Minireview

Probing the yeast proteome for RNA-processing factors

Sander Granneman* and Susan J Baserga*†‡

Addresses: *Departments of Molecular Biophysics and Biochemistry, †Genetics and ‡Therapeutic Radiology, Yale University School of Medicine, New Haven, CT 06520, USA.

Correspondence: Susan J Baserga. E-mail: susan.baserga@yale.edu

Abstract

A method has been developed to identify proteins required for the biogenesis of non-coding RNA in yeast, using a microarray to screen for aberrant patterns of RNA processing in mutant strains, and new proteins involved in the processing of ribosomal and non-coding RNAs have been found.

Functional genomics and proteomics

Yeast is widely used as a eukaryotic model system to study protein function because of its relative simplicity and the availability of powerful genetic tools. The completion of the genome sequence of the yeast *Saccharomyces cerevisiae* in 1996 [1] allowed researchers to analyze a eukaryotic organism on a genomic scale for the first time. This has greatly accelerated the development of technologies for performing large-scale proteomic and functional-genomic studies. Many of the initial studies in yeast were focused on the use of DNA microarray chips to measure expression profiles of large sets of genes in mutant strains or under varying growth conditions [2], but recent studies have mainly focused on large-scale proteomic experiments, including genome-wide two-hybrid protein-protein interaction screens [3-5], high-throughput affinity-purification of protein complexes [6,7], large-scale protein localization experiments [8] and even proteome chips [9]. Another recent study examined the growth phenotypes of yeast strains with gene deletions; approximately 96% of the annotated open reading frames (ORFs) were covered by this deletion collection [10]. Most recently, Peng *et al.* [11] have used a plethora of mutant yeast strains and microarray technology to screen for proteins involved in the synthesis and processing of ribosomal and other non-coding RNAs.

Synthesis and processing of rRNA and small non-coding RNAs

Strikingly, over 95% of the nucleic acid in yeast cells is non-coding RNA [12]. Most of these RNAs are ribosomal RNAs (mostly cytoplasmic rRNAs but including some mitochondrial rRNAs); indeed, a large portion of the cell's energy is devoted to the synthesis of ribosomes and rRNA, a process that requires hundreds of *trans*-acting factors [13].

Ribosome biogenesis takes place in a subnuclear cellular compartment, the nucleolus. Here, three of the four rRNAs are transcribed by RNA polymerase I as a single precursor or pre-rRNA. The nascent pre-rRNA is processed in a series of cleavage reactions to produce the mature 18S, 5.8S and 25S-28S rRNAs. Interestingly, processing of the nascent pre-rRNA in yeast has recently been shown to require the assembly of a pre-rRNA ribonucleoprotein (RNP) complex (the small subunit (SSU) processome, or 90S complex) that is about the size of a ribosome itself [14-16], underscoring the complexity of ribosome biogenesis. Using affinity-tag purification procedures several laboratories have isolated a number of other large pre-rRNA RNP complexes [6,7,14-20]. Also, an organelle-scale proteomic analysis of the human nucleolus has revealed the human homologs of many of these proteins as well as new ones [21]. In general, much remains to be discovered about the exact function of the proteins involved in ribosome biogenesis in the nucleolus. Moreover, the precise mechanism by which the endonucleolytic steps in pre-rRNA processing occur is not yet clear. It is not even known, in most cases, whether cleavage involves the activity of (as yet unidentified) endonucleases.

Apart from rRNAs the other non-coding RNAs comprise a long list of abundant, small RNAs, including small nucleolar

RNAs (snoRNAs), small nuclear RNAs (snRNAs), transfer RNAs (tRNAs), telomerase RNA, signal-recognition-particle RNAs and the RNA components of the RNase P and RNase MRP endonucleases. Most snoRNAs are involved in cotranscriptional chemical modification of pre-rRNA, particularly 2′-*O*-ribose methylation (in the case of 'box-C/D' snoRNAs) and base pseudouridylation (for 'box-H/ACA' snoRNAs; reviewed in [22]). The snRNAs are probably the catalysts for pre-mRNA splicing, and their association with each other and the pre-mRNA leads to the formation of the spliceosome [23]. As is the case for rRNAs, the mechanism by which many small non-coding RNAs are matured is not yet completely understood. Interestingly, it appears that several components of the machinery responsible for the cleavage and polyadenylation of mRNAs are also involved in the maturation of snRNAs and snoRNAs [24-26]. This is one of many examples of the way in which processing machineries are shared by different biogenesis pathways for non-coding-RNA.

Using microarrays to probe the yeast RNA-processing proteome

Comparative bioinformatic analyses [27,28] of protein-interaction data from several studies has revealed hundreds of uncharacterized protein-coding genes that are predicted to have a role in RNA processing and/or RNP biogenesis; many of these have not been detected or validated in large-scale proteomic studies. To test these predictions experimentally, Peng and colleagues [11] set out to measure defects in the biogenesis of non-coding RNA using oligonucleotide microarrays. The microarrays contained 212 different oligonucleotides that recognized unprocessed mRNAs and partially processed and mature products of a wide array of non-coding RNA species. These arrays were hybridized to steady-state RNA harvested from a set of strains, from each of which a protein was depleted or otherwise mutated. The mutant strains tested were chosen from the yeast deletion collection [10], from mutant strains previously collected by others, or constructed by the authors [11] using the $tetO_7$ system, which allows regulation of the protein of interest by tetracycline. The microarray showed which particular RNAs were depleted or overrepresented in each strain; strains with aberrant patterns were taken to have mutations in a gene involved in RNA biogenesis. To their credit, the authors sought to validate their microarray findings individually by northern blotting, greatly strengthening their conclusions.

The authors used a variety of sources to choose which candidate ORFs to test for in the mutant strains using their new methodology. A total of 413 ORFs (making up 7% of the yeast genome) had been previously characterized as having a role in non-coding-RNA biogenesis (Table 1). From comparative analyses of other genome-wide studies (such as [4-8,21]) the authors [11] then predicted an additional 919 ORFs to be involved in non-coding RNA biogen-

Table 1

Generating ORFs to test for their involvement in non-coding-RNA metabolism

	Total	Ratio	Percentage
Total ORFs annotated*	6,200†		
Known role in RNA metabolism	413	413/6,200	7%
Essential genes	1,050†	1,050/6,200	17%
Essential RNA-metabolism genes	253	253/6,200	4%
Total non-coding-RNA biogenesis	1,332		
Previously characterized ORFs	413	413/1,332	31%
Predicted ORFs [11]	919	919/1,332	69%
GO: biological process unknown	578	578/1,332	43%
GO: unrelated function	341	341/1,332	26%
Essential ORFs	525	525/1,332	39%

Overview of the available data used by Peng *et al.* [11] from the databases. *From the *Saccharomyces* Genome Database [33]; other numbers and percentages were generated from the supplemental data to [11]. †Approximate numbers. GO, gene ontology terms.

esis (to bring the total to 1,332 ORFs). Of the 919 additional ORFs implicated in non-coding RNA biogenesis, 578 were annotated in the databases as 'biological process unknown' and 341 were annotated with unrelated functions (see Table 1). A higher proportion than expected of the 413 previously characterized ORFs was encoded by essential genes (253/413 or 61%, and these represent nearly one quarter of all the essential genes in the whole genome; Table 1). Of the 1,332 ORFs implicated in non-coding-RNA biogenesis, 39% were encoded by essential genes (Table 1), again higher than a random sampling of the yeast genome would predict.

Of the pool of proteins implicated in non-coding RNA biogenesis, 468 were selected (of which 41% are essential) and the effects of their deletion or conditional depletion were analyzed by microarray (Table 2). These included 169 strains in which the proteins could be conditionally depleted (using the $tetO_7$ system; 36% of the tested proteins). From the microarray results, a computational classification technique was used to generate a score in the range of 1-5 for each protein; a score of 5 was considered 'positive' (that is, the protein functions in the processing of non-coding RNA). Surprisingly, using this classification system only 53% the proteins known to be involved in non-coding RNA processing, 74% of the proteins known to be involved in ribosome biogenesis and 36% of the proteins involved in snRNA/snoRNA/mRNA biogenesis were considered positive (Table 2). This is probably due to the fact that very stringent criteria were used to designate a positive; perusal of the supplementary data to the article [11] suggests that many with lower scores are indeed true posi-

Table 2

Numerical overview of the microarray results from Peng et al. [11]

	Total tested	Proportion of total tested (%)	Positives (score of 5)	Proportion positive in each class (%)
Tested proteins	468			
Negative controls	26	6%	0	0%
GO: non-coding RNA metabolism	133	32%	71	53%
GO: biological process unknown	141	24%	45	32%
GO: unrelated processes	168	49%	36	21%
GO: rRNA processing	61	13%	45	74%
GO: snRNA/snoRNA/ mRNA biogenesis	39	8%	14	36%
***tetO₇* conditional genes**	169	36%		
Essential genes	213	41%		

The numbers and percentages were generated from the supplemental data to Peng *et al.* [11]. GO, gene ontology terms.

tives. Investigation of the proteins not previously implicated in non-coding-RNA biogenesis revealed that 32% of the ORFs annotated as 'biological process unknown' were positive, as were 21% of the ORFs annotated as having unrelated functions (Table 2).

Uncovering new proteins required for RNA maturation and ribosome biogenesis

The results presented by Peng *et al.* [11] clearly prove the usefulness of their methodology in assigning function to proteins required for ribosome biogenesis. Unexpectedly, 20 ORFs annotated in the databases as 'biological process unknown' appeared to be involved in pre-rRNA processing but their mutant strains did not show a recognizable alteration in the pattern of RNA-processing defects on the microarray. Unfortunately, most of the processing defects for this subset of mutants were not investigated in more detail. As the authors have themselves stated [11], these proteins are very attractive candidates for further study.

Notably, many proteins that were annotated with functions in unrelated cellular processes appeared to (also) have a primary role in RNA biogenesis (21% of the 'unrelated' class; Table 2). One example is YOR145c, otherwise referred to as Pno1p. This protein had previously been shown to be required for biogenesis of the yeast proteasome [29], but both the microarray and the subsequent northern blot analysis of pre-RNA intermediates [11] strongly suggest a role in 18S rRNA synthesis. A second example is Lrp1p, which was previously

described to be involved in non-homologous DNA end-joining [30]. Peng *et al.* [11] have shown that it is required for correct processing of the 5.8S rRNA and that it is a component of the yeast exosome complex, a protein complex that is involved in 3′-end trimming of many RNA species and involved in mRNA degradation ([31] and references therein).

One of the problems encountered by the authors [11] was that alterations in the processing of low-abundance non-coding RNAs (such as many snoRNAs and snRNAs) were difficult to detect with their methods. Indeed, only about 36% of the proteins already known to be involved in the biogenesis of tRNA, snoRNA or snRNA were classified as positive in their screen (Table 2). The analyses did identify Bcd1p, a protein that is essential for stable accumulation of box-C/D-type snoRNAs, however. *In vivo* depletion of Bcd1p resulted in a dramatic reduction of box-C/D snoRNA steady-state levels, while box-H/ACA snoRNA levels appeared to be unaffected [11]. Thus, Bcd1p is likely to be involved in the biogenesis of box-C/D snoRNAs; it thus has a function similar to Naf1p, which is required for stable accumulation of box-H/ACA snoRNAs [32]. More detailed studies on Bcd1p will probably provide significant new insights into box-C/D snoRNA maturation.

Surprisingly, the methodology [11] was sufficiently sensitive to detect nucleotide modifications in pre-tRNAs. Deletion of the non-essential tRNA dihydrouridine synthetase Dus1p resulted in increased hybridization of oligonucleotides to the 5′ ends of tRNA, which was shown to be due to an increase in hybridization of the same amount of tRNA, rather than increased levels of the tRNA. This increased microarray hybridization signal correlated with a lack of covalent uridine modifications in a *dus1* deletion strain. This result represents the first time that covalent modifications have been detected in a microarray experiment.

The various genome-wide proteomic and functional genomic studies to date have provided a large amount of information that has allowed researchers to envisage connections between many protein and pathways. Peng *et al.* [11] have now developed some innovative tools to test predictions of protein function in non-coding-RNA biogenesis on a proteomic scale. There are now many new proteins to be analyzed and functions to be assigned.

References

1. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, *et al.*: **Life with 6000 genes.** *Science* 1996, **274:**546-567.

2.   Horak CE, Snyder M: **Global analysis of gene expression in yeast.** *Funct Integr Genomics* 2002, **2:**171-180.
3.   Uetz P, Hughes RE: **Systematic and large-scale two-hybrid screens.** *Curr Opin Microbiol* 2000, **3:**303-308.
4.   Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, *et al.*: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403:**623-627.
5.   Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98:**4569-4574.
6.   Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, *et al.*: **Functional organization of the yeast proteome by systematic analysis of protein complexes.** *Nature* 2002, **415:**141-147.
7.   Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, Adams SL, Millar A, Taylor P, Bennett K, Boutilier K, *et al.*: **Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry.** *Nature* 2002, **415:**180-183.
8.   Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, *et al.*: **Subcellular localization of the yeast proteome.** *Genes Dev* 2002, **16:**707-719.
9.   Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T, *et al.*: **Global analysis of protein activities using proteome chips.** *Science* 2001, **293:**2101-2105.
10. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Veronneau S, Dow S, Lucau-Danila A, Anderson K, Andre B, *et al.*: **Functional profiling of the *Saccharomyces cerevisiae* genome.** *Nature* 2002, **418:**387-391.
11. Peng WT, Robinson MD, Mnaimneh S, Krogan NJ, Cagney G, Morris Q, Davierwala AP, Grigull J, Yang X, Zhang W, *et al.*: **A panoramic view of yeast noncoding RNA processing.** *Cell* 2003, **113:**919-933.
12. Sherman F: **Getting started with yeast.** *Methods Enzymol* 2002, **350:**3-41.
13. Warner JR: **Nascent ribosomes.** *Cell* 2001, **107:**133-136.
14. Dragon F, Gallagher JE, Compagnone-Post PA, Mitchell BM, Porwancher KA, Wehner KA, Wormsley S, Settlage RE, Shabanowitz J, Osheim Y, *et al.*: **A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis.** *Nature* 2002, **417:**967-970.
15. Grandi P, Rybin V, Bassler J, Petfalski E, Strauss D, Marzioch M, Schafer T, Kuster B, Tschochner H, Tollervey D, *et al.*: **90S pre-ribosomes include the 35S pre-rRNA, the U3 snoRNP, and 40S subunit processing factors but predominantly lack 60S synthesis factors.** *Mol Cell* 2002, **10:**105-115.
16. Schafer T, Strauss D, Petfalski E, Tollervey D, Hurt E: **The path from nucleolar 90S to cytoplasmic 40S pre-ribosomes.** *EMBO J* 2003, **22:**1370-1380
17. Fatica A, and Tollervey D: **Making ribosomes.** *Curr Opin Cell Biol* 2002, **14:**313-318.
18. Fatica A, Cronshaw AD, Dlakic M, Tollervey D: **Ssf1p prevents premature processing of an early pre-60S ribosomal particle.** *Mol Cell* 2002, **9:**341-351.
19. Harnpicharnchai P, Jakovljevic J, Horsey E, Miles T, Roman J, Rout M, Meagher D, Imai B, Guo Y, Brame CJ, *et al.*: **Composition and functional characterization of yeast 66S ribosome assembly intermediates.** *Mol Cell* 2001, **8:**505-515.
20. Saveanu C, Namane A, Gleizes PE, Lebreton A, Rousselle JC, Noaillac-Depeyre J, Gas N, Jacquier A, Fromont-Racine M: **Sequential protein association with nascent 60S ribosomal particles.** *Mol Cell Biol* 2003, **23:**4449-4460.
21. Andersen JS, Lyon CE, Fox AH, Leung AK, Lam YW, Steen H, Mann M, Lamond AI: **Directed proteomic analysis of the human nucleolus.** *Curr Biol* 2002, **12:**1-11.
22. Kiss T: **Small nucleolar RNA-guided post-transcriptional modification of cellular RNAs.** *EMBO J* 2001, **20:**3617-3622.
23. Collins CA, Guthrie C: **The question remains: is the spliceosome a ribozyme?** *Nat Struct Biol* 2000, **7:**850-854.
24. Nedea E, He X, Kim M, Pootoolal J, Zhong G, Canadien V, Hughes T, Buratowski S, Moore CL, Greenblatt J: **Organization and function of APT, a sub-complex of the yeast cleavage and polyadenylation factor involved in the formation of mRNA and snoRNA 3' ends.** *J Biol Chem* 2003, **278:**33000-33010.
25. Dheur S, Vo le TA, Voisinet-Hakil F, Minet M, Schmitter JM, Lacroute F, Wyers F, Minvielle-Sebastia L: **Pti1p and Ref2p found in association with the mRNA 3' end formation complex direct snoRNA maturation.** *EMBO J* 2003, **22:**2831-2840.
26. Morlando M, Greco P, Dichtl B, Fatica A, Keller W, Bozzoni I: **Functional analysis of yeast snoRNA and snRNA 3'-end formation mediated by uncoupling of cleavage and polyadenylation.** *Mol Cell Biol* 2002, **22:**1379-1389.
27. Wu LF, Hughes TR, Davierwala AP, Robinson MD, Stoughton R, Altschuler SJ: **Large-scale prediction of *Saccharomyces cerevisiae* gene function using overlapping transcriptional clusters.** *Nat Genet* 2002, **31:**255-265.
28. Bader GD, Hogue CW: **Analyzing yeast protein-protein interaction data obtained from different sources.** *Nat Biotechnol* 2002, **20:**991-997.
29. Tone Y, Toh EA: **Nob1p is required for biogenesis of the 26S proteasome and degraded upon its maturation in *Saccharomyces cerevisiae*.** *Genes Dev* 2002, **16:**3142-3157.
30. Erdemir T, Bilican B, Cagatay T, Goding CR, Yavuzer U: ***Saccharomyces cerevisiae* C1D is implicated in both non-homologous DNA end joining and homologous recombination.** *Mol Microbiol* 2002, **46:**947-957.
31. Allmang C, Kufel J, Chanfreau G, Mitchell P, Petfalski E, Tollervey D: **Functions of the exosome in rRNA, snoRNA and snRNA synthesis.** *EMBO J* 1999, **18:**5399-5410.
32. Fatica A, Dlakic M, Tollervey D: **Nafl p is a box H/ACA snoRNP assembly factor.** *RNA* 2002, **8:**1502-1514.
33. ***Saccharomyces* Genome Database** [http://www.yeastgenome.org]