METHOD

RNA-ID, a highly sensitive and robust method to identify *cis*-regulatory sequences using superfolder GFP and a fluorescence-based assay

KIMBERLY M. DEAN and ELIZABETH J. GRAYHACK¹

Department of Biochemistry and Biophysics, University of Rochester Medical School, Rochester, New York 14642, USA

ABSTRACT

We have developed a robust and sensitive method, called RNA-ID, to screen for *cis*-regulatory sequences in RNA using fluorescence-activated cell sorting (FACS) of yeast cells bearing a reporter in which expression of both superfolder green fluorescent protein (GFP) and yeast codon-optimized mCherry red fluorescent protein (RFP) is driven by the bidirectional *GAL1,10* promoter. This method recapitulates previously reported progressive inhibition of translation mediated by increasing numbers of CGA codon pairs, and restoration of expression by introduction of a tRNA with an anticodon that base pairs exactly with the CGA codon. This method also reproduces effects of paromomycin and context on stop codon read-through. Five key features of this method contribute to its effectiveness as a selection for regulatory sequences: The system exhibits greater than a 250-fold dynamic range, a quantitative and dose-dependent response to known inhibitory sequences, exquisite resolution that allows nearly complete physical separation of distinct populations, and a reproducible signal between different cells transformed with the identical reporter, all of which are coupled with simple methods involving ligation-independent cloning, to create large libraries. Moreover, we provide evidence that there are sequences within a 9-nt library that cause reduced GFP fluorescence, suggesting that there are novel *cis*-regulatory sequences to be found even in this short sequence space. This method is widely applicable to the study of both RNA-mediated and codon-mediated effects on expression.

Keywords: method; fluorescent reporters; *cis*-regulatory; codons; RNA; translation

INTRODUCTION

Identification of *cis*-regulatory sequences that modulate gene expression is a critical step in the definition of the networks and mechanisms that coordinate regulatory responses. Identification of the genomic set of *cis*-regulatory sequences is challenging in all organisms because regulatory sequences are generally small in comparison to entire protein-coding regions, and because neither their location nor their precise identity is conserved even between highly related organisms. Identification of the *cis*-elements in DNA, which control transcription, has been made easier and faster by the development of both computational

(GuhaThakurta 2006; Das and Dai 2007; Brohee et al. 2011) and experimental tools (Gertz et al. 2009; Schlabach et al. 2010) to find these elements, as well as by the development of methods to link DNA sequences and transcription factors (Harbison et al. 2004; Barrera and Ren 2006; Walhout 2006; Wang et al. 2007). In contrast, although *cis*-elements in RNA mediate many aspects of post-transcriptional regulation, including RNA splicing, folding, cellular location, degradation, and translation, the large-scale discovery of these elements has been complicated by at least three factors.

First, many *cis*-regulatory elements in RNA are not yet identified because the set of RNA-binding proteins (RBP) that mediate some types of regulation is both large and incomplete. The number of RBPs appears to be large, far exceeding the number of DNA-binding proteins, based on emerging experimental methods that use quantitative proteomics to find the RBPs (Butter et al. 2009; Tsvetanova et al. 2010; Castello et al. 2012; Gebauer et al. 2012). Moreover, the set of RBPs is incomplete because RBPs are difficult to identify, since only a subset of the proteins that

Abbreviations: GFP, green fluorescent protein; RFP, red fluorescent protein; RBP, RNA binding protein; rCV, robust coefficient of variation; FACS, fluorescence-activated cell sorting; LIC, ligation-independent cloning; FSC, forward scatter; SSC, side scatter; arg, arginine; nt, nucleotide

¹Corresponding author

E-mail elizabeth_grayhack@urmc.rochester.edu

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.035907.112>.

interact with RNA contain easily recognizable RNA-binding motifs (Tsvetanova et al. 2010; Castello et al. 2012).

Second, regulatory elements in RNAs are sometimes difficult to distinguish by computational methods since the essential features of an RNA regulatory element may be either its direct sequence or the RNA structure that it forms, either in the presence or absence of protein cofactors (Macdonald and Kerr 1998; Rabani et al. 2008; Goodarzi et al. 2012). For example, both distinct structural motifs and linear sequence motifs are important for the RNA localization of the zebrafish *squint/nodal*-related 1 RNA (Gilligan et al. 2011), as well as for human mRNA stability (Goodarzi et al. 2012). In fact, discovery of *cis*-elements involved in human mRNA stability required not only two different computational analyses to find the sequence and structural elements, but also a correlation with experimentally derived whole-genome mRNA decay measurements (Goodarzi et al. 2012).

Third, some regulatory elements in RNA are not recognized as targets of RNA-binding proteins since their effects are mediated either via interactions with other RNA molecules, such as miRNAs, siRNAs, or lncRNAs (Bartel 2009; Gong and Maquat 2011), or via translation of the mRNA, for example, by codon choice or by uORFs (Hinnebusch 2005; Ingolia et al. 2009; Letzring et al. 2010; Plotkin and Kudla 2011). Estimates of the number and extent of all of these types of regulation continue to increase. For example, experimental analysis of the translated sequences in *Saccharomyces cerevisiae*, using ribosome profiling, resulted in the discovery of large numbers of translated uORFs, although their regulatory effects remain largely unknown (Ingolia et al. 2009). Moreover, we have evidence that combinations of adjacent codons modulate translation efficiency in a manner not predictable based on their individual properties (Letzring et al. 2010), thus making it necessary to sort through the 3721 codon pairs and possibly the 226,981 codon triplets for inhibitory combinations.

A general experimental method to find those sequences, structures, or functional elements within RNAs that modulate expression would facilitate the delineation of post-transcriptional regulation and might uncover novel mechanisms of regulation. Since synthetic promoters have been useful in the experimental analysis of promoter function, both for definition of individual elements and for the analysis of interactions between sites (Gertz et al. 2009; Schlabach et al. 2010), we considered that similar approaches would facilitate discovery of RNA elements.

We have developed a method that enables the sensitive, simultaneous, and highly reproducible analysis of the expression effects in yeast of a large number of sequences, inserted within or adjacent to the protein-coding region of a reporter gene. The method, called RNA-ID, uses a synthetic reporter to identify *cis*-regulatory elements in RNA, and is also effective for coding sequence variants, because sequences are inserted into the superfolder GFP variant

(Pedelacq et al. 2006) to minimize effects of altered amino acid sequence. Furthermore, noise from promoter effects is minimized by coincidental expression and measurement of RFP. We demonstrate here that this system is robust with a signal that is 250-fold above background, and allows reproducible and nearly complete (>99.6%) separation of variants in which the GFP median of one variant is ~28% that of the other. This method recapitulates both codon-mediated inhibition of expression (Letzring et al. 2010) and paromomycin-mediated, context-dependent read-through of stop codons (Bonetti et al. 1995). Finally, we demonstrate that this method can be used to screen a library of GFP variants with three randomized codons to identify variants with reduced GFP expression, relative to RFP (GFP/RFP). We provide evidence that most of these variants recapitulate their low expression phenotype upon rescreening and upon reconstruction.

RESULTS

A reporter for RNA *cis*-regulatory elements that uses the fluorescent reporters superfolder GFP and RFP

We set out to design a system to detect sequences within the RNA or coding region that modulate expression. To measure these effects rapidly and quantitatively in yeast, we developed a fluorescence-based assay that uses an integrated vector simultaneously expressing two fluorescent reporters from a bidirectional regulated promoter (*GAL1,10*) (Fig. 1A). To minimize problems caused by changes in protein sequence that could affect the activity or stability of the reporter, we built a vector in which test sequences are inserted into the superfolder GFP mutant, since this mutant was selected for and tested for robust fluorescence when fused to several insoluble proteins (Pedelacq et al. 2006). To reduce noise between cells, we coexpressed both GFP and the yeast codon-optimized RFP variant of mCherry (Fig. 1A; Keppler-Ross et al. 2008) from the bidirectional galactose-inducible promoter *GAL1,10*, which allows coordinate and controlled expression of the two reporters.

To facilitate insertion of test sequences and to set up a versatile system for testing sequences either inside or outside of the coding region, we designed the vector to allow ligation-independent cloning (LIC) (Aslanidis and de Jong 1990) into either GFP or RFP, and did not include a start codon for GFP. Thus, in the absence of a sequence insertion, GFP is not expressed. To allow LIC cloning, the vector is designed such that a single specific nucleotide is absent in the first 12–18 nt of the 3' strand adjacent to each of the restriction-cut ends. In the presence of this single specific nucleotide, the exonuclease activity of T4 DNA polymerase cleaves nucleotides from the 3' end of the duplex DNA until it encounters that specific nucleotide. This procedure efficiently generates the 12–18-nt

FIGURE 1. Design and function of the RNA-ID reporter for RNA *cis*-regulatory elements. (A) Diagram of the dual GFP and RFP vector. Expression of superfolder GFP (Pedelacq et al. 2006) and yeast-optimized mCherry RFP (Keppler-Ross et al. 2008) are under the control of the bidirectional, galactose-inducible promoter *GAL1,10*. The *MET15* gene is used for selection in yeast, and integration is directed to the *ADE2* locus. (B) Schematic illustrating LIC cloning into the sites upstream of GFP. Single-stranded sequences on the 5' and 3' ends of the vector pEKD1024 (green lines) are created by digestion of the vector with restriction endonucleases *PacI* and *BbrPI*, followed by treatment with T4 DNA polymerase in the presence of dGTP to generate 17 and 12 base single-stranded ends. Two overlapping oligonucleotides with homology with the single-stranded vector ends suffice for cloning, and allow use of a single oligonucleotide containing a randomized sequence at a defined position without requiring a full-length base-paired complement. The *top* oligonucleotide (blue line) contains a sequence complementary to the 5' LIC site, the sequence of interest, including the ATG, and a 12-base sequence complementary to the *bottom* oligonucleotide. The *bottom* oligonucleotide (brown line) minimally contains a sequence that is complementary to the 3' LIC site sequence, followed by a sequence that base pairs with the indicated complementary sequence of the *top* oligonucleotide. Sequences can also be inserted near the 5' end of the RFP gene after digestion with the restriction endonuclease *Swal* and resection with T4 DNA polymerase to create different single-stranded ends (see Materials and Methods). (C,D) Comparison of fluorescence outputs from a reporter on a multicopy plasmid versus an integrated reporter. To express GFP, an in-frame ATG is inserted upstream of GFP, as described above. (C) Histogram of GFP fluorescence profile versus cell number from yeast cells bearing the GFP and RFP genes on a 2 μ plasmid (orange), the identical GFP/RFP construct integrated at the *ade2* locus (blue), and an integrated plasmid lacking the GFP and RFP genes (gray). (D) Scatter plot of cells expressing GFP and RFP. Cells and the colors are identical to those in C. (E) Comparison of the signal and noise, with or without the RFP cutoff, from multicopy versus integrated GFP constructs.

single-stranded regions at each end. After annealing of oligonucleotides with complementary single-stranded ends to the vector (see Fig. 1B), the reaction is transformed into *Escherichia coli* without additional ligation. This design permits highly efficient cloning with two annealed oligonucleotides, one of which can contain a randomized sequence (Fig. 1B). Thus, both the length and position of a randomized sequence can be varied, as long as an in-frame ATG is included in the GFP insert.

To achieve separation of populations of cells with sequences that cause differential expression, we minimized noise between cells bearing a single sequence in three ways, as illustrated in Figure 1, for reporters bearing an ATG start codon for GFP, called ATG-GFP. First, the reporter is integrated into the chromosome, which not only results in a single unique reporter in each cell, but also reduces the noise in the GFP signal. As can be seen in Figure 1C, in which the GFP histograms from integrated and plasmid-borne copies of ATG-GFP are compared, the integrated reporter yields a much tighter GFP signal (blue trace) than that from the same reporter on a multicopy plasmid (orange trace). Furthermore, a significant fraction of the cells containing the multicopy plasmid trail into the low GFP region, which would undoubtedly complicate identification of sequences that cause low GFP expression. This same conclusion is evident from statistical analysis: The robust coefficients of variation (rCV) ($100 \times \frac{1}{2} \times [\text{Intensity (at 84.13 percentile)} - \text{Intensity (at 15.87 percentile)}] / \text{Median intensity}$), is much lower for the integrated sample (60.8), compared with that for the plasmid-borne sample (128.0) (Fig. 1E).

Second, the GFP fluorescence was normalized to RFP fluorescence in each cell (see the scatter plot in Fig. 1D), which has been shown to reduce the effects of extrinsic noise due to differential activation of the promoter in different cells, the major source of noise for the *GAL1* promoter (Raser and O'Shea 2004). Thus, as expected, GFP fluorescence (assessed with the 515/20-nm filter from cells excited at 488 nm), and yeast codon-optimized mCherry RFP fluorescence (assessed with a 610/20 nm filter from cells excited at 532 nm) are strongly correlated ($r = 0.96 - 0.98$; data not shown), and this analysis results in a tighter signal in our system (Fig. 1, cf. C and D).

Third, cells in which RFP fluorescence is less than 5×10^3 were eliminated from the evaluation to remove cells that failed to effectively induce expression. Although this step does not affect the rCV of the integrated reporter, it does significantly improve the rCV of the plasmid-borne reporter, from 128.0 to 89.3, enabling its use for some applications. However, since only 64% of the cells with the multicopy plasmid reporter pass the RFP cutoff, nearly half of the sample is unusable. Furthermore, cells bearing the integrated reporter exhibit the expected correlation between cell size and fluorescent protein, as evidenced by the correlation of GFP fluorescence with both the forward

scatter (FSC) ($r = 0.90$) and the side scatter (SSC) ($r = 0.92$), but this correlation is not observed with the multi-copy plasmid sample FSC ($r = 0.23$) and SSC ($r = 0.26$) (data not shown). As expected, strains in which the integrated vector does not contain either GFP or RFP (gray trace) fail to exhibit significant GFP fluorescence (Fig. 1C) and do not pass the RFP cutoff (Fig. 1D).

Translation inhibition due to wobble decoding of arg CGA codons is seen with the RNA-ID reporter

To verify that this RNA-ID GFP/RFP reporter system can be used to study translation regulation by *cis*-regulatory elements, we compared translation regulation with this reporter to that obtained with another reporter. We previously demonstrated that arginine (arg) CGA codon pairs (relative to arg AGA codons) reduce translation efficiency of a luciferase reporter in a dose-dependent manner (Letzring et al. 2010). In that study, we found that the reduction in luciferase activity caused by insertion of CGA codons was due to I-A wobble decoding of the CGA, since luciferase levels were nearly completely restored by expression of a mutant tRNA, tRNA^{Arg(UCG)*}, in which the anticodon of the native tRNA^{Arg(ICG)} was mutated to base pair with all three bases in the CGA codon (Letzring et al. 2010). Therefore, we examined the effects of incorporating CGA codons on expression of the GFP reporter to determine whether the CGA codons would similarly reduce GFP fluorescence.

We find that incorporation of increasing numbers of inhibitory arg CGA codons results in progressive and substantial inhibition in GFP fluorescence (Fig. 2A). We compared the median GFP/RFP fluorescence ratio from yeast cells bearing integrated reporters with insertions of increasing numbers of inhibitory CGA codons from zero to four, and found that the median ratio is steadily reduced by each CGA codon addition (Fig. 2A,B). GFP/RFP levels in cells bearing (AGA)₃-GFP are 3.6-fold greater than the levels in cells with the (CGA)₂-GFP reporter (Fig. 2A,B), remarkably similar to the threefold effect of a single CGA codon pair on expression of luciferase (Letzring et al. 2010). Likewise, the (AGA)₃-GFP signal is 16-fold greater than that from (CGA)₃-GFP and 29-fold greater than that from (CGA)₄-GFP (Fig. 2A,B). Furthermore, as reported previously for the luciferase reporter (Letzring et al. 2010), the reduced expression of (CGA)₂(AGA)-GFP and (CGA)₃-GFP

FIGURE 2. Translation regulation by arg CGA codon pairs is recapitulated with RNA-ID. (A,B) Insertion of increasing numbers of inhibitory CGA codons into the RNA-ID reporter results in progressively reduced GFP/RFP fluorescence. (A) Scatter plot of GFP fluorescence versus RFP fluorescence of yeast cells with constructs bearing the indicated sequences inserted in GFP. ((AGA)₃-GFP, teal; (CGA)₂-GFP, red; (CGA)₃-GFP, orange; (CGA)₄-GFP, purple; no ATG -GFP, brown). (B) Comparison of the median GFP/RFP value and the percentage of cells in each gate for each construct. The values reported (and the standard deviation) are the average of the median value obtained for each of four independent transformants. (C,D) Inhibition of translation by CGA codons is substantially suppressed by coexpression of a mutant tRNA^{Arg(UCG)*} that base pairs with CGA. (C) A bar graph of the GFP/RFP median of cells bearing integrated GFP constructs, and 2 μ plasmids that express no tRNA (vector), tRNA^{Arg(ICG)}, the mutant tRNA^{Arg(UCG)*}, or tRNA^{Arg(UCU)}. (D) Quantification of expression of data in C. Each value is the average of the median values obtained for four independent transformants with each plasmid. The tRNA^{Arg} species are indicated by their anticodons in the legend and table.

is substantially suppressed by introduction of tRNA^{Arg(UCG)*} to >75% of the expression of the (AGA)₃-GFP strain expressing tRNA^{Arg(UCG)*} (Fig. 2C,D). As expected, expression of (AGA)₃-GFP remains high and nearly constant ($\pm 10\%$) in strains expressing any of the tRNA species, including the major arg isoacceptor, which decodes AGA (Fig. 2C,D). Also as reported for luciferase constructs, expression of (CGA)₂(AGA)-GFP and of (CGA)₃-GFP is increased only modestly (approximately twofold) in strains expressing tRNA^{Arg(ICG)}, which uses wobble base-pairing to decode CGA, and is not affected by expression of tRNA^{Arg(UCU)}, which decodes AGA (Fig. 2C,D). Thus, both CGA inhibition and suppression of CGA inhibition by tRNA overexpression can be detected in this fluorescence-based system in a similar manner as reported for luciferase.

Four features of RNA-ID contribute to efficient separation of cells with sequences that affect expression

Four characteristics of the RNA-ID system that are important for its use in obtaining new *cis*-regulatory sequences emerge from the examination of CGA-mediated inhibition with the GFP reporter. First, there is a 250-fold dynamic range over which expression can be monitored, sufficient for most regulatory events. We found that the GFP/RFP

signal from a chromosomally integrated reporter with three arg AGA codons starting at amino acid 6 is 262-fold above the signal from an integrated construct in which GFP gene lacks an ATG codon, and thus should not be expressed efficiently (Fig. 2A,B). Similarly, we observe that the median GFP fluorescence from the integrated ATG-GFP is 283-fold greater than the median GFP fluorescence from an integrant that has neither the GFP nor the RFP gene (26,300 compared with 93) shown in Figure 1. The wide dynamic range between cells that exhibit high expression and no expression should enable detection and separation of *cis*-acting sequences that mediate various levels of expression at either end of the spectrum.

Second, the system is quantitatively sensitive to small differences in expression throughout the dynamic range. Even a single inhibitory codon pair is efficiently discriminated using this system: thus, (AGA)₃-GFP expression is 3.6-fold greater than that of (CGA)₂-GFP, which, in turn, is 4.5-fold greater than that of (CGA)₃-GFP, which has two overlapping CGA codon pairs (Fig. 2A,B). Even the dramatically reduced expression of a strain with (CGA)₄-GFP is ninefold above the background (Fig. 2A,B). Furthermore, distinct expression profiles are seen for strains in which the median GFP/RFP values differ by as little as 1.8-fold, for example, the scatter profiles from cells with (CGA)₃-GFP and (CGA)₄-GFP are clearly quantitatively distinct (Fig. 2A, compare the orange (CGA)₃ and purple (CGA)₄ scatter plots), although the median fluorescence signals are close (4.4 vs. 7.9).

Third, the resolution between yeast strains with moderately different expression levels is nearly perfect. Cells with an (AGA)₃-GFP construct can be nearly completely separated from cells with a (CGA)₂-GFP construct; that is, over 99.6% of these cells are easily separated, although the median GFP/RFP of cells with the (AGA)₃-GFP reporter is less than fourfold more than that of the (CGA)₂-GFP-containing cells. Similarly, 99.9% of cells bearing (CGA)₃-GFP can be distinguished from the population of cells with (CGA)₂-GFP, although their expression differs by only 4.5-fold. We used the remarkable separation of these distinct populations to assign expression gates, in which >99% of the indicated population resides within the indicated gate. Thus, Gate 1, the high expression gate, includes over 99% of cells with (AGA)₃-GFP or ATG-GFP; Gate 2, the intermediate expression gate includes over 99% of cells with (CGA)₂-GFP construct; Gate 3, the low expression gate, includes over 99% of cells with (CGA)₃-GFP or (CGA)₄-GFP constructs; and Gate 4, the no expression gate, includes over 99% of cells with a GFP insert that lacks a start codon (Fig. 2B).

Fourth, the GFP/RFP fluorescence is highly reproducible, even among independent transformants. We measured the median ratio for four independent transformants bearing (AGA)₃-GFP, (CGA)₂-GFP, (CGA)₃-GFP, and (CGA)₄-GFP. The median GFP/RFP values for these strains are

very similar, with an average percent standard deviation of 4.9% (Fig. 2B). Similarly, even among strains with an integrated copy of the reporter and plasmid-borne copies of tRNA genes, a condition that generally results in greater variability, the average percent standard deviation is 5.6% (Fig. 2D).

Effects of paromomycin and codon context on stop codon read-through are seen with the RNA-ID reporter

The efficiency of translation termination can be modulated by factors that increase misreading by the ribosome (Salas-Marco and Bedwell 2005; Fan-Minogue and Bedwell 2008) as well as by the sequence context surrounding the stop codon (Bonetti et al. 1995; Tork et al. 2004). To determine whether we could observe paromomycin-induced misreading and context-dependent stop codon read-through with the RNA-ID system, we examined a set of GFP reporters bearing stop codons at amino acid 7 that differ in the codons that flank the stop codons. In the absence of paromomycin, we found that all nonsense codon-containing constructs have a GFP/RFP ratio that is equivalent to a background construct lacking a start codon (Fig. 3B). However, in the presence of 100 µg/mL of paromomycin, we observed an increase in GFP/RFP expression for each of three constructs in which a stop codon is in a poor context (CAA at position 6 and CAA at position 8), but observe much less fluorescence from two constructs bearing the stop codons in a good context (CAA at position 6 and GCA or GAC at position 8) that is not expected to cause read-through (Fig. 3A,B). Both the dependence upon the concentration of paromomycin and the dependence upon the flanking sequences are consistent with results in the literature (Bonetti et al. 1995; Tork et al. 2004; Salas-Marco and Bedwell 2005; Fan-Minogue and Bedwell 2008), providing further evidence for the sensitivity and versatility of the system for translational regulation. The paromomycin-stimulated expression of the (CAA-TAA-CAA)-GFP is more than half that of the (CGA)₃-GFP construct measured in the absence of paromomycin. This paromomycin-stimulated expression of the nonsense-containing constructs is specific to nonsense codons, since growth in 100 µg/mL paromomycin results in a decrease in GFP/RFP median to ~60% from cells bearing a construct that is normally highly expressed, that is (AGA)₃-GFP (data not shown). Thus, the RNA-ID reporter system also responds to a known effector of misreading, as well as codon context.

A test library shows the existence of new inhibitory sequences

To determine whether there are inhibitory sequences similar to CGA codon pairs that can be identified with a short sequence insert, we constructed two small-scale

FIGURE 3. Paromomycin- and context-dependent stop codon read-through is recapitulated using the reporter for RNA *cis*-regulatory elements. (A) Effects of paromomycin on GFP fluorescence of strains with the TAA stop codon in good or poor contexts. Stop codons are flanked by sequences previously reported to cause low read-through (good context) or high read-through (poor context) (Bonetti et al. 1995). Scatter plots of GFP versus RFP fluorescence are shown in each set for a single strain grown in 0 µg/mL (red), 25 µg/mL (blue), and 100 µg/mL (orange) paromomycin. Strains on the *left* contain the insertion CAA-TAA-GCA beginning at codon 6 of GFP (good context), while strains on the *right* contain the sequence CAA-TAA-CAA at the same position (poor context). (B) Effects of paromomycin on median GFP/RFP levels from GFP constructs bearing stop codons in different sequence contexts. In the poor context, each stop codon is flanked by CAA on both its 5' and 3' side, while in the good context for TGA, the sequence is CAA-TGA-GAC, and for TAA, it is CAA-TAA-GCA. GFP/RFP medians were determined for four independent transformants of each construct at each concentration of paromomycin.

libraries, inserting three random or partially random codons at amino acid 6 of GFP. One library, called (NNN)₃, contains completely randomized sequences specifying codons 6, 7, and 8, and could specify 262,144 different sequence combinations, including stop codons. Since 3/64 codons are stop codons and since a stop codon in any one of the three positions should reduce expression to background levels, we expect 14% of the members of this library to migrate into Gate 4. To eliminate stop codons, the other library, called (VNN)₃, lacks T in the first position of the three semi-randomized codons 6, 7, and 8, and thus could specify 110,592 different sequence combinations. For each library, we obtained more than 100,000 *E. coli* transformants, prepared DNA, and obtained between 18,000 and 20,000 yeast transformants for analysis of expression. To examine the GFP expression profile from these libraries of transformants, 500,000 cells were subjected to fluorescence-activated cell sorting (FACS).

The libraries behave exactly as expected. We found that while most cells in the (VNN)₃ library exhibit high levels of expression (92%), there is far more heterogeneity in expression among cells from the libraries than is found with a single sequence, based on comparison to cells with an (AGA)₃-GFP reporter (Fig. 4A,B; see also Fig. 2A). The 3.4% of the cells from the (VNN)₃ library that are found in Gate 4, indicative of cells that have undetectable GFP levels expression, might have arisen due to cloning problems or problems with the oligonucleotide. On the other hand, the increase to 17.8% of cells from the (NNN)₃ library that migrated into Gate 4, nearly exactly matches the 14% increase predicted by the inclusion of stop codons in their composition. We also find that both libraries are substantially enriched in cells that migrate into the Gate 3 (low), (0.13% and 0.19% for (VNN)₃ and (NNN)₃, respectively) (Fig. 4C), compared with 0.04% of the (AGA)₃-GFP cells. Both libraries are likewise enriched in cells that migrate into Gate 2 (intermediate) (4.54% and 3.82% for (VNN)₃ and (NNN)₃, respectively) (Fig. 4C), compared with 0.2% of the (AGA)₃-GFP cells.

To determine whether the low expression is a reproducible phenotype, we isolated the population of cells that migrate in each gate, grew the cells in media to induce the reporter, and subjected this population to analytical flow cytometry. As shown in Figure 4, D and E, the low expressing population from the (VNN)₃ library (from Gate 3) is highly enriched for cells that exhibit poor expression in subsequent analysis: 89% of the cells isolated from Gate 3 migrate into either the low or intermediate gates (74% into Gate 3 and 15% into Gate 2), which is exactly the expected behavior, since much of the original population of cells from Gate 3 is found at the intersection of Gates 2 and 3 (Fig. 4B).

Furthermore, we individually cultured and reanalyzed 92 single yeast cells from Gate 3, 46 each from the (VNN)₃ and (NNN)₃ libraries, and found that 83 of these strains migrated into either the low or intermediate gates (Gate 3 or Gate 2), while nine strains migrated into the high-expression Gate 1. The 83 strains that exhibited intermediate or low expression are enriched for specific sequences, in that these 83 strains have 40 distinct sequences that are found in one to five copies, with a median copy number of 2. This suggests that the GFP expression level is mediated by that sequence rather than by the state of the cell. In contrast, the nine strains that migrate into the high-expression gate 1 each have a different sequence inserted upstream of GFP. The cause of poor expression for 13 of the 40 sequences is easily deduced, since these encode sequences with frameshifts (six sequences), with stop codons (six sequences), or with a strong secondary structure occluding the ATG (calculated ΔG of <-17 kcal/mole) (Mathews et al. 1999), which we had previously concluded was inhibitory (Letzring et al. 2010). To determine whether reduced GFP expression was caused by the inserted

FIGURE 4. Libraries of three randomized codon inserts at residue 6 of GFP results in a fraction of cells with reduced GFP expression. (A) Scatter plot from FACS of 500,000 yeast cells bearing a library of three randomized codons, (NNN)₃, inserted into GFP at residues 6–8. (B) Scatter plot from FACS of 500,000 yeast cells bearing a library of three semi-randomized codons, (VNN)₃, inserted into GFP at residues 6–8. V is A, G, or C. (C) Quantification of the fraction of the total population of each library that migrates in each gate. (D,E) Yeast strains that migrate into Gate 3, the low-expression gate, are strongly enriched for strains that exhibit low expression when regrown. (D) Scatter plot of GFP versus RFP fluorescence of cells from the (VNN)₃ library that migrated in Gate 3 in B, and were then regrown and re-examined by flow cytometry. (E) Quantification of the fraction of cells regrown from Gate 3 that migrate into each gate when reanalyzed.

sequences in the remaining 27 cases, we transformed yeast with a freshly made construct and found that in 17 of 27 cases the GFP/RFP fluorescence was nearly identical to that of the original construct. To determine whether the amino acid sequence is responsible for inhibition, we created each of the 17 amino acid sequences using optimal codons. We found that all 17 of the optimal codon variants exhibit at least sixfold greater GFP/RFP values compared with the remade nonoptimal codon variant. In 14 of the 17 cases, yeast bearing the optimal codon insertion clearly migrate into the high-expression gate 1. Thus, we infer that inhibition of GFP expression is due to either the nucleotide sequence or the codon composition, but not the amino acid insertions.

DISCUSSION

We have described a robust and sensitive method, RNA-ID, to identify RNA sequences that regulate gene expression, either due to their nucleotide sequence or structure, or because of the use of particular codons during translation. This integrated reporter for *cis*-regulatory elements in RNA uses a bidirectional *GAL1,10* promoter to drive expression of superfolder GFP (Pedelacq et al. 2006) and yeast codon-optimized mCherry GFP (Keppler-Ross et al. 2008), which provides the basis for nearly homogeneous

behavior of different cells with any particular insertion in GFP.

In this report, we demonstrate seven features of this system that make it outstanding for selection of regulatory sequences. First, the dynamic range is 250-fold, based on comparison of the signal from an integrated highly expressed GFP to the signal from a cell containing the GFP gene without an in-frame start codon. Second, measurements are quantitative and dose dependent, such that an increase in the number of inhibitory codons results in progressive reduction in the signal. Third, the resolution between strains with moderate differences in GFP expression is nearly complete, that is >99.7% of cells expressing one GFP construct are distinguishable and can be physically separated from cells expressing a construct with 25% of the activity of the first construct. Fourth, the signal is reproducible and dependent only upon the inserted sequence. The average percent standard deviation among different yeast transformants bearing the same sequence is <5%. Fifth, the system recapitulates known aspects of transla-

tional control from CGA codon-mediated inhibition and its suppression by an exact base-pairing tRNA^{Arg(UCG)*} to paromomycin-dependent stop codon read-through and context-dependent differences between stop codon efficiency. Sixth, the system is easily adaptable to screen for regulatory sequences upstream of or within the coding sequence, since the ATG start codon is absent from the vector and brought in during the LIC cloning step with the inserted sequence. Seventh, as was anticipated with the choice of the superfolder GFP, most inhibitory effects are not due to amino acid changes in the protein that might disrupt the fluorescent properties or cause instability of the protein. In addition, we provide evidence that there are indeed novel inhibitory sequences to be found in a small 9-nt (3 codon) random insert into the coding sequence of superfolder GFP, since we have found 17 sequences that replicate their inhibitory effects when remade in a new GFP clone, and have no obvious inhibitory feature such as a frameshift, secondary structure, or nonsense codon.

There are at least three direct applications for RNA-ID, since even in the well-studied, simple eukaryote yeast there is substantial evidence that translation regulation is extensive, with a wealth of *cis*-acting sequences implicated in regulating expression in diverse ways. First, RNA-ID can be used to identify combinations of codons, like CGA codon pairs (Letzring et al. 2010), which impair translation and

are not predicted based on the properties of the individual codons, whose identity is unknown. The importance of codon pairs as regulators of expression is also supported by early observations of codon pair bias and its correlation with expression in *E. coli* (Gutman and Hatfield 1989; Irwin et al. 1995; Boycheva et al. 2003) and by the reduced expression of poliovirus capsid genes recoded with underused codon pairs (Coleman et al. 2008). Second, RNA-ID can be applied to define the role of RNA sequences upstream of the translation start that modulate expression. Indeed, regulation of translation is likely to be far more complex and pervasive than earlier estimates, since analysis of translation in yeast with ribosome profiling indicated that there is significant translation upstream of the known start site for nearly one-quarter of the genes, and that there are changes in the relative translational efficiency for nearly one-third of genes during a short acute amino acid starvation (Ingolia et al. 2009). Moreover, there is massive translation regulation during meiosis, with evidence of novel ORFs and more than 10,000 meiotic-specific uORFs (Brar et al. 2012). Third, RNA-ID can be used to identify RNA sequences that interact with RNA-binding proteins to regulate expression. Estimates of the number of RNA-binding proteins in yeast and mammals continue to expand with development of methods to define the RNA-binding proteome (Butter et al. 2009; Tsvetanova et al. 2010; Castello et al. 2012; Gebauer et al. 2012).

In addition to its use in finding *cis*-regulatory sequences in RNA, there are three additional uses of this reporter method. First, the RNA-ID method can be used to quantify the effects of genes or mutations that impact specific aspects of translation, from stop codon read-through to no-go decay or nonsense-mediated decay. Second, the system might also be used to screen for amino acid sequences that promote stability or decay of the GFP protein. Third, RNA-ID could be used to identify RNA or codon sequences that confer conditional regulation. Thus, one might begin with a pool of high-expressing GFP variants and screen for a specific subset with reduced expression under particular conditions. Furthermore, if modified for use in other organisms or human cell lines, one could envision using RNA-ID as a screen for sequences that confer regulation dependent upon the cell type, for instance, the transformed cell phenotype or during the differentiation process.

MATERIALS AND METHODS

Strains, plasmids, and oligonucleotides

The yeast strain BY4741 (MATa *his3Δ1*, *leu2Δ0*, *met15Δ0*, *ura3Δ0*) was the parent strain for both linear and plasmid transformations. Plasmids containing the tRNAs *tR(UCU)K* [pDL866], *tR(ACG)D* [pDL867], and *tR(UCG)*[pDL869] on *LEU2* 2μ plasmids were obtained from D. Letzring (Letzring et al. 2010). The GFP plasmids

used here, and the oligonucleotides used to create these GFP variants, are shown in Supplemental Tables S1 and S2.

The vector pEKD1024, the integrating GFP–RFP reporter for *cis*-regulatory elements, was made by inserting the SacII–SphI fragment from pJE875, which contains the *GAL*-regulated GFP–RFP genes flanked by the *ADHI* and *CYC1* terminators, respectively, into the SacII–SphI sites in integrating vector pKD0999. pKD0999 was derived from JW132 (Whipple et al. 2011), which contains the components necessary for integration in the *ADE2* locus and the *MET15* gene, by insertion of oligonucleotides KD0315 and KD0316, into the BamHI and BbrPI sites to remove the tRNA gene, generate a multicloning site (BamHI–SphI–SacII–KpnI), destroy the BbrPI site, and add 31 nucleotides from the *MET15* promoter sequence to improve expression of *MET15*, which was reduced by the mutation in the BbrPI site (Thomas et al. 1989).

The plasmid pJE875 was derived in four steps from the plasmid BG2483, a 2μ, *URA3* vector with the *GAL1* promoter, which has been described previously (Quartley et al. 2009). First, a 672-bp fragment containing the *GAL10* promoter, a *SwaI* site designed to permit LIC cloning, and the *CYC1* terminator downstream from the *GAL10* promoter, were inserted into the *AgeI* and *SacI* sites, resulting in plasmid BG2596. Second, the *ADHI* terminator was inserted downstream from the *GAL1* promoter at the *KpnI* site, creating the plasmid BG2794. Third, superfolder GFP (Pedelacq et al. 2006) was inserted between the *EagI* and *KpnI* sites; and fourth, yeast codon-optimized mCherry RFP (Keppler-Ross et al. 2008) was inserted into the *SwaI* site with oligonucleotides HWIP2 and HWIP5, designed to regenerate a *SwaI* sequence for additional LIC cloning. The sequences inserted at each step in the construction of pJE875 are reported in the Supplemental Material. The sequence of the integrated DNA, from the *ADE2* ends, through the *MET15* gene, and the GFP and RFP genes is reported and annotated in the Supplementary Material (Supplemental Fig. 1).

Construction of GFP variants

To insert sequences into superfolder GFP using LIC cloning, 0.03–0.1 pmol of the vector pEKD1024 that had been digested with BbrPI and *PacI* and gel purified, was treated with 1 unit of T4 DNA polymerase (Novagen) in the presence of 2.5 mM dGTP (Roche), 5 mM DTT, and 1× T4 DNA Polymerase Buffer (Novagen) in a 20-μL reaction for 40 min at room temperature, followed by heat inactivation for 40 min at 75°C. The T4-treated vector (1 μL) was mixed with 0.067 pmole of annealed oligonucleotides (2 μL) for 5 min at room temperature, followed by addition of 1 μL of 25 mM EDTA (pH 8.0), and an additional incubation for 5 min at room temperature (Aslanidis and de Jong 1990; Alexandrov et al. 2004).

To anneal to the single-stranded regions of the vector, the top and bottom oligonucleotide each contain distinct single-stranded common sequences on their 5' ends (5'-AATTCCATCAACCTTAAT-3' for the top oligonucleotide, and 5'-CTTCCAAACCAC-3' for the bottom oligonucleotide). Constructs were transformed into *E. coli* cells (Novablue), then confirmed by sequence analysis. Similarly to insert sequences upstream of RFP, the *SwaI* cut and gel-purified vector is treated with T4 DNA polymerase in the presence of dGTP and mixed with oligonucleotides with a common single-stranded sequences on the 5' ends (5'-TAATCCAT

CAACCATTT-3' for the top oligonucleotide, and 5'-AAAC CATTCTCCTATTT-3' for the bottom oligonucleotide). To transform yeast, reporter constructs were digested with *Stu*I, releasing a 5737-nucleotide fragment with flanking *ADE2* sequences, which was then gel purified and used to transform yeast, selecting on SD-met plates.

Analytical flow cytometry

Strains were grown overnight at 30°C in liquid cultures in YP media containing 2% raffinose + 2% galactose + 80 mg/L Ade, then diluted in the morning to an OD₆₀₀ between 0.1 and 0.2 in the same media, and growth was continued for 4.5–5 h, to a final OD₆₀₀ from 0.5 to 1.5. In experiments involving strains with plasmids, cells were grown in appropriate synthetic drop-out media; paromomycin at 25 µg/mL or 100 µg/mL (MP Bio-medicals) was added as indicated for the experiment in Figure 3. After growth, $\sim 1.5 \times 10^6$ cells were added to a chilled 2058 tube (BD Falcon) containing cold 1× PBS (Diluted from 10× PBS [BioRad]), to a final volume of 500 µL. These tubes/cells were kept on ice until tested.

Flow cytometry was performed on an LSRII (BD Biosciences) at the URM C Flow Cytometry Core facility (FCC). GFP was excited with a 488-nm laser and detected using the 515/20-nm filter. RFP was excited by a 532-nm laser and detected using the 610/20-nm filter. Events were initially selected using the forward-scatter and side-scatter plot to exclude debris and budding yeast. Fluorescence measurements were standardized to expression from the ATG–GFP strain (derived from the *E. coli* plasmid pEKD1163) by setting the filter voltages so that both the GFP and RFP fluorescence intensities were $\sim 26,000$ and were similar to each other. For analytical flow cytometry, 10,000 events were collected for each sample containing a single sequence.

Construction and growth of libraries

To construct the (NNN)₃ and (VNN)₃ libraries, the LIC cloning reaction was performed with pEKD1024 plasmid and the oligonucleotides KD0404 and KD0350 for the (NNN)₃ library and oligonucleotides KD0437 and KD0350 for the (VNN)₃ library to produce 157,216 and 128,086 *E. coli* transformants, respectively, which were scraped and saved. DNA prepared from these cells was digested with *Stu*I, and 4.2 µg of gel-purified linear DNA was transformed into 140 mL of BY4741 yeast cells to produce 20,604 yeast transformants from the (NNN)₃ library and 18,721 yeast transformants from the (VNN)₃ library after selection on SD-met media. Transformants were replica plated once to SD-met, pooled by scraping, and aliquots were grown through two generations in selective media (S-met + 2% raffinose), diluted into 5 mL YP + 80 mg/L Ade + 2% raffinose + 2% galactose, at a starting OD₆₀₀ of 0.01, grown overnight at 30°C, then diluted into the same media at an OD₆₀₀ of 0.3, followed by 4 h of growth. Cell samples were prepared for flow cytometry as described above, except that 3×10^6 cells were diluted into a final volume of 1 mL with 1× PBS to ensure that 500,000 cells of each library could be sorted.

Fluorescence-activated cell sorting

Fluorescence-activated cell sorting (FACS) was performed on an AriaII (BD Biosciences) at the URM C FCC facility, as described for flow cytometry, except that GFP was detected with a 525/50-nm

filter. Fluorescence measurements were standardized to expression from the ATG–GFP strain by setting the filter voltages so that relative fluorescence of both GFP and RFP peaked between 1×10^4 and 5×10^4 . For the library populations, data from 500,000 events were evaluated, and cells in Gates 1, 2, and 3 were collected and saved.

Data analysis

Data analysis was performed using FlowJo software (Tree Star). Only events with an RFP fluorescence greater than 5×10^3 (RFP cutoff) were used for analysis. For data comparison, the GFP/RFP ratio was calculated by dividing the median GFP by the median RFP of a sample, and then multiplied by 100. The standard deviation is based on the median GFP/RFP ratio of four samples with identical sequence inserts that were individually grown and analyzed.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Eric Phizicky for advice during the course of the work and for comments on the manuscript and Dan Letzring and members of the Grayhack and Phizicky laboratories for helpful discussions. We are grateful to Erin Quartley (University of Rochester) for the pJE875 plasmid and to Geoffrey Waldo, Carolyn Bell (both of Los Alamos National Laboratory), and Neda Dean (Stony Brook University) for plasmids encoding superfolder GFP and the yeast codon-optimized mCherry. We also thank Mark Dumont and Jeffrey Zuber, as well as Tim Bushnell and members of the flow cytometry core for advice and discussion on flow cytometry and FACS. This work was supported by NSF grant MCB-0919658 awarded to E.J.G.

Received August 6, 2012; accepted September 14, 2012.

REFERENCES

- Alexandrov A, Vignali M, LaCount DJ, Quartley E, de Vries C, De Rosa D, Babulski J, Mitchell SF, Schoenfeld LW, Fields S, et al. 2004. A facile method for high-throughput co-expression of protein pairs. *Mol Cell Proteomics* **3**: 934–938.
- Aslanidis C, de Jong PJ. 1990. Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* **18**: 6069–6074.
- Barrera LO, Ren B. 2006. The transcriptional regulatory code of eukaryotic cells—insights from genome-wide analysis of chromatin organization and transcription factor binding. *Curr Opin Cell Biol* **18**: 291–298.
- Bartel DP. 2009. MicroRNAs: Target recognition and regulatory functions. *Cell* **136**: 215–233.
- Bonetti B, Fu L, Moon J, Bedwell DM. 1995. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in *Saccharomyces cerevisiae*. *J Mol Biol* **251**: 334–345.
- Boycheva S, Chkoderov G, Ivanov I. 2003. Codon pairs in the genome of *Escherichia coli*. *Bioinformatics* **19**: 987–998.
- Brar GA, Yassour M, Friedman N, Regev A, Ingolia NT, Weissman JS. 2012. High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* **335**: 552–557.

- Brohee S, Janky R, Abdel-Sater F, Vanderstocken G, Andre B, van Helden J. 2011. Unraveling networks of co-regulated genes on the sole basis of genome sequences. *Nucleic Acids Res* **39**: 6340–6358.
- Butter F, Scheibe M, Morl M, Mann M. 2009. Unbiased RNA–protein interaction screen by quantitative proteomics. *Proc Natl Acad Sci* **106**: 10626–10631.
- Castello A, Fischer B, Eichelbaum K, Horos R, Beckmann BM, Strein C, Davey NE, Humphreys DT, Preiss T, Steinmetz LM, et al. 2012. Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**: 1393–1406.
- Coleman JR, Papamichail D, Skiena S, Fitcher B, Wimmer E, Mueller S. 2008. Virus attenuation by genome-scale changes in codon pair bias. *Science* **320**: 1784–1787.
- Das MK, Dai HK. 2007. A survey of DNA motif finding algorithms. *BMC Bioinformatics* **8**: S21. doi: 10.1186/1471-2105-8-S7-S21.
- Fan-Minogue H, Bedwell DM. 2008. Eukaryotic ribosomal RNA determinants of aminoglycoside resistance and their role in translational fidelity. *RNA* **14**: 148–157.
- Gebauer F, Preiss T, Hentze MW. 2012. From *cis*-regulatory elements to complex RNPs and back. *Cold Spring Harb Perspect Biol* **4**: a012245. doi: 10.1101/cshperspect.a012245.
- Gertz J, Siggia ED, Cohen BA. 2009. Analysis of combinatorial *cis*-regulation in synthetic and genomic promoters. *Nature* **457**: 215–218.
- Gilligan PC, Kumari P, Lim S, Cheong A, Chang A, Sampath K. 2011. Conservation defines functional motifs in the squint/nodal-related 1 RNA dorsal localization element. *Nucleic Acids Res* **39**: 3340–3349.
- Gong C, Maquat LE. 2011. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. *Nature* **470**: 284–288.
- Goodarzi H, Najafabadi HS, Oikonomou P, Greco TM, Fish L, Salavati R, Cristea IM, Tavazoie S. 2012. Systematic discovery of structural elements governing stability of mammalian messenger RNAs. *Nature* **485**: 264–268.
- GuhaThakurta D. 2006. Computational identification of transcriptional regulatory elements in DNA sequence. *Nucleic Acids Res* **34**: 3585–3598.
- Gutman GA, Hatfield GW. 1989. Nonrandom utilization of codon pairs in *Escherichia coli*. *Proc Natl Acad Sci* **86**: 3699–3703.
- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104.
- Hinnebusch AG. 2005. Translational regulation of *GCN4* and the general amino acid control of yeast. *Annu Rev Microbiol* **59**: 407–450.
- Ingolia NT, Ghaemmaghami S, Newman JR, Weissman JS. 2009. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**: 218–223.
- Irwin B, Heck JD, Hatfield GW. 1995. Codon pair utilization biases influence translational elongation step times. *J Biol Chem* **270**: 22801–22806.
- Keppeler-Ross S, Noffz C, Dean N. 2008. A new purple fluorescent color marker for genetic studies in *Saccharomyces cerevisiae* and *Candida albicans*. *Genetics* **179**: 705–710.
- Letzring DP, Dean KM, Grayhack EJ. 2010. Control of translation efficiency in yeast by codon–anticodon interactions. *RNA* **16**: 2516–2528.
- Macdonald PM, Kerr K. 1998. Mutational analysis of an RNA recognition element that mediates localization of *bicoid* mRNA. *Mol Cell Biol* **18**: 3788–3795.
- Mathews DH, Sabina J, Zuker M, Turner DH. 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**: 911–940.
- Pedelacq JD, Cabantous S, Tran T, Terwilliger TC, Waldo GS. 2006. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**: 79–88.
- Plotkin JB, Kudla G. 2011. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* **12**: 32–42.
- Quartley E, Alexandrov A, Mikucki M, Buckner FS, Hol WG, DeTitta GT, Phizicky EM, Grayhack EJ. 2009. Heterologous expression of *L. major* proteins in *S. cerevisiae*: A test of solubility, purity, and gene recoding. *J Struct Funct Genomics* **10**: 233–247.
- Rabani M, Kertesz M, Segal E. 2008. Computational prediction of RNA structural motifs involved in posttranscriptional regulatory processes. *Proc Natl Acad Sci* **105**: 14885–14890.
- Raser JM, O'Shea EK. 2004. Control of stochasticity in eukaryotic gene expression. *Science* **304**: 1811–1814.
- Salas-Marco J, Bedwell DM. 2005. Discrimination between defects in elongation fidelity and termination efficiency provides mechanistic insights into translational readthrough. *J Mol Biol* **348**: 801–815.
- Schlabach MR, Hu JK, Li M, Elledge SJ. 2010. Synthetic design of strong promoters. *Proc Natl Acad Sci* **107**: 2538–2543.
- Thomas D, Cherest H, Surdin-Kerjan Y. 1989. Elements involved in S-adenosylmethionine-mediated regulation of the *Saccharomyces cerevisiae* *MET25* gene. *Mol Cell Biol* **9**: 3292–3298.
- Tork S, Hatin I, Rousset JP, Fabret C. 2004. The major 5' determinant in stop codon read-through involves two adjacent adenines. *Nucleic Acids Res* **32**: 415–421.
- Tsvetanova NG, Klass DM, Salzman J, Brown PO. 2010. Proteome-wide search reveals unexpected RNA-binding proteins in *Saccharomyces cerevisiae*. *PLoS ONE* **5**: e12671. doi: 10.1371/journal.pone.0012671.
- Walhout AJ. 2006. Unraveling transcription regulatory networks by protein–DNA and protein–protein interaction mapping. *Genome Res* **16**: 1445–1454.
- Wang H, Johnston M, Mitra RD. 2007. Calling cards for DNA-binding proteins. *Genome Res* **17**: 1202–1209.
- Whipple JM, Lane EA, Chernyakov I, D'Silva S, Phizicky EM. 2011. The yeast rapid tRNA decay pathway primarily monitors the structural integrity of the acceptor and T-stems of mature tRNA. *Genes Dev* **25**: 1173–1184.

RNA

A PUBLICATION OF THE RNA SOCIETY

RNA-ID, a highly sensitive and robust method to identify *cis*-regulatory sequences using superfolder GFP and a fluorescence-based assay

Kimberly M. Dean and Elizabeth J. Grayhack

RNA 2012 18: 2335-2344 originally published online October 24, 2012
Access the most recent version at doi:[10.1261/rna.035907.112](https://doi.org/10.1261/rna.035907.112)

Supplemental Material <http://rnajournal.cshlp.org/content/suppl/2012/10/11/rna.035907.112.DC1>

References This article cites 41 articles, 18 of which can be accessed free at:
<http://rnajournal.cshlp.org/content/18/12/2335.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *RNA* go to:
<http://rnajournal.cshlp.org/subscriptions>