Introns regulate RNA and protein abundance in yeast

Kara Juneau*,†, Molly Miranda†, Maureen E. Hillenmeyer†‡, Corey Nislow†§ and Ronald W. Davis*,†

*Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305.
†Stanford Genome Technology Center, Stanford University, Palo Alto, CA 94304.
‡Biomedical Informatics, Stanford University School of Medicine, Stanford, CA 94305.
§Department of Biochemistry, Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, Canada M5S 3E1
Running title: Introns regulate gene expression

Keywords: Intron, Saccharomyces cerevisiae, RNA processing

Corresponding author:

Kara Juneau
Stanford Genome Technology Center
855 California Avenue, Palo Alto, CA 94304-1103
Phone: 650-812-1976
Fax: 650-812-1975
E-mail: kjuneau@stanford.edu
ABSTRACT

The purpose of introns in the architecturally simple genome of *Saccharomyces cerevisiae* is not well understood. To assay the functional relevance of introns, a series of computational analyses and several detailed deletion studies were completed on the intronic genes of *Saccharomyces cerevisiae*. Mining existing data from genome-wide studies on yeast, it was discovered that intron containing genes produce more RNA, more protein and are more likely to be haploinsufficient than non-intronic genes. These observations for all intronic genes held true for distinct subsets of genes including ribosomal, non-ribosomal, duplicated and non-duplicated. Corroborating the result of computational analyses, deletion of introns from three essential genes decreased cellular RNA levels and caused measurable growth defects. These data provide evidence that introns improve transcriptional and translational yield and are required for competitive growth of yeast.
INTRODUCTION

The genes of complex organisms depend on introns to provide regulatory sequences that allow for accurate pre-mRNA processing and alternative splicing. In multi-cellular organisms most genes contain at least one intron, usually more. In human, for instance, 94% of the genes are interrupted by, on average, seven introns (LANDER et al. 2001; VENTER et al. 2001). Although splicing is closely coupled to several other processes during gene expression, it is still widely thought that the primary fitness benefits that introns confer to a species are through improved evolution via exon shuffling, and increased proteome complexity by alternative splicing. Based on our observations we propose that introns confer an additional advantage; they improve the transcriptional and translational output of the genes they populate.

The spliceosome, which removes introns from the coding mRNA, is a large cellular complex containing hundreds of proteins and at least five small nuclear RNAs. It is closely coupled to, and in some cases directly interacts with, the proteins responsible for transcription, capping, polyadenylation, RNA export, and nonsense-mediated decay (MANIATIS and REED 2002). Given the extensive coupling of splicing with mRNA metabolism, it is not surprising that removing the introns from genes in higher eukaryotes (where intron-containing genes predominate) disrupts mRNA synthesis and often lowers cytoplasmic mRNA levels. The question arises: are the introns directly responsible for increasing gene expression or does their removal act indirectly, by simply derailing the mRNA synthesis assembly line? There are some examples in metazoans that support a direct role in expression; introns containing transcriptional
enhancers have been identified (Sleckman et al. 1996) and one group showed that removing introns from a gene disrupts nucleosome binding (Liu et al. 1995). There is however, no consensus that introns serve to increase gene expression. To investigate the role that introns may play in cellular fitness we studied their genetic contribution to the fitness of *Saccharomyces cerevisiae*.

In contrast to multi-cellular organisms, only 5% of *Saccharomyces cerevisiae* genes are interrupted by introns (most by a single intron) and all are constitutively removed during gene expression (Ast 2004; Balakrishnan et al.). Evolutionarily, hemiascomycetous yeast have experienced a massive reduction in introns (as well as numerous genes involved in splicing) as compared to *Schizosaccharomyces pombe* and other ancient ascomycetes (Aravind et al. 2000; Bon et al. 2003). It could be interpreted that the introns in *S. cerevisiae* are nucleic acid relics that have yet to be removed by evolution (Fink 1987). This view is mitigated by the observations that the majority (71%) of *S. cerevisiae* ribosomal genes contain introns, and these intron-containing ribosomal genes produce ~24% of cellular RNA (Ares et al. 1999). Thus, arguments have been made that introns may somehow be integral to ribosome biogenesis in yeast (Bon et al. 2003).

In this paper we present data that intron-containing genes produce more RNA and more protein than single exon genes in yeast. We further show that genetic deletion of introns from yeast genes decreases mRNA production, and in two cases of three we show that intron removal causes a phenotypic growth defect. We conclude from these
observations that introns confer fitness to an organism by improving transcriptional and translational output and suggest that they are required for competitive growth of yeast in their natural environment.

MATERIALS AND METHODS

Media and growth conditions. Standard YPD (yeast extract/peptone/dextrose) media and 30°C growth conditions were used as described in (GUTHRIE and FINK 1991). Cantharidin and latrunculin A were obtained from Biomol (catalog nos.PR-105 and T-119 respectively). Concentrated stocks were dissolved in DMSO and stored at -20°C until use.

Strains and plasmids. For construction of intron-minus strains (Δi) we used the isogeneic S288c strains BY4741 (MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0), BY4742 (MATα, his3Δ1, leu2Δ0, lys2Δ0, ura3Δ0), BY4743 (MATα/α, his3Δ1/Δ1, leu2Δ0/Δ0, lys2Δ0/Δ0, MET15/Δ0, ura3Δ0/Δ0, 4741/4742)(Open Biosystems, Huntsville, Alabama, United States).

We used the kanMX4-URA3 module of the pCORE plasmid (STORICI et al. 2001) (provided by M. Resnick, National Institutes of Health) to create intron-minus act1Δi and pre3Δi strains. We used the pAG36 plasmid, from the European Saccharomyces cerevisiae Archive for Functional Analysis (EUROSCRF), for construction of the intron-minus glc7Δi strains.
**Strain Construction.** All intron-minus strains were constructed using PCR-based gene replacement (WACH 1996). Intron-minus glc7Δi strains were created by deleting the wild type gene and replacing it with the coding sequence of GLC7 tagged with a nourseothricin marker (Nat+). A plasmid was made where the GLC7 coding sequence from start to stop was inserted into the pAG36 plasmid between the HindIII and BamHI sites, respectively. Linear PCR products were amplified from the plasmid using primers (all listed 5’ to 3’): ATGGACTCACAACCAGTTGA and CGCACTTAACTTCGCATCTG, which were preceded, 5-prime with 60 nucleotides comprising the GLC7 5’ and 3’UTR sequence, respectively. The PCR products were transformed into yeast strains BY4741, BY4742 (S288c), and transformants were selected on YPD plates containing 100ng/ml nourseothricin/clonNAT (Werner BioAgents, catalog no. 5.0100). Homozygous diploid intron-minus glc7Δi/glc7Δi strains were created by mating a and α intron-minus strains and diploids were selected on media lacking methionine and lysine. Nourseothricin resistant control strains containing wild type GLC7 genomic sequence were created in a similar fashion.

Intron-minus act1Δi and pre3Δi strains were constructed using the counter-selectable marker URA3 such that perfect, marker-free deletions of the introns were created. In the diploid yeast hybrid background BY4743 the intron of one copy of either ACT1 or PRE3 was replaced by the kanMX4-URA3 module. Strains were selected on YPD containing 200ug/ml geneticin (Agri-Bio, catalog no. 3000) before being transformed a second time with either ACT1 or PRE3 coding sequence. Transformants were recovered for two days on YPD plates before being replica plated onto plates containing 2.5mg/ml 5-Fluoroorotic Acid (to select for loss of the URA3 marker). All
intron-minus strains were confirmed with PCR and verified by sequencing. Strains were sporulated, dissected and back-crossed to wild type yeast to remove or dilute deleterious mutations.

**Quantitative PCR.** RNA was extracted from log-phase yeast growing in YPD as described (SCHMITT et al. 1990). RNA (20ng/ul) was reverse transcribed into cDNA using Taqman Reverse Transcription Reagents (Applied Biosystems, catalog no. N808-0234) and poly-dT primers. Primers for quantitative PCR were designed using primer3 (Rozen and Skaletsky 2000); the program’s default settings were used to select primers that produced 40-60 basepair amplicons. Quantitative PCR reaction mixes (25ul) contained 1x SYBR Green PCR Master Mix (Applied Biosystems catalog no. 4309155), 400nM each primer, and 2.5-250pg cDNA (estimated as 1% of the quantity of total RNA used in the cDNA synthesis reaction). Data were collected and analyzed on a 7700 Sequence Detection System (Applied Biosystems, Foster City, California, United States). Results for each primer set were reported as a cycle-threshold value (Ct). Ct values report the cycle at which SYBR fluorescence crosses a threshold; the threshold was manually set at a point within the exponential phase of the PCR reactions. Gene specific differences in Ct values were calculated by subtracting \( \Delta Ct_{\text{mutant}} - \Delta Ct_{\text{wildtype}} \) (\( \Delta Ct_{\text{gene}} \)), and normalized against the average \( \Delta Ct \) for three control genes (\( ACT1 \), \( TSA1 \), and \( ARO4 \))(average: \( \Delta Ct_{\text{control}} \)). Fold-differences reported were calculated from this equation: \( 1/(2^{((\text{average}\cdot\Delta Ct_{\text{gene}}) - (\text{average}\cdot\Delta Ct_{\text{control}}))}) \). P-values were calculated using a paired T-test.
**Phenotypic growth analysis.** Yeast strains were grown overnight in YPD at 30°C to saturation then diluted to a final concentration of OD_{600} 0.0625 in a volume of 100ul YPD with or without cantharidin or latrunculin A. Normalized cultures were grown in 96-well, flat-bottomed plates (Nunc, Rochester, New York, United States), in Tecan GENios microplate readers (Tecan, Zurich, Switzerland) with constant shaking at 30°C for up to 24 hours. Growth rates were determined by comparing average doubling times as previously described (LEE et al. 2005).

**RESULTS**

**Summary:** Computational analysis on intronic versus non-intronic genes revealed several pieces of evidence that suggest introns are required for increasing gene expression:

- Intrinsic genes produce more RNA and protein than non-intrinsic genes.
- Sub-categories of genes, including ribosomal, non-ribosomal, duplicated, non-duplicated all show the same bias for intronic genes being more highly expressed than single exon genes.
- Intrinsic ribosomal genes are more likely to be haploinsufficient and duplicated than non-intronic ribosomal genes.
- Intron position and length affect gene expression.

Genetic experimentation corroborated the computational results, demonstrating that introns are essential to wild type gene expression in yeast:
• Deletion of introns from three essential genes (ACT1, GLC7, and PRE3) decreases the RNA expression of all three genes and causes a growth defect for act1 and glc7 intronless mutants.

Intron-containing genes in *Saccharomyces cerevisiae* produce more RNA and more protein than intronless genes: The *Saccharomyces cerevisiae* genome consists of 5749 Open Reading Frames (ORFs) (BALAKRISHNAN et al.), 285 (5.0%) of the nuclear encoded ORFs contain introns (BALAKRISHNAN et al.; LOPEZ and SERAPHIN 2000; SPINGOLA *et al.* 1999). When we compared the mean transcriptional level of intronic genes (i-genes) to single exon genes (e-genes), as measured by microarray analysis (DEUTSCHBAUER *et al.* 2005), we found i-genes produce 3.9-fold more RNA on average than their non-intronic counterparts (Table 1, Figure 1A). All RNA abundance analyses were carried out on three datasets, two microarray expression data sets (DEUTSCHBAUER *et al.* 2005; HOLSTEGE *et al.* 1998) and one SAGE dataset (Serial Analysis of Gene Expression) (VELCULESCU *et al.* 1997); all three analyses provided essentially identical RNA abundance results (data not shown). Additionally, when we looked at mean protein abundance (GHAEMMAGHAMI *et al.* 2003) we found that the mean level of protein produced from i-genes to be 3.3-fold higher than for e-genes (Table 1, Figure 1B).

The differences in transcription and translation of i-genes verses e-genes could be biased by the fact that a high percentage of intronic genes are also ribosomal (37% of i-genes verses 2.5% of e-genes (Table 1). To address this we sub-divided the genes
into ribosomal and non-ribosomal categories, and then compared the protein and RNA abundance of i-genes and e-genes (genes were defined as “ribosomal” if they were part of the ribosome cellular component (as defined by the SGD (BALAKRISHNAN et al.))). The ribosomal subset of genes includes all the proteins that make up the cytosolic ribosome \((RPL, RPP \text{ and } RPS)\) as well as, many related genes (e.g. the mitochondrial ribosome, translation initiation and elongation factors, etc.). The results for the ribosomal genes were striking; we found that i-genes produced 3.7-fold more RNA and 4.1-fold more protein than ribosomal e-genes. We also analyzed a subset of genes containing only the cytosolic ribosome and found that it exhibited most of the trends seen for the larger “ribosomal” subset (Table 1); however, intronic abundance trends within the cytosolic ribosome were not observed, probably due to strong stoichiometric constraints on ribosomal proteins. Notably, non-ribosomal i-genes were also found to be more highly expressed than their intronless counterparts.

These data suggest that introns are evolutionarily retained in those genes that must be expressed at high levels and that introns may be directly or indirectly involved in increasing the RNA and protein production of yeast genes.

**Ribosomal intron-containing genes are more likely to be duplicated and haploinsufficient:** If introns increase protein output to improve yeast fitness, we would predict that i-genes would be enriched in those gene categories that display haploinsufficiency and those that are functionally-related duplicated genes. Our working hypothesis regarding gene duplication is consistent with the following: when genes are duplicated they usually follow one of two paths, functional divergence or gene loss. In
yeast some duplicated ribosomal genes appear to have followed a third path in that they retain their function (KELLIS et al. 2004). A simple explanation for this functional retention is that yeast benefit from increased protein production via this form of gene duplication. Substantiating this, we found that duplicated genes (as defined by KELLIS et al. 2004) produce, on average, 1.2-fold more protein and 1.5-fold more RNA than non-duplicated genes (Table 1). If, in a manner analogous to functionally-related gene duplication, introns serve to increase protein yield, we should see introns populating duplicated genes. In general, we found a larger percentage of duplicated genes contain introns compared to non-duplicated genes (8.6% verses 3.7%) and when we examined the ribosomal subset of genes we found 55% of ribosomal i-genes are duplicated whereas only 17% of ribosomal e-genes are duplicated (Table 1). In contrast, no significant portion of i-genes partitioned with duplicated genes for non-ribosomal cellular components, which could be expected given the greater functional divergence of duplicated non-ribosomal genes.

Our previous survey of genome-wide haploinsufficiency provides an insight into the role of introns, protein abundance and fitness (DEUTSCHBAUER et al. 2005). Haploinsufficient genes are defined as those that display growth defects when their copy number is halved in diploid organisms. In yeast the majority of haploinsufficiency can be ascribed to genes that have high (presumably maximal) levels of protein production (DEUTSCHBAUER et al. 2005). Because introns appear to modulate protein expression and affect fitness, we asked if introns predominate in genes known to affect growth due to haploinsufficiency. Our studies show that the majority (58%) of ribosomal i-genes are haploinsufficient compared to only 16% of ribosomal e-genes (Table 1). We
propose that introns act to increase the protein manufactured from haploinsufficient genes in a manner similar to duplicating the locus; that is, in both cases deleterious effects from limiting protein concentration would be lessened, compared to equivalent non-intronic or non-duplicated haploinsufficient genes.

Consistent with the results for haploinsufficiency, homozygous deletions of non-essential ribosomal i-genes are more likely to cause a slow growth phenotype than ribosomal e-genes (71% verses 47%, Table 1) (DEUTSCHBAUER et al. 2005). Haploid and diploid deletions both decrease protein output from the gene (diploid deletions abolish it). Yeast strains that are sensitive to single and double deletions (haploinsufficient and slow-growing homologous deletion (slow-hom) strains) are related such that all haploinsufficient genes are either essential or slow-homs. We conclude that introns improve the translational output of the genes they occupy and thus, i-genes are significantly more sensitive to both heterozygous and homozygous deletion.

**Deleting the intron from GLC7 creates a phenotypic growth defect in rich media:**
The results from our genomic analyses suggested introns positively contribute to the fitness of *Saccharomyces cerevisiae*. To further substantiate these results we deleted the introns from three essential genes: *ACT1, GLC7* and *PRE3* and measured growth rates as an indicator of fitness.

We precisely deleted the introns from *ACT1* and *PRE3* using the counter-selectable marker *URA3*, which left no exogenous sequence in our strains that might complicate analysis. We removed the intron from *GLC7* by replacing the wild type gene with an intronless *glc7Δi* gene, tethered at the 3-prime end to a nourseothricin
selectable marker. We compared our \textit{glc7\textasciitilde{i}}-\textit{Nat} strain to both wild type (BY4743) and a control strain where the \textit{GLC7} gene retains its intron and contains the nourseothricin marker. Wild type yeast and our \textit{GLC7-Nat} control strains behaved identically (data not shown). Intron-minus strains were mated to wild type yeast, sporulated, and the resulting tetrads were dissected. This single backcross was performed to remove or dilute any potentially deleterious mutations unrelated to intron removal. All three intron-minus strains were tested as haploids and as diploids; the homozygous diploid results are presented here (haploid cells behaved the same as diploids, data not shown).

In rich YPD media a growth defect was seen for \textit{glc7\textasciitilde{i}} but not for \textit{act1\textasciitilde{i}} or \textit{pre3\textasciitilde{i}}. Yeast strains were grown for more then 24 hours with optical density measurements collected every fifteen minutes, providing very accurate and reproducible growth curves and doubling times for 5-6 population doublings. Homozygous deletion of the \textit{GLC7} intron resulted in a prominent growth defect as compared to wild type (Figure 2). The doubling time for \textit{glc7\textasciitilde{i}} yeast was measured as 2.74 hours, a 2.2-fold increase over wild type. There was no similar growth defect observed for either \textit{act1\textasciitilde{i}} (Figure 2) or \textit{pre3\textasciitilde{i}} (data not shown) in YPD. In fact, we extended our studies on \textit{act1\textasciitilde{i}} and \textit{pre3\textasciitilde{i}} to see if any growth defects manifested after 20 generations using an automated growth assay (Deutschbauer et al. 2005) and saw no difference in growth compared to wild type yeast (data not shown).

**Intron-minus \textit{glc7\textasciitilde{i}} and \textit{act1\textasciitilde{i}} exhibit increased sensitivity to drugs that target their gene products:** The lack of a measurable growth defect for \textit{act1\textasciitilde{i}} and \textit{pre3\textasciitilde{i}}
strains could indicate an overabundance of protein during growth in rich media, which could buffer any deleterious effects of intron removal. More explicitly, intron deletion may lower protein production, but not below the threshold necessary to unveil a defect for growth in rich media. To address this, we challenged our intron-minus constructs with drugs specifically targeting their protein products. Our expectation was that intron-minus strains would be more sensitive to drug exposure than wild type yeast.

We challenged our yeast cultures with either cantharidin, which targets Glc7 (among other protein phosphatases, C.Nislow unpublished) or latrunculin, which de-polymerizes actin filaments (no known drugs target Pre3). Cultures of YPD were inoculated with 0.0625 OD_{600} of cells and either mock treated or treated with increasing concentrations of drug. Data were collected as described above. It was immediately apparent that the intron-minus strains were much more sensitive to drug treatment than wild type cells. This held true even in the case of act1Δi, where no growth defect had been observed previously (Figure 2). When treated with 10μM of cantharidin the doubling time for glc7Δi increased 94% (5.31hr/2.74hr) while wild type increased 4.8% (1.31hr/1.25hr). This works out to a stunning 43-fold increase in drug sensitivity ([5.31hr-2.74hr]/[1.31hr-1.25hr]). Similarly, the growth rate for act1Δi decreased 200% compared to 70% for wild type when exposed to 3.34μM latrunculin, which translates to 5-fold increase in drug sensitivity at this drug concentration.

The dramatic phenotypic effect we observed upon intron removal underscores the point that, for many highly expressed genes that contain introns, growth in YPD without perturbation may fail to reveal the significant fitness advantage conferred by introns. Consequently, the fitness benefit from introns can only be realized when yeast
are targeted by drugs or challenged by adverse environmental factors. Here we show that when intron-containing genes are challenged they appear to buffer against deleterious effects by improving protein production.

**Intron knockouts produce less RNA than wild type genes:** The results from our genomic studies suggest that introns serve to increase both RNA and protein abundance. Because transcription and translation are directly related, we hypothesized that the deleterious effects on growth rate resulting from intron deletion could, in part, be caused by decreased transcription of the intron-minus genes. Thus, we measured gene transcript levels using quantitative real-time PCR (qPCR).

We designed five sets of primers to interrogate different positions along the exons of *GLC7, ACT1* and *PRE3*. We used qPCR and these primers to measure the concentration of cDNA that was reverse transcribed from RNA extracted from wild type, *glc7Δi/glc7Δi, act1Δi/act1Δi, and pre3Δi/pre3Δi* cultures. Gene specific samples were normalized against data from seven primers constructed for three housekeeping genes: *ACT1, TSA1, ARO4* (*act1Δi/act1Δi* data were normalized against *TSA1* and *ARO4* only). Assays were repeated three times for *ACT1*, six for *PRE3* and >12 for *GLC7*.

Results showed a modest but reproducible decrease in mRNA levels for the three loci that we interrogated (Table 2). In each case, we found a highly significant decrease in mRNA for the homozygous intron-minus strains such that expression decreased at least 26.5% (73.5% of normal), with 95% confidence intervals that ranged between 1.9% and 4.0%. These data demonstrate the importance of intron sequences for maintaining wild type levels of mRNA, regardless of the locus.
Intron position and length correlate with gene expression: The question arises, how do introns increase RNA and protein output? We conducted a few simple correlation studies, searching for a more direct link between splicing and gene expression. We compared exon length, which dictates five-prime intron positioning, with RNA and protein abundance. We found weak, yet significant, anti-correlations between five-prime exon length and both RNA and protein abundance, suggesting that the closer an intron is to the transcriptional start of a gene the more positive effect it has on gene expression (Table 3); this characteristic has been observed repeatedly for metazoan introns (FURGER et al. 2002; PALMITER et al. 1991). Equally interesting, we found highly significant correlations between intron length and both protein and RNA abundance, which showed that genes with longer introns appear to be more highly expressed than genes with shorter introns (Table 3). The idea that intron length is linked to gene expression is further supported by the fact that ribosomal genes contain larger introns than their non-ribosomal counterparts (BON et al. 2003). These correlations strengthen the argument that introns and/or the act of their removal help up-regulate gene expression for the genes they occupy.

DISCUSSION

We have shown that intron containing genes in Saccharomyces cerevisiae produce more RNA and protein than their non-intronic counterparts. Furthermore, we have demonstrated that increased i-gene expression cannot simply be attributed to a
preponderance of ribosomal and duplicated genes within the i-gene subset; the same abundance biases hold true even when ribosomal or duplicated sub-categories of i- and e-genes are analyzed separately. In addition to our computational analyses, we have conducted a rigorous genetic examination of three essential i-genes in yeast and showed that intron removal will cause dramatic phenotypic growth defects possibly resulting from modest decreases in transcriptional output of intronless genes. Our results confirm previous genetic studies where intron-minus mutants expressed from exogenous plasmids in yeast revealed small transcriptional perturbations compared to wild type constructs (Furger et al. 2002). Moreover, we expanded upon those studies and showed that intron deletions cause growth defects, which have gone unseen in previous yeast intron-knockout experiments (Ng et al. 1985). We conclude from our investigations that introns subsist in part, to increase the transcriptional and translational output of their resident genes and thus, are integral to the fitness and competitive growth of Saccharomyces cerevisiae.

The observation that introns are evolutionarily retained in those genes that must be expressed at high levels is puzzling in light of the fact that no intronic transcriptional activators have been reported in yeast and no known exon-exon junction complexes have been described, which could enhance export, stability or translation of the processed mRNA (Nott et al. 2004). It is still possible that yeast introns contain unidentified regulatory sequences that could affect RNA expression levels and some circumspect evidence exists to support this supposition. (1) We identified a weak but statistically significant anti-correlation between five-prime exon length and both RNA protein abundance for i-genes, which suggests that introns closer to the transcriptional
start enhance overall gene expression. These anti-correlations are corroborated by mutational studies on DYN2 that showed removal of the five-prime intron is more deleterious to nuclear RNA abundance than removal of the three-prime intron (FURGER et al. 2002). (2) Our computational studies on introns have also revealed an interesting correlation between intron length and increased protein and RNA abundance. This observation is substantiated by the fact that ribosomal i-genes, which manufacture large amounts of RNA have been shown to contain larger introns than non-ribosomal i-genes (BON et al. 2003). It is tempting to posit that longer introns provide more “real estate” for potential regulatory sequences although no such sequences have yet been identified.

Given the observations that length and position of introns affect gene expression, future studies should include a re-examination of sequences in and surrounding yeast intronic genes. Improved pattern recognition algorithms or better sub-categorizing of i-genes into related groups (i.e. abundant i-genes with long introns, less abundant i-genes with short introns, or ribosomal and non-ribosomal i-genes) may help separate out degenerate sites from the noise. Another possibility is that existing splice-site signals may be sufficient for up-regulating gene expression directly or indirectly through protein interactions. Consequently, it would be reasonable to also search for proteins that preferentially co-purify with various sub-categories of introns and/or intronic genes, similar to what has been done to identify protein complexes in yeast (GAVIN et al. 2006; GAVIN et al. 2002).

In an effort to advance what is known of intronic effects on yeast survival we suggest that a complete library of sequence-tagged intron deletions be made for every intronic gene in Saccharomyces cerevisiae. Our data, combined with what has been
previously published, provides two examples of growth defects caused by intron deletions (act1 and glc7) and five examples where intron deletion decreases transcriptional output (act1, asc1, dyn2, glc7, pre3) (FURGER et al. 2002). Additional intronic deletions would likely show similar defects in transcription and could display deleterious phenotypic effects caused by decreased or mis-regulated protein production. Creating the deletions would allow for in-depth analysis of each intron; barcoding them with unique nucleic acid sequence tags would permit parallel high-throughput analysis of all ~300 intronic genes simultaneously (WINZELER et al. 1999).

We conclude that one main function of introns in S. cerevisiae is to enhance the transcriptional and translational output of the genes they occupy. We do not suggest that intron function is limited to boosting gene transcription; indeed, compelling evidence exists for regulated splicing among several yeast genes (K.Juneau unpublished; ENGBRECHT et al. 1991; Li et al. 1996; NAKAGAWA and OGAWA 1999). Instead, we propose that our understanding of intronic function is far from complete and much can be gained from continued examination of intronic function in Saccharomyces cerevisiae.

Acknowledgments: We thank Adam Deutschbauer and Guri Giaever for sharing data and scientific insight. This work was supported by National Institutes of Health grant RR020000 to R.W.D. and K.J.

REFERENCES

ARES, M., J.R., L. GRATE and M. H. PAULING, 1999 A handful of intron-containing genes produces the lion's share of yeast mRNA. Rna 5: 1138-1139.
GAVIN, A.-C., P. ALOY, P. GRANDI, R. KRAUSE, M. BOESCHE et al., 2006 Proteome survey reveals modularity of the yeast cell machinery. Nature advanced online publication.


ROZEN, S., and H. J. SKALETSKY, 2000 Primer3 on the WWW for general users and for biologist programmers. Humana Press, Totowa, NJ.


### TABLE 1

Correlating gene expression with gene classification

<table>
<thead>
<tr>
<th>Gene subsets</th>
<th>RNA Abundance</th>
<th>Protein Abundance</th>
<th>Haploinsufficient</th>
<th>Slow hom</th>
<th>Ribosomal</th>
<th>CytoRibosome</th>
<th>Duplicated</th>
<th>Essential</th>
</tr>
</thead>
<tbody>
<tr>
<td>All intronic</td>
<td>5,220</td>
<td>7,640</td>
<td>24.8</td>
<td>33.8</td>
<td>36.9</td>
<td>34.5</td>
<td>26.6</td>
<td>18.3</td>
</tr>
<tr>
<td>All nonintrinsic</td>
<td>1,350</td>
<td>2,340</td>
<td>3.5</td>
<td>13.4</td>
<td>2.5</td>
<td>0.5</td>
<td>12.9</td>
<td>16.3</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>2.3E-23</td>
<td>2.7E-17</td>
<td>7.3E-36</td>
<td>7.0E-18</td>
<td>1.0E-80</td>
<td>1.1E-113</td>
<td>1.1E-09</td>
<td>3.7E-01</td>
</tr>
<tr>
<td>Ribosomal-intronic</td>
<td>23,300</td>
<td>27,500</td>
<td>57.9</td>
<td>71.0</td>
<td>100</td>
<td>93.5</td>
<td>55.1</td>
<td>10.3</td>
</tr>
<tr>
<td>Ribosomal-nonintrinsic</td>
<td>6,220</td>
<td>6,750</td>
<td>16.0</td>
<td>46.9</td>
<td>100</td>
<td>21.0</td>
<td>16.7</td>
<td>22.2</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>2.3E-28</td>
<td>5.9E-10</td>
<td>1.2E-12</td>
<td>9.8E-05</td>
<td>n/a</td>
<td>6.7E-35</td>
<td>5.5E-11</td>
<td>1.4E-02</td>
</tr>
<tr>
<td>Cyto-ribosome intronic</td>
<td>24,800</td>
<td>29,400</td>
<td>62.0</td>
<td>74.0</td>
<td>100</td>
<td>100</td>
<td>58.0</td>
<td>9.0</td>
</tr>
<tr>
<td>Cyto-ribosome nonintrinsic</td>
<td>24,000</td>
<td>44,600</td>
<td>44.1</td>
<td>47.1</td>
<td>100</td>
<td>100</td>
<td>35.3</td>
<td>32.4</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>7.4E-01</td>
<td>3.2E-01</td>
<td>7.5E-02*</td>
<td>5.8E-03</td>
<td>n/a</td>
<td>n/a</td>
<td>2.9E-02</td>
<td>3.7E-03</td>
</tr>
<tr>
<td>Nonribo intronic</td>
<td>2,140</td>
<td>2,710</td>
<td>5.5</td>
<td>12.0</td>
<td>0</td>
<td>0</td>
<td>9.8</td>
<td><strong>23.0</strong></td>
</tr>
<tr>
<td>Nonribo-nonintrinsic</td>
<td>1,290</td>
<td>2,250</td>
<td>3.2</td>
<td>12.5</td>
<td>0</td>
<td>0</td>
<td>12.8</td>
<td>16.1</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>5.0E-04</td>
<td>1.4E-01</td>
<td>9.2E-02*</td>
<td>1.0E+00</td>
<td>n/a</td>
<td>n/a</td>
<td>3.1E-01</td>
<td>1.9E-02</td>
</tr>
<tr>
<td>Dup-intronic</td>
<td>13,100</td>
<td>20,000</td>
<td>46.8</td>
<td>59.7</td>
<td>76.6</td>
<td>75.3</td>
<td>100</td>
<td>3.9</td>
</tr>
<tr>
<td>Dup-nonintrinsic</td>
<td>1,790</td>
<td>2,430</td>
<td>3.3</td>
<td>8.5</td>
<td>3.3</td>
<td>1.5</td>
<td>100</td>
<td>4.6</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>1.9E-17</td>
<td>7.0E-15</td>
<td>5.5E-26</td>
<td>2.8E-25</td>
<td>7.6E-55</td>
<td>3.0E-62</td>
<td>n/a</td>
<td>1.0E+00</td>
</tr>
<tr>
<td>Nondup-intronic</td>
<td>3,720</td>
<td>4,850</td>
<td>16.9</td>
<td>24.4</td>
<td>22.5</td>
<td>19.7</td>
<td>0</td>
<td>23.5</td>
</tr>
<tr>
<td>Nondup-nonintrinsic</td>
<td>1,280</td>
<td>2,320</td>
<td>3.6</td>
<td>14.1</td>
<td>2.4</td>
<td>0.4</td>
<td>0</td>
<td>18.0</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>8.2E-12</td>
<td>8.7E-07</td>
<td>9.0E-14</td>
<td>9.1E-05</td>
<td>4.3E-29</td>
<td>3.8E-46</td>
<td>n/a</td>
<td>4.7E-02</td>
</tr>
<tr>
<td>Duplicated</td>
<td>2,060</td>
<td>2,960</td>
<td>7.0</td>
<td>12.9</td>
<td>9.6</td>
<td>7.8</td>
<td>100</td>
<td><strong>4.6</strong></td>
</tr>
<tr>
<td>Non-duplicated</td>
<td>1,330</td>
<td>2,390</td>
<td>4.1</td>
<td>14.5</td>
<td>3.2</td>
<td>1.1</td>
<td>0</td>
<td>18.2</td>
</tr>
<tr>
<td><strong>significance</strong></td>
<td>3.5E-15</td>
<td>6.0E-03</td>
<td>1.7E-04</td>
<td>2.2E-01</td>
<td>1.2E-15</td>
<td>2.6E-27</td>
<td>n/a</td>
<td>3.8E-31</td>
</tr>
</tbody>
</table>
TABLE 1. – Intron containing genes are more highly expressed than single exon genes. Protein (GHAEMMAGHAMI et al. 2003) and RNA abundance (DEUTSCHBAUER et al. 2005) data are shown as raw fluorescent values and the significance was calculated using a paired t-test; the remaining data are reported as percentages of assayed genes and the significance was calculated using a two-tailed Fisher’s exact test. Slow hom, are non-essential genes that when deleted from the diploid yeast cause a slow growth phenotype (DEUTSCHBAUER et al. 2005). Haploinsufficient, are genes that when present as only a single copy in a diploid organism show a slow growth phenotype (DEUTSCHBAUER et al. 2005). Ribosomal, are genes annotated as being part of the “ribosome cellular component” as defined by the SGD’s Gene Ontology (GO) terminology (BALAKRISHNAN et al.). CytoRibosome, are genes that make up the cytosolic ribosome (i.e. RPL, RPP, RPS genes). Duplicate, are genes having duplicate synteny within yeast (KELLIS et al. 2004). Significant values (the largest of the pair) are bolded, significant values slightly greater-than 0.05 are marked with an asterisk (*).
<table>
<thead>
<tr>
<th>Gene</th>
<th>Percent of normal</th>
<th>Confidence Interval (95%)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>glc7Δi</em></td>
<td>73.5</td>
<td>1.9</td>
<td>4.8E-33</td>
</tr>
<tr>
<td><em>pre3Δi</em></td>
<td>73.0</td>
<td>4.0</td>
<td>2.8E-12</td>
</tr>
<tr>
<td><em>act1Δi</em></td>
<td>72.5</td>
<td>2.0</td>
<td>1.8E-12</td>
</tr>
</tbody>
</table>
TABLE 3

Gene expression correlation with intron and exon length

<table>
<thead>
<tr>
<th>biomolecule</th>
<th>intron length</th>
<th></th>
<th></th>
<th>5' exon length</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>corr. coeff</td>
<td>p-value</td>
<td></td>
<td>corr. coeff</td>
<td>p-value</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>0.662</td>
<td>2.20E-16</td>
<td></td>
<td>-0.160</td>
<td>4.39E-02</td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>0.569</td>
<td>2.20E-16</td>
<td></td>
<td>-0.360</td>
<td>1.94E-07</td>
<td></td>
</tr>
</tbody>
</table>

TABLE 3 – Genes with larger introns and genes with shorter five-prime exons exhibit increased gene expression. Correlation values (corr.coeff) and p-values (p-value) are reported for protein abundance (Protein) and RNA abundance (RNA).
FIGURE LEGENDS

FIGURE 1. – Intronic genes are highly expressed. (A) RNA abundance; (B) protein abundance. The abundance (log₂) of intronic genes (blue) and non-intronic genes (red) are shown. The Density (y-axis) corresponds to the relative number of genes at a given abundance (the distribution is smoothed with a Gaussian kernel). The protein abundance data were from, (GAEMMAHAMI et al. 2003); and the RNA abundance data were from (DEUTSCHBAUER et al. 2005). Genes containing introns (blue) are more highly expressed at both the protein and RNA level.

FIGURE 2. – Deletion of introns results in a drug-induced fitness defect for both act1 and glc7. The x-axes denote time in hours; the y-axes denote optical density (OD) of the cell cultures. Top panels show growth of wild type diploids (red) verses glc7Δ/i/glc7Δ/i (black). Bottom panels show growth of wild type diploids (red) verses act1Δ/i/act1Δ/i (black). Left-side, untreated YPD media. Right-side, YPD media plus drug; glc7 experimental cultures contained 10uM cantharidin, act1 experimental cultures contained 3.34uM latrunculin. The doubling-time of each strain is expressed in hours and appears color-coded under corresponding curves.