



## Small open reading frames: Not so small anymore

Richelle Sopko and Brenda Andrews

*Genome Res.* 2006 16: 314-315

Access the most recent version at doi:[10.1101/gr.4976706](https://doi.org/10.1101/gr.4976706)

---

**References** This article cites 19 articles, 8 of which can be accessed free at:  
<http://genome.cshlp.org/content/16/3/314.full.html#ref-list-1>

### License

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

Cold Spring Harbor Laboratory Press

# Small open reading frames: Not so small anymore

Richelle Sopko<sup>1</sup> and Brenda Andrews<sup>1,2,3</sup>

<sup>1</sup>Department of Medical Genetics and Microbiology, University of Toronto, Toronto, Ontario, Canada M5S 1A8; <sup>2</sup>Banting and Best Department of Medical Research, University of Toronto, Toronto, Ontario, Canada M5G 1L6

Today, nearly 10 years after the publication of the complete sequence of the *Saccharomyces cerevisiae* genome, the total number of genes in this organism is largely considered resolved and currently stands at 5782 (<http://www.yeastgenome.org/cache/genomeSnapshot.html>). This gene sequence information has led to construction of numerous yeast strain and plasmid collections, including the two-hybrid (Ito et al. 2000, 2001; Uetz et al. 2000) viable haploid deletion (Winzeler et al. 1999; Giaever et al. 2002), titratable promoter allele (Mnaimneh et al. 2004), and chromosomally tagged green fluorescent protein and TAP fusion libraries (Ghaemmaghami et al. 2003; Huh et al. 2003). Because creation of these collections has relied on the accuracy of gene annotation at the time of library construction, each library has limitations. One major issue for most is the fact that initial annotation of the *S. cerevisiae* genome included only those regions consisting of at least 100 contiguous codons, and therefore small open reading frames (sORFs) encoding functional proteins were largely missed (Goffeau et al. 1996) and only considered later following detection of expression (Olivas et al. 1997; Velculescu et al. 1997; Kumar et al. 2002; Oshiro et al. 2002; Kessler et al. 2003). As such, the phenotypic consequences of gene disruption of more than half of these sORFs has not been assessed, and sORFs are underrepresented in genomic libraries and other collections.

In this issue of *Genome Research*, Kastenmayer et al. (2006) provide the first systematic analysis of the prevalence of sORFs in a eukaryotic genome. Of the 299 currently recognized sORFs in the *S. cerevisiae* genome, they discovered that more than half (170) have been annotated since the genome was sequenced. Of these, 12% were identified by homology, and an even greater fraction (74%) was identified by combining both homology and empirical evidence such as detection of a transcribed and translated product. This is extremely significant given the variability in gene models that makes genome comparisons among species difficult and comparison between genera even more complex. The comparison of sORFs is additionally complicated given that a change in a single amino acid can create a large bias when assessing percent conservation. Also, prokaryotic genome sequence comparisons suggest that sORFs may be particularly susceptible to over-annotation (Nielsen and Krogh 2005). The significant apparent conservation of sORFs inspired Kastenmayer et al. to extend their analysis to include the entire list (299) of currently annotated sORFs. Using BLAST and analysis of the HomoloGene database, Kastenmayer et al. established that 184 *S. cerevisiae* sORFs have potential orthologs in other organisms and likely encode bona fide proteins with conserved functions. Like other large-scale functional genomic studies, this sORF survey affirms the importance of continued confirmation and refine-

ment of genome annotation (for recent review, see Dolinski and Botstein 2005).

To initiate characterization of those 170 sORFs discovered since the completion of the sequencing of the *S. cerevisiae* genome, Kastenmayer et al. (2006) constructed 140 deletion mutants, bringing the total number of sORF deletion strains to 247 (the rest were previously constructed by the deletion consortium, Winzeler et al. 1999). Examination of these newly constructed strains in an array of conditions uncovered growth phenotypes for 22 mutants, some of which were overlapping. For example, three new sORFs are essential, six mutants had a significant growth defect, and several exhibited sensitivities to a range of cell stressors, including DNA-damaging agents and suboptimal growth conditions. What other information about sORF function might we expect to extract from experiments with the new sORF deletion collection? The standard yeast deletion mutant collection enabled a paradigm shift in yeast genetics—the collection of deletion mutants has been systematically scrutinized for a variety of phenotypes, providing rosters of genes that may contribute to a particular biological process (Scherens and Goffeau 2004). For example, the cell morphology of each mutant strain in the deletion collection has been examined by microscopy, providing the first comprehensive list of morphogenetic defects for any organism (Giaever et al. 2002; Saito et al. 2005). Also, the deletion mutants were designed with molecular “barcodes”—unique DNA sequences that identify each mutant strain—enabling parallel phenotypic analysis in pools of deletion mutants. The fitness contribution of individual strains can then be quantitatively assessed using a barcode oligonucleotide microarray readout (Giaever et al. 2002). The sORF collection has been similarly bar-coded, and these mutant strains can now be included in all future functional profiling experiments.

Twenty-one of the 247 sORFs (8%) for which deletion mutants exist are essential for haploid viability (including three new sORFs uncovered from the Kastenmayer study). This is much lower than the incidence of essentiality in the entire genome (22%), suggesting (1) condition-dependent essentiality for these sORFs or (2) much genetic redundancy involving sORFs. Given that the sORF collection is arrayed and compatible with methods developed for the larger mutant collection, genetic redundancy can be tackled using methods of automated yeast genetics that have been recently invented (Tong et al. 2004). In particular, the so-called Synthetic Genetic Array or SGA method can be applied to systematically construct sORF double mutants, allowing assessment of genetic background in which sORFs are required for viability. A synthetic lethal interaction network has already been constructed by screening for genetic interactions involving well over 100 mutant strains and the ~5000 strains in the original deletion mutant collection (Tong et al. 2004). This project revealed that, on average, each gene will have about 26 synthetic lethal interactions. Using this average as a guide, we might expect sORFs to contribute 7774 (26 × 299) new genetic interactions to the *S. cerevisiae* genetic interaction network (Tong et al.

### <sup>3</sup>Corresponding author.

E-mail [Brenda.andrews@utoronto.ca](mailto:Brenda.andrews@utoronto.ca); fax (416) 946-8253.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.4976706>.

2004). Will sORFs have more or fewer than the average 26 genetic interactions? To date, only 28 sORFs have been reported to participate in a genetic interaction (*Saccharomyces* Genome Database; <http://www.yeastgenome.org/>). Even from the extensive genetic interaction survey of the haploid deletion collection, only 16 sORFs (of the 107 sORFs represented in this version of the collection) were found to genetically interact with any of 132 query genes used in the large genetic network project (Tong et al. 2004). Since the query mutants were biased toward genes with roles in actin-based cell polarity, cell wall biosynthesis, microtubule-based chromosome segregation, and DNA synthesis and repair, this may suggest alternative functions for most sORFs. Because the range in the number of genetic interactions identified for each of these 16 sORFs was quite large (between 1 and 18), we expect that only by screening individual sORF deletion mutants will we determine the exact number of genetic interactions per sORF.

Like genetic interactions, relatively few physical interactions with sORFs have been described. According to the *Saccharomyces* Genome Database, only 75 sORFs are reported to participate in a physical interaction with one or more other proteins. Intuitively we might expect smaller proteins possessing fewer protein domains to have fewer physical interactions. There exists no correlation, though, between the size of these sORFs and their number of reported physical interactions. Consistent with this, the smallest characterized ORFs in the *S. cerevisiae* genome encode 25 amino acid ribosomal proteins that spend their functional life in large protein complexes (*Saccharomyces* Genome Database; <http://www.yeastgenome.org/>). Future efforts to include sORFs in affinity purification and other proteomic experiments should help determine their full spectrum of physical interactions. These data will also help immeasurably in assignment of function to sORFs.

Might the characterization of sORFs be plagued by specific problems? We expect a similar spectrum of obstacles as encountered for larger ORFs, such as variability in RNA and protein isolation. However, at this stage, our limited examination of sORFs suggests that it is unlikely that this group of genes will exhibit unique group-specific properties with respect to their levels of expression or the abundance of their protein products. For example, a view of sORF protein abundance has come from quantification of natively expressed TAP-tagged sORFs (Ghaemmaghami et al. 2003). From this study it is clear that sORF expression levels vary substantially; of the ~200 C-terminally tagged sORFs, only ~100 were detectable, and of these, protein abundance varied from 99 to 1,590,000 molecules per cell, a range comparable to that seen for larger ORFs. As for other genes, functional analysis of each sORF will require its own specific investigation emphasizing the importance of creating the collection of unique sORF deletion mutants and of including these genes in various libraries for detailed study.

The prevalence of sORFs in eukaryotic genomes is only now becoming apparent. The high degree of conservation of sORFs and their prevalence in the *S. cerevisiae* genome (5% of all annotated genes) suggests that the small proteins that they encode are likely significant contributors to cell biology. Consistent with this are established roles for sORFs in a range of processes including transport, transcription, protein synthesis, DNA metabolism, and protein folding, among others. Interestingly, Kastenmayer et al. found that higher eukaryotes (including *Homo sapiens*) appear to have a similar proportion of sORFs to *S. cerevisiae*. That is, ~5% of annotated ORFs are smaller than 100 amino acids regardless of

the size of the genome. By providing a first glimpse of the prevalence of sORFs in the eukaryotic genome, Kastenmayer et al. (2006) have revealed the extent to which sORFs may have been overlooked in our quest to systematically define gene function. This class of genes may be small in size, but not in biological significance.

## References

- Dolinski, K. and Botstein, D. 2005. Changing perspectives in yeast research nearly a decade after the genome sequence. *Genome Res.* **15**: 1611–1619.
- Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., and Weissman, J.S. 2003. Global analysis of protein expression in yeast. *Nature* **425**: 737–741.
- 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.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274**: 546, 563–567.
- Huh, W.K., Falvo, J.V., Gerke, L.C., Carroll, A.S., Howson, R.W., Weissman, J.S., and O'Shea, E.K. 2003. Global analysis of protein localization in budding yeast. *Nature* **425**: 686–691.
- Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. 2000. Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad. Sci.* **97**: 1143–1147.
- 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.* **98**: 4569–4574.
- Kastenmayer, J.P., Ni, L., Chu, A., Kitchen, L.E., Au, W.-C., Yang, H., Carter, C.D., Wheeler, D., Davis, R., Boeke, J.D., et al. 2006. Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*. *Genome Res.* (this issue).
- Kessler, M.M., Zeng, Q., Hogan, S., Cook, R., Morales, A.J., and Cottarel, G. 2003. Systematic discovery of new genes in the *Saccharomyces cerevisiae* genome. *Genome Res.* **13**: 264–271.
- Kumar, A., Harrison, P.M., Cheung, K.H., Lan, N., Echols, N., Bertone, P., Miller, P., Gerstein, M.B., and Snyder, M. 2002. An integrated approach for finding overlooked genes in yeast. *Nat. Biotechnol.* **20**: 58–63.
- Mnaimneh, S., Davierwala, A.P., Haynes, J., Moffat, J., Peng, W.T., Zhang, W., Yang, X., Pootoolal, J., Chua, G., Lopez, A., et al. 2004. Exploration of essential gene functions via titratable promoter alleles. *Cell* **118**: 31–44.
- Nielsen, P. and Krogh, A. 2005. Large scale prokaryotic gene prediction and comparison to genome annotation. *Bioinformatics* **21**: 4322–4329.
- Olivas, W.M., Muhrad, D., and Parker, R. 1997. Analysis of the yeast genome: Identification of new non-coding and small ORF-containing RNAs. *Nucleic Acids Res.* **25**: 4619–4625.
- Oshiro, G., Wodicka, L.M., Washburn, M.P., Yates III, J.R., Lockhart, D.J., and Winzler, E.A. 2002. Parallel identification of new genes in *Saccharomyces cerevisiae*. *Genome Res.* **12**: 1210–1220.
- Saito, T.L., Sese, J., Nakatani, Y., Sano, F., Yukawa, M., Ohya, Y., and Morishita, S. 2005. Data mining tools for the *Saccharomyces cerevisiae* morphological database. *Nucleic Acids Res.* **33**: W753–W757.
- Scherens, B. and Goffeau, A. 2004. The uses of genome-wide yeast mutant collections. *Genome Biol.* **5**: 229.
- Tong, A.H., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. 2004. Global mapping of the yeast genetic interaction network. *Science* **303**: 808–813.
- 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.
- Velculescu, V.E., Zhang, L., Zhou, W., Vogelstein, J., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B., and Kinzler, K.W. 1997. Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- Winzler, E.A., Shoemaker, D.D., Astromoff, A., Liang, H., Anderson, K., Andre, B., Bangham, R., Benito, R., Boeke, J.D., Bussey, H., et al. 1999. Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**: 901–906.