Cell, Vol. 113, 919–933, June 27, 2003, Copyright ©2003 by Cell Press

A Panoramic View of Yeast Noncoding RNA Processing

Wen-Tao Peng,[1,5] Mark D. Robinson,[1,5] Sanie Mnaimneh,[1,5] Nevan J. Krogan,[1,5] Gerard Cagney,[1] Quaid Morris,[2] Armaity P. Davierwala,[1] Jörg Grigull,[1] Xueqi Yang,[1] Wen Zhang,[1] Nicholas Mitsakakis,[1] Owen W. Ryan,[1] Nira Datta,[1] Vladimir Jojic,[2] Chris Pal,[2] Veronica Canadien,[3] Dawn Richards,[3] Bryan Beattie,[3] Lani F. Wu,[4] Steven J. Altschuler,[4] Sam Roweis,[2] Brendan J. Frey,[2] Andrew Emili,[1] Jack F. Greenblatt,[1] and Timothy R. Hughes[1,*]

[1]Banting and Best Department of Medical Research
University of Toronto
112 College Street
Toronto, Ontario M5G 1L6
Canada

[2]Probabilistic and Statistical Inference Group
Departments of Electrical and Computer Engineering and Computer Science
University of Toronto
10 King's College Road
Toronto, Ontario M5S 3G4
Canada

[3]Affinium Pharmaceuticals
100 University Avenue
10th Floor
Toronto, Ontario M5J 1V6
Canada

[4]Bauer Center for Genomics Research
Harvard University
7 Divinity Avenue
Cambridge, Massachusetts 02138

Summary

Predictive analysis using publicly available yeast functional genomics and proteomics data suggests that many more proteins may be involved in biogenesis of ribonucleoproteins than are currently known. Using a microarray that monitors abundance and processing of noncoding RNAs, we analyzed 468 yeast strains carrying mutations in protein-coding genes, most of which have not previously been associated with RNA or RNP synthesis. Many strains mutated in uncharacterized genes displayed aberrant noncoding RNA profiles. Ten factors involved in noncoding RNA biogenesis were verified by further experimentation, including a protein required for 20S pre-rRNA processing (Tsr2p), a protein associated with the nuclear exosome (Lrp1p), and a factor required for box C/D snoRNA accumulation (Bcd1p). These data present a global view of yeast noncoding RNA processing and confirm that many currently uncharacterized yeast proteins are involved in biogenesis of noncoding RNA.

*Correspondence: t.hughes@utoronto.ca
[5]These authors contributed equally to this work.

Introduction

Over 95% of the nucleic acid in growing yeast cells is noncoding RNA (Sherman, 2002), most of which is functionally conserved over evolution and plays a role in basic cellular processes (Jeffares et al., 1998). Yeast noncoding RNAs include ribosomal RNA (rRNA), tRNA, spliceosomal RNA (snRNA), small nucleolar RNA (snoRNA, which specify cleavage and modification sites of other RNAs), telomerase RNA, signal recognition particle RNA, the RNA components of RNase P and RNase MRP (which process tRNA and rRNA, respectively), and mitochondrial rRNA (see Eddy, 2001 for review). In other organisms, some noncoding RNAs (e.g., microRNAs) play regulatory roles (Eddy, 2001), and recent experimental and computational screens suggest the existence of many more noncoding RNAs than have been identified so far (Huttenhofer et al., 2001; Eddy, 2002; Kapranov et al., 2002).

rRNA typically constitutes the vast majority of the nucleic acid content of an organism. The basic steps in yeast rRNA processing and ribosome biogenesis are well-documented: following transcription of the ~8 kb primary transcript by RNA polymerase I, the rRNA is extensively modified, cleaved into the 18S, 5.8S, and 25S fragments, and assembled into 40S and 60S subunits (Kressler et al. 1999; Venema and Tollervey, 1999; Fatica and Tollervey, 2002; and references therein). These steps involve a large assortment of factors: in addition to the 75 ribosomal proteins, several dozen small RNAs and at least 150 accessory proteins are involved in rRNA transcription, processing, and ribosome assembly. However, many details in the process of ribosome biogenesis remain unresolved, including the precise mechanism(s) by which most of the rRNA endonucleolytic steps occur. Additional rRNA-processing factors continue to be reported (e.g., Bassler et al., 2001; Harnpicharnchai et al., 2001; Dragon et al., 2002), indicating that the rRNA processing machinery has not yet been completely identified.

Mechanisms for processing noncoding RNAs other than rRNA are also not completely understood (e.g., the enzyme that trims yeast tRNA 3′ ends has not yet been identified unambiguously; Morl and Marchfelder, 2001), and there are many instances of shared processing mechanisms among different noncoding RNAs and even mRNAs (Pederson, 1998; Fatica et al., 2000). For instance, the exosome, a protein complex composed primarily of 3′→5′ exonucleases that is involved in cytoplasmic mRNA degradation (Mitchell et al., 1997), also functions in biogenesis of rRNA, snoRNA, and snRNA (Allmang et al., 1999), and RNase III cleaves not only the 3′ETS (*E*xternal *T*ranscribed *S*equence) of pre-rRNA, but is also involved in processing snoRNAs and some spliceosomal RNAs (Elela et al., 1996; Chanfreau et al., 1998; Kufel et al., 1999).

A principle objective of functional genomics and proteomics is to determine the functions of genes and proteins discovered by genome sequencing. Because large-scale experimental and computational systems for ascribing potential protein functions are often tested in yeast, it is now possible to draw hypotheses regarding cellular functions or biochemical properties of many of the ~2,000 uncharacterized or poorly characterized protein-encoding yeast genes (Mewes et al., 2002; Issel-Tarver et al., 2002). Initial analyses suggest that several hundred currently uncharacterized yeast proteins may be involved in RNA processing or RNP biogenesis activities (Bader and Hogue, 2002; Wu et al., 2002; Milkereit et al., 2003). These proteins not only include components of recently described large nucleolar protein complexes (Bassler et al., 2001; Harnpicharnchai et al., 2001; Dragon et al., 2002; Fatica and Tollervey, 2002; Grandi et al., 2002; Nissan et al., 2002) but also many other proteins that are not contained in these complexes.

With the goal of determining whether these functional genomic- and proteomic-based predictions are accurate, we have performed a large-scale survey of the functions of known and potential yeast RNA processing proteins. Our data support the prediction that many uncharacterized yeast proteins are involved in noncoding RNA biogenesis, predominantly synthesis of ribosomes.

Results

Compilation of Known and Predicted RNA Processing Proteins

In order to generate a list of proteins known and predicted to be involved in noncoding RNA and RNP biogenesis, we first compiled a catalog of the factors known to be involved in these processes. Combination of noncoding-RNA-specific categories from two online databases (Issel-Tarver et al., 2002; Mewes et al., 2002) identified a total of 413 proteins, including 253 that are essential for viability (Supplemental Figure S1 available at http://www.cell.com/cgi/content/full/113/7/919/DC1). Our collection did not include any ribosomal proteins, translation-specific proteins, or tRNA-aminoacylases.

We next applied several methods to predict potential RNA processing factors. A number of publicly available yeast data sets are amenable to making predictions, including two-hybrid (Uetz et al., 2000; Ito et al., 2001), localization (Kumar et al., 2002; Andersen et al., 2002), affinity-tagged protein complexes (Gavin et al., 2002; Ho et al., 2002), transcriptional coregulation (Wu et al., 2002), and growth phenotypes of deletion mutants (Giaever et al., 2002). Methods for refining or filtering these data in order to predict gene functions have been described (Brown et al., 2000; Wu et al., 2002). We applied different combinations of data inputs to different predictive systems (Supplemental Figure S1). In total, the resulting set of 919 predicted RNA processing proteins contained 578 proteins annotated as "GO (gene ontology) biological process unknown" (Issel-Tarver et al., 2002), including 75 predictions that arose independently from more than one data source. In addition, 341 proteins that carry other GO annotations were identified by one or more of the predictive strategies as being potentially associated with noncoding RNA biosynthesis.

Most of the predictive methods are less than 50% accurate, on the basis of how precisely they identify known RNA processing proteins (Brown et al., 2000; Wu et al., 2002; data not shown). This makes it necessary to test experimentally whether individual proteins are required for specific RNA processing events.

tet-Promoter Alleles for Phenotypic Analysis

Among the 413 known and 919 predicted RNA processing proteins, 525 are essential for cell viability. We created tetracycline-regulatable ($tetO_7$-promoter) alleles (Gari et al., 1997; Hughes et al., 2000) for 169 of these. An additional set of 26 negative control strains was obtained in which the promoter replacement was in a gene not known or predicted to be involved in RNA processing or RNP biogenesis (see Strains in Supplemental Data).

Measuring Noncoding RNA Biogenesis with a Custom DNA Oligonucleotide Microarray

Previous studies have used oligonucleotide microarrays to detect alternative or aberrant splicing of mRNA (Shoemaker et al., 2001; Clark et al., 2002). In order to facilitate simultaneous analysis of a variety of noncoding RNA processing events in our yeast mutant strains, we constructed microarrays with 212 different oligonucleotides designed to hybridize specifically to primary transcripts, processed fragments and junctions, and/or final products of a representative group of coding and noncoding RNAs (Figure 1A).

RNA processing defects typically result in accumulation of one or more precursors, which contain unprocessed flanking sequences. At the same time, the relative abundance of the downstream product(s) is often reduced. To assay for RNA processing defects, total RNA from simultaneously grown wild-type and mutant cultures was coupled directly to fluorescent dyes and hybridized to the array using a ratio-based two-color system (see Experimental Procedures). When the microarray data were represented in clustergrams (Eisen et al., 1998), processing defects were often signaled by a green color of the final product (i.e., reduced in abundance relative to wild-type) and red color of one or more flanking regions (i.e., increased in abundance relative to wild-type).

For example, Figure 1B includes an analysis of *TSR1*, a gene known to be required for normal processing of 20S pre-rRNA to 18S rRNA (Figure 1A). The $tetO_7$-*TSR1* mutant accumulated 20S pre-rRNA (i.e., 18S with the 3′ flanking sequence still attached), and displayed a concomitant reduction in the 18S product (Gelperin et al., 2001; Wu et al., 2002), which in wild-type cells is normally much more abundant than the 20S precursor. The result on the array is boxed in pink in Figure 1B: the oligonucleotides complementary to the 18S product appeared green (since they detect both the 18S and 20S rRNAs) and those specific to the flanking sequences contained only in the 20S precursor appeared red.

We initially tested the general utility of our array with a panel of 15 mutations in genes with well-established roles in a variety of RNA processing events (Figure 1B; RNA segments expected to show processing defects are boxed). In all of these examples, patterns corresponding to the expected RNA processing defect(s)

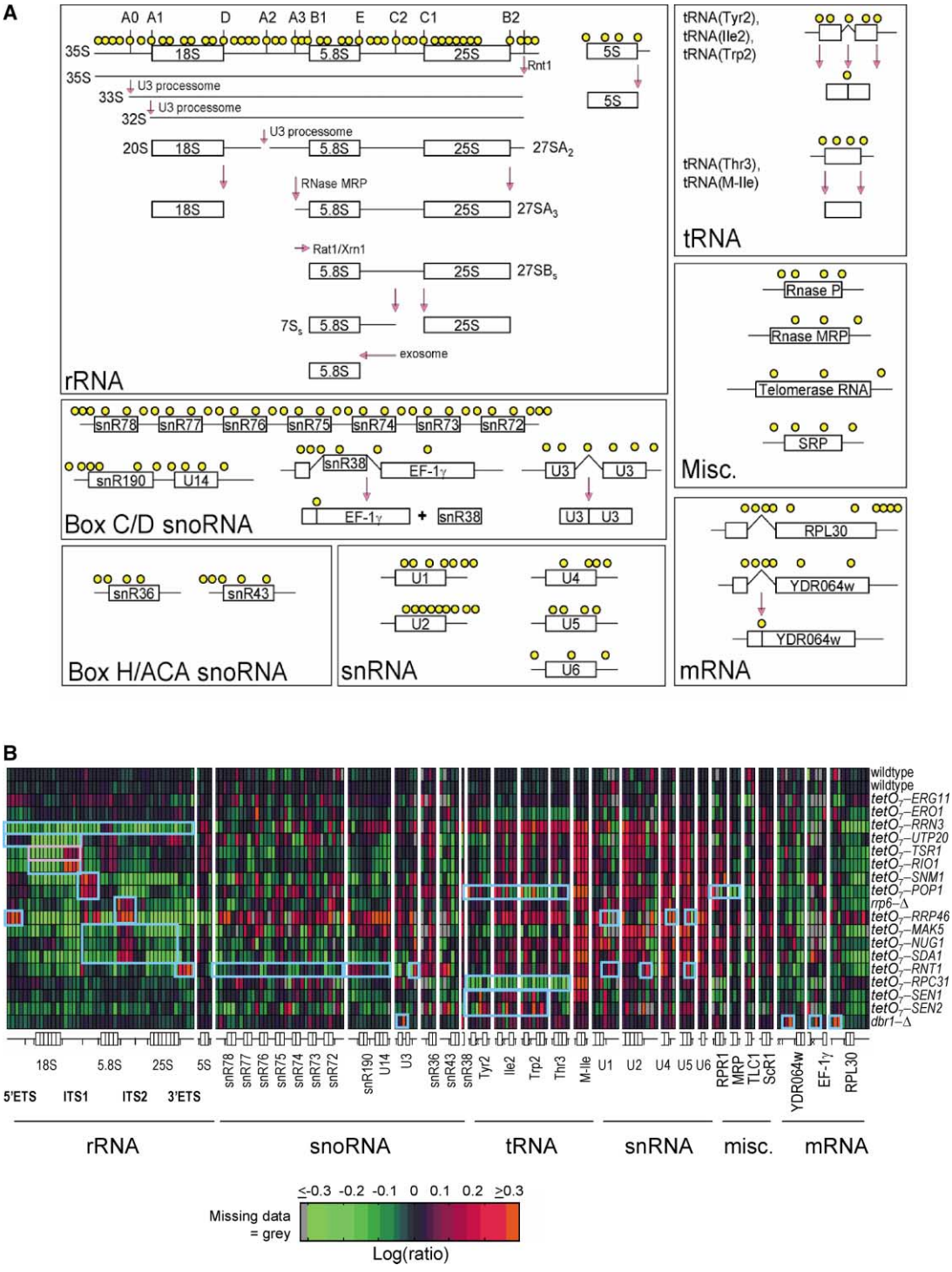

Figure 1. A Microarray that Measures Abundance and Processing of Noncoding RNA

(A) Schematic representation of RNA molecules, junctions, and processing intermediates represented by complementary oligonucleotides on the RNA processing microarray. Boxes represent the final product RNA. Lines represent parts of the primary transcript that are removed during processing. Introns are represented by a jointed line. Processing sites on the rRNA follow conventional naming (e.g., Kressler et al., 1999). Yellow circles represent sequences detected by complementary oligonucleotides on the array.

(B) Experiments demonstrating efficacy of the RNA processing microarray. Oligonucleotides are ordered according to the schematic at the bottom, which reflects final products (boxes), flanking or intervening sequences (lines), and exon-spanning probes (X's). Red indicates increased relative abundance in the mutant and green represents decreased relative abundance. Superimposed blue and pink boxes indicate anticipated processing defects. *tetO$_7$-ERG11* and *tetO$_7$-ERO1* are negative control mutants chosen to reflect the same growth rates as the positive control mutants.

were observed. For example, accumulation of intron sequences was discerned clearly in the *dbr1-*Δ mutant strain, which encodes the intron lariat debranching enzyme (Chapman and Boeke, 1991). Similarly, elevation of the 3′ETS of pre-rRNA, as well as flanking sequences of some small RNAs, was prominent in the *tetO*$_7$*-RNT1* strain (Figure 1B), consistent with the established function of Rnt1p in the maturation of these transcripts (Elela et al., 1996; Kufel et al., 1999). These examples supported the efficacy of both the *tetO*$_7$-promoter approach and the microarray methodology and showed that data could be obtained from array spots corresponding to normally low-abundance RNA fragments (e.g., introns, snoRNAs, 5′ETS and 3′ETS of pre-rRNA), even though measurements of these fragments are prone to error due to signal noise. One potential difficulty in interpreting these data is that, due to normalization, reduction of one RNA species often resulted in apparent increase in other species (e.g., tRNA appears more abundant in the *tetO*$_7$*-RRN3* mutant because the rRNA is reduced; Figure 1B). However, this can be overcome by (1) focusing on examples where the final product is green and flanking sequences are red, (2) comparing array phenotypes of mutant strains to each other, or (3) applying pattern-recognition algorithms (see below).

Phenotypic Analysis of 468 Mutants

Four hundred sixty-eight different mutant strains were assayed using the RNA processing microarray. Of these 468, 133 were genes with noncoding RNA-related annotations, 141 were genes predicted to function in RNA biogenesis but currently listed as "biological process unknown" in the *Saccharomyces* Genome Database (Issel-Tarver et al., 2002), and 168 were genes predicted by our analysis to function in RNA biogenesis, but having GO annotations related to other processes (e.g., protein synthesis, nucleobase metabolism, cell-cycle, etc.). Twenty-six mutants in genes that have firmly established and direct roles in pathways unrelated to RNA or RNP function (e.g., sterol biosynthesis) were included as negative controls. An additional 129 negative control experiments were performed in which two isogenic, wild-type cultures (grown in parallel with a set of mutant versus wild-type pairs) were compared, to establish a baseline of measurement error and variation between two different cultures. Clustering analysis (Figure 2A, top) shows that a wide variety of phenotypes were obtained.

Classifying Mutant Phenotypes Measured on the RNA Processing Microarray

We next asked whether mutations affecting different RNA classes could be accurately and objectively identified. Manual inspection of clustering diagrams (i.e., looking for expected patterns, as in Figure 1B), while effective for the identification of specific defects (see below), was difficult to apply to all mutants and RNA types objectively. Conventional 2D clustering analysis (Eisen et al., 1998) was confounded by the fact that many RNA processing factors are involved in biogenesis of multiple RNA species.

To enable impartial statistical analysis of individual RNA processing classes, we assigned each of the 468 mutants as "positive," "negative," or "unknown" for biosynthesis of rRNA, snoRNA, tRNA, snRNA, RNase P/MRP RNAs, and/or mRNA, on the basis of Medline abstracts and other compiled online databases (Issel-Tarver et al., 2002; Mewes et al., 2002; see Supplemental Data). For each of the six classes, there were at least ten positive mutants (i.e., known to affect processing of the particular RNA(s) in question) among the 468. For each RNA class, we applied a computational classification technique (based on kernel density estimation; see Supplemental Data), which automatically distinguishes whether the microarray data from a mutant is more similar to positives than it is to negatives. This technique generated a single discriminant value (i.e., score reflecting relative belief that a mutant is a positive) for each mutant in each RNA class (regardless of whether the mutant was positive, negative, or unknown). Since the score of each mutant was determined without knowledge of whether it was a positive or negative itself, the results provide an objective measure of how well de novo classifications can be made from the microarray data. Since the scores for each RNA class were derived using only the oligonucleotides on the array that detect that specific RNA class, this analysis also provided a measure of how well the processing of the individual RNA classes was measured.

ROC curves (Figure 2B) provide a summary of how well positives and negatives could be distinguished from one another in each RNA class. To produce these curves, the number of true-positives and false-positives was plotted at varying discriminant values (unknowns are not shown). For example, the curve for rRNA indicates that most of the mutants with the highest discriminant values are true-positives (i.e., bona fide rRNA processing mutants), since the curve proceeds almost vertically from the origin. The entire curve is well above the expected result for random discriminant values (dashed diagonal line), showing that most of the positive rRNA mutants can be objectively distinguished from most of the negatives. A subset of the mutants known to affect tRNA transcription and/or processing are also readily distinguished from the vast majority of the other mutants by this technique, as are mutants that are known to affect transcription and/or processing of the RNA subunits of RNases P and MRP. Mutants in snoRNA, snRNA, and mRNA biogenesis factors were identified less reliably (Figure 2B), possibly because the abundance of these RNAs (and hence signal-to-noise on the array) is lower. In the case of snoRNA and snRNA, the relatively small number of positive examples also makes it difficult for the technique to extract patterns that distinguish them from negatives. Together, these results indicate that the array data can be used to objectively identify mutants with processing defects in a variety of noncoding RNA classes. It is not surprising that the classifications are less than perfect, since not all of the expected defects in known RNA processing mutants could be detected by our array, and some of our mutant alleles did not display complete loss of function (since many *tetO*$_7$-promoter alleles of essential genes do not completely cease growth).

The discriminant values generated for the unknowns can be used to guide and prioritize more detailed analysis of specific mutants in specific RNA classes. The six lower images in Figure 2A show the data from all of the

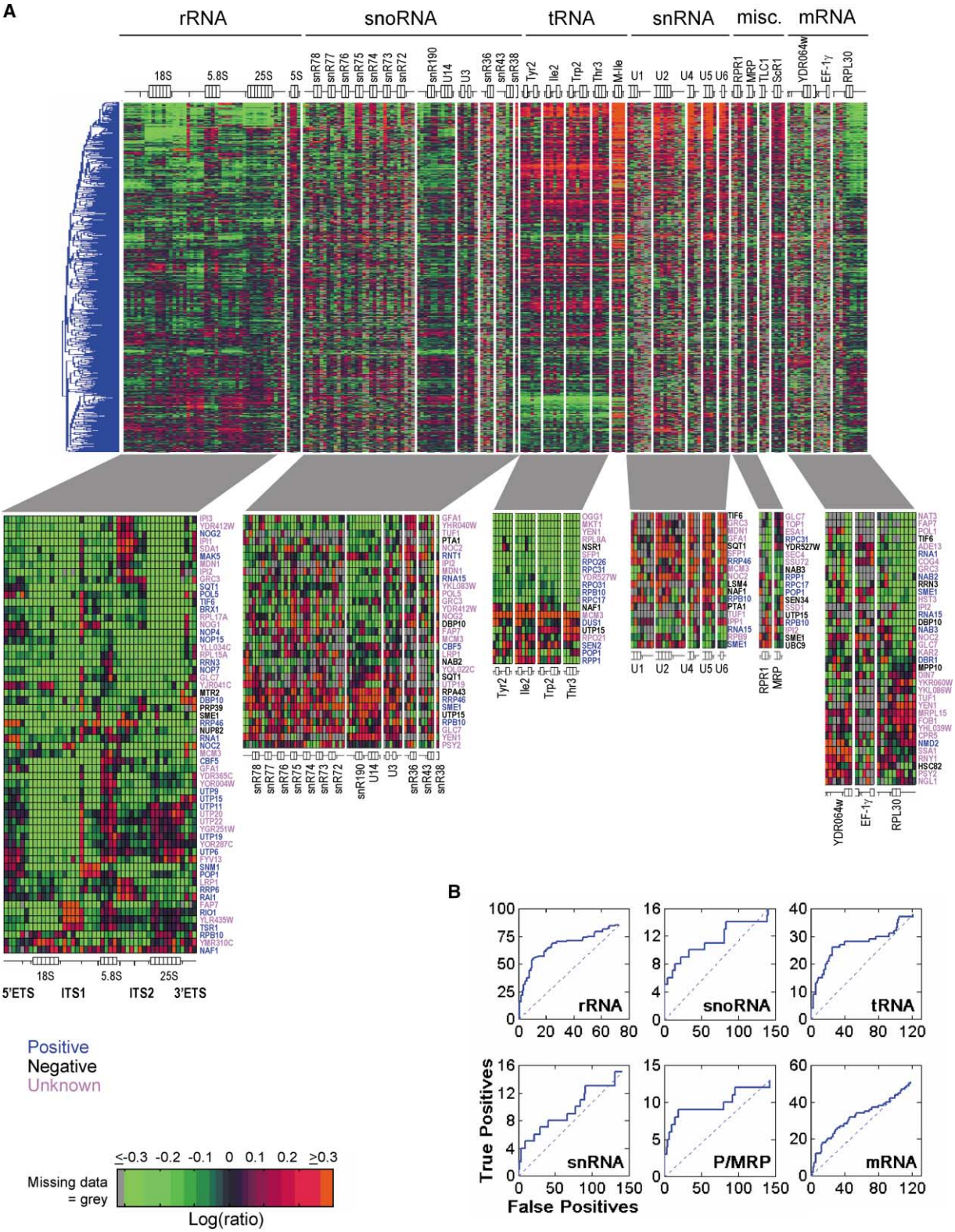

Figure 2. Classification of 597 Experiments Assaying Yeast Mutants with the RNA Processing Microarray

(A) Clustering analysis. Oligonucleotides were ordered according to the schematic at the top, which is identical to that in Figure 1B. The six pullouts at the bottom are the mutants in each of the six RNA classes in (B) with discriminant values higher than 90% of the negative mutants in that RNA class.

(B) The six curves were generated by plotting the number of true-positives and false-positives at varying discriminant values. These curves are scaled versions of ROC curves. See Supplemental Data for details and discriminant values.

mutants with discriminant values that are above 90% of the negatives from the relevant RNA class. Unknowns are labeled in purple. The largest number of unknowns, and the largest variety of easily interpreted phenotypes, were found among mutants affecting rRNA.

Proteins Involved in rRNA Processing

We next selected mutant strains with specific rRNA processing phenotypes detected on the microarray for further examination by other assays, in order to confirm the microarray phenotypes and discover gene functions. Figure 3 shows 115 mutants with microarray phenotypes that correspond to patterns anticipated from defects in one or more specific steps in processing of the 35S rRNA transcript, i.e., accumulation of one or more flanking or intervening precursor sequences. Since most of the rRNA-related array phenotypes could be recognized and interpreted manually, the different groups were derived by clustering all of the mutants only on the basis of the rRNA data, and manually extracting and reordering the clusters so that the order from top to bottom of Figure 3 reflects the known sequence of processing events (shown in Figure 1A). The specific processing defects that characterize each group of mutants are boxed in blue in Figure 3. Mutant strains that are currently annotated as "biological process unknown" but were classified as positives by one or more statistical techniques we applied (Supplemental Data), yet did not obviously correspond to any easily identifiable class of rRNA processing defect, are shown at the bottom of Figure 3.

Genes with U3 Processome-Like Mutant Phenotypes

Figure 3 (top) shows 37 mutants that have apparent defects in 5′ETS removal, a phenotype suggested by a decrease in 18S and 20S hybridization, but elevated levels of 5′ETS relative to 18S. Most of these also have elevated hybridization to the A2 junction probe, indicating defective or delayed A2 cleavage and consistent with a role in U3 function (Kressler et al., 1999). Indeed, at least eight mutants correspond to components of the recently described "U3 processome" (Dragon et al., 2002). We examined three of the strains mutated in uncharacterized genes (*ygr272c*-Δ, *fyv7*-Δ, and *tetO*$_7$-*YGR251w*) by Northern blotting (Figure 4A), which confirmed accumulation of uncleaved 35S precursor and the aberrant 23S species, and reduction of both 27SA2 and 20S pre-rRNA. These defects are very similar to those for strains with mutations in the established U3 processome components Utp9p and Bud21p, the unpublished Utp20 and Utp22 proteins (UTP, *U T*hree *P*rocessome; Issel-Tarver et al., 2002), and also a C-terminal deletion in *NSR1*, which is required for 18S biogenesis and proper snoRNA localization (Verheggen et al., 2001; Lee et al., 1992) (Figure 4A). 27SB pre-rRNA abundance is normal in these mutants, again consistent with the known U3 processome components (Figure 4A). The three uncharacterized proteins are all required for normal growth: Ygr272cp and Fyv7p are nonessential (but deletion mutants grow slowly) while Ygr251wp is essential (Giaever et al., 2002). Both Ygr272cp and Fyv7p contain coiled-coil motifs, which typically function as protein-protein interaction domains, and are highly enriched among the list of 413 known noncoding RNA biogenesis proteins we assembled [90 of the 413 (21.8%) contain a coiled-coiled domain; in comparison to 776 among all 6267 yeast ORFs (12.4%); the probability of attaining such a high proportion by random draws is $P \leq 2 \times 10^{-8}$ assuming random draws] (Issel-Tarver et al., 2002; Robinson et al., 2002). However, none of these three proteins was previously linked to rRNA processing in any large-scale two-hybrid or affinity-tagging studies of yeast protein complexes, including the description of the U3 processome (Uetz et al., 2000; Ito et al., 2001; Dragon et al., 2002; Gavin et al., 2002; Ho et al., 2002).

An additional mutant, *tetO*$_7$-*RNA1*, displayed a previously unreported accumulation of 5′ETS-A0 sequences (Figure 3). This fragment is likely the cleaved but undegraded 5′ETS-A0 (Figure 4A). The Rna1 protein has previously been implicated in a variety of other RNA-related processes (reviewed in Hopper et al., 1990). In addition to showing a 5′ETS degradation defect, our *tetO*$_7$-*RNA1* mutant was severely compromised for overall rRNA biosynthesis (Figure 4A).

Mutation of Pol II Processing Factors Delays A2 Cleavage

Figure 3 shows 22 mutant strains that display increased hybridization to the A2 junction, in the absence of any other striking pattern aside from most of the other probes reporting decreased hybridization, suggesting that delayed A2 cleavage is the sole defect. Analogous to the *TSR1* 20S/18S example above, this pattern is explained by the fact that in the wild-type cell the uncleaved 35S has much lower abundance than the cleavage products, 20S and 27S, to clarify that 20S and 27S are themselves the cleavage products (e.g., Figure 4A, lanes 1 and 2). Accumulation of 35S was confirmed by Northern blotting for two of these strains (*tetO*$_7$-*SSU72* and *tetO*$_7$-*GLC7*; data not shown). This pattern is distinct from all other mutants, including the U3-like mutants and also mutants affecting processing of ITS2 (*I*nternal *T*ranscribed *S*equence *2*; see Figures 1, 3, and below), which also accumulate 35S precursor uncleaved at A2 as has been previously noted (Grandi et al., 2002). Seven of the 22 mutated genes in this group (*LSM5*, *SSU72*, *GLC7*, *PRP5*, *PRP39*, *PRP42*, and *EPL1*) have activities related to Pol II transcription, particularly mRNA splicing and 3′-end formation. This is a significant enrichment in this group of 22 mutants ($P < 0.0057$, assuming random draws of the 51 Pol II mutants among the 468 mutants). The delay in A2 cleavage may be a consequence of either altered snoRNA processing and/or reduction in the abundance of ribosomal proteins, many of which are produced from spliced mRNA transcripts.

Fap7 and Tsr2 Are Required for 20S Processing

Two proteins not previously known to be involved in rRNA processing displayed prominent accumulation of the 20S precursor RNA when mutated (Figure 4B). One of them, Fap7p, was previously described as an essential nuclear protein involved in oxidative stress response (Juhnke et al., 2000); however, it has also been identified in association with Krr1p (Ito et al., 2001) and Utp13p (Gavin et al., 2002), supporting a role for Fap7p in small-subunit ribosome biogenesis (Gromadka and Rytka, 2000; Sasaki et al., 2000; Dragon et al., 2002).

The second protein, Ylr435wp (referred to hereafter as Tsr2, for *T*wenty *S R*NA accumulation), is nonessential, but deletion resulted in slow growth (doubling time ~2.5 hr) in addition to a prominent 20S accumulation and a corresponding 18S deficit (Figure 4B). We de-

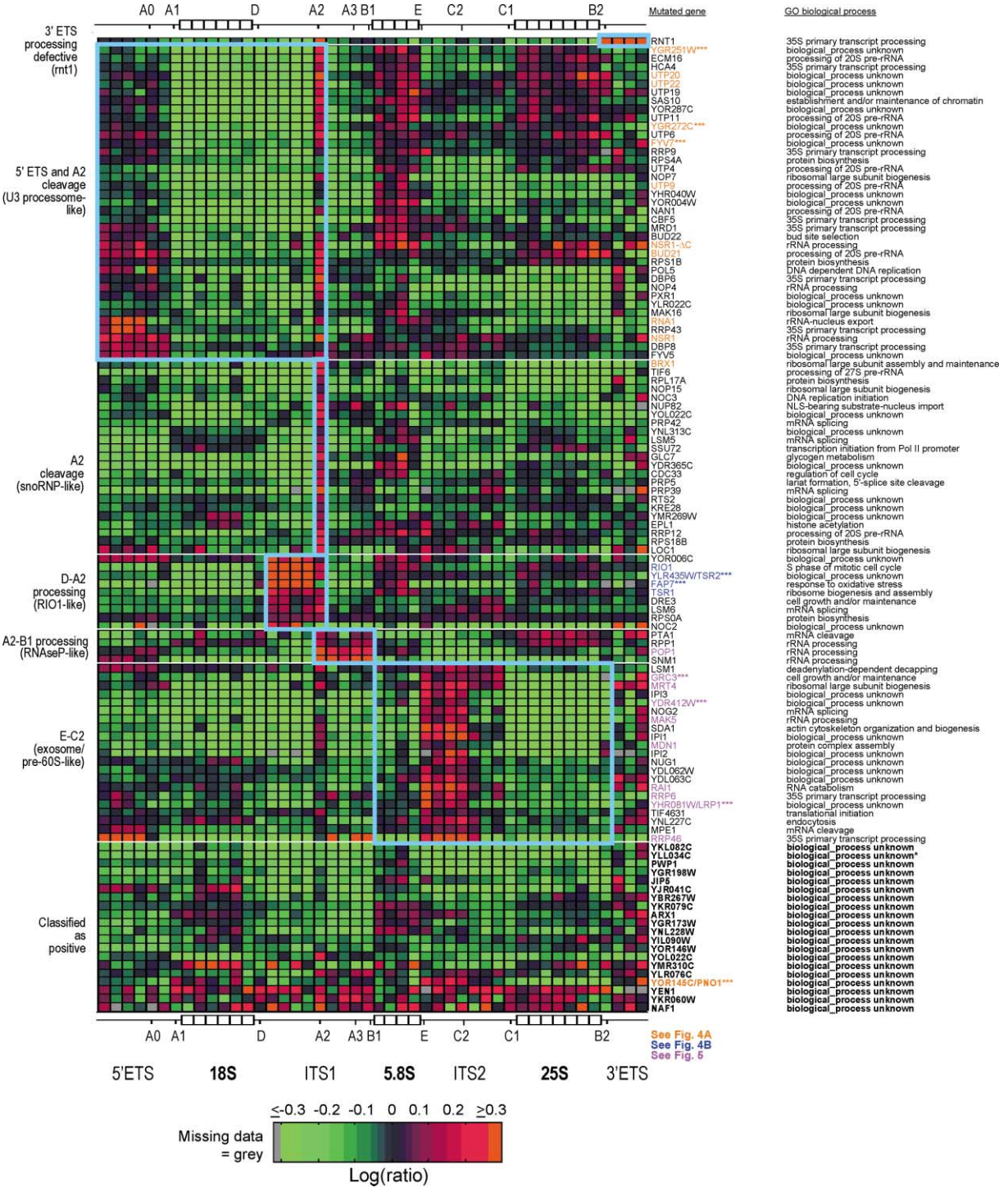

Figure 3. Mutant Strains with rRNA Biogenesis Defects Detected Using Microarrays

GO biological process was taken from SGD (Issel-Tarver et al., 2002). See text for details of mutant selection. The mutated genes analyzed further in Figures 4 and 5 are colored.

tected Tsr2p in association with Rps26p, using TAP-tagging (Figure 4C; Rigaut et al., 1999). This association has also been detected by two-hybrid analysis (Uetz et al., 2000). In view of these observations, the role of Tsr2p in small subunit biogenesis is likely to be direct.

ITS2-Processing Factors

A large number of proteins are already known or suspected to be involved in the synthesis of the large ribosomal subunit (Kressler et al., 1999; Fatica and Tollervey, 2002), and several variants of a large pre-60S complex

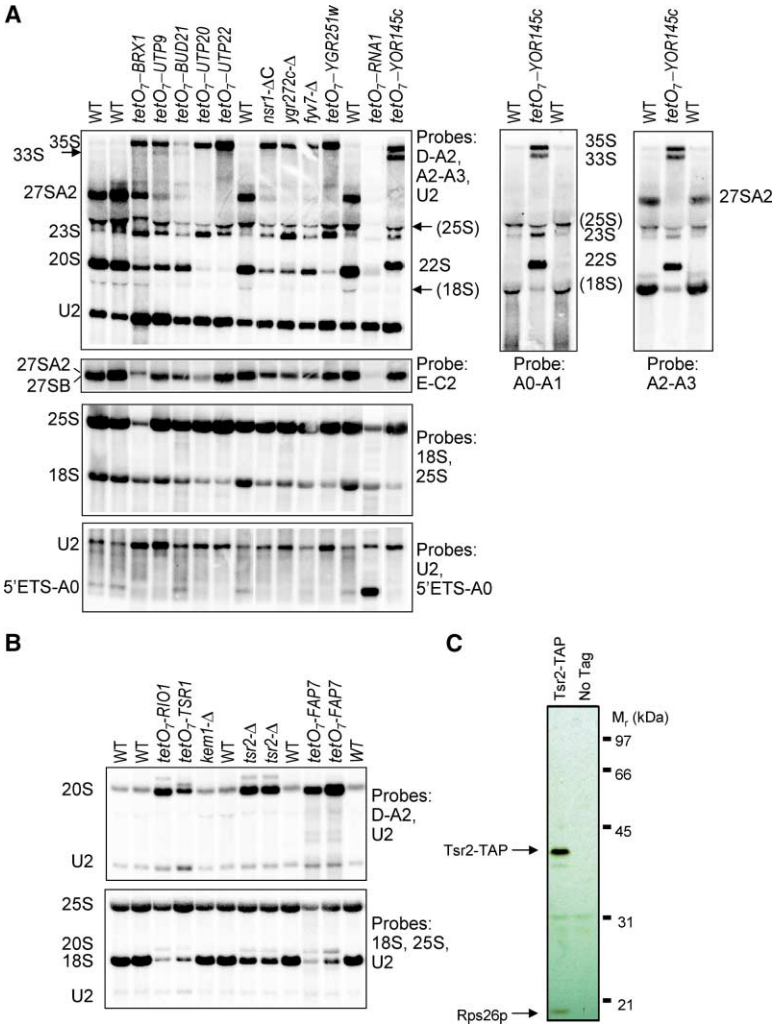

Figure 4. Northern Blot Analysis of Mutants with 18S rRNA Biogenesis Defects

(A) Strains with apparent defects in 5′ETS removal or degradation. *nsr1*ΔC is the *ygr160w*-Δ mutant from Giaever et al., 2002; *YGR160w* overlaps the C-terminal half of *NSR1*. *tetO7-BRX1* is shown for comparison. The bands indicated as (18S) and (25S) are presumably crosshybridization of one or more of the probes to the highly abundant 18S and 25S rRNAs.

(B) Strains with apparent defects in 20S pre-rRNA processing.

(C) Affinity purification of TAP-tagged Tsr2p.

involved in ITS2 processing and large subunit export have been described recently (Bassler et al., 2001; Harnpicharnchai et al., 2001; Saveanu et al., 2001; Fatica et al., 2002a; Grandi et al., 2002, Nissan et al., 2002). Hallmarks of defects in the synthesis of large ribosomal subunits include a reduction in 25S and 5.8S rRNA, and in many cases altered abundance of ITS2-containing 27S and 7S pre-rRNAs. At least 20 mutant strains displayed specific elevation of ITS2-containing sequences, among them mutants in pre-60S components *MAK5*, *SDA1*, and *NUG1* (Figure 3). The *tetO7-MDN1* strain, mutated in a poorly characterized component of pre-60S complexes (Bassler et al., 2001) mimicked the *tetO7-MAK5* phenotype (Figures 3 and 5A). Deletion of *MRT4*, a nonessential gene previously implicated in mRNA turnover and also present in a published pre-60S complex (Harnpicharnchai et al., 2001), also caused defects closely resembling those of *tetO7-MAK5*, as did *tetO7-YDR412w* (Figures 3 and 5A). Ydr412wp has not been previously reported to be involved in 60S biogenesis (Issel-Tarver et al., 2002; Mewes et al., 2002), although it has been detected in a protein complex with affinity-tagged 60S biogenesis factor Erb1p (Pestov et al., 2001) along with 37 other proteins, many of which are involved in processes not related to RNA (Ho et al., 2002).

tetO7-GRC3, in contrast, had a distinctive impact on ITS2 processing, including reduced levels of 27SA2 and 7S in addition to the more typical reduction in 25S and increase in 35S levels (Figure 5A). Furthermore, *tetO7-GRC3* displayed the 3′ extended 5.8S species also observed in *rrp6*-Δ (Figure 5A). The *tetO7-GRC3* strain had a very pronounced 27SB accumulation, visible even with a 25S probe (Figure 5A). *GRC3* is almost completely uncharacterized (Issel-Tarver et al., 2002) and Grc3p is not found in any of the published pre-60S complexes and is also not represented in any large-scale protein interaction analyses (Uetz et al., 2000; Ito et al., 2001; Gavin et al., 2002; Ho et al., 2002).

An Exosome-Associated Protein Required for 5.8S Formation

The strain carrying a deletion in *YHR081w* (referred to hereafter as *LRP1*, for *L*ike *Rrp*6) displayed a prominent phenotype strongly resembling that of *rrp6*-Δ, whether assayed by microarray (Figures 3 and 6A) or Northern blotting (Figures 5B and 6B). *RRP6* encodes an exosome-associated protein required for complete formation of the E junction at the 3′ end of the 5.8S rRNA (Briggs et al., 1998). Rrp6p and Lrp1p are both nuclear proteins (Kumar et al., 2002), and Rrp6p associates with exosome components with near 1:1 stoichiometry (Fig-

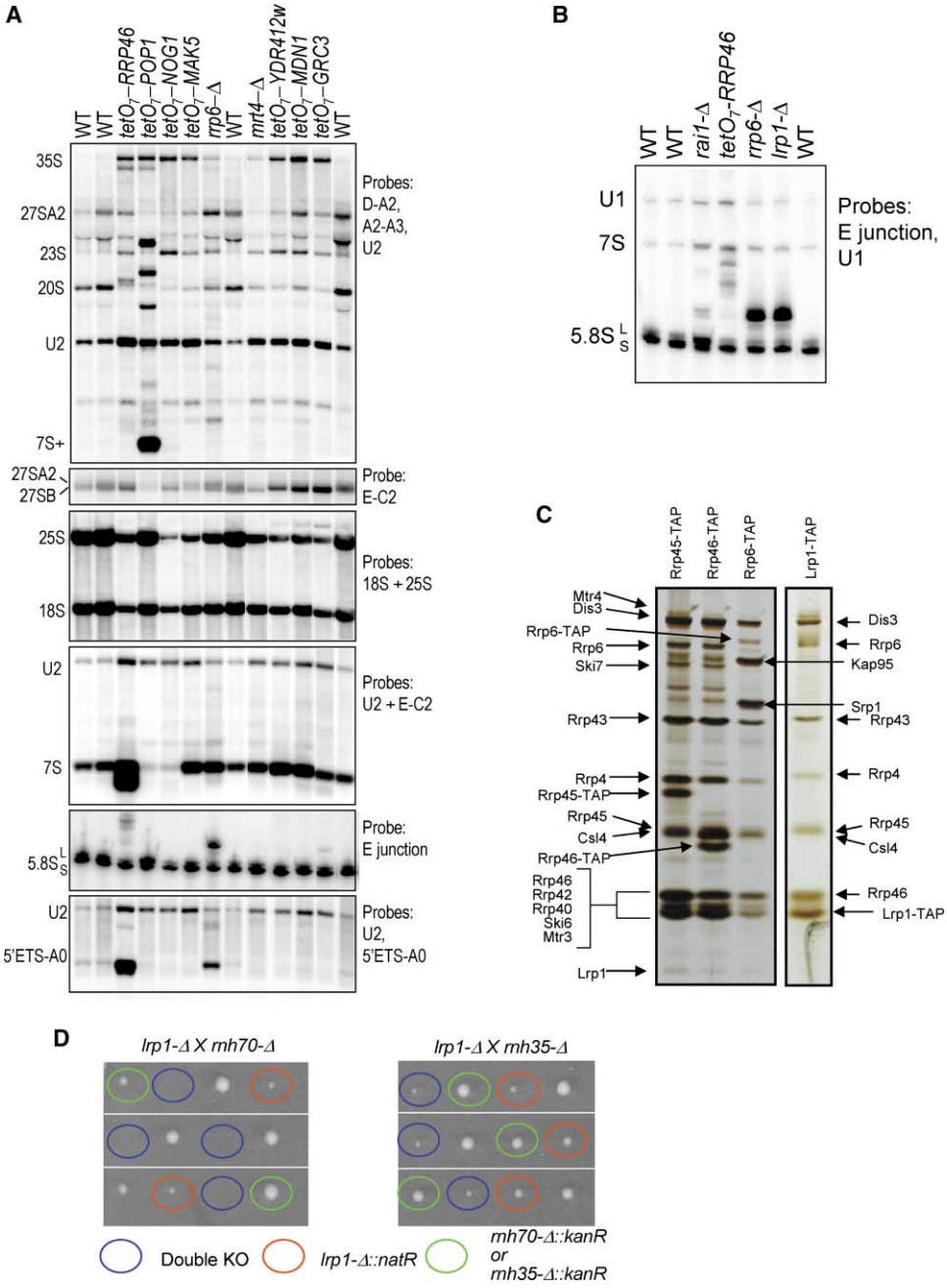

Figure 5. Northern Blot Analysis of Mutant Strains with 27S Pre-rRNA-Related Processing Defects

(A) Equal amounts of total RNA from each strain were resolved on 1% agarose/glyoxal (or 8% polyacrylamide/urea, for the E-junction probe), blotted to nylon, and probed sequentially with oligonucleotides as indicated. *tetO7-POP1* and *tetO7-NOG1* are shown for comparison.

(B) Strains with defects specifically in E junction processing, resolved on 8% polyacrylamide/urea, blotted to nylon, and probed as indicated.

(C) Association of Lrp1p with the exosome in affinity-purified complexes. Each band indicated was identified by MALDI MS.

(D) Representative tetrads showing genetic interaction of *LRP1* with *RNH70* (*REX1*) and *RNH35*.

ure 5C). Lrp1p was also found in TAP-purifications of exosome components Rrp45 and Rrp46 (Figure 5C; Gavin et al., 2002), although in a lower proportion than Rrp6p, suggesting that it is either weakly associated or that it is a component of only a subset of exosome particles. Affinity purification of TAP-tagged Lrp1p recovered components of the exosome with near 1:1 yield (Figure 5C), supporting the latter hypothesis and suggesting that most or all Lrp1p may be exosome-associated.

Both *RRP6* and *LRP1* are nonessential genes, but deletion of either gene causes a slow-growth phenotype (Briggs et al., 1998; Giaever et al., 2002; Erdemir et al., 2002; Figure 5D). Deletion of either gene was synthetically lethal with deletion of an RNase D-encoding gene, *rnh70-Δ* (Figure 5D; van Hoof et al., 2000; Rnh70p is

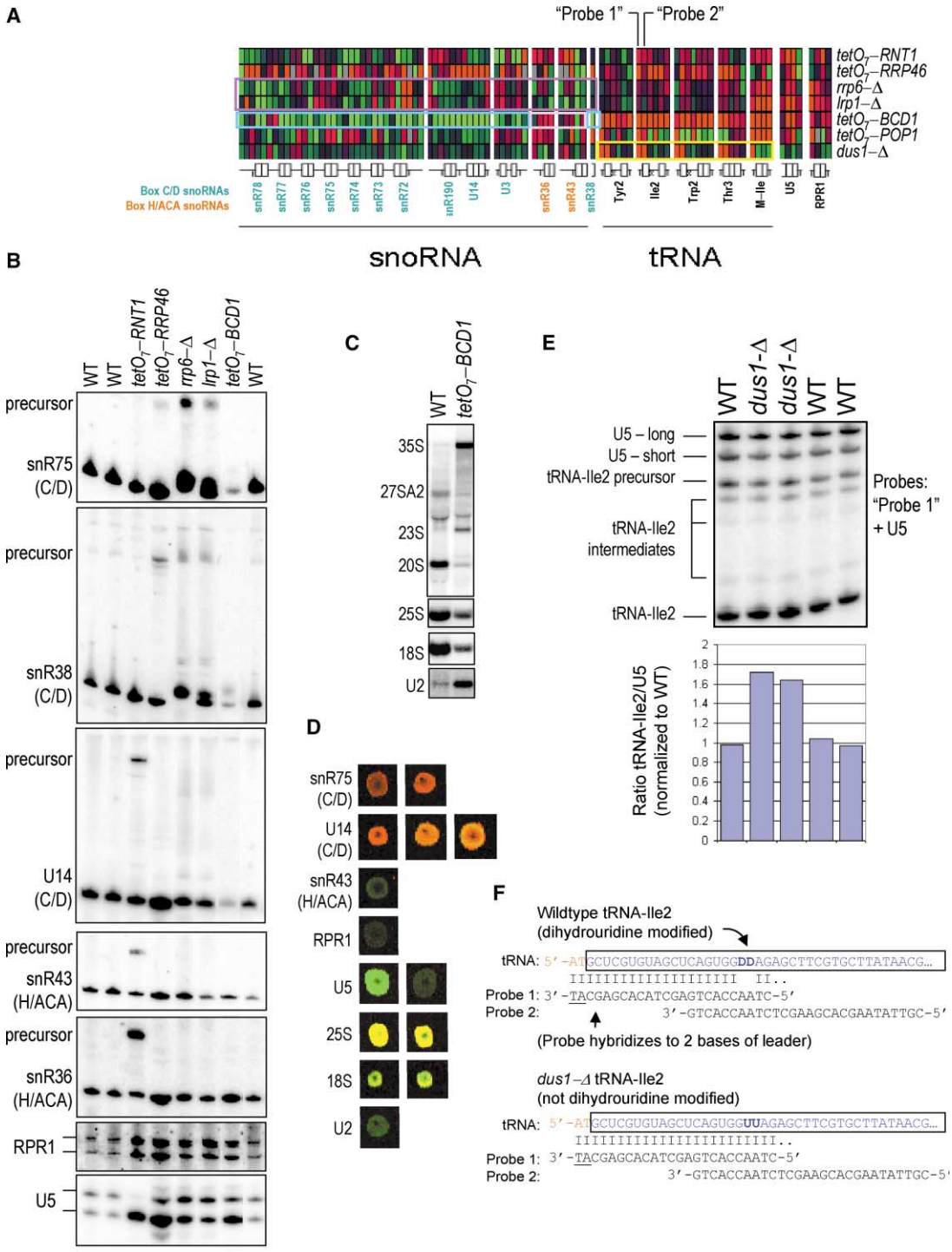

Figure 6. Alterations in Nonribosomal RNAs

(A) Microarray data from selected mutants affecting snoRNA and tRNA. Color scale is the same as in Figure 1. *RPR1* is the genetic name for the RNase P RNA subunit.

(B) Northern blotting analysis on 8% polyacrylamide/urea gels, confirming alterations in snoRNA.

(C) Northern blotting analysis of *tetO*$_7$*-BCD1* on a 1% agarose/glyoxal gel, confirming rRNA defects. Probes are top image: D-A2, A2-A3, and U2 simultaneously; bottom three images: 25S, 18S, U2 simultaneously.

(D) Microarray spots from an array that was hybridized with total yeast RNA in the green channel and RNA extracted from the *BCD1*-TAP purification in the red channel. The selected spots were taken from a single, unadjusted false-color image saved directly from the scanning software. Yellow indicates strong hybridization in both channels.

(E) Alterations in tRNA dihydrouridine modification assayed by hybridization. Top: equal amounts of total RNA from each strain were resolved on 8% polyacrylamide/urea, blotted to nylon, and probed with radiolabeled oligonucleotide "Probe 1" and a U5-specific oligonucleotide. Bottom: the ratio of the tRNA-Ile2 final product (bottom band) to that of U5 (top two bands, taken together) was determined by phosphorimager analysis, and normalized to the average of the three wild-type lanes.

(F) Schematic explanation of the *dus1*-Δ results observed in (A) and (E). Loss of dihydrouridine modification presumably results in higher binding affinity.

also known as Rex1p). *lrp1*-Δ also displayed a synthetic interaction with *rnh35*-Δ (Figure 5D). Lrp1p does not appear to be redundant with Rrp6p; there is no obvious sequence similarity between the two proteins, and no synthetic genetic interaction between them was obtained (data not shown). The Lrp1p-TAP purification contained Rrp6p (Figure 5C), suggesting that the two proteins associate with the exosome simultaneously. In addition, *lrp1*-Δ has a less pronounced effect than *rrp6*-Δ on processing of snoRNAs, snR38 and snR75 (Figure 6B).

A *YOR145c (PNO1) Mutant Bypasses A1 and A2 Cleavage*

Some of the mutants in uncharacterized genes displayed alterations in rRNA that could not be easily classified (Figure 3, bottom). *tetO*$_7$-*YOR145c*, for example, displayed an rRNA profile on the microarray that did not match any of the common processing defects. By Northern blotting, *tetO*$_7$-*YOR145c* showed complete or nearly complete loss of 20S and 27SA2 precursors, a defect in 18S accumulation, and what appears to be the presence of 22S (A0–A3) pre-rRNA (Figure 4A, right). In large-scale affinity-tagging studies of protein complexes, Yor145cp associates primarily with U3 components and 20S processing factors (e.g., Utp22p, Utp18p, Tsr1p, and Rio2p) (Gavin et al., 2002; Ho et al., 2002), consistent with a primary role in rRNA biogenesis. *YOR145c* was recently designated *PNO1*, for "*P*artner of *No*b1," a protein required for biogenesis of the proteasome (Tone and Toh-E, 2002). It is difficult to reconcile a proteasome association with our data and other published data (Gavin et al., 2002; Ho et al., 2002); we propose that the acronym be changed to "*P*rotein *N*eeded for *O*rdinary A1–A2 cleavage."

Alterations in Nonribosomal RNAs

BCD1, a Zinc Finger-Containing Protein Required for Box C/D snoRNA Accumulation

Mutations affecting nonribosomal RNAs were also identified (Figures 2A and 2B), although these data were often less straightforward to interpret than rRNA data. A striking exception is shown in Figure 6A. The *tet*-promoter allele of *YHR040w* (referred to hereafter as *BCD1*, for *B*ox *C*/*D* snoRNA accumulation) contained much lower amounts of Box C/D snoRNAs than wild-type (Figure 6A, boxed in blue), yet retained higher levels of Box H/ACA snoRNAs (Figure 6A, boxed in pink). This observation was confirmed by Northern blotting (Figure 6B). Consistent with a role in biogenesis or function of Box C/D snoRNAs (most of which are involved in ribose methylation of rRNA; see Weinstein and Steitz, 1999 for review), this strain was also compromised for rRNA biogenesis (Figure 6C). Furthermore, a *BCD1*-TAP purification was enriched for Box C/D snoRNAs, which could be detected by phenol-extracting RNA from the purification and hybridizing it to the microarray (Figure 6D), even when the amount of Bcd1-TAP protein and any associated proteins was too low for identification by mass spectrometry (Bcd1-TAP was visible by Western blotting against the TAP tag) (data not shown).

We cannot distinguish from these data whether Bcd1p is constitutively associated with Box C/D snoRNAs, or whether it is an assembly or localization factor analogous to Naf1p for Box H/ACA snoRNPs (Fatica et al., 2002b). However, the latter alternative is supported by the fact that Bcd1p was not detected in purified, enzymatically active Box C/D snoRNPs (Galardi et al., 2002). Bcd1p contains a predicted nuclear localization signal as well as a potential zinc finger domain, which together suggest a role in binding and transport of nucleic acid.

Detection of Covalent Modification of RNA Using a Microarray

Deletion of *DUS1*, which encodes a tRNA dihydrouridine synthetase (Xing et al., 2002), resulted in increased hybridization of RNA to array spots complementary to 5′-ends of pre-tRNA (Figure 6A, boxed in yellow). We reasoned that tRNA 5′-end processing by RNase P might be dependent on dihydrouridines in these tRNAs; however, no enrichment of any unprocessed precursor species in the *dus1*-Δ mutant was observed by Northern blotting (Figure 6E). Instead, we detected a relative increase in signal from the mature tRNA species (Figure 6E). This indicates that the increased signal on the pre-tRNA 5′-end microarray spots in *dus1*-Δ is not due to altered tRNA 5′-end processing, but rather to increased hybridization of the unmodified nucleotides in the *dus1*-Δ strain to the array spots in question, all of which are complementary to the dihydrouridine-containing part of the tRNA D-loop (which is near the 5′-end, and is so named because it contains *d*ihydrouridines) (Figure 6F). This is supported by the fact that an oligonucleotide that overlaps the dihydrouridines ("Probe 2" in Figures 6A and 6F) but not the 5′ leader also has an increase in signal (i.e., appears "red") on the array (Figure 6A). To our knowledge, this is the first demonstration that alteration of in vivo RNA base modifications can be detected using a microarray.

Discussion

Roughly one third of the ~6,200 yeast genes, including ~200 of the ~1,050 required for cell viability, have eluded detailed genetic or biochemical characterization (Issel-Tarver et al., 2002; Mewes et al., 2002; Giaever et al., 2002). Predictions based on functional genomics and proteomics efforts have suggested RNA processing as a functional category that may account for a substantial fraction of these uncharacterized genes (Wu et al., 2002; Bader and Hogue, 2002). Since these predictions are error-prone, validation studies with mutant strains and precise phenotypic assays are a key aspect of functional genomics and proteomics as well as bioinformatic efforts. For essential genes, creation of mutants is an initial hurdle. The *tetO*$_7$ system, like any transcriptional shutoff, has the disadvantage that phenotypes are manifested gradually, making it difficult to distinguish primary from secondary effects; however, it has the advantage that doxycycline at low concentrations is physiologically innocuous to yeast (Hughes et al., 2000). We are currently creating a complete set of *tetO*$_7$ shutoff alleles for all essential yeast genes (S.M. and T.R.H., unpublished data).

Often, assays that measure precise molecular defects are time-consuming, labor-intensive, and/or expensive. Microarrays are used widely for measuring nucleic acid abundance, particularly mRNA, and have previously

Table 1. Ten Genes Required for Normal Noncoding RNA Processing

ORF	Gene Name	Defect	Coregulation	Physical Association	Sequence Features	Mutant Growth (Giaever)
YGR251w	-	UTP-like	Yes	-	-	Essential
YGR272c	-	UTP-like	Yes	-	coiled-coil	Slow
YLR068w	FYV7	UTP-like	Yes	-	coiled-coil	Essential
YDL166c	FAP7	20S	Yes	KRR1 (Gavin), UTP13 (Ito)	- -	Essential
YLR435w	TSR2	20S	Yes	RPS26 (Ito, this study)	-	Slow
YDR412w	-	ITS2	Yes	-	coiled-coil	Essential
YLL035w	GRC3	ITS2, 5.8S	Yes	-	-	Essential
YHR081w	LRP1	5.8S	Yes	Exosome (Gavin, this study)	coiled-coil	Slow
YOR145c	PNO1	A1, A2 skipped	Yes	RIO2 (Gavin) UTP22 (Gavin) TSR1 (Gavin) UTP18 (Ho)	KH domain	Essential
YHR040w	BCD1	Box C/D snoRNAs	Yes	snR75, U14 (this study)	zinc finger, NLS	Essential

been used to monitor mRNA splicing (Shoemaker et al., 2001; Clark et al., 2002). Our work extends this technique, in that we have simultaneous queried the relative abundance of multiple fragments and junctions in noncoding RNAs. In addition, we show that covalent modifications of RNA can be detected by oligonucleotide microarrays (Figure 6A). Although we demonstrated this for dihydrouridine, the altered binding energy of other modified nucleotides is also known to affect hybridization efficiency (e.g., Nguyen et al., 2002). The recent commercial availability of high-density custom oligonucleotide arrays (Hughes et al., 2001; Nuwaysir et al., 2002) should facilitate creation of assays like ours on a more comprehensive scale, perhaps simultaneously allowing widespread detection of modified nucleotides in RNA.

Among the RNA-related factors we confirmed by Northern blotting, we believe ten are previously undescribed. Biochemical evidence for a direct role in RNA processing (i.e., unambiguous protein-protein or protein-RNA associations) exists for only five (Lrp1p, Tsr2p, Fap7p, Pno1p, and Bcd1) (Table 1). For the remainder (Ygr272cp, Fyv7p, Ygr251wp, Ydr412wp, and Grc3p) the only supporting evidence is that all five are coregulated with established rRNA processing factors at the transcriptional level (Table 1, Wu et al., 2002). We have also verified by Northern blotting the rRNA phenotypes of two genes whose products were previously associated with pre-60S particles (*MRT4* and *MDN1*).

Although we focused primarily on confirming defects in mutants with easily interpreted array phenotypes, those with unusual patterns may ultimately prove the most interesting. *GRC3*, for example, is of particular interest because it appears to be both genetically and biochemically distinct from any other known ITS2 processing factor. The array phenotype of our *tetO*$_7$-*YOR145c* (*PNO1*) strain did not resemble that of any of the other mutant strains apparently because it was the only strain in our collection that bypassed A1–A2 cleavage. Our present classification systems tend to identify mutants that are similar to positive control mutants. It may be worthwhile to consider selecting mutants for Northern blotting analysis on the simple basis that the array phenotype is different from that of all other mutant strains.

It is also noteworthy that some of our RNA processing mutants are already annotated as being involved in other cellular processes: Lrp1p, for example, has previously been described as a protein involved in nonhomologous DNA end-joining (Erdemir et al., 2002), and Fap7p has been described as a protein involved in oxidative stress response (Juhnke et al., 2000). One possibility is that these proteins may have dual/multiple functions. Alternatively, one phenotype may be a secondary consequence of the other.

To our knowledge, it is not widely appreciated that nearly one quarter of all yeast essential genes (253/$\sim$1,050) are already known to be involved in biogenesis or function of noncoding RNA and RNPs. In contrast, annotations for the entire process of cell division and DNA replication currently encompass only 207 genes that are required for viability (Mewes et al., 2002). The yeast research community has not yet analyzed many of the best candidate genes for noncoding RNA processing (e.g., those that are predicted on the basis of multiple data types) because they are essential for viability, and no conditional alleles are yet available. We have also not fully characterized many of the interesting mutants in the collection of 468 described here. Hence, the number of proteins known to be required for production of noncoding RNA and RNPs, particularly the ribosome, will almost certainly continue to increase, underscoring both the value of genome-scale research and the complexity of these processes.

Experimental Procedures

Array Construction

Oligonucleotide sequences are contained in the Supplemental Data. Oligonucleotides were diluted to a final concentration of 1 μg/μl in a solution of 50% DMSO, 0.1%SDS, and 8 copies of each were spotted onto poly-L-lysine slides with 16 pins using a robotic spotter (Virtek, Toronto, Canada) following procedures outlined in Hegde et al., 2000.

Strains

For a full description of strains, see Supplemental Data. Homozygous deletion mutants (Giaever et al., 2002) were acquired from Research Genetics. *tetO*$_7$-promoter alleles were constructed in strain R1158 (Hughes et al., 2000; a gift from C. Roberts) by replacing the 100 bases upstream of the start codon with a cassette (*kan*^{R}-*tetO*$_7$-TATA$_{CYC1}$) from plasmid RP188 via one-step homologous integration. Individual *tetO*$_7$-promoter mutants are available upon request; the full collection will shortly be available from Open Biosystems (Huntsville, AL).

RNA Extraction and Array Analysis

Isogenic wild-type and mutant strains were grown in parallel in SC medium at 30°C with vigorous shaking in baffled flasks (Bellco) to final cell concentrations matched as closely as possible to 10^7 cells/ml. *tetO*$_7$-promoter strains were exposed to 10 μg/ml doxycycline for a total of ~24 hr. Cells were harvested by 2 min centrifugation at 3000 RPM in a table-top centrifuge before freezing in liquid N_2. RNA was extracted by hot acidic phenol, ethanol precipitated, DNase I-treated, phenol extracted, and ethanol precipitated. 5 μg of resulting total RNA from wild-type and mutant were coupled with Alexa Fluor 546 or 647 according to protocol of the manufacturer (Molecular Probes "Ulysis" kit), ethanol precipitated, and hybridized to the array following procedures outlined in Hegde et al, 2000. All hybridizations were performed in duplicate, with fluors reversed on the second array. Arrays were scanned on an Axon 4000B instrument with GenePix software.

Image Processing and Array Normalization

All image processing, normalization, and subsequent analyses were implemented on Matlab (Mathworks). Initial log (ratio) for each oligonucleotide was determined by subtracting the ratio of each spot on the fluor-reversed array from the corresponding spot on the fluor-forward array, then removing outliers among the eight replicates and taking the mean. Normalization followed Yang et al. (2002) whereby a lowess smoother is applied to the ratios of each experiment over intensity.

Northern Blotting

For high-molecular-weight species, 5 μg of total RNA was separated on 1% agarose-glyoxal gels, and transferred to a Hybond N^+ membrane (Amersham) by capillary transfer. For low-molecular-weight species, 5 μg of total RNA was separated on either 8% or 10% polyacrylamide/TBE/urea gels, and electroblotted to a Hybond N^+ membrane (Amersham) using a semi-dry apparatus (Bio-Rad) with 0.5 × TBE as the transfer buffer. The membranes were UV cross-linked and hybridized in Church buffer using 5′-^{32}P-end-labeled oligonucleotide probes, with sequential probing as indicated in figures. Results were analyzed using a Phosphorimager (Bio-Rad Personal FX). Oligonucleotide probe sequences are found in the Supplemental Data.

Protein Purification and Mass Spectrometry

Complexes were purified on IgG and calmodulin columns from extracts of yeast cells (3 liters) grown in YPD medium to an OD_{600} of 1.0–1.5 using the TAP tagging procedure as previously described (Rigaut et al., 1999; Krogan et al., 2002).

Statistical Classification

Classification methods followed conventional statistical techniques (Bishop, 1995; Brown et al., 2000; Duda et al., 2001). Details are found in the Supplemental Data.

Data Availability

Oligonucleotide sequences on the array, probe sequences for Northern blots, all microarray data, a strain table, and information underlying Figures 1 and 2 are found at ftp://hugheslab.med.utoronto.ca/Pengetal.

Acknowledgments

We thank Naveed Mohammad, Richard Chang, Luan Chuen, Nina Enriquez, and Andrew Tsao for technical contributions; Carl Edward Rasmussen for MATLAB code for kernel density estimation; Ata Ghavidel for stimulating discussion of results; and Ben Blencowe, Charlie Boone, Jim Friesen, and Jim Ingles for critical evaluation of the manuscript. This work was supported by Genome Canada and a CIHR Operating grant to T.R.H. We apologize to colleagues whose work could not be cited due to space limitations and refer the reader to SGD (Issel-Tarver et al., 2002) and MIPS (Mewes et al., 2002), which curate the yeast literature on a gene-by-gene basis.

Received: March 18, 2003
Revised: June 6, 2003
Accepted: June 6, 2003
Published: June 26, 2003

References

Allmang, C., Kufel, J., Chanfreau, G., Mitchell, P., Petfalski, E., and Tollervey, D. (1999). Functions of the exosome in rRNA, snoRNA and snRNA synthesis. EMBO J. *18*, 5399–5410.

Andersen, J.S., Lyon, C.E., Fox, A.H., Leung, A.K., Lam, Y.W., Steen, H., Mann, M., and Lamond, A.I. (2002). Directed proteomic analysis of the human nucleolus. Curr. Biol. *12*, 1–11.

Bader, G.D., and Hogue, C.W. (2002). Analyzing yeast protein-protein interaction data obtained from different sources. Nat. Biotechnol. *20*, 991–997.

Bassler, J., Grandi, P., Gadal, O., Lessmann, T., Petfalski, E., Tollervey, D., Lechner, J., and Hurt, E. (2001). Identification of a 60S preribosomal particle that is closely linked to nuclear export. Mol. Cell *8*, 517–529.

Bishop, C.M. (1995). Neural Networks for Pattern Recognition (New York: Oxford University Press)

Briggs, M.W., Burkard, K.T., and Butler, J.S. (1998). Rrp6p, the yeast homologue of the human PM-Scl 100-kDa autoantigen, is essential for efficient 5.8 S rRNA 3′ end formation. J. Biol. Chem. *273*, 13255–13263.

Brown, M.P., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares, M., Jr., and Haussler, D. (2000). Knowledge-based analysis of microarray gene expression data by using support vector machines. Proc. Natl. Acad. Sci. USA *97*, 262–267.

Chanfreau, G., Legrain, P., and Jacquier, A. (1998). Yeast RNase III as a key processing enzyme in small nucleolar RNAs metabolism. J. Mol. Biol. *284*, 975–988.

Chapman, K.B., and Boeke, J.D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. Cell *65*, 483–492.

Clark, T.A., Sugnet, C.W., and Ares, M., Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science *296*, 907–910.

Dragon, F., Gallagher, J.E., Compagnone-Post, P.A., Mitchell, B.M., Porwancher, K.A., Wehner, K.A., Wormsley, S., Settlage, R.E., Shabanowitz, J., Osheim, Y., et al. (2002). A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. Nature *417*, 967–970.

Duda, R., Hart, R., and Stork, D. (2001). Pattern Classification, 2nd Edition. (New York: John Wiley & Sons).

Eddy, S.R. (2001). Non-coding RNA genes and the modern RNA world. Nat. Rev. Genet. *2*, 919–929.

Eddy, S.R. (2002). Computational genomics of noncoding RNA genes. Cell *109*, 137–140.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA *95*, 14863–14868.

Elela, S.A., Igel, H., and Ares, M., Jr. (1996). RNase III cleaves eukaryotic preribosomal RNA at a U3 snoRNP-dependent site. Cell *85*, 115–124.

Erdemir, T., Bilican, B., Cagatay, T., Goding, C.R., and Yavuzer, U. (2002). Saccharomyces cerevisiae C1D is implicated in both non-homologous DNA end joining and homologous recombination. Mol. Microbiol. *46*, 947–957.

Fatica, A., and Tollervey, D. (2002). Making ribosomes. Curr. Opin. Cell Biol. *14*, 313–318.

Fatica, A., Morlando, M., and Bozzoni, I. (2000). Yeast snoRNA accumulation relies on a cleavage-dependent/polyadenylation-independent 3′-processing apparatus. EMBO J. *19*, 6218–6229.

Fatica, A., Cronshaw, A.D., Dlakic, M., and Tollervey, D. (2002a). Ssf1p prevents premature processing of an early pre-60S ribosomal particle. Mol. Cell *9*, 341–351.

Fatica, A., Dlakic, M., and Tollervey, D. (2002b). Naf1p is a box H/ACA snoRNP assembly factor. RNA *8*, 1502–1514.

Galardi, S., Fatica, A., Bachi, A., Scaloni, A., Presutti, C., and Bozzoni, I. (2002). Purified box C/D snoRNPs are able to reproduce site-specific 2′-O-methylation of target RNA in vitro. Mol. Cell. Biol. *22*, 6663–6668.

Gari, E., Piedrafita, L., Aldea, M., and Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. Yeast *13*, 837–848.

Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature *415*, 141–147.

Gelperin, D., Horton, L., Beckman, J., Hensold, J., and Lemmon, S.K. (2001). Bms1p, a novel GTP-binding protein, and the related Tsr1p are required for distinct steps of 40S ribosome biogenesis in yeast. RNA *7*, 1268–1283.

Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the Saccharomyces cerevisiae genome. Nature *418*, 387–391.

Grandi, P., Rybin, V., Bassler, J., Petfalski, E., Strauss, D., Marzioch, M., Schafer, T., Kuster, B., Tschochner, H., Tollervey, D., et al. (2002). 90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors. Mol. Cell *10*, 105–115.

Gromadka, R., and Rytka, J. (2000). The KRR1 gene encodes a protein required for 18S rRNA synthesis and 40S ribosomal subunit assembly in Saccharomyces cerevisiae. Acta Biochim. Pol. *47*, 993–1005.

Harnpicharnchai, P., Jakovljevic, J., Horsey, E., Miles, T., Roman, J., Rout, M., Meagher, D., Imai, B., Guo, Y., Brame, C.J., et al. (2001). Composition and functional characterization of yeast 66S ribosome assembly intermediates. Mol. Cell *8*, 505–515.

Hegde, P., Qi, R., Abernathy, K., Gay, C., Dharap, S., Gaspard, R., Hughes, J.E., Snesrud, E., Lee, N., and Quackenbush, J. (2000). A concise guide to cDNA microarray analysis. Biotechniques *29*, 548–556.

Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. (2002). Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature *415*, 180–183.

Hopper, A.K., Traglia, H.M., and Dunst, R.W. (1990). The yeast RNA1 gene product necessary for RNA processing is located in the cytosol and apparently excluded from the nucleus. J. Cell Biol. *111*, 309–321.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., et al. (2000). Functional discovery via a compendium of expression profiles. Cell *102*, 109–126.

Hughes, T.R., Mao, M., Jones, A.R., Burchard, J., Marton, M.J., Shannon, K.W., Lefkowitz, S.M., Ziman, M., Schelter, J.M., Meyer, M.R., et al. (2001). Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. Nat. Biotechnol. *19*, 342–347.

Huttenhofer, A., Kiefmann, M., Meier-Ewert, S., O'Brien, J., Lehrach, H., Bachellerie, J.P., and Brosius, J. (2001). RNomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. EMBO J. *20*, 2943–2953.

Issel-Tarver, L., Christie, K.R., Dolinski, K., Andrada, R., Balakrishnan, R., Ball, C.A., Binkley, G., Dong, S., Dwight, S.S., Fisk, D.G., et al. (2002). Saccharomyces Genome Database. Meth. Enzymol. *50*, 329–346.

Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA *98*, 4569–4574.

Jeffares, D.C., Poole, A.M., and Penny, D. (1998). Relics from the RNA world. J. Mol. Evol. *46*, 18–36.

Juhnke, H., Charizanis, C., Latifi, F., Krems, B., and Entian, K.D. (2000). The essential protein Fap7 is involved in the oxidative stress response of Saccharomyces cerevisiae. Mol. Microbiol. *35*, 936–948.

Kapranov, P., Cawley, S.E., Drenkow, J., Bekiranov, S., Strausberg, R.L., Fodor, S.P., and Gingeras, T.R. (2002). Large-scale transcriptional activity in chromosomes 21 and 22. Science *296*, 916–919.

Kressler, D., Linder, P., and de La Cruz, J. (1999). Protein trans-acting factors involved in ribosome biogenesis in Saccharomyces cerevisiae. Mol. Cell. Biol. *19*, 7897–7912.

Krogan, N.J., Kim, M., Ahn, S.H., Zhong, G., Kobor, M.S., Cagney, G., Emili, A., Shilatifard, A., Buratowski, S., and Greenblatt, J.F. (2002). RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. Mol. Cell. Biol. *22*, 6979–6992.

Kufel, J., Dichtl, B., and Tollervey, D. (1999). Yeast Rnt1p is required for cleavage of the pre-ribosomal RNA in the 3' ETS but not the 5' ETS. RNA *5*, 909–917.

Kumar, A., Agarwal, S., Heyman, J.A., Matson, S., Heidtman, M., Piccirillo, S., Umansky, L., Drawid, A., Jansen, R., Liu, Y., et al. (2002). Subcellular localization of the yeast proteome. Genes Dev. *16*, 707–719.

Lee, W.C., Zabetakis, D., and Melese, T. (1992). NSR1 is required for pre-rRNA processing and for the proper maintenance of steady-state levels of ribosomal subunits. Mol. Cell. Biol. *12*, 3865–3871.

Mewes, H.W., Frishman, D., Guldener, U., Mannhaupt, G., Mayer, K., Mokrejs, M., Morgenstern, B., Munsterkotter, M., Rudd, S., and Weil, B. (2002). MIPS: a database for genomes and protein sequences. Nucleic Acids Res. *30*, 31–34.

Milkereit, P., Kuhn, H., Gas, N., and Tschochner, H. (2003). The pre-ribosomal network. Nucleic Acids Res. *31*, 799–804.

Mitchell, P., Petfalski, E., Shevchenko, A., Mann, M., and Tollervey, D. (1997). The exosome: a conserved eukaryotic RNA processing complex containing multiple 3'→5' exoribonucleases. Cell *91*, 457–466.

Morl, M., and Marchfelder, A. (2001). The final cut. The importance of tRNA 3′-processing. EMBO Rep. *2*, 17–20.

Nguyen, A., Zhao, C., Dorris, D., and Mazumder, A. (2002). Quantitative assessment of the use of modified nucleoside triphosphates in expression profiling: differential effects on signal intensities and impacts on expression ratios. BMC Biotechnol.*2*, 14.

Nissan, T.A., Bassler, J., Petfalski, E., Tollervey, D., and Hurt, E. (2002). 60S pre-ribosome formation viewed from assembly in the nucleolus until export to the cytoplasm. EMBO J. *21*, 5539–5547.

Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. (2002). Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res. *12*, 1749–1755.

Pederson, T. (1998). The plurifunctional nucleolus. Nucleic Acids Res. *26*, 3871–3876.

Pestov, D.G., Stockelman, M.G., Strezoska, Z., and Lau, L.F. (2001). ERB1, the yeast homolog of mammalian Bop1, is an essential gene required for maturation of the 25S and 5.8S ribosomal RNAs. Nucleic Acids Res. *29*, 3621–3630.

Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. Nat. Biotechnol. *17*, 1030–1032.

Robinson, M.D., Grigull, J., Mohammad, N., and Hughes, T.R. (2002). FunSpec: a web-based cluster interpreter for yeast. BMC Bioinformatics *3*, 35.

Sasaki, T., Toh-E, A., and Kikuchi, Y. (2000). Yeast Krr1p physically and functionally interacts with a novel essential Kri1p, and both

proteins are required for 40S ribosome biogenesis in the nucleolus. Mol. Cell. Biol. *20*, 7971–7979.

Saveanu, C., Bienvenu, D., Namane, A., Gleizes, P.E., Gas, N., Jacquier, A., and Fromont-Racine, M. (2001). Nog2p, a putative GTPase associated with pre-60S subunits and required for late 60S maturation steps. EMBO J. *20*, 6475–6484.

Sherman, F. (2002). Getting started with yeast. In Methods in Enzymology, Vol. 350, Guide to Yeast Genetics and Molecular Biology, C. Guthrie and G.R. Fink, eds. (New York: Academy Press).

Shoemaker, D.D., Schadt, E.E., Armour, C.D., He, Y.D., Garrett-Engele, P., McDonagh, P.D., Loerch, P.M., Leonardson, A., Lum, P.Y., Cavet, G., et al. (2001). Experimental annotation of the human genome using microarray technology. Nature *409*, 922–927.

Tone, Y., and Toh-E, A. (2002). Nob1p is required for biogenesis of the 26S proteasome and degraded upon its maturation in Saccharomyces cerevisiae. Genes Dev. *16*, 3142–3157.

Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. (2000). A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. Nature *403*, 623–627.

van Hoof, A., Lennertz, P., and Parker, R. (2000). Three conserved members of the RNase D family have unique and overlapping functions in the processing of 5S, 5.8S, U4, U5, RNase MRP and RNase P RNAs in yeast. EMBO J. *19*, 1357–1365.

Venema, J., and Tollervey, D. (1999). Ribosome synthesis in Saccharomyces cerevisiae. Annu. Rev. Genet. *33*, 261–311.

Verheggen, C., Mouaikel, J., Thiry, M., Blanchard, J.M., Tollervey, D., Bordonne, R., Lafontaine, D.L., and Bertrand, E. (2001). Box C/D small nucleolar RNA trafficking involves small nucleolar RNP proteins, nucleolar factors and a novel nuclear domain. EMBO J. *20*, 5480–5490.

Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. Curr. Opin. Cell Biol. *11*, 378–384.

Wu, L.F., Hughes, T.R., Davierwala, A.P., Robinson, M.D., Stoughton, R., and Altschuler, S.J. (2002). Large-scale prediction of Saccharomyces cerevisiae gene function using overlapping transcriptional clusters. Nat. Genet. *31*, 255–265.

Xing, F., Martzen, M.R., and Phizicky, E.M. (2002). A conserved family of Saccharomyces cerevisiae synthases effects dihydrouridine modification of tRNA. RNA *8*, 370–381.

Yang, Y.H., Dudoit, S., Luu, P., Lin, D.M., Peng, V., Ngai, J., and Speed, T.P. (2002). Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Res. *30*, e15.

Accession Numbers

The data reported in this paper have been deposited at http://www.ebi.ac.uk/arrayexpress/ under Accession number E-MEXP-12 and at http://www.ncbi.nlm.nih.gov/geo/ under Accession number GSE457.