Representing Transcriptional Heterogeneity and Inter-Strain Variation at the *Saccharomyces* Genome Database

Patrick C. Ng, Edith D. Wong, Stacia R. Engel, Robert S. Nash, Marek S. Skrzypek, Suzi Aleksander, Kevin A. MacPherson, Travis Sheppard, Shuai Weng, Kalpana Karra, J. Michael Cherry and the SGD Project

Department of Genetics, School of Medicine, Stanford University, Palo Alto, CA

One of missions of the *Saccharomyces* Genome Database (SGD; [https://www.yeastgenome.org](https://www.yeastgenome.org)) is to catalog genome information about *Saccharomyces cerevisiae* that includes variations in DNA sequence and RNA expression. To present a high resolution view of the transcriptional dynamics in *S. cerevisiae*, SGD utilized data generated by collaborators (Pelechano et al., 2013) through Transcript IsoForm Sequencing (TIF-seq) technology. This technique captures the heterogeneity of RNA transcription in *S. cerevisiae*, which can vary widely across experimental perturbations, within different environments, and between cells. The YeastMine data warehouse ([https://yeastmine.yeastgenome.org](https://yeastmine.yeastgenome.org)) provides downloadable tables mapping these isoforms to the locations of transcripts identified in a curated group of other genomic datasets. Transcript abundance and outer boundaries can be visualized in the JBrowse genome navigator ([https://browse.yeastgenome.org](https://browse.yeastgenome.org)), and annotation files of all transcripts that fully encompass annotated genes are available from SGD through Amazon Web Services (AWS). The AWS cloud server also hosts variant genotyping files for 25 *S. cerevisiae* strains (Song et al., 2015) assembled by the Automated Genome Analysis Pipeline (AGAPE), and these single nucleotide polymorphisms and small insertions/deletions are also shown in JBrowse. SGD’s Variant Viewer ([https://www.yeastgenome.org/variant-viewer/](https://www.yeastgenome.org/variant-viewer/)) presents the molecular consequence of variants within annotated genes and infers evolutionary relatedness across SGD’s 12 reference strains. With these resources in place, we seek to incorporate more transcriptome and strain variant data in concert with improving technical capabilities and a growing collection of datasets. This work is supported by a grant from the NHGRI (U41 HG001315).