>71</sup> (Figure 4a). We place the full 5' UTR before the chromosomal GFP gene (as we did for the *RPS22B* UTR described above) and measured about a 10-fold reduction in expression after induction of the cre-less gene (Supplementary figure S2, green graphs). This indicates that the 5' UTR is sufficient to confer feedback regulation by Rpl1. We also tested different versions of the gene on the plasmid fused to mCherry. Besides the cre-less version (*RPL1B-cl*), we tested three alternative constructs (Supplemental figure S3): wild-type for both the 5'UTR and coding region (*RPL1B-wt*); wild-type for the 5'UTR with a shuffled coding region (*RPL1B-cd*); and mutant 5'UTR but wild-type coding region (*RPL1B-mt*). After induction the two constructs containing the wild-type 5'UTR (*RPL1B-wt* and *RPL1B-cd*) produced about 2-fold less mCherry fusion protein than those with the mutant 5' UTR (*RPL1B-mt* and *RPL1B-cl*), consistent with the regulatory element being within the 5' UTR (Supplementary figure S4). Although the effect is smaller, the constructs with the wild-type 5' UTR also reduce the repression of GFP-tagged *RPL1B*, presumably as a result of elevated level of Rpl1-mCherry protein. The wild-type and cre-less versions of *RPL1B*-mCherry both reduce *RPL1A* expression equivalently, about 3-fold after 10 and 20 hours of induction (Supplementary figure S5). To further test the evidence for whether Rpl1 directly interacts with the 5' UTR sequence of its own message we designed an EMSA experiment with purified Rpl1 (see Methods) and a synthetic RNA, containing the UTR sequence. The presence of a distinct slower migrating band in lanes 2-3 (Supplementary figure S6) indicates direct binding between Rpl1 and the 5'-UTR of its own message.

Having verified that there is a post-transcriptional cis-regulatory element in the 5' UTR of *RPL1B* that responds to the Rpl1 concentration, we sought to identify the sequence features required for binding by Rpl1. The RNAstructure web server<sup>72</sup> (see Methods) predicts the same minimum free energy secondary structure and maximum expectation secondary structure for the 64 base long 5' UTR (Figure 4b). The structure has two primary stems, with hairpin stem 1 (HP1) shown in orange and hairpin stem 2 (HP2) shown in blue. Remarkably, every known binding site for Rpl1 and its orthologs in other species contains a common sequence and structure motif, shown in Figure 4c. This includes the binding sites for *E. coli*

ribosomal protein L1 on both the L11 mRNA and the 23SrRNA<sup>73</sup>, and ribosomal proteins from several other bacteria and archaea on mRNA and rRNA binding sites<sup>49,74</sup>. It also includes the binding sites on stems H77 and H78 of 28SrRNA of both human and yeast<sup>75,76</sup>. The *RPL1B* UTR structure contains the same sequence and structure motif except for a single G to A change shown by the arrows in Figures 4b,c.

To test if those sequence and structure features are required for regulation by Rpl1, we made several variants of the *RPL1B* 5' UTR and placed them upstream of GFP driven by the *TEF2* promoter. Figure 4d shows the wild-type 5'UTR with the Hairpin1 and Hairpin2 stems marked. Below that are 12 different variants and the logarithm of the GFP expressions from cells both uninduced (-Gal) and induced (+Gal) for the expression of the cre-less Rpl1 protein. The log expression values and the differences between induced and uninduced are also shown in the Table 1. The wild-type UTR is repressed by 0.70 (log reduction in GFP), verifying that the 5' UTR is the regulatory region. Removing the bases at either the 5' or 3' sides of the structured region reduces but does not eliminate repression, but removing both together does eliminate repression, perhaps by altering the structure of the mRNA. To test the importance of the hairpin structures we modified the 5' half of each stem to eliminate the structure, and then compensated by modifying the 3' half to make a complementary sequence and recover the secondary structure. HP1-mut modifies the 5' half of the HP1 sequence to eliminate the structure and it completely abolishes repression. HP1-cmp restores the predicted secondary structure and regains regulatory activity, in fact to a slightly higher level than the wild-type sequence.

We mutated HP2 in two different ways, first altering only the upper three bases in the 5' half (HP2-mut1) and then by altering the upper four bases in the 5' half (HP2-mut2). Both eliminate repression, demonstrating the importance of HP2 for the regulatory site. When the upper three bases are compensated (HP2-cmp1), repression is restored. However, when the upper four bases of HP2 are compensated (HP2-cmp2), there is still no regulation. This highlights the importance of the first G in the conserved sequence GGGAG shown in Figure 4c. To further test the importance of the bases in the conserved motif we altered the GGAAG to GGTTG (3muts, which also contains an additional mutation to maintain the wild-type structure), and the repression is eliminated. In another variant we changed the GGAAG to GCAAC, along with two additional changes to maintain the secondary structure (4muts), and again the repression is eliminated. Finally, we tested the importance of the base that differs between the *Rpl1B* 5' UTR (the A with the arrow in Figure 4b) and the conserved G in the binding site motif (the arrow in Figure 4c). The A to G variant (A43G) is still repressed, indicating that either base is acceptable at that position for regulatory activity. However, the expression of GFP in the non-induced state is lower for the mutant A43G than for the wild-type sequence, perhaps because it provides a higher affinity binding site for the intrinsic Rpl1 protein in the cell.

## Conclusions

Many recent studies use high-throughput methods to identify protein-mRNA interactions, some of which may be regulatory interactions. However, most of those approaches will only identify common events, where a protein regulates many genes. Discoordination of protein and mRNA levels adds to the complexity of these studies and any inferences that can be drawn. Additionally, cases of autoregulation, such as in bacteria, would be missed because the proteins have only a single target and are not, primarily, regulatory proteins. Non-coding segments of RNA can evolve rapidly and acquire various roles

including regulation of gene expression. They can be highly sensitive and autonomous sensors of the cellular environment and determine their own fate, such as to be translated or not. Riboswitches are an especially compelling example of such autonomous regulation because no protein is involved in the feedback response. Sensing protein concentrations can also be accomplished by RNAs and is likely to be much more wide spread than is currently known, but because of the single target limitation to detection, directed searches are needed to identify such cases. We have used the GFP-fusion collection in yeast as a means of rapidly screening for examples of feedback regulation. We observed that more than 20% of ribosomal genes appear to be autoregulated, and that is likely an underestimate of the true number due to some limitations of our approach. We find that *RPS22B* regulates its own splicing, which is required for translation and is similar to a few other known autoregulatory examples in yeast. *RPL1B*, on the other hand, appears to regulate its own translation via binding to a sequence and structure motif in the 5' UTR that is remarkably conserved in bacteria, archaea and eukaryotes from yeast to mammals. We know that motif is necessary but it doesn't appear to be sufficient as changes outside of the conserved region can also affect regulation. More work is required to identify the complete mechanism, and to uncover the regulatory domains of the remaining examples we identified. The GFP-fusion collection in yeast provides an outstanding resource for identifying feedback regulation of all types, but the development of more flexible approaches will be necessary to do similar searches in more complex genomes.

## Materials and Methods

### GAL1 vector and yeast strains

The synthetic genes were cloned into a custom plasmid, MBJ1-mod5 (Supplementary figure S7). The vector backbone was pMW102-empty-MORF. The main features of this vector are a  $\beta$ -lactamase gene, replication origin for selection in *E. coli*, the URA3 gene for selection in *ura*- yeast strains and GAL1 promoter for induction with galactose and the mCherry coding region that can be fused to the cre-less gene. The His5 terminator site with 5'-NheI and 3'-XhoI restriction sites was synthesized by G-blocks from IDT. The mCherry sequence was cloned out from pMVS124-pACT1 (a generous gift from Max Staller) with a 3'-NheI and 5'-AvrII restriction sites by PCR using. pMW102-empty-MORF was linearized using primers to incorporate PacI and AvrII restriction sites downstream of the GAL1 promoter. One microgram of purified mCherry, His5 terminator and linearized vector backbone were digested with NheI, AvrII, XhoI and PacI for 15 min and gel purified. The digested DNAs were ligated with T4 DNA ligase for 15 min at room temperature and transformed into DH5 $\alpha$  cells. The sequences of the selected clones were verified by Sanger sequencing.

The cre-less genes were synthesized and integrated directly into the MBJ1-mod5 vector by Twist Biosciences (San Francisco, CA). The genes were designed to be lacking any cis-regulatory elements (cre-less) by eliminating any introns, replacing the 5' UTR with an alternative sequence, and shuffling the synonymous codons of the gene using the program CodonShuffle<sup>64</sup>. The sequences of the shuffled coding regions are provided in the supplementary file2.

The parental GFP-tagged yeast strains (S288C) were taken from the Yeast GFP library<sup>48</sup> (a gift from Heather True-Kro). GFP-tagged strains were transformed with the plasmid containing the corresponding cre-less gene driven by the GAL1 promoter. Two independent clones were assayed for each sample.

### **Yeast transformation with cre-less plasmid**

Yeast strains from the GFP collection (*MAT $\alpha$  SUC2 gal2 mal2 mel flo1 flo8-1 hap1 ho bio1 bio6*) were grown in YPD medium (1% yeast extract, 1% bacto-peptone and 2% glucose) at 30 °C, overnight in 96 well plate format or in individual culture tubes. The cells were inoculated in 1 ml fresh YPD (10% v/v) and grown to OD<sub>600</sub> = 1. The cells were collected by centrifugation at 2000 Xg for 2 min and then mixed with 0.6-1  $\mu$ g of the cre-less plasmid in buffer containing 100 mM LiOAc, 50 % PEG (MW 3350). The resulting mixture was incubated for 5 min and subjected to heat shock at 40 °C for 20 min. The cells were mixed with 200  $\mu$ l of fresh YPD, incubated at 30 °C with shaking and plated on selective medium (SD-URA). Several colonies from each plate were collected after 2-3 days for galactose induction.

### **Galactose induction and flow cytometry assay**

Liquid cultures were inoculated and grown overnight in 400ul SD-URA with 2% raffinose in 96 deep well plates at 30 °C. Overnight cultures were diluted into both SD-URA with 2% raffinose (uninduced) and SD-URA with 2% raffinose and 0.2% galactose (induced). Cells were grown at 30C for 10 hours.

200  $\mu$ L cultures were transferred to 96-well plates and assayed on a CytoFLEX (Beckman Coulter). The live cells were gated and 10,000 events were acquired.

### ***RPL1B* Variants**

Three additional *RPL1B* variants were synthesized (Supplementary figure S3) beside the cre-less version of the gene (*RPL1B-cl*). The *RPL1B-wt* retained the wild type sequence of native *RPL1B* with the 5' UTR. The *RPL1B-cd* retained the 5' UTR of the wild type sequence but the internal mRNA sequence was shuffled and was identical to the cre-less sequence. The *RPL1B-mt* was identical to the wild type coding sequence but lacked the 5' UTR. All of these variations of *RPL1B* sequences were fused with mCherry sequence and cloned into MJB1 plasmid as described above.

### **Reporter gene assay**

To construct yeast strains expressing GFP with either the *RPS22B* UTR or *RPL1B* UTR variants a new plasmid (BAC690-TEF) was designed. BAC690-TEF (Supplementary figure S8) was designed based on the BAC690\_Euroscarf vector, which has an eGFP ORF 5' to an ADH1 terminator<sup>77</sup>. The TEF2 promoter sequenced was amplified using PCR from purified yeast *Saccharomyces cerevisiae* (S288C) genomic DNA. The TEF2 promoter was cloned into the BAC690\_Euroscarf vector 5' of the eGFP sequence. UTR variants were cloned upstream of eGFP using NEB's HiFi DNA Assembly. A PCR amplicon was made containing the TEF2 promoter, UTR variant, eGFP, and ADH1 terminator as well as a kanamycin phosphotransferase cassette for integration in the yeast genome. The integration location was a dubious ORF (YBR032W).

Transformed cells were grown on YPD plates containing 200ug/ml G418 to select for the integrated kanamycin cassette. Correct integrations were verified by colony PCR of the 5' and 3' junctions. Integrated strains were then transformed with their corresponding cre-less plasmid. Galactose induction and flow cytometry measurements were done as described above.

### **RNA Structure prediction**



We used the Matthews lab [RNAstructure](https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html) prediction web server <sup>72</sup> (<https://rna.urmc.rochester.edu/RNAstructureWeb/Servers/Predict1/Predict1.html>) version 6.0.1 with default parameters. It produces both a minimum free energy predicted secondary structure and a maximum expectation prediction secondary structure, which may be different. The structures shown in Figures 4b,c were created using the StructureEdit program from the same package.

## Acknowledgements

Funding for this work came from NIH grants R01HG000249 and R21GM126307. FB was supported by R25HG006687. We thank Dr. Max Staller for the gift of a plasmid used in this work and Dr. Heather True-Krob for the gift of the yeast GFP-fusion collection.

## References

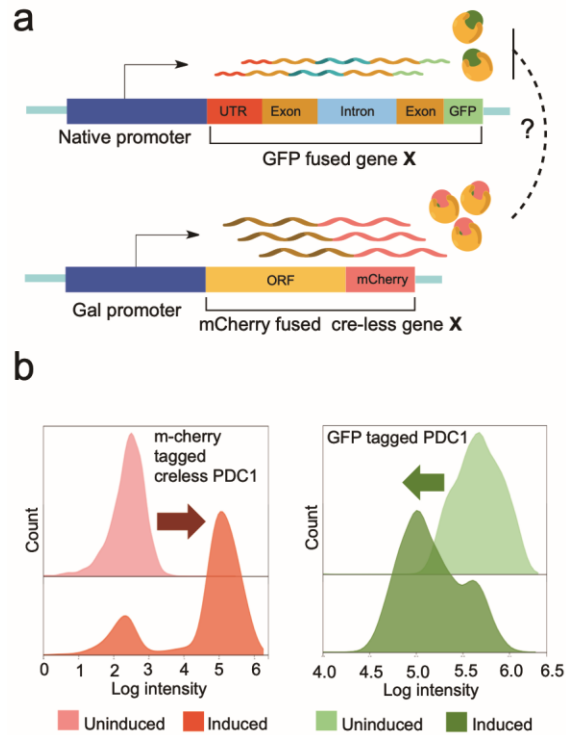
- 1 Atkinson, D. E. Regulation of enzyme function. *Annu Rev Microbiol* **23**, 47-68, doi:10.1146/annurev.mi.23.100169.000403 (1969).
- 2 Pardee, A. B. Molecular-Basis of Biological Regulation - Origins from Feedback Inhibition and Allosterity. *Bioessays* **2**, 37-40, doi:DOI 10.1002/bies.950020112 (1985).
- 3 Ptashne, M. *et al.* Autoregulation and function of a repressor in bacteriophage lambda. *Science* **194**, 156-161 (1976).
- 4 Ptashne, M., Johnson, A. D. & Pabo, C. O. A genetic switch in a bacterial virus. *Sci Am* **247**, 128-130, 132, 134-140 (1982).
- 5 Alon, U. Network motifs: theory and experimental approaches. *Nat Rev Genet* **8**, 450-461, doi:10.1038/nrg2102 (2007).
- 6 Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet* **31**, 64-68, doi:10.1038/ng881 (2002).
- 7 Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104, doi:10.1038/nature02800 (2004).
- 8 Hughes, T. R. & de Boer, C. G. Mapping yeast transcriptional networks. *Genetics* **195**, 9-36, doi:10.1534/genetics.113.153262 (2013).
- 9 Maclsaac, K. D. *et al.* An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* **7**, 113, doi:10.1186/1471-2105-7-113 (2006).
- 10 Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323**, 785-793, doi:10.1016/s0022-2836(02)00994-4 (2002).
- 11 Savageau, M. A. Comparison of classical and autogenous systems of regulation in inducible operons. *Nature* **252**, 546-549, doi:10.1038/252546a0 (1974).
- 12 Wall, M. E., Hlavacek, W. S. & Savageau, M. A. Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5**, 34-42, doi:10.1038/nrg1244 (2004).
- 13 Lareau, L. F. & Brenner, S. E. Regulation of splicing factors by alternative splicing and NMD is conserved between kingdoms yet evolutionarily flexible. *Mol Biol Evol* **32**, 1072-1079, doi:10.1093/molbev/msv002 (2015).
- 14 Pervouchine, D. *et al.* Integrative transcriptomic analysis suggests new autoregulatory splicing events coupled with nonsense-mediated mRNA decay. *Nucleic Acids Res* **47**, 5293-5306, doi:10.1093/nar/gkz193 (2019).
- 15 Sun, Y. *et al.* Autoregulation of RBM10 and cross-regulation of RBM10/RBM5 via alternative splicing-coupled nonsense-mediated decay. *Nucleic Acids Res* **45**, 8524-8540, doi:10.1093/nar/gkx508 (2017).

- 16 Nagengast, A. A., Stitzinger, S. M., Tseng, C. H., Mount, S. M. & Salz, H. K. Sex-lethal splicing autoregulation in vivo: interactions between SEX-LETHAL, the U1 snRNP and U2AF underlie male exon skipping. *Development* **130**, 463-471, doi:10.1242/dev.00274 (2003).
- 17 Penalva, L. O., Lallena, M. J. & Valcarcel, J. Switch in 3' splice site recognition between exon definition and splicing catalysis is important for sex-lethal autoregulation. *Mol Cell Biol* **21**, 1986-1996, doi:10.1128/MCB.21.6.1986-1996.2001 (2001).
- 18 Ascano, M., Gerstberger, S. & Tuschl, T. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Curr Opin Genet Dev* **23**, 20-28, doi:10.1016/j.gde.2013.01.003 (2013).
- 19 Hannigan, M. M., Zagore, L. L. & Licatalosi, D. D. Mapping transcriptome-wide protein-RNA interactions to elucidate RNA regulatory programs. *Quant Biol* **6**, 228-238, doi:10.1007/s40484-018-0145-6 (2018).
- 20 Matia-Gonzalez, A. M., Laing, E. E. & Gerber, A. P. Conserved mRNA-binding proteomes in eukaryotic organisms. *Nat Struct Mol Biol* **22**, 1027-1033, doi:10.1038/nsmb.3128 (2015).
- 21 Mitchell, S. F., Jain, S., She, M. & Parker, R. Global analysis of yeast mRNPs. *Nat Struct Mol Biol* **20**, 127-133, doi:10.1038/nsmb.2468 (2013).
- 22 Ramanathan, M., Porter, D. F. & Khavari, P. A. Methods to study RNA-protein interactions. *Nat Methods* **16**, 225-234, doi:10.1038/s41592-019-0330-1 (2019).
- 23 Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393-1406, doi:10.1016/j.cell.2012.04.031 (2012).
- 24 Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829-845, doi:10.1038/nrg3813 (2014).
- 25 Mitchell, S. F. & Parker, R. Principles and properties of eukaryotic mRNPs. *Mol Cell* **54**, 547-558, doi:10.1016/j.molcel.2014.04.033 (2014).
- 26 Nomura, M. Journey of a molecular biologist. *Annu Rev Biochem* **80**, 16-40, doi:10.1146/annurev-biochem-082808-091843 (2011).
- 27 Nomura, M., Gourse, R. & Baughman, G. Regulation of the synthesis of ribosomes and ribosomal components. *Annu Rev Biochem* **53**, 75-117, doi:10.1146/annurev.bi.53.070184.000451 (1984).
- 28 Nomura, M., Yates, J. L., Dean, D. & Post, L. E. Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. *Proc Natl Acad Sci U S A* **77**, 7084-7088 (1980).
- 29 Deiorio-Haggar, K., Anthony, J. & Meyer, M. M. RNA structures regulating ribosomal protein biosynthesis in bacilli. *RNA Biol* **10**, 1180-1184, doi:10.4161/rna.24151 (2013).
- 30 Meyer, M. M. rRNA Mimicry in RNA Regulation of Gene Expression. *Microbiol Spectr* **6**, doi:10.1128/microbiolspec.RWR-0006-2017 (2018).
- 31 Romby, P. & Springer, M. Bacterial translational control at atomic resolution. *Trends Genet* **19**, 155-161, doi:10.1016/S0168-9525(03)00020-9 (2003).
- 32 Springer, M. *et al.* Autogenous control of Escherichia coli threonyl-tRNA synthetase expression in vivo. *J Mol Biol* **185**, 93-104 (1985).
- 33 Torres-Larios, A. *et al.* Structural basis of translational control by Escherichia coli threonyl tRNA synthetase. *Nat Struct Biol* **9**, 343-347, doi:10.1038/nsb789 (2002).
- 34 Frugier, M. & Giege, R. Yeast aspartyl-tRNA synthetase binds specifically its own mRNA. *J Mol Biol* **331**, 375-383, doi:10.1016/s0022-2836(03)00767-8 (2003).
- 35 Levi, O. & Arava, Y. mRNA association by aminoacyl tRNA synthetase occurs at a putative anticodon mimic and autoregulates translation in response to tRNA levels. *PLoS Biol* **17**, e3000274, doi:10.1371/journal.pbio.3000274 (2019).
- 36 Castello, A., Hentze, M. W. & Preiss, T. Metabolic Enzymes Enjoying New Partnerships as RNA-Binding Proteins. *Trends Endocrinol Metab* **26**, 746-757, doi:10.1016/j.tem.2015.09.012 (2015).

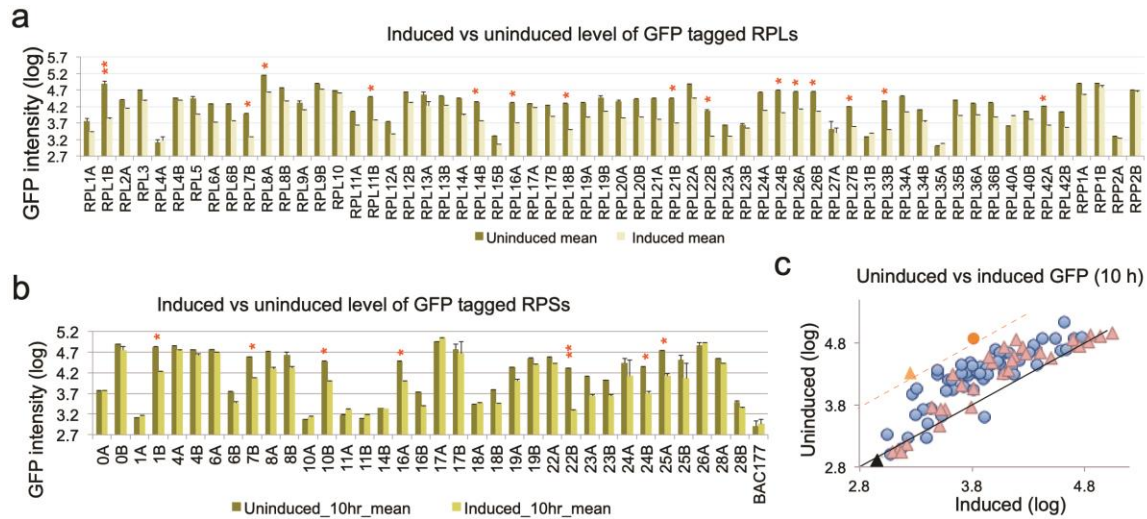
- 37 Hentze, M. W., Castello, A., Schwarzl, T. & Preiss, T. A brave new world of RNA-binding proteins. *Nat Rev Mol Cell Biol* **19**, 327-341, doi:10.1038/nrm.2017.130 (2018).
- 38 Andrade, M. *et al.* DNA polymerase of bacteriophage T4 is an autogenous translational repressor. *Proc Natl Acad Sci U S A* **85**, 7942-7946 (1988).
- 39 Gold, L. Posttranscriptional regulatory mechanisms in Escherichia coli. *Annu Rev Biochem* **57**, 199-233, doi:10.1146/annurev.bi.57.070188.001215 (1988).
- 40 McPheeters, D. S., Stormo, G. D. & Gold, L. Autogenous regulatory site on the bacteriophage T4 gene 32 messenger RNA. *J Mol Biol* **201**, 517-535 (1988).
- 41 Russel, M., Gold, L., Morrisett, H. & O'Farrell, P. Z. Translational, autogenous regulation of gene 32 expression during bacteriophage T4 infection. *J Biol Chem* **251**, 7263-7270 (1976).
- 42 Tuerk, C., Eddy, S., Parma, D. & Gold, L. Autogenous translational operator recognized by bacteriophage T4 DNA polymerase. *J Mol Biol* **213**, 749-761, doi:10.1016/S0022-2836(05)80261-X (1990).
- 43 Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol Cell* **43**, 867-879, doi:10.1016/j.molcel.2011.08.024 (2011).
- 44 Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463, doi:10.1038/nrm1403 (2004).
- 45 McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995-1011, doi:10.1261/rna.061234.117 (2017).
- 46 Stormo, G. D. & Ji, Y. Do mRNAs act as direct sensors of small molecules to control their expression? *Proc Natl Acad Sci U S A* **98**, 9465-9467, doi:10.1073/pnas.181334498 (2001).
- 47 Vitreschak, A. G., Rodionov, D. A., Mironov, A. A. & Gelfand, M. S. Riboswitches: the oldest mechanism for the regulation of gene expression? *Trends Genet* **20**, 44-50, doi:10.1016/j.tig.2003.11.008 (2004).
- 48 Huh, W. K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691, doi:10.1038/nature02026 (2003).
- 49 Fu, Y., Deiorio-Haggar, K., Anthony, J. & Meyer, M. M. Most RNAs regulating ribosomal protein biosynthesis in Escherichia coli are narrowly distributed to Gammaproteobacteria. *Nucleic Acids Res* **41**, 3491-3503, doi:10.1093/nar/gkt055 (2013).
- 50 Dean, D. & Nomura, M. Feedback regulation of ribosomal protein gene expression in Escherichia coli. *Proc Natl Acad Sci U S A* **77**, 3590-3594 (1980).
- 51 de la Cruz, J. *et al.* Feedback regulation of ribosome assembly. *Curr Genet* **64**, 393-404, doi:10.1007/s00294-017-0764-x (2018).
- 52 Warner, J. R. The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* **24**, 437-440, doi:10.1016/s0968-0004(99)01460-7 (1999).
- 53 Warner, J. R. & McIntosh, K. B. How common are extraribosomal functions of ribosomal proteins? *Mol Cell* **34**, 3-11, doi:10.1016/j.molcel.2009.03.006 (2009).
- 54 Chaker-Margot, M. Assembly of the small ribosomal subunit in yeast: mechanism and regulation. *RNA* **24**, 881-891, doi:10.1261/rna.066985.118 (2018).
- 55 Chaker-Margot, M. & Klinge, S. Assembly and early maturation of large subunit precursors. *RNA* **25**, 465-471, doi:10.1261/rna.069799.118 (2019).
- 56 Fewell, S. W. & Woolford, J. L., Jr. Ribosomal protein S14 of Saccharomyces cerevisiae regulates its expression by binding to RPS14B pre-mRNA and to 18S rRNA. *Mol Cell Biol* **19**, 826-834, doi:10.1128/mcb.19.1.826 (1999).
- 57 Gabunilas, J. & Chanfreau, G. Splicing-Mediated Autoregulation Modulates Rpl22p Expression in Saccharomyces cerevisiae. *PLoS Genet* **12**, e1005999, doi:10.1371/journal.pgen.1005999 (2016).
- 58 Macias, S., Bragulat, M., Tardiff, D. F. & Vilardell, J. L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. *Mol Cell* **30**, 732-742, doi:10.1016/j.molcel.2008.05.002 (2008).

- 59 Vilardell, J., Chartrand, P., Singer, R. H. & Warner, J. R. The odyssey of a regulated transcript. *RNA* **6**, 1773-1780, doi:10.1017/s135583820000145x (2000).
- 60 Vilardell, J., Yu, S. J. & Warner, J. R. Multiple functions of an evolutionarily conserved RNA binding domain. *Mol Cell* **5**, 761-766, doi:10.1016/s1097-2765(00)80255-5 (2000).
- 61 Plocik, A. M. & Guthrie, C. Diverse forms of RPS9 splicing are part of an evolving autoregulatory circuit. *PLoS Genet* **8**, e1002620, doi:10.1371/journal.pgen.1002620 (2012).
- 62 Badis, G., Saveanu, C., Fromont-Racine, M. & Jacquier, A. Targeted mRNA degradation by deadenylation-independent decapping. *Mol Cell* **15**, 5-15, doi:10.1016/j.molcel.2004.06.028 (2004).
- 63 He, F., Li, C., Roy, B. & Jacobson, A. Yeast Edc3 targets RPS28B mRNA for decapping by binding to a 3' untranslated region decay-inducing regulatory element. *Mol Cell Biol* **34**, 1438-1451, doi:10.1128/MCB.01584-13 (2014).
- 64 Katz, L. & Burge, C. B. Widespread selection for local RNA secondary structure in coding regions of bacterial genes. *Genome Res* **13**, 2042-2051, doi:10.1101/gr.1257503 (2003).
- 65 Eberhardt, I. *et al.* Autoregulation of yeast pyruvate decarboxylase gene expression requires the enzyme but not its catalytic activity. *Eur J Biochem* **262**, 191-201 (1999).
- 66 Hooks, K. B., Naseeb, S., Parker, S., Griffiths-Jones, S. & Delneri, D. Novel Intronic RNA Structures Contribute to Maintenance of Phenotype in *Saccharomyces cerevisiae*. *Genetics* **203**, 1469-1481, doi:10.1534/genetics.115.185363 (2016).
- 67 Danin-Kreiselman, M., Lee, C. Y. & Chanfreau, G. RNase III-mediated degradation of unspliced pre-mRNAs and lariat introns. *Mol Cell* **11**, 1279-1289, doi:10.1016/s1097-2765(03)00137-0 (2003).
- 68 Parenteau, J. *et al.* Introns within ribosomal protein genes regulate the production and function of yeast ribosomes. *Cell* **147**, 320-331, doi:10.1016/j.cell.2011.08.044 (2011).
- 69 Peng, B., Williams, T. C., Henry, M., Nielsen, L. K. & Vickers, C. E. Controlling heterologous gene expression in yeast cell factories on different carbon substrates and across the diauxic shift: a comparison of yeast promoter activities. *Microb Cell Fact* **14**, 91, doi:10.1186/s12934-015-0278-5 (2015).
- 70 Ban, N. *et al.* A new system for naming ribosomal proteins. *Curr Opin Struct Biol* **24**, 165-169, doi:10.1016/j.sbi.2014.01.002 (2014).
- 71 Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71-76, doi:10.1126/science.1084337 (2003).
- 72 Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129, doi:10.1186/1471-2105-11-129 (2010).
- 73 Said, B., Cole, J. R. & Nomura, M. Mutational analysis of the L1 binding site of 23S rRNA in *Escherichia coli*. *Nucleic Acids Res* **16**, 10529-10545, doi:10.1093/nar/16.22.10529 (1988).
- 74 Kohrer, C., Mayer, C., Neumair, O., Grobner, P. & Piendl, W. Interaction of ribosomal L1 proteins from mesophilic and thermophilic Archaea and Bacteria with specific L1-binding sites on 23S rRNA and mRNA. *Eur J Biochem* **256**, 97-105, doi:10.1046/j.1432-1327.1998.2560097.x (1998).
- 75 Anger, A. M. *et al.* Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80-85, doi:10.1038/nature12104 (2013).
- 76 Bernier, C. R. *et al.* RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss* **169**, 195-207, doi:10.1039/c3fd00126a (2014).
- 77 Sheff, M. A. & Thorn, K. S. Optimized cassettes for fluorescent protein tagging in *Saccharomyces cerevisiae*. *Yeast* **21**, 661-670, doi:10.1002/yea.1130 (2004).

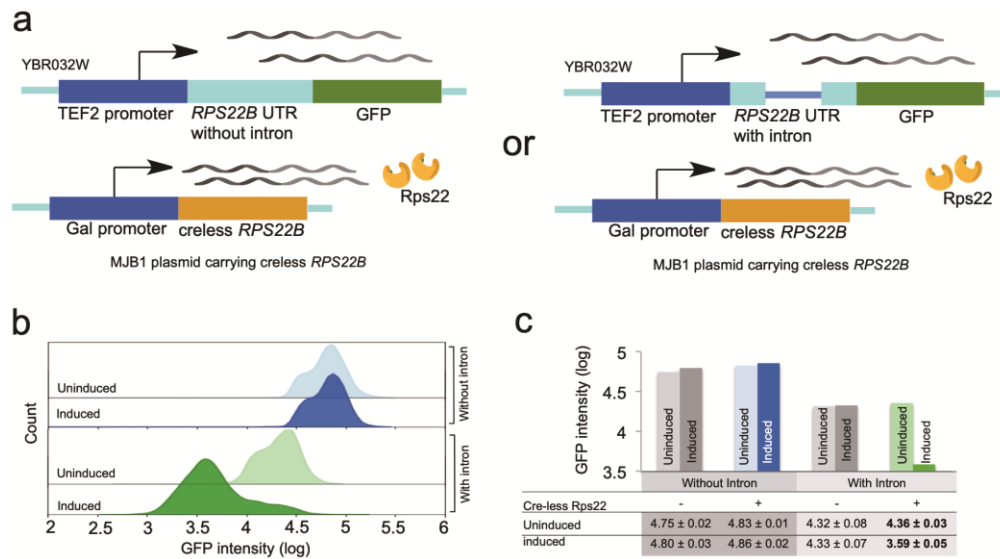
## Figures



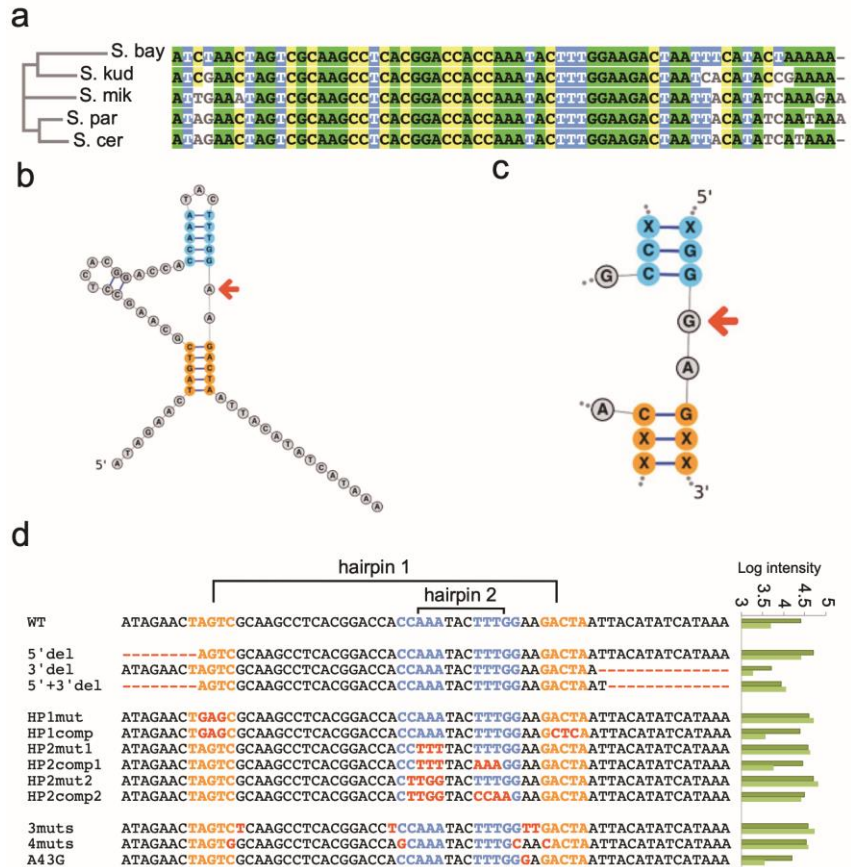
**Figure 1:** Screening for protein autoregulation. **(a)** One endogenous gene is fused to GFP and a cre-less (cis-regulatory element-less) version of the same gene is fused to mCherry. The cre-less version lacks the native 5' and 3' UTRs and any introns and has shuffled codons in order to prevent the protein autoregulating the cre-less mRNA. After inducing the cre-less gene, possible autoregulation is detected by a decrease in GFP levels. **(b)** An example of autoregulation with the gene PDC1. Overexpression of the cre-less PDC1 (seen as an increase in mCherry fluorescence) leads to a decrease of the endogenous PDC1 (seen as a decrease in GFP fluorescence).



**Figure 2:** Screening of ribosomal protein autoregulation. **(a)** 60 large ribosomal subunit proteins were screened for autoregulation. Genes with more than a 10-fold reduction of their endogenous protein after induction of their cre-less protein are highlighted with two red stars (\*\*). Genes with a 3-fold to 10-fold reduction are highlighted by one red star (\*). **(b)** 35 small ribosomal subunit proteins were screened for autoregulation. BAC177 is a negative control with no GFP-tagged proteins. **(c)** Scatterplot showing the uninduced vs induced log<sub>10</sub> GFP levels for large ribosomal subunit proteins (triangles) and small ribosomal subunit proteins (circles). *RPL1B* (red circle) and *RPS22B* (orange triangle) showed the highest levels of autoregulation. The black line passing through the center represents equal GFP levels for induced and uninduced, while the orange dotted line represents a 10-fold reduction in the induced compared to uninduced. The black triangle is BAC177 with no GFP-tagged proteins.



**Figure 3: Regulation of RPS22B.** (a) Two reporter constructs were made to test the sequence requirements for autoregulation by overexpressed cre-less Rps22. One construct had the wild-type *RPS22B* 5' UTR (containing a 557 bp intron) placed upstream of GFP and driven by the *TEF2* promoter. A second construct was identical except that the intron was removed. Both constructs were separately integrated into the YBR032W locus and transformed with the plasmid for expressing cre-less Rps2. (b) Histograms of log GFP intensity from flow cytometry of cells containing either the UTR-with-intron reporter (green) or the UTR-without-intron reporter (blue). Cells were measured as either uninduced (lighter-shaded colors) or after induction of cre-less Rps22 (darker-shaded colors). Each histogram curve represents cells picked from a single colony after transformation of the cre-less plasmid. Multiple colonies for each reporter construct were picked and measured this way, and these replicates were used to generate the bar plot in 3c. (c) Bar plot comparing the log GFP levels of the two reporter constructs in cells with or without the plasmid for expression of cre-less Rps22 and with or without galactose induction. Only the UTR-with-intron construct (green) shows autoregulation after induction of cre-less Rps22.



**Figure 4: RPL1B autoregulation.** (a) Alignment of the 5' UTR sequence of *RPL1B* across different species of saccharomyces shows a high level of conservation. (b) The predicted structure of the *RPL1B* 5' UTR with hairpin stem 1 (HP1) in orange and hairpin stem 2 (HP2) in blue. The red arrow indicates A at position 43 which differs from the G nucleotide seen in the consensus motif bound by Rpl1 in other species. (c) The common sequence and structure motif bound by Rpl1 in other species. The red arrow indicates the G nucleotide that is changed to A in *RPL1B*'s 5' UTR sequence. (d) Variant sequences of the *RPL1B* 5' UTR that were used to probe the sequence and structure requirements for autoregulation by Rpl1. Variants included deletions from the 5' and 3' ends, mutations to disrupt the hairpin stem structures as well as compensating mutations to restore them, and mutations affecting the core GGAAG of the motif. The log GFP expression for uninduced (dark green) and induced (light green) cells are shown to the right for each sequence.



**Table 1. GFP expression level (log) with or without induction in the presence of *RPL1B* UTR variants.**

	Uninduced (-Gal)	Induced (+Gal)	Difference ( $\Delta$ )
WT	4.4	3.7	0.7
5'del	4.7	4.4	0.3
3'del	3.71	3.26	0.45
5'+3'del	3.95	4.04	-0.09
HP1mut	4.6	4.7	-0.1
HP1comp	4.39	3.57	0.82
HP2mut1	4.57	4.63	-0.06
HP2comp1	4.46	3.75	0.71
HP2mut2	4.71	4.81	-0.1
HP2comp2	4.5	4.41	0.09
3mut	4.59	4.72	-0.13
4mut	4.53	4.57	-0.04
A43G	4.19	3.54	0.65