Published in final edited form as:

Methods Enzymol. 2016 ; 572: 237–253. doi:10.1016/bs.mie.2016.02.003.

RNA-ID, a Powerful Tool for Identifying and Characterizing Regulatory Sequences

Christina E. Brule, Kimberly M. Dean, Elizabeth J. Grayhack¹

Department of Biochemistry and Biophysics, Center for RNA Biology, University of Rochester
School of Medicine and Dentistry, Rochester, New York, USA

Abstract

The identification and analysis of sequences that regulate gene expression is critical because regulated gene expression underlies biology. RNA-ID is an efficient and sensitive method to discover and investigate regulatory sequences in the yeast *Saccharomyces cerevisiae*, using fluorescence-based assays to detect GFP relative to an RFP control in individual cells. Putative regulatory sequences can be inserted either in-frame or upstream of a superfolder GFP fusion protein whose expression, like that of RFP, is driven by the bidirectional *GAL1,10* promoter. In this chapter, we describe the methodology to identify and study *cis*-regulatory sequences in the RNA-ID system, explaining features and variations of the RNA-ID reporter, as well as some applications of this system. We describe in detail the methods to analyze a single regulatory sequence, from construction of a single GFP variant to assay of variants by flow cytometry, as well as modifications required to screen libraries of different strains simultaneously. We also describe subsequent analyses of regulatory sequences.

Keywords

RNA-ID; flow cytometry; translation; yeast

1. Introduction

Identification of elements in DNA or RNA that regulate gene expression is crucial to mapping regulatory pathways and elucidating the mechanisms of such regulation. There is no comprehensive list of RNA or DNA regulatory sequences, despite the advances in technology that have made it easier to identify regulatory elements, such as high-throughput methods to identify motifs recognized by RNA binding proteins or sequences targeted by miRNAs (Campbell and Wickens, 2015; Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., et al., 2010). However, methods to identify sequences that modulate expression *in vivo* are also needed, since comprehensive identification of sequences regulated by RNA binding proteins will require mapping the targets of over 700 RNA binding proteins even in the single celled yeast *Saccharomyces cerevisiae* (Matia-Gonzalez, A.M., Laing, E.E., and Gerber, A.P., 2015), and since some regulatory sequences, such as synonymous codons, are not bound by proteins. RNA-ID is a method to identify and

¹Corresponding author: elizabeth_grayhack@urmc.rochester.edu.

study *cis*-regulatory sequences, based on their effects on expression of superfolder GFP (Fig. 1A) (Dean and Grayhack, 2012); putative regulatory sequences are inserted either in-frame with GFP or upstream of GFP and are assayed by flow cytometry.

RNA-ID is an effective method that has been particularly useful in analysis of sequences whose effects are mediated by translation. Several features contribute to the efficacy of the RNA-ID system. First, the signal is robust, allowing differences in expression to be measured over a 250-fold range. Second, noise is minimized by normalization of GFP to an RFP that is controlled by the same promoter and by integration of the reporter into the chromosome (eliminating copy number effects). Third, results are reproducible over biological replicates (standard deviations are typically less than 5% for 4 independent biological replicates) (Dean and Grayhack, 2012). This method can also be coupled to deep sequencing for a high-throughput method, which has been used to identify features of tRNAs required for a functional tRNA (Guy, M.P., Young, D.L., Payea, M.J., Zhang, X., Kon, Y., Dean, K.M., et al., 2014).

2. Features of the RNA-ID System

The RNA-ID system is a fluorescence-based assay where the bidirectional *GALI,10* promoter drives expression of both a GFP fusion protein and RFP (Fig. 1A). Both the GFP and RFP genes were designed to carry unique restriction sites flanked by sequences that allow ligation-independent cloning (LIC) (Aslanidis and de Jong, 1990) to insert novel sequences into either GFP or RFP. Furthermore, the reporter is integrated into the chromosome to eliminate copy number effects and expression of GFP or RFP variants is assayed using flow cytometry.

We used superfolder GFP to diminish expression effects from different amino acid insertions, since superfolder GFP has robust fluorescence when insoluble proteins are fused to its N-terminus (Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S., 2006). Thus, GFP expression from RNA-ID should be relatively insensitive to inserted amino acids. Indeed, the difference in GFP expression from six different fusion reporters with varying lengths of *GLN4* (from 5 amino acids to 99 amino acids) fused to GFP is small, from 60 to 87 GFP/RFP units (Wolf and Grayhack, 2015). In the RNA-ID reporter, superfolder GFP is preceded by a site for 3C protease, an HA epitope, and His6, all of which facilitate analysis and purification of the resulting protein (Quartley, E., Alexandrov, A., Mikucki, M., Buckner, F.S., Hol, W.G., DeTitta, G.T., 2009). Antibody detection of the HA epitope, which allows analysis of the polypeptide size in SDS-PAGE gels, was used to demonstrate the presence of a full-length frameshifted protein (Wolf and Grayhack, 2015). Moreover, sequences are inserted outside of the essential GFP regions just upstream of the HA epitope, and can even be inserted upstream of the start codon AUG, outside of the coding sequence.

Normalization of GFP to RFP reduces noise between cells of different sizes and reduces extrinsic noise that occurs due to heterogeneity or stochastic events (Raser and O'Shea, 2004), since the expression of these two proteins is driven by the same regulatory elements in the bidirectional *GALI,10* promoter. Additionally, RFP is used to impose a cutoff for

cells in which GAL induction is not effective, since RFP expression is presumably independent of the regulatory sequences in GFP. The RFP in the RNA-ID reporter is a version of mCherry that has been recoded with optimal codons for high expression in yeast (Keppler-Ross, S., Noffz, C., and Dean, N., 2008).

Derivatives of the RNA-ID reporter have been constructed to analyze the effects of regulatory sequences at internal locations (Fig. 1B, 1C). A specific reporter was constructed to allow analysis of effects on sequences both upstream and downstream of the regulatory sequence (Letzring, D.P., Wolf, A.S., Brule, C.E., and Grayhack, E.J. 2013). Thus, in a vector with a *Renilla* luciferase-GFP fusion protein, the regulatory sequence is inserted at amino acid 318 downstream of *Renilla* luciferase and upstream of GFP (Fig. 1B). Inhibitory CGA codons result in the recruitment of a quality control apparatus that targets the upstream nascent polypeptide, *Renilla* luciferase, for degradation (Letzring et al., 2013). Deletion of *LTN1*, which encodes the E3 ubiquitin ligase component of the quality control apparatus, restores *Renilla* luciferase activity, but does not affect readthrough of the codons and thus does not affect GFP (Letzring et al., 2013). The ability to measure both the upstream and downstream consequences of a regulatory sequence are invaluable. In another example, varying lengths of the N-terminal domain (NTD) of *GLN4* were placed upstream of the regulatory sequence to measure the effects of internal gene location on GFP expression (Fig. 1C) (Wolf and Grayhack, 2015).

LIC cloning in the RNA-ID system allows efficient and robust cloning of large numbers of different sequences without regard to the restriction sites present in the inserted sequences. In LIC cloning (Fig. 2), the linearized vector is incubated with T4 DNA polymerase in the presence of a single dNTP, which inhibits its exonuclease activity at the complementary nucleotide, thus generating specific single stranded ends of 12–18 nucleotides. A major advantage of using LIC cloning is that all PCR products generated with the appropriate LIC complementary ends can be efficiently cloned in this system, even if they do not contain any (or contain multiple) sites for the restriction endonucleases used to cleave the vector. Indeed, we used LIC cloning methods to clone and express more than 4,000 ORFs from *Leishmania major*, *Trypanosoma brucei*, *Trypanosoma cruzi*, *Plasmodium falciparum*, and closely related organisms (Quartley et al., 2009), and developed a system to co-express multiple genes without requirement for PCR amplification using LIC to join two vectors (Alexandrov, A., Vignali, M., LaCount, D.J., Quartley, E., de Vries, C., De Rosa, D. et al., 2004). Furthermore, LIC cloning has been used to ligate multiple PCR products in a single reaction; these constructs also contained unique sequences (added during PCR) for identification of the particular construct (Quivey, R.G., Jr., Grayhack, E.J., Faustoferri, R.C., Hubbard, C.J., Baldeck, J.D., Wolf, A.S., 2015).

3. Applications of the RNA-ID System

There are several applications of the RNA-ID system; here we describe examples of three major applications and suggest potential applications. RNA-ID has been used to study translation (e.g. inhibitory codons, chemical inhibitors, frameshifting), trans-acting factors (e.g. tRNA functionality), and the effects of mutations on regulation (e.g. *asc1* on inhibitory codons and met22 on tRNAs).

The RNA-ID reporter has been used extensively to measure translation inhibition, particularly the effects of synonymous codons on expression, (Dean and Grayhack, 2012). It has long been known that the choice of synonymous codons modulates translation efficiency, but the identity of codons or codon combinations that impede translation are unknown (Letzring, D.P., Dean, K.M., and Grayhack, E.J., 2010; Plotkin and Kudla, 2011). Such inhibitory codon-mediated signals are *cis*-acting regulators of translation that are almost certainly not recognized by RNA or DNA binding proteins. Thus, their detection is entirely dependent on a functional assay for expression, like RNA-ID.

In a systematic analysis, we had previously demonstrated that the Arg CGA codon is poorly translated due to a defect in I•A wobble decoding (Letzring et al., 2010). To examine effects of the Arg CGA codons in RNA-ID, GFP/RFP from reporters with CGA codons was compared to GFP/RFP from synonymous reporters with optimal AGA codons to control for amino acid effects. The sensitivity of the RNA-ID system is illustrated by the demonstration that the effects from a single CGA-CGA codon pair are easily detected by flow cytometry (Fig. 3). Either overproducing the native tRNA, tRNA^{Arg(ICG)}, or expressing a non-native, exact base-pairing tRNA, tRNA^{Arg(UCG)*}, resulted in increased GFP/RFP from a reporter with CGA codons, but not from a synonymous reporter with AGA codons (Dean and Grayhack, 2012). This demonstrated that reduced expression was specifically due to translation of CGA codons. The RNA-ID system was also used to measure the effects of chemicals on translation of reporters with stop codons, as well as reporters with CGA codons. As expected based on previous analysis of stop codon readthrough (Salas-Marco and Bedwell, 2005), paromomycin enhanced readthrough of several stop codons in GFP, while cycloheximide had no effect (Dean and Grayhack, 2012). We also found that expression of CGA-containing reporters increased in the presence of paromomycin, suggesting that CGA codons are recognized as near-cognate mismatches in the A site of the ribosome (Letzring et al., 2013). Frameshifting events can also be detected using the RNA-ID system because of the large dynamic range, that allows detection of expression levels as low as 1% of the in-frame construct (Wolf and Grayhack, 2015).

RNA-ID has also been used to study *trans*-acting regulatory factors, specifically to identify the determinants of tRNA function in a high-throughput analysis of >25,000 tRNA variants (Guy et al., 2014). tRNA function requires a similarity in shape between all tRNAs for their role in translation, but differences in their individual sequence to allow specific aminoacylation by tRNA synthetases and unique modifications (Phizicky and Hopper, 2010). To study the features of tRNAs that determine tRNA function and susceptibility to Rapid tRNA Decay (RTD), Guy et al. (Guy et al., 2014) generated a library of *SUP4_{oc}* tRNA variants with a UUA anticodon that can readthrough a UAA stop codon. The stop codon is inserted into GFP, thus GFP expression reports on the functionality of a given tRNA variant. *SUP4_{oc}* tRNA function was remarkably tolerant of mutations with 70 single mutations that were highly functional although over half of these mutations resulted in RTD of the tRNA (Guy et al., 2014). Furthermore, this method of analyzing tRNA variants could be applied to any regulatory protein, where GFP expression reports on the functionality of protein variants.

RNA-ID system is a powerful genetic tool that can be used to find components in a regulatory system or to study the effects of mutations on regulation. Specifically, RNA-ID has been used to study the effects of mutations on codon-mediated regulation of translation (Letzring et al., 2013; Wolf and Grayhack, 2015) and to study the determinants of tRNA susceptibility to the RTD pathway (Guy et al., 2014). *ASC1* is a ribosomal small subunit protein involved in quality control pathways related to stalled ribosomes (Brandman, O., Stewart-Ornstein, J., Wong, D., Larson, A., Williams, C.C., Li, G.W., et al., 2012; Kuroha, K., Akamatsu, M., Dimitrova, L., Ito, T., Kato, Y., Shirahige, K., et al., 2010). Since ribosomes are likely to stall at CGA codons (Letzring et al., 2013), we measured GFP/RFP in wild-type and *asc1* strains from reporters with either CGA or AGA codons (Letzring et al., 2013; Wolf and Grayhack, 2015); indeed, deletion of *ASC1* improved readthrough of CGA codons at amino acid 318 (relative to the AGA control), but not of CGA codons at amino acid 6 (Letzring et al., 2013; Wolf and Grayhack, 2015). The synonymous control is very important in measurements in mutant strains because some mutations affect expression of the GFP or RFP reporters themselves. Similarly, Guy et al. (Guy et al., 2014) defined RTD substrates by comparing tRNA function in wild-type and *met22* strains, and then verified these results by reconstruction of variants and direct measurement of levels of *SUP4_{oc}* tRNA variants.

4. Analysis of *Cis*-Regulatory Sequences in the RNA-ID Reporter

In this section, we describe the steps to construct RNA-ID reporters with specific sequences, integrate these reporters into the yeast genome and analyze GFP/RFP expression by flow cytometry. A schematic of the steps involved in this entire process is shown in Fig. 4.

4.1 Cloning sequences into the RNA-ID reporter

To clone sequences into the RNA-ID reporter requires that both the vector and the insert have complementary single stranded ends, and thus the first steps in this process are designed to generate those ends. After this preparation, annealing the insert and vector is sufficient for transformation. In the RNA-ID vector, GFP lacks an initiating AUG and thus there is no GFP expression from the base vector (pEKD1024). This allows insertion of inserts with the variable region either upstream or downstream of the start codon.

Preparation of vector for cloning into GFP.

1. To obtain vector with appropriate ends, incubate 5–20 µg of pEKD1024 DNA with two restriction enzymes, PacI (NEB) and BbrPI (Roche), in a 100–400 µl reaction. Purify the vector by electrophoresis in a 1% LMP agarose (Invitrogen) gel in 1X TAE buffer (40 mM Tris–acetate, 1 mM EDTA, pH 8.1, from a 50X stock) at 4 °C. Excise the 10,184 bp band, and extract DNA from the gel using the QIAQuick Gel Extraction Kit (Qiagen). The concentration of the gel purified DNA (ng/ul), measured on a NanoDrop Spectrophotometer at 260λ using an extinction coefficient of 50 ng-cm/µl, is needed to calculate the pmoles/µl, which is essential for the resection with T4 DNA polymerase. (*Note: use conversion factor of 660 g/mol•bp).

2. To generate the single stranded ends, incubate 0.03 to 0.10 pmol of gel purified vector DNA with T4 DNA polymerase (Roche) in the presence of dGTP (Roche) (Fig. 5A). On ice, mix the following in a 1.5 ml microcentrifuge tube: 2.0 μ l of T4 buffer (Roche), 2.0 μ l of 25 mM dGTP, 1.0 μ l of 100 mM DTT, 0.03 to 0.10 pmol of gel purified vector, 0.8 μ l of T4 DNA Polymerase (Roche, CAT# 70099), ddH₂O up to 20.4 μ l. Incubate at 22 °C (room temperature) for 40 minutes, then inactivate the T4 DNA polymerase by incubating the reaction at 75 °C for 40 minutes. Store at –20 °C.

Preparation of inserts.: Here, we describe three different inserts for LIC cloning into GFP: a PCR product, oligonucleotides and a partially randomized sequence.

1. To insert a large DNA fragment, it is necessary to amplify the desired DNA by PCR using oligonucleotide primers with appropriate sequences added to their 5' end (Fig. 5B). Design oligonucleotides to PCR amplify the desired DNA fragment and to the 5' end of the forward oligonucleotide add the following sequence: 5'- AATTCCATCAACCTTAAT, to the 5' end of the reverse oligonucleotide add the following sequence: 5'-CTTCCAAACCAC. After PCR amplification, purify the DNA on a 1% LMP agarose gel and incubate 0.12–0.48 pmoles with T4 DNA polymerase (in a 20.4 μ l reaction, as described for the vector) in the presence of dCTP, the dNTP complementary to that used on the vector.
2. Sequences up to 42 nucleotides can be easily inserted with 2 complementary oligonucleotides that contain the test sequence and the appropriate sequences for LIC added to their 5' ends (5'- AATTCCATCAACCTTAAT to the forward oligonucleotide and 5'-CTTCCAAACCAC to the 5' end of the reverse oligonucleotide: (Fig. 5C). Make a 25 μ M working stock of each of the forward and reverse oligonucleotides by diluting in TE 8 (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), then mix 4 μ l of the forward oligonucleotide, 4 μ l of the reverse oligonucleotide, 3 μ l of 0.5 M NaCl, and 19 μ l ddH₂O in a 1.5 ml microcentrifuge tube. Boil for 2–3 minutes using either a beaker of water over an open flame or a heat block (be sure to pre-boil the water or pre-heat the heat block before starting the time). Then, let the reaction cool slowly to 55–65 °C (either by setting the entire beaker or the heat block on the bench). For each set of annealed oligonucleotides, make a 1:50 dilution in ddH₂O and store at –20 °C.
3. To generate a library of variants or insert a longer sequence, we use two oligonucleotides, one of which only contains the non-randomized portion of the sequence. We use a minimum of 13 nucleotides that form a complementary double stranded region, but the remaining sequences can be single stranded (Fig. 5D). For a longer sequence, simply overlap the two oligonucleotides by 13 nucleotides and extend the insert with the reverse oligonucleotide. Subsequent preparation of the insert is the same as described for annealing two oligonucleotides.

Annealing the insert and vector.: To complete the RNA-ID reporter, the vector and insert are annealed together without the need for ligases.

1. To a 1.5 ml microcentrifuge tube, add 1 μ l of T4 treated vector (be sure the 1 μ l drop is at the bottom of the microcentrifuge tube). Add 2 μ l of the 1:50 dilution of the annealed oligonucleotides or T4-treated PCR reaction (slowly eject 2 μ l into the 1 μ l drop and physically mix with pipette tip). Incubate the reaction at 22 °C (room temperature) for 5 minutes.
2. Add 1 μ l of 25 mM EDTA (slowly eject into mix and physically mix with pipette tip). Incubate the reaction at 22 °C (room temperature) for 5 minutes. Store at 4 °C, or immediately transform into competent *E. coli* cells.

Variation: cloning into RFP.: To clone into RFP, the vector is prepared for LIC cloning by incubating DNA with the restriction endonuclease *Swa*I, followed by gel purification and other steps described for the GFP vector. Inserts are prepared by adding common sequences (different from the GFP common sequences) to the 5' ends of the forward oligonucleotide (5'-TAATCCATCAACCATTT) and the reverse oligonucleotide (5'-AAACCATTCTCCTATTT).

4.2 Transformation of the RNA-ID reporter into yeast

Linearization of DNA.—To obtain a linear DNA fragment with the RNA-ID reporter flanked on each end by sequences homologous to *ADE2*, the DNA is incubated with *Stu*I (NEB), then subjected to phenol extraction and ethanol precipitation. Specific and efficient integration into the *ADE2* locus is facilitated by the ~400 bp of *ADE2* sequences on either end of the linear fragment of DNA (~5737 nucleotides; will vary based on insertion length).

Transformation of the RNA-ID reporter into yeast.—Each transformation requires 10 ml of cells, grown to OD₆₀₀ between 1.3 and 2.5 (log phase), ~300 ng of *Stu*I cut DNA, and yields 50–200 transformants. Directions are given for a single transformation, but step 1 can be performed in bulk.

1. To prepare competent yeast cells for transformation, grow the desired yeast strain in YPD to OD₆₀₀ between 1.3 and 2.5 (log phase). For each transformation, pellet 10 ml of cells by centrifugation at 3,000 rpm for 5 minutes. Decant supernatant. Wash twice with half volume of 100 mM LiOAc. Resuspend cells in 100 μ l of 100 mM LiOAc per transformation. Add 10 μ l salmon sperm DNA per transformation.
2. Aliquot 110 μ l of competent cells to one 2059 tube (13 ml conical tube; Falcon) per transformation. Add 300 ng of linearized DNA to each tube and tap mix. Incubate at 30 °C in a water bath for 15 minutes (do not shake). After the first incubation, add 600 μ l of a LiOAc-PEG solution (8 volumes of 60% PEG 3350 to 1 volume of 1M LiOAc to 1 volume of ddH₂O) to each tube and tap mix. Incubate at 30 °C in a water bath for 20 minutes (do not shake). After the second incubation, add 68 μ l of DMSO to each tube and tap mix. Heat shock cells at 42 °C in a water bath for 15 minutes, but do not shake. After the third incubation,

pellet cells by centrifugation at 1500 rpm for 5 minutes. Carefully remove the LiOAc-PEG solution using a pipette man.

3. To facilitate growth of transformed cells, include a recovery period where each tube receives 1 ml YPD (gently tap mix to resuspend cells) and is then incubated in a 30 °C shaker at 250 rpm for 3–4 hours. After this recovery period, plate 250 μ l of each transformation on a SD-met plate and incubate at 30 °C for 2 days. These plates have a background of non-transformed cells; thus, individual colonies are streaked onto SD-met plates to obtain single colonies and are incubated at 30 °C for 2 days.

4.3 Flow cytometry

Yeast strains carrying RNA-ID reporters are grown in YP + 2% raffinose + 2% galactose + 80 mg/L Ade (Sherman, F., Fink, G., and Hicks, J.B., 1986) for two reasons: 1. The bidirectional *GAL1,10* promoter is repressed in the presence of dextrose, but induced in the presence of galactose. 2. The RNA-ID reporter is integrated into the yeast chromosome at the *ADE2* locus, thus the cells are deficient in adenine biosynthesis causing them to turn pink; this pigment increases the background in the GFP channel.

Growth of strains for flow cytometry.—Single colonies from the transformation (or from saves) are used to inoculate 5 ml YP + 2 % raffinose + 2 % galactose + 80 mg/L Ade and grown overnight (~14h) at 30 °C. Then each culture is diluted into 5 ml of the same media to OD₆₀₀ between 0.1–0.2 and grown for 4–6 hours to OD₆₀₀ of ~0.8 (Dean and Grayhack, 2012). For each strain, 500 μ l of culture is transferred to a 5 ml polystyrene round-bottom tube (Falcon) and kept on ice (for up to an hour) until analysis by flow cytometry. Flow cytometry and downstream analysis is performed for at least 3, preferably 4, independent isolates of each strain (Dean and Grayhack, 2012).

Reference and control strains.—A reference strain carrying the RNA-ID reporter with ATG inserted into GFP is used as the primary control in all flow cytometry experiments (Dean and Grayhack, 2012). GFP and RFP voltages are adjusted until fluorescence in both channels is 26,000, resulting in a 1:1 ratio of GFP:RFP. GFP and RFP from several control strains, that have been extensively analyzed, are measured in each experiment to assure reproducibility: (AGA)₃, a reporter with high GFP/RFP; (CGA)₃, a reporter with low GFP/RFP; JW132, the base vector for the RNA-ID reporter (does not contain GFP or RFP) (Whipple, J.M., Lane, E.A., Chernyakov, I., D'Silva, S., and Phizicky, E.M., 2011); and an RNA-ID reporter lacking a GFP insert (GFP[−] RFP⁺).

4.4 Tips for analysis of a library

Generating a large library of variants.—To adequately cover the sequence scope of many libraries, it is important to obtain 50,000–100,000 yeast transformants. It is difficult to obtain this number of chromosomally integrated yeast transformants using many protocols. Yoshiko Kon and E. Phizicky (personal communication) found that using ~2 μ g of DNA to transform 10 ml of yeast at OD₆₀₀ of 1.3–2.5 resulted in >50,000 yeast transformants; this protocol was used by Guy et al. (Guy et al., 2014).

Purifying transformants.—Analyzing libraries of transformants presents a unique difficulty since each plate of yeast transformants also contains many non-transformed cells that are viable and will grow if transferred to rich media. Thus, it is necessary to impose an additional round of selection for the transformed yeast, but this must be done in bulk. We have resolved this issue in two ways, either by replica plating the transformation plates onto selective media (Dean and Grayhack, 2012) or by scraping and pooling transformation plates, followed by additional growth in media that selects for transformants (Guy et al., 2014), the latter of which seems to be the more effective approach.

FACS: Fluorescence-activated cell sorting.—The protocol for growing libraries of yeast for FACS analysis is described in Guy et al. (Guy et al., 2014). During the growth required to induce the GFP and RFP expression, cells need to be maintained in log phase, which necessitated a series of dilutions prior to sorting. To maintain the complex relative composition of the library, it is extremely important to inoculate each dilution series with 10x or more cells corresponding to the original number of transformants. FACS instrumentation and the gate set-up is described in Dean and Grayhack, 2012.

5. Verification of Regulatory Sequences

RNA-ID is a useful tool for verifying the effects of regulatory sequences on translation and has been used extensively to distinguish between codon-mediated regulation and sequence/structure-mediated regulation. For example, inhibition caused by CGA codons is suppressed by expressing either tRNA^{Arg(ICG)} or tRNA^{Arg(UCG)*}, indicating that the inhibitory defect of CGA codons occurs in translation.

A second method to assess whether or not codon use is important for the observed regulation is to determine if the reading frame is required for inhibition. In this case, the inserted sequence is moved out of frame and strains carrying the reporter with the frame-shifted insert are analyzed by flow cytometry. An example of this type of analysis is shown in Fig. 6. The inhibitory sequence GGA CCT CGG was identified by sequencing GFP variants isolated from a low expression pool of cells obtained by FACS (Dean and Grayhack, 2012). To determine if this sequence or translation of these codons is responsible for low GFP, we reconstructed the original variant and a frame-shifted derivative (Fig. 6A), both of which yield low GFP/RFP (Fig. 6B). Thus, inhibition by this sequence is not due to synonymous codon effects.

6. Conclusions and Extensions of the RNA-ID reporter system

While the methodology described above provides useful tools for identification and characterization of regulatory sequences, three limitations of the current RNA-ID system could be addressed in the future. First, re-engineering the promoter and UTR regions of the RNA-ID vector would allow analysis of regulatory elements in different growth conditions, in different stresses, with different expression levels, and with different transcription start sites, and could also allow analysis of uORFs. Use of the *GAL1,10* promoter to direct GFP and RFP expression allows good regulation of expression, but is not compatible with growth of yeast in its preferred carbon source, glucose. Since the *GAL1,10* regulatory region is

flanked by unique restriction sites (PacI and SpeI), it would be relatively simple to replace the promoter and 5' UTR sequences in the current vector with other bi-directional yeast promoters, such as the *RPL13A*, *RPL16B* promoter (Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Munster, S., Camblong, J., et al., 2009). This might facilitate detection of sequences that only regulate expression in particular conditions. Second, the vector could be converted to allow cloning with standard restriction enzyme technology, simply by introducing restriction sites with 5' or 3' overhangs at the PacI and BbrPI sites. Third, since many regulatory signals are found in the 3' UTR, we could modify the vector to allow cloning at the 3' end of the GFP. Thus, RNA-ID is a robust and sensitive method to identify regulatory sequences that can be expanded in new directions.

Acknowledgements

We thank Eric Phizicky, Andrew Wolf, Yoshiko Kon, and Michael Guy for their important contributions in developing the approach described in this chapter. This work was supported by NSF grant MCB-1329545 to E.J.G. C.E.B. was partially supported by NIH Training Grant in Cellular, Biochemical, and Molecular Sciences 5T32 GM068411.

ABBREVIATIONS

FACS	fluorescence-activated cell sorting
LIC	ligation independent cloning
GFP	green fluorescent protein
RFP	red fluorescent protein
NTD	N-terminal domain
RTD	rapid tRNA decay

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. [PubMed: 15240823]
- Aslanidis C, and de Jong PJ (1990). Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res* 18, 6069–6074. [PubMed: 2235490]
- Brandman O, Stewart-Ornstein J, Wong D, Larson A, Williams CC, Li GW, Zhou S, King D, Shen PS, Weibezahn J, et al. (2012). A ribosome-bound quality control complex triggers degradation of nascent peptides and signals translation stress. *Cell* 151, 1042–1054. [PubMed: 23178123]
- Campbell ZT, and Wickens M (2015). Probing RNA-protein networks: biochemistry meets genomics. *Trends Biochem Sci* 40, 157–164. [PubMed: 25636997]
- Dean KM, and Grayhack EJ (2012). RNA-ID, a highly sensitive and robust method to identify cis-regulatory sequences using superfolder GFP and a fluorescence-based assay. *RNA* 18, 2335–2344. [PubMed: 23097427]
- Guy MP, Young DL, Payea MJ, Zhang X, Kon Y, Dean KM, Grayhack EJ, Mathews DH, Fields S, and Phizicky EM (2014). Identification of the determinants of tRNA function and susceptibility to rapid tRNA decay by high-throughput in vivo analysis. *Genes Dev* 28, 1721–1732. [PubMed: 25085423]
- Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr., Jungkamp AC, Munschauer M, et al. (2010). Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* 141, 129–141. [PubMed: 20371350]

- Methods Enzymol.* Author manuscript; available in PMC 2019 September 18.

Figure 1. The RNA-ID reporter and its derivatives.

(A) In the RNA-ID vector, expression of two reporters GFP and RFP is driven by the bidirectional *GAL1,10* promoter. The GFP reporter is a fusion protein encoding a 3C protease site, an HA epitope, His6, (marked by dark box) followed by superfolder GFP. Putative regulatory sequences are inserted into the *PacI*, *BbrPI* sites (GFP) or into the *Swal* site (RFP) using LIC cloning. The GFP lacks a start codon, allowing insertion of sequences upstream or within the coding sequence. (B) The *Renilla* luciferase-GFP fusion protein allows for analysis of the upstream nascent polypeptide and uses the same restriction sites for cloning. (C) In the *GLN4*-GFP fusion protein, varying lengths of *GLN4* can be PCR amplified to allow for the insertion of sequences at different locations relative to the start codon.

Figure 2.
An illustration of the general principles of LIC.

Figure 3. Demonstration of the sensitivity of the RNA-ID system, showing a flow cytometry scatter plot in which each cell is represented by a dot. $(AGA)_3$ has a GFP/RFP value of 124 and $(CGA)_2\text{-AGA}$ has a GFP/RFP value of 26.

Figure 4. Flowchart of the steps involved in analysis of cis-regulatory sequences in the RNA-ID system.

Figure 5.

Preparation of vector and inserts for LIC. (A) A schematic of the vector pEKD1024 showing the major steps to generate single stranded DNA ends for LIC using T4 DNA polymerase and dGTP. Restriction enzyme sites are shaded and the nucleotide where the T4 DNA polymerase stops is indicated by white text over a dark background. (B, C, D) Schematics of various types of inserts that can be cloned into the vector pEKD1024. (B) A PCR amplified insert and the product after treatment with T4 DNA polymerase in the presence of dCTP. (C, D) Inserts formed with oligonucleotides. (C) Complete overlap of inserted sequences. (D) Partially randomized sequence with 13 nucleotide overlap.

Figure 6.

Analysis to determine if reading frame is important for regulation. (A) The inhibitory sequence in frame and in the +1 frame and the optimal sequence in frame are shown. To frameshift the inhibitory sequence, a single nucleotide (white text, shaded background) is inserted upstream of the Val GUU and a single nucleotide is deleted downstream (the third nucleotide of the in frame ACC codon). (B) GFP/RFP values for the three sequences described. Error bars represent $\pm$ SD.