Reply to Chen and Zhang: On interpreting genome-wide trends from yeast mutation accumulation data

We are pleased to learn that our mutation accumulation (MA) dataset of 864 single nucleotide mutations (SNMs) in the budding yeast *Saccharomyces cerevisiae* (1) has attracted interest and is being applied to other analyses.

MA mutations in large numbers can provide an unbiased picture of genome-wide patterns. One previously observed pattern is a positive correlation between transcription rate and mutation rate, inferred to be the result of transcription-associated mutagenesis (TAM) (2). In our paper, we used a conservative subset of 181 SNMs to analyze nascent transcription rate (TR) microarray data (3) and noted that the SNM rates observed were not significantly different between genes with varying TRs. Chen and Zhang note that when mRNA sequencing data (mRNA-seq) or nascent transcript sequencing data (NET-seq) are applied to the same question using the full set of 559 exonic SNMs, positions with SNMs have significantly higher mean transcript coverage than the rest of the genome (2). They therefore conclude that the SNM data support TAM. As noted by Chen and Zhang, our discrepant conclusions could be caused by both the use of all 559 SNMs vs. a conservative subset of SNMs and the use of transcript sequencing vs. microarray data (2). Supporting the latter point, the correlation between the mRNA-seq and NET-seq datasets is strong ($R^2 = 0.6344$), whereas the correlations between these datasets and the TR microarray dataset we used are not (mRNA-seq~TR: $R^2 = 0.1745$, NET-seq~TR: $R^2 = 0.2483$).

Although RNA sequencing data have nucleotide resolution and microarray data do not, the overall coverage distribution is highly skewed, with 82–86% of the coding sequences assayed having 0–1× transcript coverage. The average sequencing read depth of a group of sites could therefore be elevated by a handful of nucleotide positions with especially high coverage. Removing the five SNM positions with the highest transcript coverage reduces average SNM mRNA-seq coverage from 5.49 to 1.29 and NET-seq coverage from 3.37 to 1.54. The corresponding random expectations are 1.20 and 1.43, respectively (2).

To reduce the effects of specific nucleotides, we can instead first use per-nucleotide read depth from each dataset to segregate the coding sequences into regions with high (>1×), medium (1×), and low (0×) transcription. The mutation densities of MA SNMs within each category are thus measures of mutation rates relative to transcript coverage. We found that for mRNA-seq data, mutations at GC nucleotides were more common at positions with higher steady-state mRNA levels but not significantly so (GC: low = $1.08 \pm 0.07 \times 10^{-4}$/bp, medium = $1.11 \pm 0.16 \times 10^{-4}$/bp, high = $1.46 \pm 0.19 \times 10^{-4}$/bp; $P = 0.11$, χ^2 test). AT positions did not display a similar pattern ($P = 0.48$, χ^2 test). In NET-seq data, no trend was observed (GC: $P = 0.64$, χ^2 test; AT: $P = 0.55$, χ^2 test).

Although not a definitive answer, we agree that the trends observed by Chen and Zhang may lend some support to the hypothesis that mutation rate increases with transcription rate, although it is unclear from these data if a few specific sites drive the signal or if this is a more widespread phenomenon.

***Yuan O. Zhu*[a,b,c,1], *Mark L. Siegal*[d], *David W. Hall*[e], *and Dmitri A. Petrov*[b]**

[a]*Department of Genetics, Stanford University, Stanford, CA 94305;* [b]*Department of Biology, Stanford University, Stanford, CA 94305;* [c]*Genome Institute of Singapore, Singapore 138672;* [d]*Center for Genomics and Systems Biology, Department of Biology, New York University, New York, NY 10003; and* [e]*Department of Genetics, University of Georgia, Athens, GA 30602*

1 Zhu YO, Siegal ML, Hall DW, Petrov DA (2014) Precise estimates of mutation rate and spectrum in yeast. *Proc Natl Acad Sci USA* 111(22):E2310–E2318.

2 Chen X, Zhang J (2014) Yeast mutation accumulation experiment supports elevated mutation rates at highly transcribed sites. *Proc Natl Acad Sci USA* 111:E4062.

3 Pelechano V, Chávez S, Pérez-Ortín JE (2010) A complete set of nascent transcription rates for yeast genes. *PLoS ONE* 5(11): e15442.

Author contributions: Y.O.Z., M.L.S., D.W.H., and D.A.P. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. Email: yuanzhu@stanford.edu.