Rapid identification of mRNA processing defects with a novel single-cell yeast reporter

MATTHEW R. SORENSON1 and SCOTT W. STEVENS2,3,4
1Graduate Program in Microbiology, 2Department of Molecular Biosciences, and 3Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA

ABSTRACT

It has become increasingly evident that gene expression processes in eukaryotes involve communication and coordination between many complex, independent macromolecular machines. To query these processes and to explore the potential relationships between them in the budding yeast Saccharomyces cerevisiae, we designed a versatile reporter using multicolor high-throughput flow cytometry. Due to its design, this single reporter exhibits a distinctive signature for many defects in gene expression including transcription, histone modification, pre-mRNA splicing, mRNA export, nonsense-mediated decay, and mRNA degradation. Analysis of the reporter in 4967 nonessential yeast genes revealed striking phenotypic overlaps between chromatin remodeling, histone modification, and pre-mRNA splicing. Additionally, we developed a copper-inducible reporter, with which we demonstrate that 5-fluorouracil mimics the mRNA decay phenotype of cells lacking the 3′–5′ exonuclease Rrp6p. Our reporter is capable of performing high-throughput, rapid, and large-scale screens to identify and characterize genetic and chemical perturbations of the major eukaryotic gene expression processes.

Keywords: pre-mRNA splicing; mRNA processing; flow cytometry; drug screening; gene expression

INTRODUCTION

Eukaryotic gene expression is orchestrated by a complex, dynamic, and regulated network of protein–protein and protein–nucleic acid interactions. These interactions require several multiprotein/ribonucleoprotein complexes working in concert to achieve and maintain proper levels of transcribed mRNA and translated protein (Komili and Silver 2008), and to avoid a disease state. From the organization of chromatin to the degradation of mRNAs, the ability to study the individual processes within gene expression, as well as the connections between them, has been aided by technological advances including the use of reporter genes that are sensitive to defects in a specific gene expression process. Many of these techniques and reporters have been developed for the model eukaryotic organism Saccharomyces cerevisiae.

To query the pre-mRNA splicing pathway in yeast, initial efforts involved the fusion of intron-containing fragments to various reporter genes such as lacZ (Larson et al. 1983; Teem and Rosbash 1983; Jacquier et al. 1985; Rain and Legrain 1997), HIS4 (Parker and Guthrie 1985), and a thymidine kinase (Fouser and Friesen 1986). These relied on the spliceosome to excise the intron from reporter gene pre-mRNA to produce a translation-competent mRNA, ultimately generating the measured signal. These reporters often had limitations of low signal and a narrow dynamic range. A complementary reporter with splicing-dependent signals was also created in which a synthetic intron was cloned into the lacZ gene such that β-galactosidase activity resulted from translation of the unspliced pre-mRNA (Legrain and Rosbash 1989), resulting in increased signal under suboptimal pre-mRNA splicing conditions. This was unique because the construct reports levels of pre-mRNA splicing substrate and not product. Although this reporter improved upon some limitations, and has additional utility as an mRNA export assay, it still required extract preparation and biochemical analysis originating from one or more heterogeneous populations of cells.

Further progress was made with the development of the ACT1-CUP1 reporter that was designed so that the degree of copper resistance positively correlated with pre-mRNA splicing efficiency (Lesser and Guthrie 1993). Although this reporter has proved to be instrumental in examining cis-pre-mRNA splicing elements and transacting factors, it does have limitations. Most notably the endogenous...
nonessential CUP1 gene is often present in two or more copies present in most laboratory strains (Karin et al. 1984; Tohoyama et al. 1996), requiring that all genomic copies of this gene be removed in order to specifically query pre-mRNA splicing efficiency. Additionally, after the high-copy reporter is introduced into cells, the data derive from growth on solid media containing an ensemble population of cells. More recently a series of lacZ splicing reporters with competing splice sites was developed in yeast that enables the identification of factors that have a propensity to regulate alternative splicing (Dreumont and Séraphin 2013). There are also many alternative splicing reporters for use in higher eukaryotes that utilize a combination of reporter genes, but as alternative splicing is far less prevalent in budding yeast, these are beyond the scope of this work.

Reporters that query other gene expression events in yeast such as transcription have also been developed. One such system uses the synthetic genetic array methodology (Tong et al. 2001, 2004) along with a dual fluorescent reporter system as a genome-wide approach to discover regulators of a promoter of interest (Fillingham et al. 2009; Kainth et al. 2009; Kainth and Andrews 2010). This system greatly improved upon early hybrid yeast promoter reporters that utilized lacZ (Guarente and Ptashne 1981; Guarente et al. 1982) or genes for nutritional requirements (Titz et al. 2006). It is important to note that the product of some reporter genes can be very stable, confounding gene-induction studies and requiring approaches such as exploiting a destabilized GFP as a transcription reporter (Li et al. 1998).

Another challenge with reporters, especially in high-throughput studies when using β-galactosidase activity, is the need for cell lysis. This can be alleviated using reporter genes/signals that do not require lysis, or other approaches such as a secreted luciferase (Tochigi et al. 2010).

In surveying gene expression, one also needs to consider the role of pre-mRNA and mRNA degradation, not just their biogenesis, processing, and export. For example, two groups developed plasmid-based nonsense-mediated decay (NMD) reporters that use lacZ gene fusions with premature termination codons (Leeds et al. 1991; Wang et al. 2001). In these systems, there is an increase in β-galactosidase activity in the absence of functional NMD. A similar approach was used with a lacZ construct that was used to assay the effects of cis-elements on transcript stability (Muhlrad et al. 1995). There are also NMD reporters that use GFP (Paillusson et al. 2005) and chemiluminescence (Boelz et al. 2006) in mammalian cells, one for Giardia lamblia that uses chemiluminescence (Chen et al. 2008), and one utilizing GFP and/or DsRed in Drosophila melanogaster (Metzstein and Krasnow 2006).

In this work we describe the development, validation, and utility of a pre-mRNA splicing reporter with improved properties, a wider range of applications, increased sensitivity, and the ability to analyze data at the level of individual living cells. Expression of our plasmid-borne gene expression reporter is driven by the constitutive TDH3 promoter or the inducible CUP1 promoter and utilizes the fluorescent proteins mCherry and GFP to generate red and green fluorescence, which reflect the in vivo levels of reporter pre-mRNA and mRNA, respectively. Using flow cytometry, we are able to rapidly observe relative levels of proteins produced from the spliced and unspliced reporter transcripts with single-cell resolution. We demonstrate that our reporter is uniquely sensitive to defects in many gene expression processes including pre-mRNA splicing, histone modification, transcription, mRNA export, nonsense-mediated decay, and other degradation processes. Additionally, our assay is amenable to high-throughput analysis, allowing us to analyze thousands of genetic, chemical, or environmental perturbations in a single day. Our analysis of the haploid deletion collection proves this principle, as well as providing a rich data set to query the function and interconnectedness between individual genes or processes, which we are currently exploring.

RESULTS

Design and validation of the eukaryotic gene expression reporter

We designed a reporter that is amenable to high-throughput single-cell analysis and simultaneously generates signals reflecting pre-mRNA and mRNA levels. To accomplish our goal, we chose to drive transcription of our reporter with the strong and constitutive TDH3 (GPD) promoter. The exons (EX1 and EX2 in Fig. 1A) and intron were based on the relatively inefficiently spliced yeast RPL28/CYH2 gene (Swida et al. 1986) to increase the chances that the unspliced transcript levels were detectable. More details, including the cloning strategy and the DNA sequence can be found in the Supplemental Material.

In this system, levels of exported pre-mRNA (intron containing) are reflected by red fluorescence via a codon-optimized open reading frame (ORF) for mCherry (Shaner et al. 2004) that was placed between the 5′ splice site and branch point sequence, in frame with the first exon and a portion of the intron (Fig. 1A, left). After nuclear export, translation of this pre-mRNA terminates at a stop codon located at the end of the mCherry ORF. To measure the level of exported spliced message, the enhanced GFP (eGFP) ORF (Heim et al. 1995) was cloned in frame with the second exon (Fig. 1A, right). When the pre-mRNA is unspliced, the second exon was designed to be out of frame with the first exon and a stop codon is present after the mCherry ORF, ensuring that GFP translation and fluorescence results only from the spliced message. The final construct was built into a series of low-copy centromeric plasmids with different auxotrophic and dominant selection markers for maximal flexibility.

To rapidly measure the levels of reporter fluorescence in individual yeast cells we utilize flow cytometry. In Figure
IB, we show flow cytometry data for an empty vector control (non-fluorescent) and cells overexpressing either GFP or mCherry alone (Fig. 1B, left). As an additional control to determine the behavior of mCherry and GFP when present at high levels and at a defined 1:1 ratio, we designed a strain in which the yeast gene \( \text{UFD2} \) was tagged at the C terminus with GFP (Huh et al. 2003) and was modified to contain the \( \text{TDH3} \) promoter followed by an N-terminal mCherry tag. This nonessential yeast gene was chosen as a linker between the fluorescence proteins as the X-ray crystal structure of this protein shows that the N- and C-termini are \( \approx 146 \) Å apart (Tu et al. 2007), eliminating potential fluorescence resonance energy transfer between GFP and mCherry (Albertazzi et al. 2009), while keeping their mRNA and protein levels equal. Expression of the dimer resulted in equivalent amounts of fluorescence in the flow cytometer, demonstrating that fluorescence operates as a suitable proxy for in vivo reporter transcript levels (Fig. 1B, right).

**Describing gene expression with a ‘phenograph’ and validating expression**

In wild-type yeast cells, our reporter exhibited 2.5-fold greater red fluorescence than green fluorescence, indicating that while translating pools of pre-mRNA and mRNA are not 1:1, as in the dimer control, our reporter produces both translated pre-mRNA and mRNA at adequate levels for detection and with sufficient dynamic range available to measure changes (Fig. 1C). We refer to the flow cytometry data presented in this work as a phenograph, as it graphically represents the reporter expression phenotype in a given yeast strain. To ensure that no GFP signal results from the pre-mRNA transcript, we mutated the intron branch point nucleotide from A to G or the 3' splice site from AG to UG. As predicted, both of these mutations abolished green fluorescence, demonstrating that intron removal is required for the production of GFP from this reporter (Fig. 1D).

We note that the elongated shape of the phenograph is due to heterogeneity of cell volume as we observed a strong correlation between fluorescence intensity and cell volume (Fig. 2A). Additionally, the fact that the cells fall near the middle of the phenographs allows for detection of an increase or decrease of at least two orders of magnitude with either signal. Reassuringly, the levels of fluorescent protein correlate with transcript levels by RT-qPCR (Fig. 2B).

As with any technology or assay, reproducibility is of critical importance, especially if one wishes to compare data from experiments performed on different days. To demonstrate this quality of our reporter, we compared data from wild-type cells containing our constitutive reporter from four separate experiments on different days. When these

---

**FIGURE 1.** Gene expression reporter design and validation. Design (A) of the gene expression reporter construct is shown. Expression of the synthetic reporter was driven by the strong constitutive \( \text{TDH3} \) promoter. Exon 1 (EX1), the intron, and exon 2 (EX2) were designed using features of the inefficiently spliced \( \text{CYH2} \) gene. Dashed lines indicate the region comprising the intron. The mCherry ORF was placed in frame with EX1 and the eGFP ORF was cloned in frame with EX2. Exported intron-containing pre-mRNAs and mature mRNAs are translated resulting in proteins that fluoresce red and green, respectively. The construct was cloned into pRS316 (see Supplemental Material for more detail). Control samples (B) for flow cytometry are shown. Phenographs for wild-type yeast harboring an empty vector (non-fluorescent), or vectors expressing GFP-only or mCherry-only are shown (left). (Right) A phenograph of the expression of the mCherry-Ufd2p-GFP dimer results in a 1:1 mCherry:GFP signal. The pseudo color represents cell density within the phenograph. Reporter expression in wild-type yeast (C) is displayed and referred to as a phenograph. Phenographs for branch point (A to G) and 3' splice site (CAG to CuG) mutant reporters (D) in wild-type yeast are shown (top and bottom, respectively). The y- and x-axis depict mCherry and GFP fluorescence, respectively.
and reporter transcript levels. Reporter expressing cells were gat-
ed (A) into four equally populated groups based on their position within
the phenograph, moving out from origin (essentially grouping cells
based on their fluorescence intensity). The inset compares the average
total fluorescence with the average cell volume (forward scatter) for
the four groups. An exponential regression line is applied ($R^2 = 0.98$).
The flow cytometry mCherry/GFP ratio is plotted for several deletion
strains vs. the ratio of unspliced and spliced reporter transcripts (B),
which was determined by RT-qPCR analysis. Flow cytometry and RT-
qPCR ratios are relative to wild type (WT) and an exponential trend
line is shown ($R^2 = 0.91$). The inset cartoon depicts the primer location
for the unspliced (top) and spliced transcripts. Phenographs from wild-
type yeast analyzed on different days (C) are visually indistinguishable
when overlaid.

| Data are overlaid they are virtually indistinguishable (Fig. 2C),
demonstrating great day-to-day reproducibility. | Deletion of genes within specific gene expression processes exhibit similar phenograph patterns |
---|---

**To assess the behavior of the reporter we generated phenographs from a small collection of yeast strains, each lacking a single gene functioning in some aspect of gene expression.** It was expected that defects in the splicing machinery would decrease the efficiency of intron removal (Vijayraghavan et al. 1989), resulting in populations of cells that are red-shifted, due to a decrease in green fluorescent signal which would result from decreased levels of spliced reporter transcript. In the first set of splicing mutants (mud2Δ, swm2Δ, and swt21Δ), we detected a relatively minor red-shift, whereas in the bud13Δ, ist3Δ, and lea1Δ strains we detect a more severe red-shift (Fig. 3, cf. A and B). This shift in the ratio of the translation products of the pre-mRNA to mRNA is also reflected in the corresponding RNA levels as determined by RT-qPCR (Fig. 2B).

Next, we considered whether our reporter was sensitive to and behaved differently in strains containing mutations in other gene expression pathways. The most striking were the nonsense-mediated decay (NMD) mutants nam7Δ, nam2Δ, and upf3Δ (Fig. 3C; He et al. 1997). This is not surprising because the long 3′ UTR of the reporter pre-mRNA, which extends from the 3′ end of the mCherry ORF through the GFP coding region, is predicted to be an ideal substrate for NMD in wild-type yeast (Muhlrad and Parker 1999). Without a functional NMD pathway, we would expect increased mCherry signal due to increased stability and translation of the reporter pre-mRNA. As expected we detected a severe red-shift which was a result of over a 10-fold increase in red fluorescence with very little change in green fluorescence intensity, supporting the suggestion that disruptions in the NMD pathway result in accumulated levels of translatable reporter pre-mRNA transcripts (He et al. 1993). RT-qPCR analysis further corroborates this, revealing a large increase in the pre-mRNA/mRNA ratio in NMD mutants compared with wild-type cells (Fig. 2B). NMD has also been shown to mask intron accumulation of poorly spliced transcripts by degrading the pre-mRNA transcripts in yeast (Sayani et al. 2008), which is also consistent with the red-shifted phenographs and RT-qPCR results. Our results indicate that the reporter pre-mRNA is inefficiently spliced and that this is masked by NMD in wild-type cells, similar to the suboptimally spliced natural introns highlighted by Sayani and colleagues. We also analyzed forward scatter values in these mutants to see whether cell volume increased compared with wild type. There was no difference seen in these mutants, supporting the conclusion that increased translation of the stabilized pre-mRNA resulted in the red-shift and not an accumulation of fluorescent proteins in a larger cell.

Defects within early steps of gene expression also altered reporter expression. We analyzed the phenographs of several mutants lacking factors involved in mRNA transcription (rtf1Δ, srb5Δ, and rtr1Δ) (Fig. 3D), chromatin remodeling (htz1Δ, arp6Δ, and vps72Δ) (Fig. 3E) and histone modification (hos3Δ, rpd3Δ, and set2Δ) (Fig. 3F). The phenographs from these mutants were primarily red-shifted, but to a lesser degree than in NMD mutants and more closely resembling some pre-mRNA splicing mutants (Fig. 3A). The observed red-shift in these mutants, which is a result of a decrease in green fluorescence (spliced signal), is not surprising due to
the large number of reports describing the interconnected nature of chromatin remodeling, histone modification, transcription, and pre-mRNA splicing. We also note that the phenographs for mutants with no direct function in gene expression do not significantly change compared with wild type (atg11Δ, cue3Δ, and pig2Δ) (Fig. 3G).

Interestingly, a subset of gene expression factor deletions produced an overall green-shifted phenograph rather than red-shifted phenographs, indicating a relative increase in the spliced reporter signal and/or a decrease in pre-mRNA signal. This was most striking in yeast strains lacking factors of the TREX-2 complex (sac3Δ, thp1Δ, sus1Δ, and sem1Δ) (Fig. 4A), which couples mRNA transcription and mRNA export (Fischer et al. 2002; Rodriguez-Navarro et al. 2004). The loss of TREX-2 core components Sac3p or Thp1p results in indistinguishable phenographs, which would be expected. Sus1p has been proposed to serve as a link between the SAGA complex and TREX-2 complex (Ellisdon et al. 2010), which results in a similar but not identical phenograph when removed from yeast. Analysis of different classes of mRNA export-related mutants, such as the THO complex and components of the nuclear pore (Stewart 2010), demonstrate a similar green-shift (Fig. 4B,C). Possible explanations for this effect are changes in mRNA stability, a preferential export of specific transcripts, or as our RT-qPCR data suggests, the defect in mRNA export likely leads to a lengthened nuclear residency and might increase the likelihood of splicing, increasing the green to red ratio (Fig. 2B). It is also important to note that others have used translationally competent reporter pre-mRNAs as mRNA export reporters, providing further support for the green-shifted effect we observe (Galy et al. 2004; Dufu et al. 2010; Teplova et al. 2011; Caporilli et al. 2013).

One interesting green-shifted phenograph was that of the mlp1Δ mutant (Fig. 4D), a result of a nearly 80% decrease in mCherry signal, with a minor change in GFP. This is in contrast to the green-shifts observed in the previously described mutants (Fig. 4A–C), which primarily result from an increase in GFP. Mlp1p has been shown to mediate retention of unspliced reporter mRNAs, resulting in increased pre-mRNA reporter translation in its absence (Galy et al. 2004; Hackmann et al. 2014), which is in contrast with the result shown here. Although both reporters are weakly spliced due to suboptimal intron architecture, our reporter contains the extended 3′ UTR that is efficiently degraded in the cytoplasm via NMD (Fig. 3C). This feature, along with the role of Mlp1p as a checkpoint for export competent messenger ribonucleoparticles (mRNPs) (Green et al. 2003; Palancade et al. 2005), leads us to believe that exported unspliced reporter mRNPs in mlp1Δ cells could be improperly packaged, leading to a decreased signal by virtue of NMD and/or other degradation pathways. Further studies are needed to confirm this. Of note, we also observe a similar effect in cells lacking the SR-like protein Gbp2 (Fig. 4E), which was recently shown to be involved with Mlp1p in a potential mRNP
quality control mechanism coupled with export (Hackmann et al. 2014).

We also note that the deletion of factors involved the turnover of mRNA transcripts, which stabilizes mRNA messages, also results in green-shifted cells (\(dhh1\Delta\) and \(ccr4\Delta\)) (Fig. 4F, G; Goldstrohm and Wickens 2008). Although the \(rrp6\Delta\) mutant results in an overall green-shift like \(dhh1\Delta\) and \(ccr4\Delta\) (Fig. 4H), it is important to note that the \(rrp6\Delta\) mutant is the only deletion mutant that results in significant increases of both GFP and mCherry fluorescence. Rrp6p is a nuclear exosome component with 3′–5′ exonuclease activity, and is thought to degrade aberrant transcripts, among other functions (Chlebowski et al. 2013). Loss of this enzyme could stabilize the pre-mRNA and in turn increase the probability of it being spliced within the nucleus. This may lead to increased signals of both reporter messages. On the other hand, the \(dhh1\Delta\) and \(ccr4\Delta\) mutants only increase in the green fluorescence spliced signal, which is most likely due to their function in decreasing the stability of mature cytoplasmic transcripts in a wild-type strain.

We also looked into the effects on total fluorescence of many of these mutants by plotting the mean mCherry and GFP values for the 21,000 cells that were analyzed by flow cytometry (Supplemental Fig. 1). As expected, we observe a decrease in the overall fluorescence in mRNA transcription mutants, as well as specific patterns for other mutant groups, such as NMD and mRNA export. The directed screen of gene expression mutants demonstrates that our reporter is uniquely sensitive to many gene expression events, increasing its utility by permitting numerous biological pathways to be queried at single-cell resolution. The similarity between phenographs resulting from mutants within a given gene expression process (cf. each panel in Figs. 3 and 4) suggests that our reporter can be used to test the relatedness within and between pre-mRNA processing events. It is interesting that when numerous nonessential translation mutants were analyzed, no appreciable effect on reporter expression was
detected suggesting no translation bias between the spliced and unspliced reporter RNAs (Supplemental Fig. 2). It is important to note that the red- and green-shifts can be a result of changes in one or both signals, as shown in Supplemental Figure 1.

**Optimization and utilization of a copper-inducible reporter**

We also explored using our reporter with conditional alleles of essential genes involved in gene expression. Since most temperature- or cold-sensitive conditional alleles exhibit rapid response to changed temperature, the existing pools of extremely long-lived fluorescent proteins (Kitsera et al. 2007) would mask changes resulting from the mutation due to the relatively short time spent at the nonpermissive temperature. To circumvent this, and to measure de novo signal after temperature shift, we placed the reporter described in Figure 1A under the control of the *CUP1* promoter, which is rapidly and highly activated upon addition of copper (Peña et al. 1998). Induction with 1 mM CuSO₄ resulted in significant fluorescence within 1 h, reaching peak levels within 3–4 h (Fig. 5A). For completeness we followed the time course into stationary phase (6+ h) and

**FIGURE 5.** Exploring conditional alleles and drug treatment with a copper-inducible reporter. A time course (A) of *CUP1*-Reporter induction is shown. Wild-type cells harboring a *CUP1*-Reporter were grown to early logarithmic phase and induced with 1 mM CuSO₄. Samples were taken at the indicated time points. Yeast with wild-type *MEX67* or *mex67-5* and the *CUP1*-Reporter were grown at permissive (31°C) temperature, then shifted to permissive or nonpermissive (34°C or 37°C) temperature with 1 mM CuSO₄ (to induce reporter) for 4 h before analysis. Data for cells at permissive temperature are shown in the background in red, with data from cells at nonpermissive temperature in the foreground in blue (B). Pseudo-color phenographs *insets* are shown for the corresponding nonpermissive conditions. Wild-type cells with *CUP1*-Reporter and an empty vector or RRP6 overexpression plasmid were grown in minimal media with galactose. Cells were treated with DMSO (vehicle only control) or 100 μM 5-FU, induced with 1 mM CuSO₄ and grown for 4 h before analysis (C). The rightmost panel of each set is an overlay of the DMSO and 5-FU data sets for comparison.
did not see any appreciable differences in reporter expression. During optimization efforts we found that induction efficiency was highly dependent on growth phase, with the highest and most consistent levels produced when cells were induced between OD 0.1 and OD 0.2 (A600, data not shown). Additionally, we have demonstrated that data from cells grown in standard culture tubes, 96-well, and 384-well plates are comparable using the CUP1-Reporter (Supplemental Fig. 3), simplifying the nature of high-throughput/high-content experiments.

Since we observed many mutants in the mRNA export pathway resulting in strong green-shifts, we analyzed the prototypical mex67-5 allele of the essential Mex67p, homolog of the human TAP mRNA export factor (Segregf et al. 1997). Yeast with the temperature-sensitive mex67-5 allele fail to export bulk mRNA at elevated temperature. We cultured MEX67 (wild-type) or mex67-5 yeast at permissive temperature (31°C) to early log phase, and then shifted to nonpermissive temperature (34°C or 37°C) while adding 1 mM CuSO4 to induce the CUP1-driven reporter. Yeast with wild-type MEX67 displayed very little difference between permissive and nonpermissive temperature, while cells with mex67-5 were green-shifted at nonpermissive temperature (Fig. 5B). The green-shifted phenotype is consistent with the previously analyzed nonessential gene deletions in this process (Fig. 4A–C), yet we observed more cell-to-cell variability with the conditional allele, which is perhaps not surprising since one would expect a more uniform phenotype with a complete gene deletion. We attempted to analyze the strong temperature-sensitive splicing allele prp2-1 but we found that the cells were not able to express sufficient levels of the reporter at nonpermissive temperature during the experiment, most likely due to the rapid growth defects conferred by the prp2-1 allele (Hartwell 1967).

We were also interested in determining the utility of the inducible reporter with chemical insults that impair a specific gene expression process. Drug treatment of cells is similar to the use of conditional alleles in that existing pools of mCherry and GFP would not significantly change during the relatively short treatment, thus the inducible reporter would add great utility to these experiments. The nucleoside analog fluorouracil (5-FU), a widely used chemotherapeutic (Longley et al. 2003), has been shown to inhibit the 3′–5′ exo-nuclease Rp6p of the nuclear exosome (Fang et al. 2004; Kammler et al. 2008), although the exact mechanism of inhibition is unknown. As expected, treatment with 100 μM 5-FU while inducing the CUP1-Reporter resulted in a green-shift, compared with a DMSO only control (Fig. 5C), and mimicked the deletion of RRP6 (Fig. 4E). Additionally, we show that overexpression of RRP6 is able to reverse the effect of 5-FU on reporter expression (Fig. 5C). Together, the mex67-5 and 5-FU data demonstrate the usefulness of the inducible CUP1-Reporter in analyzing essential genes and chemical perturbations.

**Reporter is amenable to high-throughput analysis**

The nature of our reporter also allows for rapid high-throughput flow cytometry analysis, permitting one to acquire data from hundreds of unique strains or conditions within a matter of hours, while collecting tens of thousands of data points (events or cells) for each sample. With the significant amount of data our reporter system generates, we sought out ways to compare reporter expression in many different samples in an unbiased manner. Since this type of analysis platform was lacking for flow cytometry data we developed a statistical workflow that vectorizes the data, describing the phenograph shape by reporting the distribution of fluorescent events in the two-dimensional phenomenographs (see Materials and Methods and Supplemental Fig. 4: M Sorenson and S Stevens, in prep.). These vectorized data are then compared with each other using a standard hierarchical clustering algorithm.

To demonstrate the high-throughput nature and reproducibility of our reporter assay, as well as to query all nonessential yeast genes for their role(s) in gene expression, we completed a high-throughput screen of 5103 viable strains in the yeast haploid deletion collection (Winzeler et al. 1999). Individual transformants were obtained for 5020 strains (98%), and were individually analyzed by high-throughput flow cytometry. Data conforming to our criteria (see Materials and Methods) were obtained for 4967 mutants (97% of viable strains, 99% of transformed strains). To first assay the reproducibility of our assay across the multiple days this experiment took place we observed the clustering behavior of data from biological replicates assayed on different days. In Figure 6A we show the clustering behavior of the deletion collection data with the biological replicates added. Enlarged areas of interest, shown in Figure 6, B and C, reveal the tight clustering of fluorescent controls, as well as previously described gene expression mutants such as sac3Δ, lea1Δ, and bud13Δ. The replicates clustered very tightly together and near genes of similar function within this large data set, further demonstrating the reproducibility and robustness of our assay.

**Clustering behavior of the haploid deletion collection**

Since most of the strains in the deletion collection are not related to gene expression, it was reasonable to expect that the vast majority of strains would be highly similar to wild type (Figs. 1C, 3G). We observed that 87% of the deletion mutants clustered within a very large clade (Fig. 7A, bottom). The limited variation observed between the mutants in this group likely describes biological noise or mild indirect effects on gene expression.

Within a small clade near the top of the dendrogram (Fig. 7A), a group of yeast strains (n = 16) was detected that was enriched with mutants lacking mRNA export factors (n = 9) (Fig. 7B). Note the tight clustering of sac3Δ and...
thp1Δ, which was expected due to their nearly indistinguishable phenographs (Fig. 2G) and their roles in the TREX-2 complex. Moreover, there were other export factors such as additional TREX-2 components (sus1Δ and sem1Δ), THO complex (thp2Δ), nuclear pore components (mlp1Δ and nup120Δ), and the poly(A) RNA-binding factor gbp2Δ.

Also within this clade were mutants lacking genes that influence mRNA decay or stability including dhh1Δ (decapping), ccr4Δ (deadenylation), and rrp6Δ (nuclear exosome).

Next, we investigated the clustering behavior of pre-mRNA splicing mutants and other red-shifted mutants within the larger red-shifted clade (n = 85) (Fig. 7C). The smaller groups within this clade are each enriched with factors within certain gene expression processes such as pre-mRNA splicing, transcription, chromatin remodeling, histone modification, and NMD (Fig. 3). The pre-mRNA splicing mutants clustered into two slightly offset clades within this group (Fig. 3D, G). This suggests that our reporter is differentially sensitive to pre-mRNA splicing mutations (see Discussion) and quantitatively illustrates the difference in the phenographs we reported in Figure 2, A and B.

Within this red-shifted group, there was enrichment of chromatin remodeling and mRNA transcription mutants (Fig. 3B, E, F). Adding to the confidence that the clustering of the reporter data correlates with functionally related genes, the chromatin remodeling mutants in this clade contain components of the SWR1 complex (swc3Δ and arp6Δ) and also the H2AZ histone variant (HTZ1) on which it operates (Mizuguchi et al. 2004). Similarly, the mRNA transcription clade contained deletions of constituents of the mediator complex (cse2Δ, srb5Δ, soh1Δ, and srb2Δ) (Guglielmi et al. 2004). The most tightly clustering group within this red-shifted clade contained the NMD mutants (Fig. 7H), which was expected due to their distinct phenograph shape (Fig. 3C). The remarkable degree to which the transcription, histone modification, and chromatin remodeling mutants cluster among pre-mRNA splicing mutants (Fig. 7C) is not surprising, given that most introns are believed to be removed cotranscriptionally (Brugiolo et al. 2013).

**DISCUSSION**

When compared with existing yeast pre-mRNA splicing reporters for budding yeast, the reporter described here...
expresses well from a low-copy vector, is able to simultaneously measure both pre-mRNA and mRNA levels, and is amenable to rapid, sensitive, gene expression analysis in single cells via flow cytometry without the need to prepare extracts from cells. It is important to note that the red- and green-shifts are not only due to changes in splicing efficiency, per se, but can also be a result of the localization and relative stability of either the pre-mRNA or mRNA, as exemplified by mRNA export mutants (Figs. 4A–D, 5B) and mRNA degradation and stability mutants (Figs. 3C, 4F–H). The use of flow cytometry to analyze reporter expression also permits simultaneous observation of cell volume and complexity using the forward and side scatter measurements, respectively. These features enable us to assay the noise in eukaryotic gene expression among a population of cells in a unique way (Waks and Silver 2010; Tang et al. 2011), as well as simultaneously examining effects on yeast morphology with forward and side scatter values.

In the design of our reporter, one goal was to maximize its utility compared with previously described gene expression reporter. Since our readout was single cell in nature, we also desired to use a stable low-copy centromeric plasmid to limit the amount of cell-to-cell variability due to copy-number variation. The single-cell aspect of our reporter also permits the investigator to use our reporter and fluorescence-activated cell sorting to identify clones of interest in a rapid and unbiased manner when used in conjunction with conventional genetic screens. This assay should be feasible and cost effective for researchers at research institutions that have access to core facilities with flow cytometry capabilities. Additionally, with the advent of many small-footprint tabletop flow cytometers, it is also within reach for many investigators to complete the flow cytometry analysis within their own laboratory. Also, by avoiding the use of endogenous reporter genes, and by cloning the reporter into plasmids with various auxotrophic and dominant selection markers, we have virtually eliminated genotypic restrictions. We also expanded the application of our reporter by developing the copper-inducible CUP1-reporter, and demonstrated its usefulness with conditional mutants of essential genes (mex67-5), as well as with treatment of 5-FU. During development of the CUP1-reporter we also examined reporter...
expression in the same mutants as in Figures 3 and 4 and compared expression with the TDH3-driven reporter saw patterns and shifts of similar magnitude (data not shown).

In our survey of nearly 5000 knockout strains, our novel gene expression reporter proved to be highly sensitive and specific for numerous gene expression–related processes. This data set (Supplemental Table 1), as well as binned data for the biological replicates and controls in Figure 6 (Supplemental Table 2), will allow the community to perform similar analyses in additional strains or under different conditions and directly compare their data to learn more about their genes or alleles of interest. Additionally, researchers can reanalyze these data and look for their gene of interest to identify knockout mutants with similar effects on reporter expression, as well as refer to a list of red- and green-shifted mutants (Supplemental Table 3). Our ability to directly compare the expression of the reporter between strains by means of the vectorized phenograph and the fact that genes participating in related functions behave similarly by these analyses makes this a useful tool for future studies. Within our clustering analyses of the deletion collection we noticed that mutations predicted to lead to a completely defective, nonessential pathway (e.g., NMD) cluster tightly with each other. This is in contrast to mutations of components of essential machineries (e.g., pre-mRNA splicing), which cluster less tightly and result in decreased efficiency, but not a complete loss. Lastly, we observed that mutants containing deletions of dubious or unverified ORFs regularly clustered near deletions of verified ORFs with which they overlapped, such as ynl140cΔ and tho2Δ, ydl094cΔ and pmt5Δ, and ydr290wΔ and rtl103Δ, respectively.

We also demonstrated that our reporter assay was robust and reproducible. When we included reporter data from biological replicates of wild-type cells into the deletion collection data, we observe that they cluster with deletion mutants unrelated to gene expression (cf. Figs. 2C, 3G) that have wild-type-like phenographs. The reproducible behavior of mutants with a more distinct phenograph (Fig. 6) demonstrates the robustness of our assay and analysis pipeline.

A very powerful aspect of our reporter is that it is sensitive, yet uniquely specific, to defects in many different gene expression events, including the processes involved in both biogenesis of mRNAs as well as their degradation. This allows for a researcher to simultaneously assay the effects of a wide variety of stimuli on many processes. We have found that our analysis pipeline using our binning and clustering technique, as well as our telltale phenograph patterns (Figs. 3, 4), quickly identifies potential pathways to explore further with orthogonal assays, such as the RT-qPCR we present in Figure 2. Additionally, multiplexing of reporters can also be a way to screen for effects on multiple reporters simultaneously (Chen et al. 2012; Kulak and Lum 2013), as long as their signals do not overlap or disturb each other. But as with all other reporter technologies, it is still critical to ascertain the effect on the original or endogenous targets.

Although researchers have scoured the yeast genome for decades, it is often with the advent of new technologies and techniques that new breakthroughs occur. Here, we have described the design, validation, and broad utility of our novel gene expression reporter. The reporter not only improves upon previous pre-mRNA splicing reporters in terms of speed, sensitivity, and genetic limitations, but it also is sensitive and specific to many other gene expression events. Our analysis pipeline coupled with high-throughput flow cytometry allows for rapid, unbiased evaluation of thousands of mutations or conditions. Using our gene expression reporter with standard genetic-screening approaches, in series with fluorescence activated cell sorting allows researchers to explore gene expression processes, and connections within and between them, with increased speed and sensitivity.

**MATERIALS AND METHODS**

**Strains, plasmids, and oligos**

Standard yeast methods were used (Sherman 1991). Yeast strain BY4742 was used as a wild-type comparison except where noted. For the mCherry-Uld2p-GFP strain, a PCR product containing the TDH3 promoter and the mCherry ORF was used to genomically N-terminally tag UFD2-GFP by homologous recombination (Huh et al. 2003). For directed deletion mutant screen and the deletion collection screen, the Yeast MATalpha Collection was used (Thermo/Open Biosystems; Cat# YSC1054). For m6e7-5 experiments the respective strain (Segref et al. 1997) was transformed with empty or MEX67 plasmids and a CUP1-Reporter, selecting only for the introduced plasmids not the m6e7-5 plasmid. Please refer to the Supplemental Material for plasmids and oligos.

**Flow cytometry**

For standard flow cytometry, log-phase yeast were fixed in 10% formaldehyde, washed, and resuspended in 1X PBS. For high-throughput flow cytometry, 96-well plates containing stocks of the deletion strains harboring the reporter were replicated into 150-µL selective media. When the average well in a plate reached mid-log phase, cells were harvested, washed, and resuspended in 100 µL 1X PBS and analyzed immediately with a Becton Dickinson LSRSFortessa equipped with a high-throughput sampler unit. For excitation of eGFP and mCherry, 488 nm and 561 nm lasers were used, respectively. For detection a 505 LP and 530/30 BP were used for eGFP and a 600 LP and 610/20 BP for mCherry. For each sample 21,000 events were collected and 10,000 events are shown in the phenographs. FlowJo was used to analyze data and to prepare phenographs, plotting cells based on their fluorescent intensity, which is in arbitrary units based on predetermined cytometer settings.

**Binning and clustering analysis workflow**

For nonbiased comparison of phenographs we first imported the flow cytometry data into FlowJo. The GFP and mCherry values for all 21,000 fluorescent events were exported to a .txt file. Using SAS software (Version 9.2, SAS Institute Inc.), the .txt file was...
imported and a log transformation was applied. Next, the SAS algorithm assigned the first 20,000 events into 1600 bins based on their GFp and mCherry value, in essence, dividing the fluorescence space into a linear 40 × 40 grid. The number of bins and range of acceptable values can be adjusted, allowing the user to adjust resolution and sensitivity. Data files without sufficient data are flagged for reanalysis. The number of events in each bin for each sample is reported and vectorized data are exported as a .txt file. The result is a table in which the number of events in each of the bins is recorded in individual columns for each sample listed in separate rows. Supplemental Tables 1 and 2 are examples of such data, containing the number of cells in each of the 1600 bins (columns) for the respective yeast strain (rows). Vectorized data were then clustered hierarchically with Cluster (Eisen et al. 1998) and viewed with Java TreeView (Saldanha 2004). The analysis and output is comparable to the analysis of microarray data, except that instead of comparing the expression level of many genes, we compare the degree of occupancy of the 1600 different bins, which describes phenograph shape.

RT-qPCR analysis

Two OD units of log phase cells were harvested by centrifugation in a microtube and total RNA was extracted with 266 µL acid phenol and 266 µL AES Buffer (50 mM sodium acetate at pH 5.3, 10 mM EDTA, and 1% SDS). Samples were vortexed well and incubated for 6 min at 65°C, vortexing every minute (Wise 1991). Samples were placed in ice water for 5 min then transferred to a Phase Lock Gel Heavy 2-ml tube (5 PRIME). Tubes were centrifuged for 5 min at 2500 rpm. Acid phenol (266 µL) was added; tubes were gently inverted five times and then centrifuged. Chloroform (266 µL) was added; tubes were inverted to mix and then centrifuged. The aqueous phase above the gel matrix was transferred to a new microtube. A total of 3 M sodium acetate (1:10 volumes, pH 5.3) was added, followed by isopropanol (1.1 volumes). Tubes were incubated at −20°C overnight to precipitate the RNA. RNA was harvested and washed twice with 1 mL 70% cold ethanol and resuspended in 30–50 µL DEPC-treated water.

RNA (5 µg) was DNase treated (5 U RQ1 DNase, Promega) for 20 min at 37°C. RNA Clean and Concentrator kit (Zymo) was used to purify the RNA per the manufacturer’s protocol (6.5 µL DEPC water was used to elute, resulting in a final volume of 5 µL). To DNase-treated RNA the following was added; 1 µL 10X RT Buffer (0.5 M Tris–HCl at pH 8.5), 1 µL random nonamer primers (25 µM; IDT), and 2 µL DEPC-H2O. The RNA/primer mix was incubated at 60°C for 5 min and immediately transferred to ice for 5 min. A total of 11 µL of an RT master mix (1X RT Buffer, 3 mM MgCl2, 10 mM DTT, 0.5 mM dNTPs, and 5 U/µL MultiScribe RT enzyme) was added. The RT reaction was incubated overnight at 42°C and the resulting cDNA was diluted with four volumes DEPC-H2O. Each qPCR reaction contained 5 µL cDNA, 7.5 µL Power SYBR Green PCR Master Mix (ABI), and 2.5 µL primers (1.5 µM each; see Supplemental Material for sequences). Standard thermocycling, fluorescence detection and analysis (standard curve) were completed using the VIia 7 Real-Time PCR System (Life Technologies).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank members of the Stevens laboratory, Arlen Johnson, Edward Marcotte, Andrew Ellington and Vishy Iyer for helpful discussion and advice, Jeff Pleiss for RT-qPCR protocols and advice. We thank Richard Salinas and Cecilia Harkey for flow cytometry and robotic expertise, respectively. We thank Marrianna Grenadier of UT SBS Graphics. This work was supported by grants to S.W.S. from the American Cancer Society (RSG-05-137-01-GMC), National Institute of Health (GM084246), and a training fellowship for M.R.S. from the Cancer Prevention & Research Institute of Texas (RP101501).

Received September 25, 2013; accepted February 20, 2014.

REFERENCES

Albertazzi L, Arosio D, Marchetti L, Ricci F, Beltram F. 2009. Quantitative FRET analysis with the E’GFp-mCherry fluorescent protein pair. Photochem Photobiol 85: 287–297.
Boez S, Neu-Yilik G, Gehring NH, Hentze MW, Kulozik AE. 2006. A chemiluminescence-based reporter system to monitor nonsense-mediated mRNA decay. Biochem Biophys Res Commun 349: 186–191.
Bruigolo M, Herzel L, Neugebauer KM. 2013. Counting on co-tran
terctional splicing. F1000Prime Reports 5: 9.
Caporilli S, Yu Y, Jiang J, White-Cooper H. 2013. The RNA export fac
tor, Nxt1, is required for tissue specific transcriptional regulation. PLoS Genet 9: e1003526.
Chen Y-H, Su L-H, Sun C-H. 2008. Incomplete nonsense-mediated mRNA decay in Giardia lamblia. Int J Parasitol 38: 1305–1317.
Chen J, Young SM, Allen C, Seeper A, Peli-Gulli M-P, Panchaud N, Waller A, Ursu O, Yao T, Golden JE, et al. 2012. Identification of a small molecule yeast TORC1 inhibitor with a multiplex screen based on flow cytometry. ACS Chem Biol 7: 715–722.
Chlebowski A, Lubas M, Jensen TH, Dziembowski A. 2013. RNA decay machines: the exosome. Biochim Biophys Acta 1829: 552–560.
Dreumont N, Séraphin B. 2013. Rapid screening of yeast mutants with reporters identifies new splicing phenotypes. FEBS J 280: 2712–2726.
Dufu K, Livingstone MJ, Seebacher J, Gygi SP, Wilson SA, Reed R. 2010. mRNA export proteins, Aly and CIP29, to assemble the TREX complex based on flow cytometry. ACS Chem Biol 7: 715–722.
Eisen MB, Spellman PT, Brown PO, Botstein D. 1998. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci 95: 14863–14868.
Ellison AM, Jani D, Kohler A, Hurt E, Stewart M. 2010. Structural basis for the interaction between yeast Spt-Ada-Gcn5 acetyltransferase (SAGA) complex components Sgf11 and Sus1. J Biol Chem 285: 3850–3856.
Fang F, Hoskins J, Butler JS. 2004. 5-fluorouracil enhances exosome-de
dependent accumulation of polyadenylated RNAs. Mol Cell Biol 24: 10766–10776.
Fillingham J, Kainth P, Lambert J-P, van Bakel H, Tsui K, Peña
Castillo L, Nislow C, Figyes D, Hughes TR, Greenblatt J, et al. 2009. Two-color cell array screen reveals interdependent roles for histone chaperones and a chromatin boundary regulator in histone gene repression. Mol Cell 35: 340–351.
Fischer T, Strasser K, Racz A, Rodriguez-Navarro S, Oppizzi M, Ihrig P, Lechner J, Hurt E. 2002. The mRNA export machinery requires the novel Sac3p–Peg12 complex to dock at the nucleoplasmic entrance of the nuclear pores. EMBO J 21: 5843–5852.
Fouser LA, Friesen JD. 1986. Mutations in a yeast intron demonstrate the importance of specific conserved nucleotides for the two stages of nuclear mRNA splicing. Cell 45: 81–93.
Galy V, Gadal O, Fromont-Racine M, Romano A, Jacquier A, Nehrbaus U. 2004. Nuclear retention of unspliced mRNAs in yeast is mediated by perinuclear Mlp1. Cell 116: 63–73.
Tohyama H, Shiraishi E, Amano S, Inouhe M, Joho M, Murayama T. 1996. Amplification of a gene for metallothionein by tandem repeat in a strain of cadmium-resistant yeast cells. *FEMS Microbiol Lett* **136**: 269–273.

Tong AHY, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CWV, Bussey H, et al. 2001. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**: 2364–2368.

Tong AHY, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M, et al. 2004. Global mapping of the yeast genetic interaction network. *Science* **303**: 808–813.

Tu D, Li W, Ye Y, Brunger AT. 2007. Structure and function of the yeast U-box-containing ubiquitin ligase Ufd2p. *Proc Natl Acad Sci* **104**: 15599–15604.

Vijayraghavan U, Company M, Abelson J. 1989. Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisae*. *Genes Dev* **3**: 1206–1216.

Waks Z, Silver PA. 2010. Nuclear origins of cell-to-cell variability. *Cold Spring Harb Symp Quant Biol* **75**: 87–94.

Wang W, Czaplinski K, Rao Y, Peltz SW. 2001. The role of Upf proteins in modulating the translation read-through of nonsense-containing transcripts. *EMBO J* **20**: 880–890.

Winzeler EA, Shoemaker DD, Astromoff A, Liang H, Anderson K, Andre B, Bangham R, Benito R, Boeke JD, Bussey H, et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901–906.

Wise JA. 1991. Preparation and analysis of low-molecular weight RNAs and small ribonucleoproteins. *Methods Enzymol* **194**: 405–415.