

## Gene and Protein Expression Data at the *Saccharomyces* Genome Database

Edith D. Wong, Robert S. Nash, Stacia R. Engel, Marek S. Skrzypek, Olivia Lang, Kalpana Karra, Shuai Weng, Gail A. Binkley, and J. Michael Cherry

Department of Genetics, Stanford University, Stanford, CA 94305

### **Abstract**

Knowing when, where and to what levels genes are expressed is crucial for understanding cellular biology. The budding yeast, *Saccharomyces cerevisiae*, is a model organism with extensive research literature that covers every aspect of cellular processes. By comprehensively curating the literature, the *Saccharomyces* Genome Database (SGD; [www.yeastgenome.org](http://www.yeastgenome.org)) strives toward providing a complete picture of budding yeast biological processes, and also enabling functional connections to other model and multi-cellular organisms. SGD has curated data that pertain to when, where and to what extent genes are expressed at both the RNA and protein levels. RNA expression data and metadata have been collected in a semi-automated fashion, using scripts to programmatically retrieve relevant information from the Gene Expression Omnibus (GEO; <https://www.ncbi.nlm.nih.gov/geo/>), subsequently followed by manual curation. Expression data can be explored using our expression tool, SPELL ([spell.yeastgenome.org](http://spell.yeastgenome.org)) and our instance of JBrowse ([browse.yeastgenome.org](http://browse.yeastgenome.org)). Metadata collected and manually curated from GEO and directly from the literature allow users to easily find and examine datasets of interest. SGD also provides manually curated protein subcellular localization information that uses the cellular component branch of the Gene Ontology (GO; [www.geneontology.org](http://www.geneontology.org)). In addition, SGD hosts the yeast GFP Fusion Localization Database (YeastGFP; <https://yeastgfp.yeastgenome.org/>), that provides comprehensive visual protein localization data for yeast gene products. Finally, SGD has manually curated multiple large-scale studies that determine protein abundance and decay. We continue to integrate gene and protein expression data, allowing users to examine differences in expression under varying conditions. And as part of the Alliance of Genome Resources ([www.alliancegenome.org](http://www.alliancegenome.org)), we are exploring ways to make connections with expression data from other model organisms.