

Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*

James P. Kastenmayer,^{1,6} Li Ni,^{2,6} Angela Chu,³ Lauren E. Kitchen,¹ Wei-Chun Au,¹ Hui Yang,² Carole D. Carter,¹ David Wheeler,⁴ Ronald W. Davis,³ Jef D. Boeke,⁵ Michael A. Snyder,² and Munira A. Basrai^{1,7}

¹Genetics Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20889, USA; ²Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA; ³Department of Biochemistry, Stanford University School of Medicine, Stanford, California 94305, USA; ⁴National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA; ⁵Department of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA

Genes with small open reading frames (sORFs; <100 amino acids) represent an untapped source of important biology. sORFs largely escaped analysis because they were difficult to predict computationally and less likely to be targeted by genetic screens. Thus, the substantial number of sORFs and their potential importance have only recently become clear. To investigate sORF function, we undertook the first functional studies of sORFs in any system, using the model eukaryote *Saccharomyces cerevisiae*. Based on independent experimental approaches and computational analyses, evidence exists for 299 sORFs in the *S. cerevisiae* genome, representing ~5% of the annotated ORFs. We determined that a similar percentage of sORFs are annotated in other eukaryotes, including humans, and 184 of the *S. cerevisiae* sORFs exhibit similarity with ORFs in other organisms. To investigate sORF function, we constructed a collection of gene-deletion mutants of 140 newly identified sORFs, each of which contains a strain-specific “molecular barcode,” bringing the total number of sORF deletion strains to 247. Phenotypic analyses of the new gene-deletion strains identified 22 sORFs required for haploid growth, growth at high temperature, growth in the presence of a nonfermentable carbon source, or growth in the presence of DNA damage and replication-arrest agents. We provide a collection of sORF deletion strains that can be integrated into the existing deletion collection as a resource for the yeast community for elucidating gene function. Moreover, our analyses of the *S. cerevisiae* sORFs establish that sORFs are conserved across eukaryotes and have important biological functions.

[Supplemental material is available online at www.genome.org.]

The initial *Saccharomyces cerevisiae* genome sequencing effort annotated all ORFs of at least 100 contiguous codons (including the first ATG) not contained entirely within a longer ORF (Goffeau et al. 1996). Knowledge of sORF (small open reading frame; <100 amino acids) function is limited compared to that of larger genes, although small proteins include members of important classes such as mating pheromones, proteins involved in energy metabolism, proteolipids, chaperonins, stress proteins, transporters, transcriptional regulators, nucleases, ribosomal proteins, thioredoxins, and metal ion chelators (for review, see Basrai et al. 1997). Computational discovery of sORFs is difficult because they are “buried” in an enormous pile of meaningless short ORFs that arise by chance. In addition, sORFs are not favorable targets for random mutagenesis. Similar challenges plague attempts to identify non-coding RNAs (ncRNAs), transcripts that function at the level of RNA rather than as templates for translation (for review, see Eddy 2001). Despite the challenges of sORF identification, reports since the publication of the *S. cerevisiae* genome indicate

that sORFs are quite numerous in *S. cerevisiae* and many are evolutionarily conserved from distantly related fungi to humans.

Many *S. cerevisiae* sORFs were discovered through expression-based analyses. Velculescu and colleagues used serial analysis of gene expression (SAGE) to identify, quantify, and compare global gene expression patterns in *S. cerevisiae* (Velculescu et al. 1995, 1997; Basrai and Hieter 2002). The SAGE technique is based on two principles: (1) a 9–10-bp sequence tag derived from a defined region in any poly(A)⁺ transcript that uniquely identifies the transcript; and (2) multiple sequence tags that are concatenated and sequenced in a single sequencing lane. In addition to confirming expression of annotated genes, the SAGE study provided the first evidence that hundreds of non-annotated reading frames (NORFs), including many sORFs, are transcribed in *S. cerevisiae*. We subsequently characterized one of these sORFs, *NORF5/HUG1*, and determined that it is a downstream target of the *MEC1*-mediated pathway for DNA damage and replication arrest (Basrai et al. 1999). These results validated the functional significance of sORFs found through systems biology approaches and suggested that other sORFs may have important functions.

Since the SAGE study, additional studies provided expression-based evidence for sORFs. Transcripts for potential sORFs or ncRNAs from intergenic regions were detected by Northern blotting (Olivas et al. 1997). A combined microarray and proteomics

⁶These authors contributed equally to this work.

⁷Corresponding author.

E-mail basraim@nih.gov; fax (301) 480-0380.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.4355406>.

approach confirmed transcription of many sORFs discovered by SAGE and detected peptides corresponding to numerous sORFs, including some not reported by SAGE (Oshiro et al. 2002). Additional sORFs were discovered using a gene-trap strategy based on genomic integration of a modified bacterial transposon, and their expression was confirmed by strand-specific oligonucleotide dot-blot arrays (Kumar et al. 2002). Interestingly, some of the sORFs discovered by gene-trap are antisense to coding genes (Kumar et al. 2002).

Potential sORF homologs were identified for many of the sORFs discovered in the expression-based studies, and recent comparative genomic studies have expanded the number of sORFs with potential orthologs. Conserved sORFs were reported from comparisons of the *S. cerevisiae* genome to partial genome sequences from 13 hemiascomycetes and the complete genome sequences from distantly and closely related fungi (Blandin et al. 2000; Brachat et al. 2003; Cliften et al. 2003; Kellis et al. 2003). A recent study that combined homology searching with RT-PCR identified conserved sORFs whose expression was detected at the level of RNA (Kessler et al. 2003).

Based on the published literature, at least 299 genes in *S. cerevisiae* likely encode sORFs. We discovered that a similar percentage of sORFs are annotated in multiple eukaryotes and that many of the *S. cerevisiae* sORFs have potential orthologs in other eukaryotes. We constructed gene-deletion strains for 140 sORFs, bringing the total number of sORF deletion strains to 247. We analyzed these 140 new sORF deletion strains for growth phenotypes and identified sORFs that are essential for haploid growth and for growth at high temperature. We also identified sORFs required for growth under genotoxic conditions including exposure to hydroxyurea (HU), bleomycin, methyl methane sulfonate (MMS), or ultraviolet (UV) radiation. These data highlight the value of expression analyses and comparative genomics to identify sORFs and the advantages of *S. cerevisiae* genetics in investigating sORF function.

Results and Discussion

Evidence of *S. cerevisiae* sORFs

The *S. cerevisiae* genome has 299 annotated sORFs (*Saccharomyces* Genome Database; <http://www.yeastgenome.org/>) (Fig. 1A; Supplemental Table A). By comparing the sORFs reported since the publication of the *S. cerevisiae* genome, we determined that the majority of sORFs (170) were discovered in the gene expression and homology studies mentioned above, while the remainder were previously reported in the literature (Fig. 1A, "129 previously known"). We analyzed the literature for reports of transcription, translation, or homology for the 170 new sORFs. Those that were reported by SAGE (Velculescu et al. 1997), microarrays (Kumar et al. 2002; Oshiro et al. 2002), RT-PCR (Kessler et al. 2003), Northern blot (Olivas et al. 1997), or gene-trap (Kumar et al. 2002) were considered transcribed. The sORFs detected by gene-trapping were considered transcribed and translated because the β -galactosidase assays used to detect integration require transcription and translation. The mass-spectrometry study also identified sORFs with evidence of translation (Oshiro et al. 2002). Finally, sORFs reported in homology searches were classified as supported by homology (Velculescu et al. 1997; Blandin et al. 2000; Kumar et al. 2002; Oshiro et al. 2002; Brachat et al. 2003; Cliften et al. 2003; Kessler et al. 2003).

Many of the new sORFs were detected by more than one

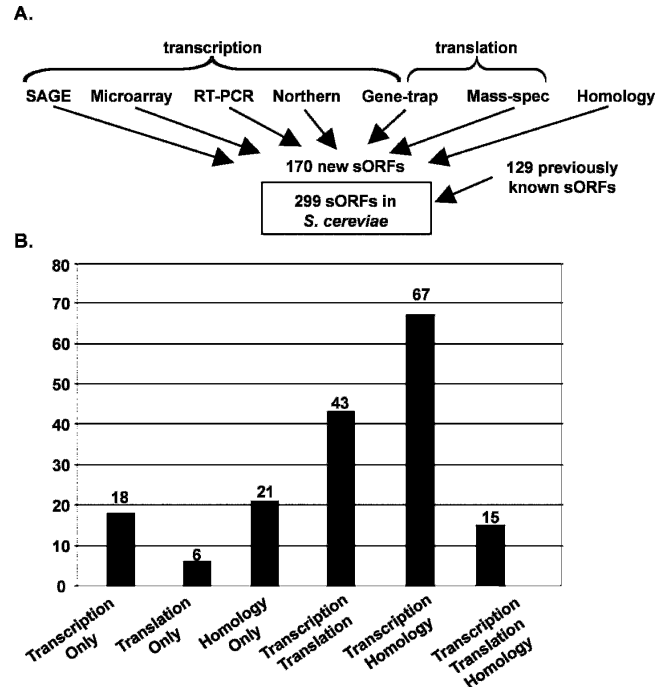


Figure 1. Evidence of *S. cerevisiae* sORFs. (A) Gene expression-based analyses and homology searching reveal 170 potential sORFs, bringing the total number of annotated sORFs in *S. cerevisiae* to 299. Reports in the literature provided empirical evidence of transcription derived from SAGE, microarray, RT-PCR, Northern blot, and gene-trap experiments, while empirical evidence of translation was derived from reports of mass-spectrometry and gene-trap experiments. Comparative genomic studies provided evidence of homology. (B) The bar graph depicts the subcategorization of the evidence for the 170 new sORFs, showing that the largest number of sORFs were identified based on evidence of transcription followed by evidence based on translation and homology.

approach (Fig. 1B; Supplemental Table A). For example, a large number of sORFs were discovered as both transcribed and translated (43 sORFs) or transcribed and with potential orthologs (67 sORFs), while several (15 sORFs) show evidence of transcription, translation, and homology. sORFs discovered only by transcription-based assays (18 sORFs) may represent ncRNAs, rather than protein-coding genes. sORFs detected at the level of RNA and homology may also be ncRNAs rather than protein-coding genes if the homology is the result of conservation of an RNA rather than protein-coding sequence. The sORFs discovered only by homology (21 sORFs) may represent genes expressed under certain conditions not used in the gene expression studies or could represent conserved sequences such as regulatory elements that are not expressed (Cliften et al. 2003). Most of the sORFs were detected by two or more techniques and likely represent bona fide genes.

Small proteins constitute a significant percentage of annotated proteins in eukaryotes

The 299 sORFs constitute ~5% of the 5865 genes annotated for *S. cerevisiae* in the NCBI RefSeq database (<http://www.ncbi.nlm.nih.gov/RefSeq/>) (Fig. 2; Pruitt et al. 2005). We determined the percentage of annotated small proteins from additional eukaryotes in the NCBI RefSeq database (see Methods). We selected representative eukaryotes including another fungus (*Schizosaccharomyces pombe*), worms (*Caenorhabditis elegans*), plants (*Arabi-*

dopsis thaliana), insects (*Drosophila melanogaster*), and mammals (*Mus musculus* and *Homo sapiens*). Interestingly, a similar percentage of sORFs are annotated for these organisms (~5%), including multicellular eukaryotes that have much larger genomes and a greater number of ORFs (Fig. 2). These results suggest that sORFs are not favored in single-celled eukaryotes or in those with smaller genomes and fewer genes. However, the evidence for the sORFs of *S. cerevisiae* comes from multiple analyses that may not have been used for all the representative eukaryotes (Fig. 2), and future experiments may reveal additional sORFs in these and other systems. Nevertheless, sORFs represent hundreds and in some cases >1000 ORFs in eukaryotes, and likely contribute significantly to the biology of eukaryotes.

sORFs are evolutionarily conserved

Many of the new sORFs were discovered based on homology (103 of 170 sORFs) (Fig. 1B), indicating that sORFs likely have fundamental functions across eukaryotes. However, the databases used to search for sORF orthologs differed between reports (e.g., Kessler et al. 2003 used the NCBI fungi sequences, while Oshiro et al. 2002 used the nonredundant sequences from all species), and a search for orthologs of the complete set of 299 sORFs had not been reported. We conducted two searches using the entire set of 299 sORFs. First, we conducted BLAST analyses to examine the conservation of the sORFs in the representative eukaryotes (Fig. 2). Second, to examine sORF conservation more broadly, we examined the data on the sORFs in the HomoloGene database, which was built with genome sequences from a wide variety of eukaryotes (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?DB=homologene>) (Wheeler et al. 2005).

For our BLAST analyses, we compared the sORFs to the annotated proteins from the representative eukaryotes (Fig. 2) and to a database derived from genomic and EST sequences of these organisms, the UniGene database (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene>) (Pontius et al. 2003; see Supplemental material). We discovered that 46 of the sORFs exhibit significant alignments with annotated proteins from two or more of the representative eukaryotes, with BLAST bit scores ranging from 48 to 147 (Table 1). In the representative eukaryotes, ~60% of the proteins that are similar to the *S. cerevisiae* sORFs are less than ~100 amino acids (data not shown). We also discovered 44 sORFs that align with transcripts from two or more representative eukaryotes in the UniGene database (Supplemental Table B; see Supplemental material).

Our analysis of HomoloGene revealed additional conserved sORFs. HomoloGene is a system that automatically detects ho-

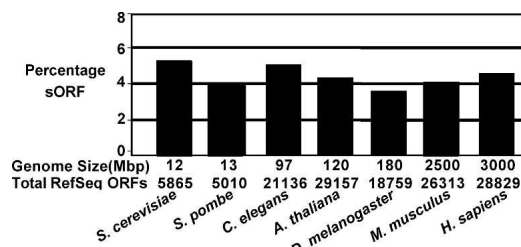


Figure 2. sORFs constitute a similar percentage of annotated ORFs in representative eukaryotes. The percentage of sORFs for *S. cerevisiae* and representative eukaryotes was calculated and is depicted in the bar graph. The genome size (megabases) and the number of RefSeq ORFs for each eukaryote are displayed below the graph.

Table 1. BLAST bit scores of *S. cerevisiae* sORFs that exhibit significant alignments with annotated proteins from other eukaryotes

sORFs	<i>S.p.</i>	<i>C.e.</i>	<i>A.t.</i>	<i>D.m.</i>	<i>M.m.</i>	<i>H.s.</i>
YLR185W	147	130	137	139	127	127
YHR021C	132	112	121	126	127	126
YKL156W	132	112	121	126	126	126
YDR500C	141	124	132	135	123	123
YOR210W	114	108	106	107	0	107
YJR094W-A	110	106	105	107	105	105
YPR043W	110	106	105	107	105	105
YDR424C	99.4	103	99	102	101	102
YNR032C-A	100	93.2	99.8	93.2	101	101
YOR167C	113	88.6	83.6	90.5	86.7	98.6
YLR264W	114	88.6	84.3	90.5	85.9	98.6
YKRO57W	136	98.6	88.6	90.5	97.1	98.2
YJL136C	136	98.2	88.2	90.5	96.7	97.8
YDR139C	91.7	93.6	91.3	94	95.1	95.5
YNL024C-A	0	0	76.6	85.9	89.7	90.1
YLR038C	105	72.4	88.6	105	86.7	88.2
YOR159C	100	89.7	91.7	87.8	88.2	87
YGR037C	109	103	85.5	102	71.6	86.7
YIL008W	89.4	0	0	78.2	84.3	82.8
YER146W	80.1	76.3	69.3	74.7	74.3	81.6
YOR298C-A	97.1	77.8	65.1	74.7	80.1	80.1
YER048W-A	0	0	0	62.8	72	79
YBL071W-A	90.1	78.6	80.5	87.4	77.4	77.8
YFR032C-A	67	64.3	67.4	56.6	74.3	74.3
YBR089C-A	102	74.3	92.8	74.3	73.2	73.2
YPR052C	94.7	72.8	89.7	76.3	70.9	70.9
YEL020W-A	92	63.5	82.4	68.2	63.5	64.7
YLR325C	69.3	63.9	54.7	60.8	62	62.8
YJR135W-A	69.7	0	49.3	56.6	61.2	60.1
YKLO53C-A	69.7	0	56.2	57	56.2	57.4
YHR143W-A	60.1	50.8	54.7	54.3	55.5	55.5
YCL057C-A	73.9	47.8	0	50.1	51.2	53.1
YLR287C-A	73.6	53.5	51.2	50.1	50.8	50.8
YOR182C	73.6	53.5	51.2	50.1	50.8	50.8
YLR388W	93.2	92.8	91.7	91.7	87.4	48.9
YFL017W-A	76.6	74.7	78.2	81.3	77.4	48.9
YBL026W	122	113	120	121	122	0
YHR072W-A	88.6	70.9	74.7	85.5	72.8	0
YHR005C-A	88.2	76.3	71.2	73.6	69.7	0
YLL018C-A	82.8	70.5	63.2	0	68.6	0
YGR183C	0	0	0	50.8	48.5	0
YLR438C-A	75.5	77.8	75.5	76.3	0	0
YNL259C	72.4	59.7	48.1	58.5	0	0

The BLAST bit scores obtained from a comparison of the sORFs to annotated proteins present in NCBI RefSeq are shown. A higher number indicates a more significant match. The BLAST bit score is a normalized version of the raw BLAST alignment score that is given in units of bits. Bit scores are independent of the scoring system used, so that, in addition to the score itself, only the size of the search space is needed to calculate their statistical significance. (*S.p.*) *Schizosaccharomyces pombe*; (*C.e.*) *Caenorhabditis elegans*; (*A.t.*) *Arabidopsis thaliana*; (*D.m.*) *Drosophila melanogaster*; (*M.m.*) *Mus musculus*; (*H.s.*) *Homo sapiens*.

mologs among the annotated genes of several completely sequenced eukaryotic genomes including *H. sapiens* and *M. musculus* (Supplemental material). Seventy-one sORFs were found in HomoloGene clusters conserved at several taxonomic levels, and 55 of the clusters have an assignment from the Conserved Domain Database (Marchler-Bauer et al. 2005), a collection of multiple sequence alignments for ancient domains and full-length proteins (Supplemental Table B). The conserved domains cover a broad spectrum; however, a conserved domain derived from SMART domain 00651, annotated as “small nuclear ribonucleoprotein particles (snRNPs) involved in pre-mRNA splicing,” was represented most frequently, occurring five times. Zinc-finger,

ubiquitin-like, and ribosomal protein domains were also encountered multiple times.

In summary, our results, combined with previously published reports, establish that 184 of the *S. cerevisiae* sORFs may have potential orthologs in other organisms (Supplemental Table B), including distantly related organisms, such as humans, and ~60% of these orthologs may themselves be sORFs (data not shown). Therefore, functional analysis of the *S. cerevisiae* sORFs has the potential to yield insight into the functions of the *S. cerevisiae* sORFs and those of other eukaryotes.

Generation of sORF deletion strains

Gene-deletion strain collections of *S. cerevisiae* have revolutionized functional analyses of genes (e.g., Winzeler et al. 1999). Since only 106 of the 299 sORFs are represented in the previous collection (version 1.0), we attempted to construct gene-deletion strains of the remaining sORFs in the same genetic background as described for the initial yeast gene-deletion strains (Winzeler et al. 1999; Supplemental Fig. 1) (see Methods).

Using homologous recombination, we constructed individual strains in which sequences from the start codon to the stop codon of the sORF were replaced by a *kanMX* cassette in a diploid strain (Methods) (Supplemental Fig. 1). Each sORF gene-deletion mutant is publicly available either as haploids (*MAT α* and *MAT α*) or as diploids (homozygous or heterozygous; see http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html). The gene-deletion strains contain “molecular barcodes” that will facilitate rapid identification and analysis of genes in genome-wide approaches to analyze gene function (Winzeler et al. 1999; for review, see Pan et al. 2004). We determined that the molecular barcodes corresponding to the sORF deletions are detectable in microarray experiments using the Tag3 arrays (Affymetrix) (data not shown). In total, we constructed 140 sORF heterozygous deletion strains (~93% of the 151 attempted), bringing the total number of sORF deletion strains available to the yeast community to 247 (Supplemental Table A). The remaining sORF deletion strains were not constructed because of technical problems including the inability to design gene-specific primers or to recover transformants.

Identification of essential sORFs

Sporulation of eight of the 140 new sORF heterozygous deletion strains resulted in two viable (gentamicin-sensitive) and two inviable (sORF deletion) spores, indicating that the corresponding sORFs are essential for viability (Table 2). Three of these genes were previously uncharacterized—*YLR099W-A*, *YNL024C-A*, and *YNL138W-A*—while the remaining five sORFs were previously shown to be essential, which we confirmed in the gene-deletion strain background. These sORFs are required for functions such as kinetochore or spindle integrity (Cheeseman et al. 2002; Li et al. 2005), ER to golgi transport (Heidtmann et al. 2005), and pseudouridine biosynthesis (Henras et al. 1998). Combined with the results from Version 1.0, 21 of the sORFs are essential, representing ~8% of sORFs analyzed. The percentage of essential sORFs differs from that of larger ORFs in which ~20% of ORFs tested are essential (Winzeler et al. 1999). This difference may reflect redundancy in sORF function, or indicate that sORFs have regulatory roles rather than essential functions.

Table 2. sORFs that result in a lethal or a slow-growth phenotype when deleted in haploid strains

Phenotype	sORF	Gene name	Size (amino acids) ^a	
Lethal	<i>YBR233W-A</i>	<i>DAD3</i>	94	
	<i>YDR320C-A</i>	<i>DAD4</i>	72	
	<i>YER074W-A</i>	<i>YOS1</i>	87	
	<i>YHR072W-A</i>	<i>NOP10</i>	58	
	<i>YKL138C-A</i>	<i>HSK3</i>	69	
	<i>YLR099W-A</i>		87	
	<i>YNL024C-A</i>		72	
	<i>YNL138W-A</i>	<i>YSF3</i>	85	
	Slow growth	<i>YBL071C-B</i>		32
		<i>YBL071W-A</i>	<i>KTI11</i>	82
<i>YGL007C-A</i>			28	
<i>YGL188C-A</i>			46	
<i>YPL096C-A</i>		<i>ERI1</i>	68	
<i>YPL189C-A</i>			68	

^aSize = size of protein encoded by sORF in amino acids.

Phenotypic analyses of haploid sORF deletion strains

Six of the new haploid sORF deletion strains exhibit slow-growth phenotypes when grown at 30°C (Table 2), including strains deleted for *YBL071W-A/KTI11* and *YPL096C-A/ERI1*, which are known to exhibit slow growth (Fichtner and Schaffrath 2002; Sobering et al. 2003). We further analyzed the growth of the new haploid sORF deletion strains in pilot screens under the following conditions: (1) at high (37°C) or low (11°C) temperatures; (2) in the presence of a sole nonfermentable carbon source; and (3) in the presence of the replication-arrest agent HU and DNA-damaging agents including MMS, bleomycin, and UV radiation. To confirm the results of the pilot screen, we sporulated heterozygous strains corresponding to the haploid strains with phenotypes and characterized the meiotic progeny. For each strain, we analyzed three independent sORF deletion spores and confirmed

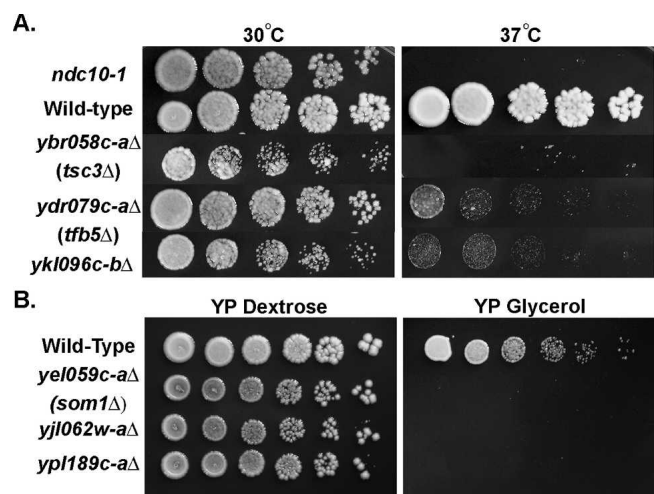


Figure 3. sORFs required for growth at 37°C (Ts) or in the presence of nonfermentable carbon source (petite phenotype). (A) sORFs required for growth at 37°C. Growth assays of 3 μ L of fivefold serial dilutions of logarithmic-phase cells of the sORF deletion strains spotted on YPD plates and incubated at 30°C or 37°C for 2 to 3 d. The wild-type strain BY4741 and the temperature-sensitive strain *ndc10-1* (JK421) served as controls. (B) Growth assays of 3 μ L of fivefold serial dilutions of logarithmic-phase cells of the sORF deletion strains spotted on YP Dextrose or YP Glycerol plates and incubated at 30°C for 2 to 3 d.

that the phenotype was linked to the sORF deletion. Upon verification of the phenotypes, we confirmed chromosomal deletion of the sORF by PCR and sequence analysis of the genomic locus at the site of integration of the *kanMX* cassette (see Methods).

We observed that three of the sORF deletion strains are temperature-sensitive (Ts) for growth at 37°C (Fig. 3A), while none of the sORF deletion strains showed a cold-sensitive growth phenotype at 11°C (data not shown). A Ts allele of the kinetochore mutant (*ndc10-1*) served as a control (Goh and Kilmartin 1993). We discovered a new gene required for growth at the nonpermissive temperature of 37°C, *YKL096C-B*, and confirmed the previously reported Ts growth phenotypes of strains with a mutation in sORFs *YBR058C-A/TSC3* and *YDR079C-A/TFB5*, which are involved in sphingolipid biosynthesis and transcription regulation, respectively (Fig. 3A; Gable et al. 2000; Ranish et al. 2004).

We tested the sORF deletion strains for a “petite” phenotype, which refers to an inability to grow in the presence of a nonfermentable carbon source and is an attribute of several mutants including mitochondrial mutants (for review, see Chen and Clark-Walker 2000). We determined that three sORF deletion strains exhibit a “petite” phenotype (Fig. 3B; Table 3). Our results confirm the role of *YEL059C-A/SOM1* for mitochondrial function (Esser et al. 1996) and suggest a similar function for the previously uncharacterized sORFs *YJL062W-A* and *YPL189C-A*. Consistent with a role in mitochondrial function, Yjl062w-ap fused to GFP has been localized to the mitochondrion (Huh et al. 2003).

To investigate the potential role of sORFs in response to genotoxic stress, we assayed the sORF deletion strains for sensitivity to the replication-arrest agent HU and to DNA-damaging agents bleomycin, MMS, and UV radiation. Sensitivity to these genotoxic agents can provide important clues about the roles of the genes in replication, transcription, cell-cycle progression, and chromosome segregation (Chang et al. 2002; Aouida et al. 2004; Parsons et al. 2004). In addition, many *S. cerevisiae* genes required for responding to DNA damage and replication arrest have human orthologs, mutations in which lead to human diseases (for review, see Zhou and Elledge 2000).

For these studies, strains grown to logarithmic phase were serially diluted, spotted on medium containing the appropriate drug, or exposed to UV, and incubated for 2–3 d at 30°C. The *S. cerevisiae* checkpoint mutant *mec1Δ sml1Δ*, which exhibits sensitivity to HU, bleomycin, MMS, and UV radiation, served as a

control (Kiser and Weinert 1996). HU inhibits ribonucleotide reductase, an enzyme that is required for synthesis of dNTPs in *S. cerevisiae* and other systems, and leads to an arrest in S-phase of the cell cycle (Elledge et al. 1993). As shown in Figure 4A, three sORF deletion strains exhibit varying degrees of sensitivity to growth on HU-containing media, with *ybr058c-aΔ/tsc3Δ* being the most sensitive. Our results suggest new roles for Tsc3p, a sphingolipid biosynthetic enzyme (Gable et al. 2000); Sus1p, a component of the SAGA and Sac3p–Tthp1p complexes (Rodriguez-Navarro et al. 2004); and the uncharacterized *YBR196C-A*, in responding to replication arrest.

Next, we tested bleomycin, a radiomimetic drug that leads to both single- and double-stranded DNA damage (Chen and Stubbe 2005), and discovered that four sORF deletion strains are sensitive to bleomycin. sORF deletion strains lacking *YBR058C-A/TSC3* showed the most sensitivity, while *ykl096c-bΔ* and *ydr524w-cΔ* strains were only moderately sensitive to bleomycin (10 μU/mL) (Fig. 4B). Our results extend the role of Tsc3p in responding to replication arrest caused by HU (Fig. 4A) to an additional role in responding to DNA damage caused by bleomycin.

We also discovered a new sORF required for growth in the presence of MMS. MMS is a DNA-alkylating agent that primarily methylates DNA on N⁷-deoxyguanine and N³-deoxyadenine (Pegg 1984). Resistance to MMS requires genes from the bypass, post-replication, recombination, base excision repair, and/or checkpoint pathways (Weinert et al. 1994; Xiao et al. 1996; Ter-cero and Diffley 2001). The sORF deletion strain *ybr111w-aΔ/sus1Δ* is sensitive to growth on MMS medium (Fig. 4C), a phenotype that, to our knowledge, has not been previously reported for strains deleted for *YBR111W-A/SUS1*. Ybr111w-ap/Sus1p is a component of the SAGA complex and the Sac3p–Tthp1p mRNA export complex (Rodriguez-Navarro et al. 2004). These results, combined with earlier results (Fig. 4A), suggest a novel role for Sus1p in response to DNA damage induced by MMS and replication arrest induced by HU. Finally, we confirmed a UV-sensitivity phenotype previously reported for the *ydr079c-aΔ/tfb5Δ* strain in a different genetic background (Fig. 4D; Ranish et al. 2004).

The sORF deletion strains exhibit overlapping and distinct phenotypes

Taken together, we observed conditional phenotypes for nine sORF deletion strains (Table 3). Not surprisingly, we observed

Table 3. sORFs with phenotypes when deleted

sORF	Gene name	Size (amino acids) ^a	Protein expression	Phenotypes of sORF deletion strains					
				TS (37°C) ^b	Glycerol ^c	HU	Bleo	MMS	UV
<i>YBR058C-A</i>	<i>TSC3</i>	80	Reported (Gable et al. 2000)	S ^e		S	S		
<i>YBR111W-A</i>	<i>SUS1</i>	96	Reported (Rodriguez-Navarro et al. 2004)			S		S	
<i>YBR196C-A</i>		49	ND ^d			S			
<i>YDR079C-A</i>	<i>TFB5</i>	72	Reported (Ranish et al. 2004)	S					S
<i>YDR524W-C</i>		29	Figure 6				S		
<i>YEL059C-A</i>	<i>SOM1</i>	74	Reported (Jan et al. 2000)		S				
<i>YJL062W-A</i>		85	Figure 6		S				
<i>YKL096C-B</i>		49	ND	S			S		
<i>YPL189C-A</i>		68	Figure 6		S				

^aSize = size of protein encoded by sORF in amino acids.

^bTS = temperature-sensitive.

^cGlycerol = petite phenotype when grown on media containing glycerol and ethanol as sole carbon source.

^dND = not done.

^eS = sensitive to high temperature (37°C) or to genotoxic agents.

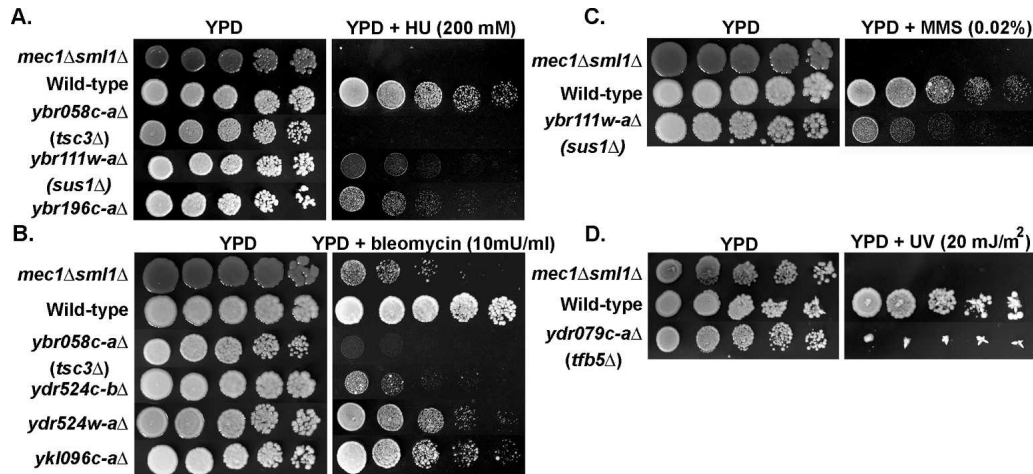


Figure 4. sORFs required for growth on media containing replication-arrest and DNA-damaging agents. Growth assays of 3 μ L of fivefold serial dilutions of logarithmic-phase cells of the sORF deletion strains spotted on YPD plates (control) or spotted on YPD plates containing (A) 200 mM HU, (B) 10 μ M/ml bleomycin, or (C) 0.02% MMS. For sensitivity to UV, we irradiated strains spotted on YPD with (D) 20 mJ/m^2 of UV radiation. All plates were incubated at 30°C for 2 to 3 d. The checkpoint mutant (*mec1* Δ *sml1* Δ , U953–61D) and the wild-type strain (BY4741) served as controls.

that several of the sORF deletion strains exhibit overlapping phenotypes when subjected to DNA damage or replication arrest, an observation made with other ORF deletion strains (Chang et al. 2002; Table 3). For example, two of the HU-sensitive strains also exhibit sensitivity to bleomycin and MMS. Interestingly, all three Ts sORF deletion strains are also sensitive to DNA-damage or replication-arrest agents. These results may suggest that the role of these genes in response to DNA damage and replication arrest may be essential for haploid growth at the nonpermissive temperature of 37°C.

Phenotypic analyses of deletion strains for genes flanking the sORFs

Six of the sORFs that exhibited phenotypes distinct from wild type when deleted (*YBR058C-A/TSC3*, *YBR111W-A/SUS1*, *YDR079C-A/TFB5*, *YEL059C-A/SOM1*, *YJL062W-A*, and *YKL096C-B*) are within 300 bp of larger ORFs. The phenotypes we observed may be due to altered expression of the neighboring ORFs caused by disruptions in their promoters or 5'- or 3'-untranslated regions rather than loss of function of the deleted sORFs. We therefore examined the phenotypes of strains with deletions of genes that are within 300 bp of the sORFs, a conservative approach, as ~60% of ORFs, both large and small, are within 300 bp of another ORF. In all but two cases (*YBR111W-A/SUS1*, *YJL062W-A*), deletion of the neighboring genes did not produce the phenotypes we observed for the *sorf* Δ strain (Supplemental Table C). For these two deletions strains, their phenotypes could be due to interference of expression of a neighboring ORF, loss of the sORF, or both.

We determined that the deletion strain for *YGR271C-A* showed slow growth, Ts, and an HU-sensitivity phenotype and that a deletion strain for *YGR272C*, which is 51 bp away from *YGR271C-A*, also exhibits such phenotypes (Fig. 5A). Sequence analysis of the genomic locus revealed that *YGR271C-A* is contiguous with *YGR272C*, forming a single ORF, consistent with the similarity of these two predicted ORFs to a single ORF (*PABR143C*) from *Ashbya gossypii* (Fig. 5B; Brachat et al. 2003). We constructed a new gene-deletion strain for the larger ORF,

which we denote as *ygr271c-a* Δ /*ygr272c* Δ , and determined that this strain showed a more severe slow growth, Ts, and HU-sensitivity phenotype compared to the *ygr271c-a* Δ or *ygr272c* Δ strains (Fig. 5A). Further analysis in a cell cycle arrest–release experiment showed that the *ygr271c-a* Δ /*ygr272c* Δ strain exhibits a significant delay of at least 40 min in exiting from the G₁ phase of the cell cycle after an arrest with α -factor (Fig. 5C). Our results, combined with the analysis of protein expression described below, establish that *YGR271C-A* and *YGR272C* constitute a single ORF, which we have named *EFG1* (Exit from G₁).

Protein expression analysis of the sORFs

Recent evidence of expression at the protein level for sORFs has come from genome-wide TAP- and GFP-tagging experiments (Ghaemmaghani et al. 2003; Huh et al. 2003; Supplemental Table A). Protein expression for some of the sORFs detected in our screens has been reported in these (Supplemental Table A) and other studies (Table 3). We epitope-tagged a subset of sORFs identified in our phenotypic analyses by introducing a haemagglutinin epitope (HA) at the C-terminus in their chromosomal context and examined expression of the tagged protein by Western blot analysis. We detected expression of proteins from strains expressing HA-tagged *YJL062W-A*, *YPL189C-A*, and *YDR524W-C* (Fig. 6, lanes 1,2,4). We also detected a band of expected size for *EFG1*, further confirming that *YGR271C-A* and *YGR272C* constitute a single ORF (Fig. 6, lane 3).

Conclusions

In the past, the function of sORFs has been elusive owing to inherent difficulties in identifying them based on genetic, biochemical, or solely computational approaches. *S. cerevisiae* represents one of the few systems with a wealth of data derived from several functional genomic and comparative genomic studies. Using the strengths of *S. cerevisiae* as a model, we provide the first systematic investigation of sORF function in any system. Our analysis of the literature combined with our genetic analyses for sORF function presents a comprehensive database for the 299 sORFs in *S. cerevisiae*. Of the *S. cerevisiae* sORFs, 184 are related to

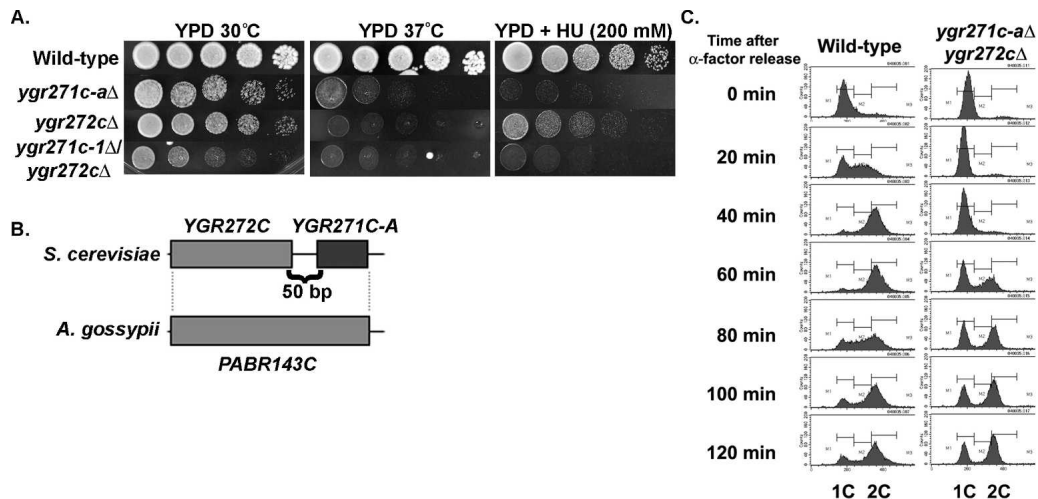


Figure 5. *YGR271C-A* and *YGR272C* constitute a contiguous ORF required for growth at 37°C, growth on HU-containing media and cell cycle progression. (A) The *ygr271c-aΔ*, *ygr272cΔ* and *ygr271c-aΔ/ygr272cΔ* strains exhibit slow growth, Ts, and HU-sensitive phenotypes. Growth assays were carried out as in Figures 3 and 4. (B) *YGR271C-A* and *YGR272C* exhibit similarity to the *A. gossypii* ORF *PABR143C* (Brachat et al. 2003). By sequencing the genomic locus in *S. cerevisiae*, we determined that *YGR271C-A* and *YGR272C* constitute a single larger ORF. (C) Cell cycle analysis was done by arresting wild-type (BY4741) and *ygr271c-aΔ/ygr272cΔ* strains with α -factor (0 min) and then releasing them into pheromone-free media. Samples were analyzed for DNA content at 20-min intervals as indicated by flow cytometry. The 1C and 2C peaks denote cells with a 1N and 2N DNA content corresponding to cells in G₁ or G₂/M phase of the cell cycle, respectively.

sequences in other eukaryotes, suggesting the evolutionary conservation of the structure and perhaps function of these sORFs. Although relatively little is known about sORF functions, they have been implicated in key cellular processes including transport, intermediary metabolism, chromosome segregation, genome stability, and other functions. The sORF gene-deletion collection should lead to the discovery of additional functions for sORFs in *S. cerevisiae*. Moreover, our results, which emphasize the biological significance of sORFs in *S. cerevisiae* that are conserved across eukaryotes, should provide an impetus for the identification and characterization of sORFs in other systems, including humans.

Methods

Analysis of sORF percentage in representative eukaryotes

The number of sORFs coding for proteins of 100 amino acids in length or less, annotated on the transcripts of model organisms in the NCBI RefSeq database, was determined using a query of the Entrez Protein database of the form: `srcdb refseq[prop] AND homo sapiens[orgn] AND 0:100[slen]`. The total number of ORFs in each set was counted using a query of the form: `srcdb refseq[prop] AND homo sapiens[orgn]`. The version of RefSeq used was that present in Entrez on 3/15/2005 corresponding to RefSeq release 10, available on 3/6/2005 with updates from 3/6/2005 to 3/15/2005.

Homology searches

HomoloGene

For this study, HomoloGene build 38.1, dating from November 23, 2004, was used ([ftp://ftp.ncbi.nih.gov/pub/HomoloGene/](http://ftp.ncbi.nih.gov/pub/HomoloGene/)).

BLAST of sORFs with annotated proteins and UniGene

Single sequence representatives of the UniGene clusters, the “seq-uniques” described in Supplemental material, were downloaded for each organism from the NCBI FTP site ([ftp://ftp.ncbi.nih.gov/](http://ftp.ncbi.nih.gov/)) and compared to the sORFs using BLAST (see Supplemental material; Altschul et al. 1997). The best BLAST hit was extracted for each sORF only if the hit spanned at least one-third of the translated ORF with an amino acid identity of at least 40%; otherwise, no hit was extracted. The results are summarized in Supplemental Table B.

Media and yeast strains

The media and methodology for yeast growth were as described (Gietz et al. 1992, 1995; Adams et al. 1997; Brachmann et al. 1998). The deletion strain was generated in diploid strain BY4743 (*MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 lys2Δ0/LYS2 MET15/met15Δ0 ura3Δ0/ura3Δ0*), and the haploid spores isogenic with

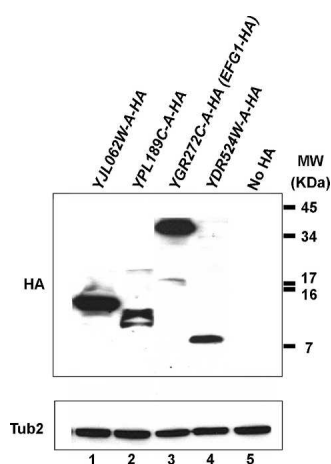


Figure 6. Protein expression analysis of HA-tagged sORFs. Western blot analysis of protein extracts prepared from strains expressing HA-tagged ORFs (*YJL062W-A*, *YPL189C-A*, *YGR271C-A/YGR272C*, *YDR524W-C*) and the wild-type strain (BY4741) not expressing a HA-tagged protein (no HA tag). Proteins of the expected size were detected (ORF-HA), and Tub2p served as a loading control (Tub2).

BY4741 (*MAT α his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0*) and BY4742 (*MAT α his3 Δ 1 leu2 Δ 0 lys2 Δ 0 ura3 Δ 0*) were identified (Winzeler et al. 1999). Other strains include the temperature-sensitive control strain JK421 (*MAT α ade2-1 ura3-1 his3-11,1 trp1-1 leu2-3,112 can1-100 ndc10-1*) (Goh and Kilmartin 1993) and the checkpoint mutant U953-61D (*MAT α leu2-3,112 ade2-1 can1-100 his3-11,15 ura3-1 trp1-1 RAD5 mec1 Δ ::TRP1 sml1 Δ ::HIS3*) (Zhao et al. 1998).

Gene-deletion strain construction and confirmation

A PCR-generated (Baudin et al. 1993; Wach et al. 1994) deletion strategy was used to systematically replace each sORF from its start to its stop codon with a *kanMX* module and two unique 20-mer molecular barcodes as done previously for the gene-deletion strain collection (Winzeler et al. 1999; Giaever et al. 2002; Supplemental Fig. 1; Supplemental material; barcode sequences are given in Supplemental Table A). Each sORF gene-deletion mutant is publicly available either as haploids (*MAT α* and *MAT α*) or as diploids (homozygous or heterozygous; see http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html).

Phenotypic analyses of sORF deletion strains

For sensitivity to HU, MMS, bleomycin, UV, and nonpermissive growth temperatures, we assayed serial dilutions of the sORF strains on YPD or YPD containing 200 mM hydroxyurea (HU; H8627; Sigma), 0.02% methane methylsulfonate (MMS; 64294; Fluka Chemika), or 10 μ M/ml bleomycin (BLM; 3154-01; Bristol-Myers Squibb Co.). For sensitivity to UV-radiation, we irradiated strains spotted on YPD with 20 mJ/m^2 using a Stratalinker (Stratagene). For growth at the nonpermissive temperatures, we incubated plates at either 11°C or 37°C. A “petite” phenotype was determined by plating strains on a modified YPD medium in which dextrose was substituted with 2% glycerol and 2% ethanol.

Protein expression analysis of HA-tagged ORFs

ORFs were fused in-frame at the genomic locus with three copies of the HA epitope at their C-terminus as previously described (Longtine et al. 1998; Supplemental material). Protein extracts of ORF-HA-expressing strains were analyzed by Western blot analysis as described previously (Crotti and Basrai 2004). The primary antibody was anti-HA (clone 12CA5-Roche) or anti-Tub2p (polyclonal antibody, Basrai lab), and the secondary antibody was HRP-conjugated sheep anti-mouse IgG (NA931V; Amersham).

α -Factor arrest/release experiments

Strains were grown overnight at 30°C in YPD medium and then diluted into fresh medium to obtain a logarithmic-phase culture. Cells were arrested in the presence of 3 μ M α -factor (T-6901; Sigma) at 30°C for 90 min, washed twice with water, and resuspended in fresh YPD medium and incubated at 30°C. DNA content was assayed every 20 min after release from the α -factor arrest for a total of 3 h as described previously (Doheny et al. 1993; Basrai et al. 1996) using a Becton-Dickinson FACSort flow cytometer and CellQuest software (BD Biosciences).

Acknowledgments

The authors thank Anand Sethuraman and Mike Cherry for help with compiling the sequences of the sORFs; Anuj Kumar for sharing unpublished data; Lucy Liu and Xiuquiong Zhou for tetrad dissections; Keith Anderson, Ana Aparicio, and Mike Jensen of the SGTC (Stanford Genome Technology Center) for assistance

with the A.M.O.S. primers; and Mark Johnston for advice and support of this work. This work was supported in part by NIH grant R01-HG02432 to J.D.B. and by the Intramural Research Program of the NIH and NCI.

References

- Adams, A., Gottschling, D.E., Kaiser, C.A., and Stearns, T. 1997. *Methods in yeast genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Aouida, M., Tounekti, O., Leduc, A., Belhadj, O., Mir, L., and Ramotar, D. 2004. Isolation and characterization of *Saccharomyces cerevisiae* mutants with enhanced resistance to the anticancer drug bleomycin. *Curr. Genet.* **45**: 265–272.
- Basrai, M.A. and Hieter, P. 2002. Transcriptome analysis of *Saccharomyces cerevisiae* using serial analysis of gene expression. *Methods Enzymol.* **350**: 414–444.
- Basrai, M.A., Kingsbury, J., Koshland, D., Spencer, F., and Hieter, P. 1996. Faithful chromosome transmission requires Spt4p, a putative regulator of chromatin structure in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**: 2838–2847.
- Basrai, M.A., Hieter, P., and Boeke, J.D. 1997. Small open reading frames: Beautiful needles in the haystack. *Genome Res.* **7**: 768–771.
- Basrai, M.A., Velculescu, V.E., Kinzler, K.W., and Hieter, P. 1999. NORF5/HUG1 is a component of the MEC1-mediated checkpoint response to DNA damage and replication arrest in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19**: 7041–7049.
- Baudin, A., Ozier-Kalogeropoulos, O., Denouel, A., Lacroute, F., and Cullin, C. 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **21**: 3329–3330.
- Blandin, G., Durrens, P., Tekaiia, F., Aigle, M., Bolotin-Fukuhara, M., Bon, E., Casaregola, S., de Montigny, J., Gaillardin, C., Lepingle, A., et al. 2000. Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett.* **487**: 31–36.
- Brachat, S., Dietrich, F.S., Voegeli, S., Zhang, Z., Stuart, L., Lerch, A., Gates, K., Gaffney, T., and Philippsen, P. 2003. Reinvestigation of the *Saccharomyces cerevisiae* genome annotation by comparison to the genome of a related fungus: *Ashbya gossypii*. *Genome Biol.* **4**: R45.
- Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and Boeke, J.D. 1998. Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* **14**: 115–132.
- Chang, M., Bellaoui, M., Boone, C., and Brown, G.W. 2002. A genome-wide screen for methyl methanesulfonate-sensitive mutants reveals genes required for S phase progression in the presence of DNA damage. *Proc. Natl. Acad. Sci.* **99**: 16934–16939.
- Cheeseman, I.M., Anderson, S., Jwa, M., Green, E.M., Kang, J., Yates III, J.R., Chan, C.S., Drubin, D.G., and Barnes, G. 2002. Phospho-regulation of kinetochore-microtubule attachments by the Aurora kinase Ipl1p. *Cell* **111**: 163–172.
- Chen, X.J. and Clark-Walker, G.D. 2000. The petite mutation in yeasts: 50 years on. *Int. Rev. Cytol.* **194**: 197–238.
- Chen, J. and Stubbe, J. 2005. Bleomycins: Towards better therapeutics. *Nat. Rev. Cancer* **5**: 102–112.
- Cliften, P., Sudarsanam, P., Desikan, A., Fulton, L., Fulton, B., Majors, J., Waterston, R., Cohen, B.A., and Johnston, M. 2003. Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**: 71–76.
- Crotti, L.B. and Basrai, M.A. 2004. Functional roles for evolutionarily conserved Spt4p at centromeres and heterochromatin in *Saccharomyces cerevisiae*. *EMBO J.* **23**: 1804–1814.
- Doheny, K.F., Sorger, P.K., Hyman, A.A., Tugendreich, S., Spencer, F., and Hieter, P. 1993. Identification of essential components of the *S. cerevisiae* kinetochore. *Cell* **73**: 761–774.
- Eddy, S.R. 2001. Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.* **2**: 919–929.
- Elledge, S.J., Zhou, Z., Allen, J.B., and Navas, T.A. 1993. DNA damage and cell cycle regulation of ribonucleotide reductase. *Bioessays* **15**: 333–339.
- Esser, K., Pratje, E., and Michaelis, G. 1996. SOM 1, a small new gene required for mitochondrial inner membrane peptidase function in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **252**: 437–445.

- Fichtner, L. and Schaffrath, R. 2002. KTI11 and KTI13, *Saccharomyces cerevisiae* genes controlling sensitivity to G₁ arrest induced by *Kluyveromyces lactis* zymocin. *Mol. Microbiol.* **44**: 865–875.
- Gable, K., Slife, H., Bacikova, D., Monaghan, E., and Dunn, T.M. 2000. Tsc3p is an 80-amino acid protein associated with serine palmitoyltransferase and required for optimal enzyme activity. *J. Biol. Chem.* **275**: 7597–7603.
- 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.
- Gietz, D., St Jean, A., Woods, R.A., and Schiestl, R.H. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Res.* **20**: 1425.
- Gietz, R.D., Schiestl, R.H., Willems, A.R., and Woods, R.A. 1995. Studies on the transformation of intact yeast cells by the LiAc/SS-DNA/PEG procedure. *Yeast* **11**: 355–360.
- Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldman, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., et al. 1996. Life with 6000 genes. *Science* **274**: 546–567.
- Goh, P.-Y. and Kilmartin, J.V. 1993. NDC10: A gene involved in chromosome segregation in *S. cerevisiae*. *J. Cell Biol.* **121**: 503–512.
- Heidtman, M., Chen, C.Z., Collins, R.N., and Barlowe, C. 2005. Yos1p is a novel subunit of the Yip1p–Yif1p complex and is required for transport between the endoplasmic reticulum and the Golgi complex. *Mol. Biol. Cell* **16**: 1673–1683.
- Henras, A., Henry, Y., Bousquet-Antonelli, C., Noaillac-Depeyre, J., Gelugne, J.P., and Caizergues-Ferrer, M. 1998. Nhp2p and Nop10p are essential for the function of H/ACA snoRNPs. *EMBO J.* **17**: 7078–7090.
- 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.
- Jan, P.S., Esser, K., Pratje, E., and Michaelis, G. 2000. Som1, a third component of the yeast mitochondrial inner membrane peptidase complex that contains Imp1 and Imp2. *Mol. Gen. Genet.* **263**: 483–491.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- 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.
- Kiser, G.L. and Weinert, T.A. 1996. Distinct roles of yeast MEC and RAD checkpoint genes in transcriptional induction after DNA damage and implications for function. *Mol. Biol. Cell* **7**: 703–718.
- 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.
- Li, J.M., Li, Y., and Elledge, S.J. 2005. Genetic analysis of the kinetochore DASH complex reveals an antagonistic relationship with the ras/protein kinase A pathway and a novel subunit required for Ask1 association. *Mol. Cell Biol.* **25**: 767–778.
- Longtine, M.S., McKenzie III, A., Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P., and Pringle, J.R. 1998. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**: 953–961.
- Marchler-Bauer, A., Anderson, J.B., Cherkuri, P.F., DeWeese-Scott, C., Geer, L.Y., Gwadz, M., He, S., Hurwitz, D.I., Jackson, J.D., Ke, Z., et al. 2005. CDD: A Conserved Domain Database for protein classification. *Nucleic Acids Res.* **33**: D192–D196.
- Olivas, W.M., Muhlrud, 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.
- Pan, X., Yuan, D.S., Xiang, D., Wang, X., Sookhai-Mahadeo, S., Bader, J.S., Hieter, P., Spencer, F., and Boeke, J.D. 2004. A robust toolkit for functional profiling of the yeast genome. *Mol. Cell* **16**: 487–496.
- Parsons, A.B., Brost, R.L., Ding, H., Li, Z., Zhang, C., Sheikh, B., Brown, G.W., Kane, P.M., Hughes, T.R., and Boone, C. 2004. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. *Nat. Biotechnol.* **22**: 62–69.
- Pegg, A.E. 1984. Properties of the O6-alkylguanine-DNA repair system of mammalian cells. *IARC Sci. Publ.* 575–580.
- Pontius, J.U., Wagner, L., and Schuler, G.D. 2003. *UniGene: A unified view of the transcriptome*. National Center for Biotechnology Information, Bethesda, MD. <http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=handbook.chapter.857>
- Pruitt, K.D., Tatusova, T., and Maglott, D.R. 2005. NCBI Reference Sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33**: D501–D504.
- Ranish, J.A., Hahn, S., Lu, Y., Yi, E.C., Li, X.J., Eng, J., and Aebersold, R. 2004. Identification of TFBS, a new component of general transcription and DNA repair factor IIH. *Nat. Genet.* **36**: 707–713.
- Rodriguez-Navarro, S., Fischer, T., Luo, M.J., Antunez, O., Brettschneider, S., Lechner, J., Perez-Ortin, J.E., Reed, R., and Hurt, E. 2004. Sus1, a functional component of the SAGA histone acetylase complex and the nuclear pore-associated mRNA export machinery. *Cell* **116**: 75–86.
- Sobering, A.K., Romeo, M.J., Vay, H.A., and Levin, D.E. 2003. A novel Ras inhibitor, Eri1, engages yeast Ras at the endoplasmic reticulum. *Mol. Cell Biol.* **23**: 4983–4990.
- Tercero, J.A. and Diffley, J.F. 2001. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* **412**: 553–557.
- Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K. 1995. Serial analysis of gene expression. *Science* **270**: 484–487.
- Velculescu, V., Zhang, L., Zhou, W., Vogelstein, B., Basrai, M.A., Bassett Jr., D.E., Hieter, P., Vogelstein, B., and Kinzler, K. 1997. Characterization of the yeast transcriptome. *Cell* **88**: 243–251.
- Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. *Yeast* **10**: 1793–1808.
- Weinert, T.A., Kiser, G.L., and Hartwell, L.H. 1994. Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair. *Genes & Dev.* **8**: 652–665.
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S., Helmberg, W., et al. 2005. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **33**: D39–D45.
- 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.
- Xiao, W., Chow, B.L., and Rathgeber, L. 1996. The repair of DNA methylation damage in *Saccharomyces cerevisiae*. *Curr. Genet.* **30**: 461–468.
- Zhao, X., Muller, E.G., and Rothstein, R. 1998. A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools. *Mol. Cell* **2**: 329–340.
- Zhou, B.B. and Elledge, S.J. 2000. The DNA damage response: Putting checkpoints in perspective. *Nature* **408**: 433–439.

Received June 28, 2005; accepted in revised form October 6, 2005.



Functional genomics of genes with small open reading frames (sORFs) in *S. cerevisiae*

James P. Kastenmayer, Li Ni, Angela Chu, et al.

Genome Res. 2006 16: 365-373

Access the most recent version at doi:[10.1101/gr.4355406](https://doi.org/10.1101/gr.4355406)

Supplemental Material <http://genome.cshlp.org/content/suppl/2006/02/14/16.3.365.DC1>

References This article cites 57 articles, 18 of which can be accessed free at:
<http://genome.cshlp.org/content/16/3/365.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>