G3 Genes | Genomes | Genetics

INVESTIGATION

Perturbations of Transcription and Gene Expression-Associated Processes Alter Distribution of Cell Size Values in *Saccharomyces cerevisiae*

Nairita Maitra, Jayamani Anandhakumar, Heidi M. Blank, Craig D. Kaplan,[1] and Michael Polymenis[1]
Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843
ORCID IDs: 0000-0002-2251-5246 (H.M.B.); 0000-0002-7518-695X (C.D.K.); 0000-0003-1507-0936 (M.P.)

ABSTRACT The question of what determines whether cells are big or small has been the focus of many studies because it is thought that such determinants underpin the coupling of cell growth with cell division. In contrast, what determines the overall pattern of how cell size is distributed within a population of wild type or mutant cells has received little attention. Knowing how cell size varies around a characteristic pattern could shed light on the processes that generate such a pattern and provide a criterion to identify its genetic basis. Here, we show that cell size values of wild type *Saccharomyces cerevisiae* cells fit a gamma distribution, in haploid and diploid cells, and under different growth conditions. To identify genes that influence this pattern, we analyzed the cell size distributions of all single-gene deletion strains in *Saccharomyces cerevisiae*. We found that yeast strains which deviate the most from the gamma distribution are enriched for those lacking gene products functioning in gene expression, especially those in transcription or transcription-linked processes. We also show that cell size is increased in mutants carrying altered activity substitutions in Rpo21p/Rpb1, the largest subunit of RNA polymerase II (Pol II). Lastly, the size distribution of cells carrying extreme altered activity Pol II substitutions deviated from the expected gamma distribution. Our results are consistent with the idea that genetic defects in widely acting transcription factors or Pol II itself compromise both cell size homeostasis and how the size of individual cells is distributed in a population.

KEYWORDS
cell size
gamma distribution
RSC
RNA polymerase
THO

Mechanisms that control cell size have long been viewed as critical for the coupling between cell growth and cell division, which in turn governs rates of cell proliferation (Turner *et al.* 2012; Pringle and Hartwell 1981; Ginzberg *et al.* 2015; Westfall and Levin 2017; Willis and Huang 2017). Hence, size control has attracted attention in many systems, from bacteria and yeasts to animals (Si *et al.* 2017; Tzur *et al.* 2009; Son *et al.* 2012; Jorgensen *et al.* 2002; Zhang *et al.* 2002). Most studies have dealt with situations where the typical size of cells in a given experimental system and condition shifts to a different value, due to genetic or environmental perturbations. Despite many rounds of cell division, proliferating cells usually maintain their size in a given nutrient environment. Considering cell size as a proxy for cell growth, then shifts to a smaller or larger size provide a convenient 'metric' to gauge alterations in biological processes that are thought to be central to the physiological coupling between cell growth and division.

Cells tune their gene expression output to their size, to maintain the proper concentrations of macromolecules as cells change in volume (Vargas-Garcia *et al.* 2018). Changes in ploidy and the well-known positive association between cell size and DNA content (Gregory 2001) perhaps illustrate a straightforward solution to this problem. Compared to smaller haploid and diploid cells, larger polyploid ones have more genomic templates from which to drive gene expression. It has also been reported that ploidy-associated increases in cell size drive transcriptional changes (Galitski *et al.* 1999; Wu *et al.* 2010). The situation appears more complex in cells of different size but of the same genome (Marguerat and Bahler 2012; Zhurinsky *et al.* 2010). In fission yeast, it has been proposed that cells of different size regulate global transcription rates regardless of cellular DNA content so that their transcriptional output per

doi: https://doi.org/10.1534/g3.118.200854
Manuscript received September 12, 2018; accepted for publication November 19, 2018; published Early Online November 21, 2018.

Supplemental material available at Figshare: https://figshare.com/s/99db4b2cb8c4deedbbcd.
[1]Corresponding authors: 300 Olsen Blvd, Biochemistry and Biophysics College Station, TX 77845 (Email: craig.kaplan@pitt.edu; polymenis@tamu.edu)

protein remains constant (Zhurinsky *et al.* 2010). Based on single molecule transcript counting in mammalian cells, a positive association between transcription burst magnitude and cell size has been reported (Padovan-Merhar *et al.* 2015). Furthermore, it appears that the doubling of the available DNA templates for transcription after DNA replication is countered by a decrease in transcription burst frequency in cells that have replicated their DNA, later in the cell cycle (Padovan-Merhar *et al.* 2015). These mechanisms, involving independent control of the frequency and the magnitude of transcription bursts, are thought to maintain the scaling of mRNA counts with the size of mammalian cells. In the budding yeast *Saccharomyces cerevisiae*, analogous single molecule experiments monitoring transcription bursts as a function of cell size and cell division have not been reported. Instead, a somewhat different mechanism has been proposed to explain the positive association of mRNA steady-state levels with cell size, due to increased stability of mRNAs in larger cells (Mena *et al.* 2017). Furthermore, it has been proposed that levels of active RNA Pol II are higher in small G1 cells with un-replicated DNA (Mena *et al.* 2017).

The genetic control of cell size has been studied extensively in *S. cerevisiae*. In this organism, systematic surveys of all single-gene deletions have been carried out to identify mutants that are bigger or smaller than the wild type (Zhang *et al.* 2002; Jorgensen *et al.* 2002). Similar size-based screens have also been carried out in other organisms (Björklund *et al.* 2006), with similar outputs, namely the identification of small or large-celled mutants. In contrast, much less attention has been placed on how the size of individual cells within a population, mutant or wild type, is distributed. We reasoned that if we first determine if yeast cells fit a particular distribution of sizes, we might then determine what types of mutants alter such a stereotypical distribution of cell sizes to understand its genetic basis.

Here we report that size in a population of *S. cerevisiae* cells is best described with a gamma distribution. We also identify genes that are required to maintain this distribution. These genes overwhelmingly encode proteins involved in global gene expression, especially in transcription. Lastly, we show that defects arising from alterations to the Pol II active site alter size homeostasis and the pattern of size distributions.

MATERIALS AND METHODS

Yeast strains and cell size measurements

For cell size measurements we report in Figure 5, the homozygous deletion strains were in the diploid S288c background of strain BY4743 (*MATa/α his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0*). The strains were constructed by the Systematic Deletion Project (Giaever *et al.* 2002; Brachmann *et al.* 1998).

For cell size measurements we report in Figure 6, genomic variants of *RPO21/RPB1* were created by CRISPR/Cas9-mediated genome editing (see Table S1 for a list of the oligonucleotides used in this study) of strain CKY3284, a derivative of the S288c strain background (see Table S2 for a list of the strains used in this study). Briefly, the sequence encoding a guide RNA was cloned into plasmid pML107 ((Laughery *et al.* 2015); see Table S3 for a list of the plasmids used in this study). This plasmid was introduced into CKY3284 by transformation, along with annealed double-stranded repair oligonucleotides or annealed overlapping oligonucleotides (Integrated DNA Technologies, Skokie, Illinois; see Table S1) filled in with Phusion DNA polymerase (New England Biolabs, Ipswich, Massachusetts) containing either a silently mutated PAM site or both a silently mutated PAM site and relevant *RPO21/RPB1* mutation. Variants were confirmed by PCR amplification of the mutated region and DNA sequencing. Oligonucleotides were annealed as follows: synthesized, lyophilized oligonucleotides were resuspended at 100 μM in 10 mM Tris pH 8.0. Perfectly matched oligonucleotide pairs were annealed as follows: 37.5 μl of each oligonucleotide was mixed with 25 μl of 1M Tris pH 8.0 and heated for 5 min at 95°, then tubes were transferred into a 70° heat block followed by removal of heat block from heater; they were moved to 4° overnight when block temperature reached room temperature. 8 μl of these annealed oligonucleotides were used as repair templates in individual transformations. Oligonucleotides that were overlapping (noted in Table S3) were annealed and extended by 5 cycles of standard PCR followed by thermal annealing as above.

All strains were cultured at 30° in standard YPD medium (1% $^{w}/_{v}$ yeast extract, 2% $^{w}/_{v}$ peptone, 2% $^{w}/_{v}$ dextrose). Cell size was measured with a Z2 Beckman Coulter channelyzer as described previously (Guo *et al.* 2004; Bogomolnaya *et al.* 2006).

DNA content analysis was done as we have described previously (Hoose *et al.* 2012; Hoose *et al.* 2013).

Statistical analysis

In all our statistical analyses we used R language packages, as indicated in each case. The cell size frequency distributions we analyzed from the literature were from (Soma *et al.* 2014) for the BY4743 dataset, and from (Jorgensen *et al.* 2002) for the haploid dataset (BY4741; *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) (see File S1; sheets 'by4743_raw', and 'Jorgensen_raw', respectively). Replicates of several strains in the Jorgensen dataset were marked as such, following their systematic ORF name. The cell size frequencies from (Soma *et al.* 2014) and (Jorgensen *et al.* 2002) were used to simulate distributions from n = 1000 cells in every case (see File S1; sheets 'by4743', and 'jorgensen', respectively). Similarly, we also generated the size distributions shown in Figures 5 and 6 (see File S1; sheets 'figure 5', and 'figure 6', respectively). To generate counts from frequencies for downstream statistical analysis, we used the R code listed in File S2.

To test for normality, we implemented the Shapiro-Wilk test (Shapiro and Wilk 1965) from the *stats* R language package, as described in detail in File S2. The corresponding p-values are shown in File S1, in the sheet columns marked as 'SW(p)'. Since normality was not observed for any of the BY4743-based mutant distributions, we then fitted them to several right-side skewed distributions, including lognormal, Gamma and Weibull. To this end, we used the goodness-of-fit function 'gofstat' of the *fitdistrplus* R language package (Delignette-Muller and Dutang 2015), implementing the Anderson-Darling test (Anderson and Darling 1952) for each of the samples shown in Table 1. Using the goodness-of-fit function 'gofstat' of the fitdistrplus R language package we also obtained the corresponding statistic values for the Kolmogorov-Smirnov and Cramér-von Mises tests (Table 1). Fitting wild type cell size distributions to more complex, three-parameter generalized gamma models only minimally improved the fit, but it increased complexity. As a result, the preferred model was the standard two-parameter gamma distribution, based on a lower value of the Bayesian Information Criterion (Schwarz 1978). To calculate the shape (α) and rate (β) parameters of the gamma-fitted distributions (see Table 1), we used the maximum-likelihood estimates approach implemented by the 'fitdistr' function of the *MASS* R language package. The same analysis was applied to the two BY4741 samples from the 'jorgensen' dataset shown in Table 1. To obtain the Anderson-Darling test p-values for gamma distribution fits for each strain, we used the 'gofTest' function of the *goft* R language package (González-Estrada and Villaseñor 2018). These p-values are shown in File S1, in the sheet columns marked as 'AD(p)'. For the 'jorgensen'

Table 1 Statistical parameters of cell size distributions of wild type strains

Strain	Carbon source	Anderson-Darling Statistic			Kolmogorov-Smirnov Statistic			Cramer-von Mises Statistic			Gamma parameters	
		Lognormal	Weibull	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull	Gamma	Shape (α)	Rate (β)
BY4743	Dextrose(2%)	3.161116	2.479761	0.769909	0.03981673	0.0200254	0.03695591	0.47005958	0.09900603	0.3672601	6.624628	0.060432
BY4743	Dextrose(2%)	2.481744	3.453834	0.907658	0.03451547	0.02340116	0.04168316	0.35931106	0.11427787	0.50807483	6.839326	0.0613
BY4743	Dextrose(2%)	2.350162	4.639409	1.397938	0.03299974	0.02592211	0.04547749	0.29286774	0.15548112	0.67758731	7.939692	0.062422
BY4743	Dextrose(2%)	9.379353	3.256118	2.080058	0.05563504	0.02812955	0.03526128	1.17880153	0.17324015	0.43802261	6.327387	0.051845
BY4743	Dextrose(0.5%)	2.400366	5.534772	1.885079	0.08902688	0.04867087	0.02641991	3.3903958	0.89513746	0.14380645	6.631854	0.056427
BY4743	Dextrose(0.5%)	2.681816	4.711678	1.737208	0.08033942	0.04140794	0.02836311	2.18084296	0.45362558	0.25145377	7.028829	0.059809
BY4743	Dextrose(0.5%)	2.554072	4.049485	1.352729	0.03971196	0.01749115	0.04803978	0.42735647	0.05621092	0.67291786	6.523879	0.056879
BY4743	Dextrose(0.5%)	2.238338	4.465921	1.296817	0.02285933	0.02726508	0.05816792	0.13284149	0.12489905	0.85713441	6.682971	0.05811
BY4743	Dextrose(0.5%)	1.657507	5.112165	1.088213	0.03371034	0.02031434	0.05064857	0.29001527	0.07296596	0.7676489	6.92371	0.059789
BY4743	Dextrose(0.5%)	2.257491	4.7403	1.40862	0.03967463	0.02448684	0.0559948	0.25676227	0.09533312	0.8832544	7.118115	0.060818
BY4743	Dextrose(0.1%)	2.093822	3.809769	0.68981	0.04676217	0.01914837	0.05218087	0.37783319	0.05085369	0.52856064	5.491663	0.051904
BY4743	Dextrose(0.1%)	3.495197	2.657985	1.090478	0.03157352	0.02077244	0.04550404	0.24346903	0.07980251	0.63674229	5.502986	0.050511
BY4743	Dextrose(0.1%)	0.998583	5.800609	0.817631	0.03476032	0.02287527	0.0470058	0.31940454	0.08175868	0.5341655	5.038751	0.050269
BY4743	Dextrose(0.1%)	1.429023	4.734133	0.725159	0.0423421	0.02390521	0.0375983	0.5340163	0.14950146	0.3783479	4.987775	0.049489
BY4743	Dextrose(0.05%)	3.090379	4.551848	0.58666	0.02497927	0.02174877	0.05687372	0.12924293	0.10453503	0.8573444	4.084743	0.043347
BY4743	Dextrose(0.05%)	1.194635	5.519737	0.903827	0.03355771	0.02326251	0.04499282	0.21631169	0.0932074	0.65946712	4.601468	0.047501
BY4743	Dextrose(0.05%)	2.240336	5.114651	0.608227	0.03175526	0.03285626	0.05452121	0.29699836	0.22149283	0.81935031	4.13688	0.046411
BY4743	Dextrose(0.05%)	1.685737	5.903186	0.693725	0.03123157	0.02902691	0.04582818	0.36362006	0.21041353	0.66669872	4.137698	0.046938
BY4743	Dextrose(0.05%)	2.300439	3.695214	0.388641	0.03437376	0.02568932	0.04441959	0.36181641	0.17975946	0.58983905	4.936219	0.052235
BY4743	Dextrose(0.05%)	1.850587	4.21797	0.692334	0.03038782	0.02735584	0.04853824	0.29098868	0.163558	0.6771316	4.86957	0.04992
BY4743	Galactose(2%)	4.48614	2.026274	1.063566	0.02452582	0.02348538	0.05047681	0.21260335	0.13107877	0.75881009	4.352128	0.049247
BY4743	Galactose(2%)	4.019618	1.957646	0.849565	0.02877741	0.02730421	0.04607656	0.30208734	0.17412821	0.69220803	4.34835	0.048053
BY4743	Galactose(2%)	2.866003	3.179497	0.85723	0.05826168	0.02860853	0.02928325	0.82117011	0.20351277	0.21822962	4.460477	0.050979
BY4743	Galactose(2%)	3.541828	2.324092	0.906612	0.04896814	0.0244765	0.02815314	0.68864672	0.142866	0.23356047	4.651173	0.053049
BY4743	Galactose(2%)	5.23192	1.70193	1.398188	0.04622281	0.03137464	0.03432644	0.51534675	0.14159861	0.37807561	4.561268	0.0503
BY4743	Galactose(2%)	4.898133	2.389382	1.858014	0.04906314	0.02468138	0.03192201	0.6025577	0.14846111	0.28277728	4.886206	0.053239
BY4743	Glycerol(2%)	4.023934	1.706263	0.484643	0.06243632	0.0343775	0.02854547	0.94857643	0.2648869	0.18764683	4.576071	0.046215
BY4743	Glycerol(2%)	4.912524	1.426551	0.417717	0.06189211	0.03498619	0.03483287	0.87430393	0.32292365	0.27798862	4.19207	0.046484
BY4743	Glycerol(2%)	3.537731	2.212626	0.166045	0.04390741	0.01948744	0.03145215	0.62067725	0.06717106	0.23822465	4.243508	0.04657
BY4741	Dextrose(2%)	8.884836	1.728113	1.143929	0.05150172	0.01837113	0.02720123	0.75401418	0.05740949	0.20737051	3.651767	0.074363
BY4741	Dextrose(2%)	9.115701	3.162732	1.503567	0.04546417	0.01249397	0.03272571	0.56067769	0.02099474	0.30962955	4.003634	0.083534

dataset we also used the 'ad.test' function of the *goftest* R language package, as follows: ad.test(strain, null = "dgamma", shape = 3.8277, rate = 0.078949). The shape and rate parameters were the average of the two wild type BY4741 samples in the 'jorgensen' dataset. We also used the same functions to obtain the test's statistic, which was used to identify the 49 genes that when deleted yield size distributions that deviate the most from a gamma pattern (shown in File S1/sheet 'Gamma_deviant_Genes').

All other R language functions and packages used to generate plots are described in the corresponding figures.

Data Availability

Table S1 lists the oligonucleotides used in this study. Strains (Table S2) and plasmids (Table S3) are available upon request. File S1 contains all the datasets, including distributions of size frequencies and associated p-values, used in this study. The code used to analyze the data is provided in File S2. We have uploaded all the supplementary material (which includes seven supplementary figures) to figshare.

Supplemental material available at Figshare: https://figshare.com/s/99db4b2cb8c4deedbbcd.

RESULTS

Rationale

The approach we followed in this study is shown schematically in Figure 1. How different values of a single parameter are distributed can be instructive for the underlying processes leading to its distribution pattern. Accurately fitting the measured variable to a univariate distribution is also necessary for its proper statistical analysis, when determining how removed a given observation (*e.g.*, a mutant) is from the most typical one (wild type) in a population. With regards to the size of individual organisms, the usual pattern is that deviations from the common type are not symmetrical. Instead, small individuals tend to be more frequent than large ones, leading to distributions which are positively skewed, with a right-side tail (Frank 2016). The sizes of bacterial and animal cells have been modeled on lognormal distributions (Hosoda *et al.* 2011). To our knowledge, although *S. cerevisiae* is a prime model system in studies of size control, how size is distributed in this organism has not been examined. Consequently, our objectives for this study were: First, determine how cell size is distributed in *S. cerevisiae* (Figure 2). Second, use the distribution model that fits the best to empirical data as a metric to identify mutants that deviate the most from that distribution (Figures 3, 4). Third, validate the outliers experimentally (Figure 5), and test the role of the corresponding biological processes in determining size distributions (Figure 6).

Cell size in S. cerevisiae fits a gamma distribution

We examined size frequency distributions from diploid cells cultured in different carbon sources, using published data from our laboratory (Soma *et al.* 2014). First, we looked at whether these distributions fit a Gaussian pattern. To test for normality, we employed the Shapiro-Wilk test (Shapiro and Wilk 1965), because it has the highest power compared to other tests (Razali and Wah 2011). In every one of the 29 wild type distributions from diploid cells we tested, the associated p-value was significantly lower than an alpha level of 0.01 (see Figure 2A, left box; the individual values are shown in File S1/sheet 'by4743_SW_AD_p'/column 'SW(p)'). Hence, the null hypothesis that these populations were normally distributed was rejected.

Given that cell size distributions were positively skewed with right-side tails, we fit the empirical data to non-Gaussian distributions that yield such patterns, such as lognormal, Weibull and gamma (Table 1).

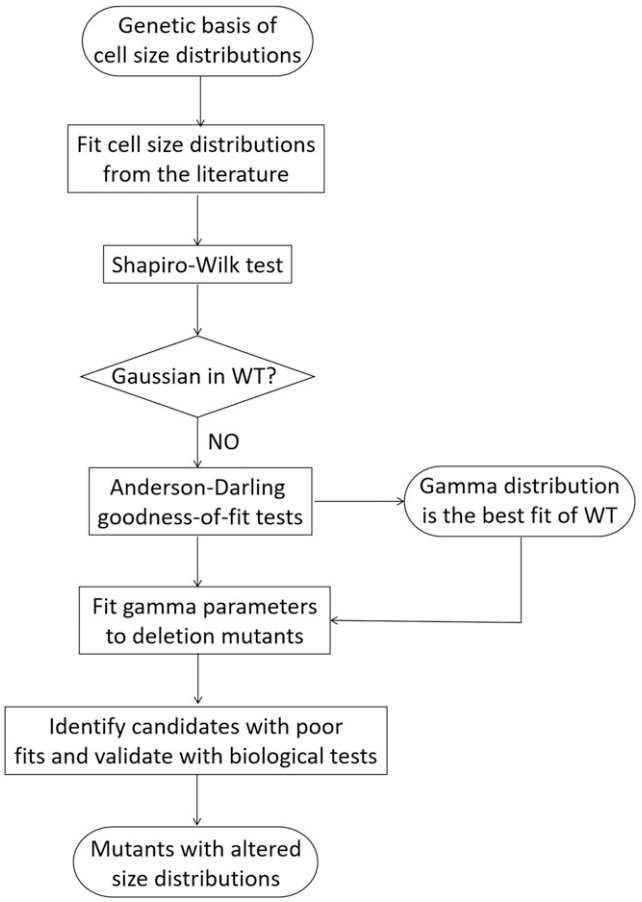

Figure 1 Flowchart of our approach to identify *S. cerevisiae* mutants with altered cell size distributions. See text for details.

To test the goodness of these fits, we primarily relied on the Anderson-Darling test (Anderson and Darling 1952), which is thought to be of higher power than other goodness-of-fit tests for non-normal distributions. In every case, the value of the test's statistic was the lowest for the gamma distribution (Table 1). To test further a lognormal distribution of these samples, we log-transformed these values and then examined if they were normally distributed, a prediction for values that are lognormally distributed. In no sample was this the case (not shown), arguing against lognormal distributions being the best fit for *S. cerevisiae* cell size values. Next, we calculated the Anderson-Darling associated p-value for a gamma distribution for each of the 29 diploid size distributions. In all but one sample, the p-value was higher than an alpha level of 0.01 (see Figure 2A, right box; the individual values are shown in File S1/sheet 'by4743_SW_AD_p'/column 'AD(p)'). Hence, the fits of all these samples are consistent with a gamma distribution. We looked at the empirical data of the one sample for which the p-value was significantly lower than the 0.01 cutoff (Figure S1). It appears that this distribution is irregular, with a shoulder on the right-side tail, perhaps explaining the poor fit (Figure S1). Nonetheless, even for this sample, the gamma distribution was the better fit, compared to lognormal or Weibull distributions (Table 1).

The suitability of a gamma distribution pattern to accurately describe *S. cerevisiae* cell size data were also evident when different (Gamma, lognormal, Weibull) theoretical fits were displayed on a histogram of continuous empirical data (Figure 2B; File S1, from the second sample in sheet 'by4743'). From the associated goodness-of-fit diagnostic plots

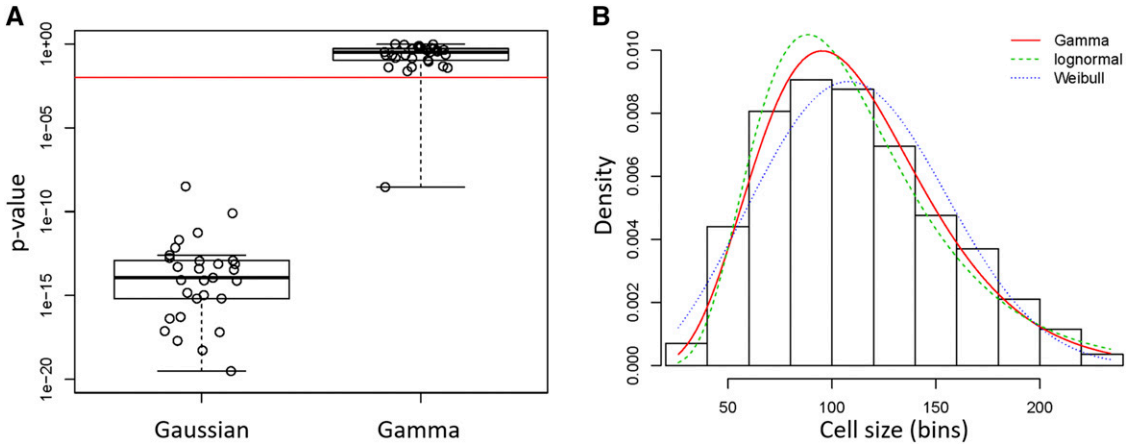

Figure 2 Cell size values of *S. cerevisiae* cells fit a Gamma distribution pattern. A, Box plots of the p-values associated with statistical tests of whether cell size values from (Soma *et al*. 2014) of wild type BY4743 cells (see Table 1) are distributed according to a normal (Gaussian), or Gamma distribution. The Shapiro-Wilk test was used to test for the Gaussian distribution, while the Anderson-Darling test was used to test for the Gamma distribution (see Materials and Methods). The red horizontal line indicates a significance level of $P = 0.01$. The density plot of the sole outlier that did not fit a Gamma distribution is shown in Figure S1. B, Histogram and theoretical densities for the indicated cell size distribution of *S. cerevisiae* cells. The distributions were fitted to continuous, empirical data depicted in the histogram from wild type diploid (BY4743 strain) cells, cultured in standard YPD medium (1% yeast extract, 2% peptone, 2% dextrose). On the y-axis are the density frequency values, while on the x-axis are the cell size bins, encompassing the cell size values shown in the corresponding spreadsheet associated with this plot (see File S1). The plots were generated with the 'denscomp' function of the *fitdistrplus* R language package. Additional goodness-of-fit plots associated with this graph are shown in Figure S2.

(Figure S2), the gamma distribution is a better fit than the related lognormal and Weibull distributions for the values in the middle of the distribution (Figure S2C). For the data in the left-side tail, lognormal and gamma are superior to Weibull, albeit Weibull performs better for the data in the right-side tail of the distribution (Figure S2B). Taken together, the sizes of *S. cerevisiae* cells best fit a gamma distribution. In the Discussion, we expand on this finding.

Next, we calculated the shape (α) and rate (β) parameters that describe gamma distributions for each of the above wild type distributions (Table 1; see Materials and Methods). Note that the samples we analyzed were from cells growing in different carbon sources (dextrose, galactose, glycerol) and, in the case of dextrose, at different concentrations (0.05–2%) of this preferred carbon source for the organism. In all cases, the best fit of the size data were a gamma distribution (Table 1), regardless of nutrient composition. We note that the shape parameter (α) was reduced from 6.3-7.9 in rich, replete medium (2% dextrose) to 4.1-4.9 in carbon-restricted medium (0.05% dextrose; see Table 1), as expected for the accompanying reduction in cell size in this medium (Soma *et al.* 2014).

Identifying mutants with cell size distributions that deviate the most from gamma

A significant outcome of our results that wild type cell size distributions from *S. cerevisiae* are best described with gamma distributions is that fitting an empirical distribution to a gamma pattern can be used as a 'metric' to identify mutants that deviate the most. To this end, we used the frequency data of cell distributions from (Jorgensen *et al.* 2002), which surveyed strains carrying single deletions in all non-essential genes in *S. cerevisiae* (Giaever *et al.* 2002). These strains were haploid, but in the same (S288c) genetic background as the strains we examined in Figure 2. They were also cultured in the same standard, dextrose-replete, YPD medium (see Materials and Methods). From the 5,052 size distributions in the (Jorgensen *et al.* 2002) dataset, not a single one fit to a normal, Gaussian distribution, based on the p-values from the Shapiro-Wilk test (Figure 3A, left box; the individual values are shown in File S1/sheet 'jorgensen_SW_AD_p'/column 'SW(p)'). The best fits of the two wild type (BY4741) samples in the (Jorgensen *et al.* 2002) dataset were also gamma distributions, compared to lognormal or Weibull (Table 1; the bottom two rows). We note that since the size of both haploid and diploid cells fit a gamma distribution pattern (Table 1), ploidy *per se* does not appear to change the pattern of cell size distributions.

Next, we calculated the Anderson-Darling associated p-value for a gamma distribution for each of the 5,052 samples. For about half of them (n = 2,527), the p-value was higher than an alpha level of 0.01 (see Figure 3A, right box; the individual values are shown in File S1/sheet 'by4743_SW_AD_p'/column 'AD(p)'). Since even small experimental irregularities in empirical distributions disturb their fit to theoretical densities (*e.g.*, see Figure S1), it is noteworthy that half of the frequencies could be adequately fitted. Furthermore, even for the samples whose p-values did not pass the alpha level of 0.01, it is clear for the overwhelming majority of them that their gamma-fitted p-values were orders of magnitude higher than their p-values for fits to Gaussian distributions (compare the right to the left plot in Figure 3A).

Next, to identify genes that may be necessary for the gamma distribution pattern of cell size in *S. cerevisiae*, we focused on the samples whose distributions deviated the most from a gamma distribution (*i.e.*, the ones with the lowest p-values shown in Figure 3A, right plot). The sample with the worst fit was from a strain lacking Rox3p, a subunit of the RNA polymerase II Mediator complex (Gustafsson *et al.* 1997). Not only cells from this mutant were large, as also identified by (Jorgensen *et al.* 2002), but their size distribution was negatively skewed, with a left-side tail from the main peak (see Figure 3B; the smaller peak to the extreme left of the distribution likely arose from small particulate debris from dead cells in the culture).

Given the severe departure from a gamma distribution for *rox3Δ* cells (Figure 3B), we next looked at the 50 samples with the worst fits. Including *rox3Δ*, these were from 49 deletion strains (one strain in this

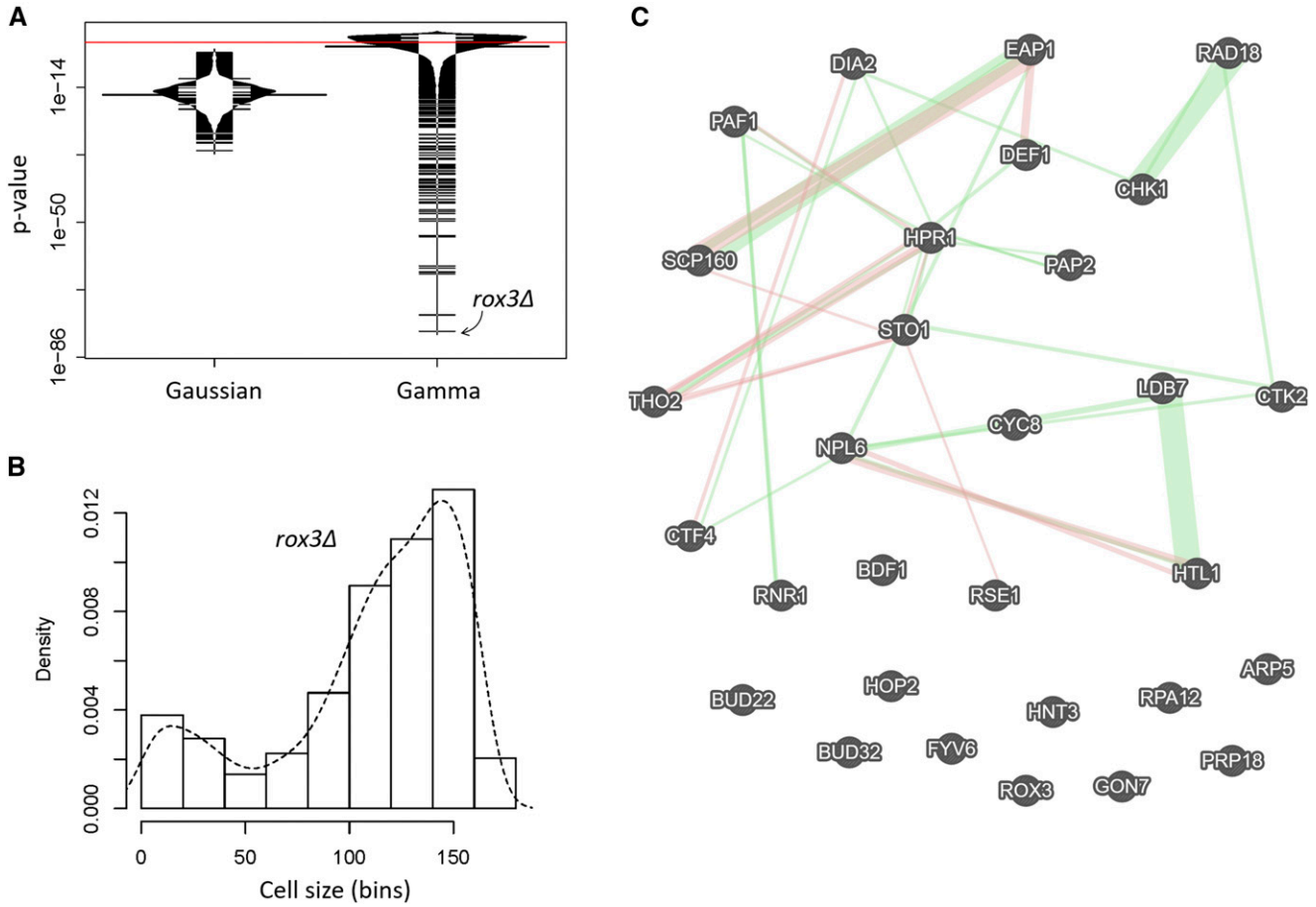

Figure 3 Deletion mutants of *S. cerevisiae* with severely altered cell size distributions. A, Bean plots of the p-values associated with statistical tests of whether cell size values of strains carrying single-gene deletions of all non-essential genes from (Jorgensen *et al.* 2002) are distributed according to a normal (Gaussian), or Gamma distribution. The statistical tests were performed and displayed as in Figure 2A. The red line indicates a significance level of $P = 0.01$. The most extreme outlier of the Gamma-fitted distributions (*rox3Δ*) is indicated with the arrow. B, Histogram and density for the indicated cell size distribution from *rox3Δ* cells, from (Jorgensen *et al.* 2002). On the y-axis are the density frequency values, while on the x-axis are the cell size bins, encompassing the cell size values shown in the corresponding spreadsheet associated with this plot (see File S1). C, Network of the interactions among the 30 genes that belonged in the gene ontology group 'nucleic acid metabolic process' [GO:0090304; $P = 0.008391$). The network was drawn on the GeneMANIA platform (Warde-Farley *et al.* 2010), with genetic interactions shown in green and physical ones in red.

set was measured twice by (Jorgensen *et al.* 2002)). The systematic names of these strains are shown in File S1/sheet 'Gamma_deviant_Genes'. Since experimental irregularities could be the reason for the extremely poor fits to a theoretical distribution (*e.g.*, see Figure S1), we relied on gene ontology enrichment as a functional, unbiased criterion to guide our identification of physiologically relevant mutants. Based on the YeastMine platform (Balakrishnan *et al.* 2012), 30 of the 49 genes belong to the ontology group 'nucleic acid metabolic process' (GO:0090304; $P = 0.008391$, after Holm-Bonferroni test correction). A smaller group of 16 genes (15 of which were also in the 'nucleic acid metabolic process' set) belonged to the ontology group 'cellular response to DNA damage stimulus' (GO:0006974; $P = 2.510022e\text{-}5$ after Holm-Bonferroni test correction). The full gene ontology output for the 30 genes of the 'nucleic acid metabolic process' is shown in File S1/sheet 'GO 0090304'.

Most of the 30 gene products of the 'nucleic acid metabolic process' have a network of previously reported genetic and physical interactions (Figure 3C), consistent with their involvement in common cellular processes. Upon closer inspection, most mutant strains in this group lack genes encoding gene products that regulate gene expression globally, especially transcription (*LDB7*, *HTL1*, *NPL6*, *ROX3*, *CYC8*, *PAF1*, *HPR1*, *BUD32*, *CTK2*, *GON7*, *RPA12*, *BDF1*, *ARP5*, *THO2*), but also splicing and RNA processing (*RSE1*, *BUD22*, *STO1*, *PRP18*, *PAP2*), or translation (*SCP160*, *EAP1*). There was some obvious coherence in this set of gene deletions, in that RSC complex submodule-encoding genes (*LDB7*, *HTL1*, *NPL6*) as well as two genes encoding members of the THO complex (*THO2*, *HPR1*) were identified. To examine if non-gamma cell size distributions were a phenotype common among deletions of each of the components of these large transcription-related complexes, we looked at their corresponding size distributions, for each of the components of the RSC, THO, PAF, and Mediator (MED) complexes interrogated by (Jorgensen *et al.* 2002). Interestingly, every gene deletion encoding a gene product that is part of the RSC complex had a cell size distribution that deviated significantly from gamma (Figure S3A). In contrast, only a subset of PAF or THO deletions had gamma-deviant size distributions (Figure S3B,C), while most of the MED deletions were similar to wild type, with the notable exception of the extreme distribution of *rox3Δ* cells (Figure S3D).

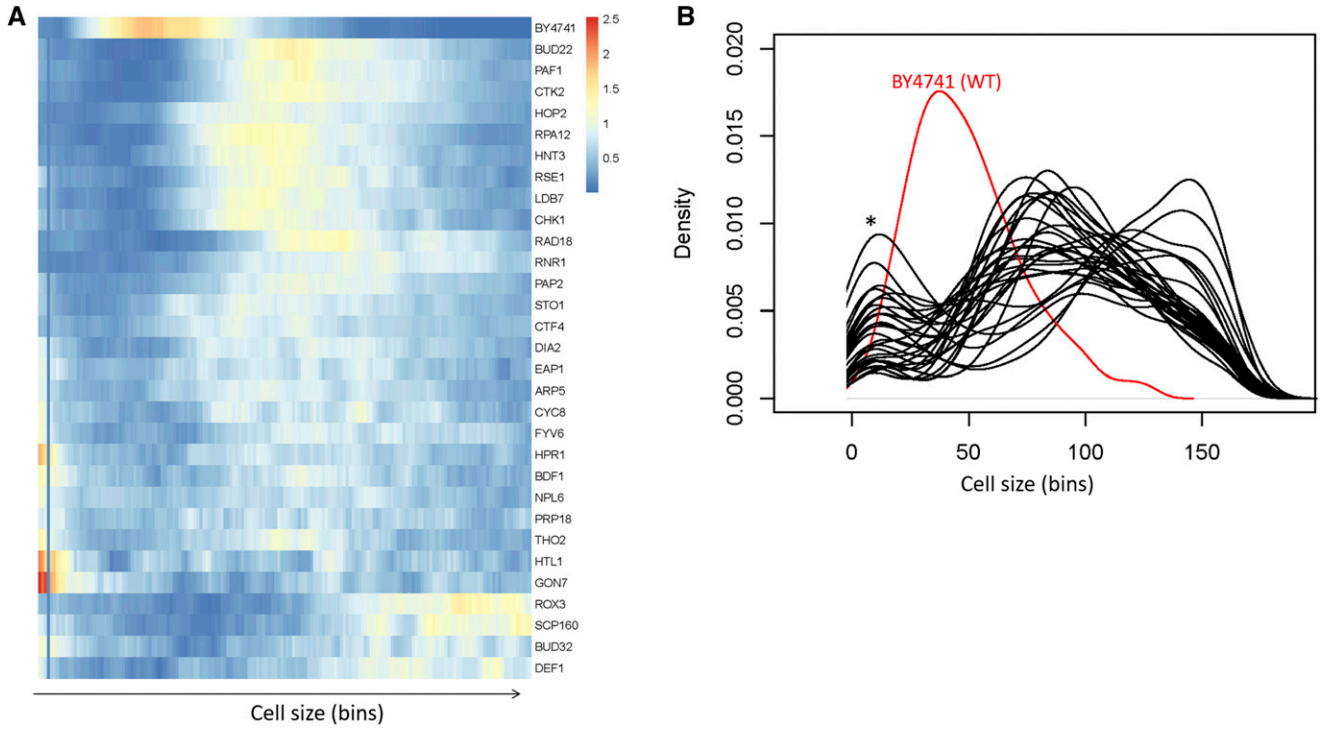

Figure 4 Cell size distributions from the gene deletions that belonged in the gene ontology group 'nucleic acid metabolic process' [GO:0090304]. A, Heatmap showing the clustering of cell size density frequencies from the outliers. The frequency values for each mutant are shown along the bins encompassing the cell size values shown in the corresponding spreadsheet associated with this plot (see File S1). The heatmap was generated with the *pheatmap* R language package. B, Density plots of cell distributions of the same deletion mutants shown in A. The deletion mutants are shown in black, while their wild type counterpart is shown in red. On the y-axis are the density frequency values, while on the x-axis are the corresponding cell size bins. The asterisk indicates a peak in the distributions that likely arose from small particle debris in the cultures.

Next, we looked at the empirical cell size distributions of the corresponding 30 deletion mutants. Every single one had a severely irregular size distribution (Figure 4). Several distributions resembled that of *rox3Δ* cells, with a pronounced negative skew (*scp160Δ*, *bud32Δ*, *def1Δ*; Figure 4A), while others were very irregular, even multimodal

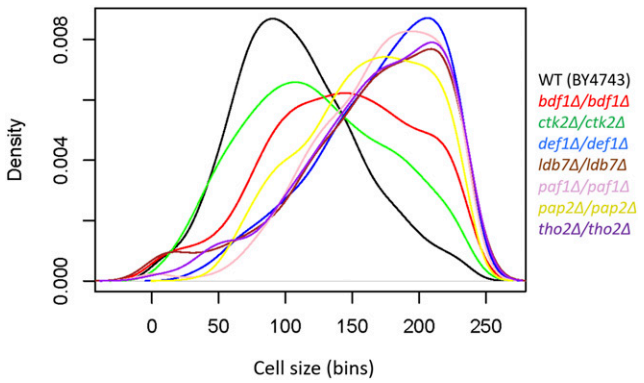

Figure 5 Density plots of cell size distributions from homozygous diploid strains. The cell size of the indicated deletion mutants was measured in standard YPD medium (1% yeast extract, 2% peptone, 2% dextrose), as described in Materials and Methods. Each strain was measured several times (see File S1; sheet 'figure5'), from which representative density plots are shown. On the y-axis are the density frequency values, while on the x-axis are the cell size bins, encompassing the cell size values shown in the corresponding spreadsheet associated with this plot (see File S1).

(*sto1Δ*, *ctf4Δ*, *dia2Δ*, *eap1Δ*, *arp5Δ*, *cyc8Δ*, *fyv6Δ*, *hpr1Δ*, *bdf1Δ*, *npl6Δ*, *prp18Δ*, *tho2Δ*, *htl1Δ*, *gon7Δ*; Figure 4A). Strikingly, all outliers were also abnormally large cell size mutants (Figure 4A,B). Most had already been identified as such by (Jorgensen *et al.* 2002) and others (Zhang *et al.* 2002; Manukyan *et al.* 2008), but some had not. These previously unidentified large cell size mutants lacked the following genes: *CHK1*, encoding a serine/threonine kinase and DNA damage checkpoint effector (Liu *et al.* 2000); *RAD18*, encoding an E3 ubiquitin ligase (Bailly *et al.* 1997); *RPA12*, encoding RNA polymerase I subunit A12.2 (Van Mullem *et al.* 2002); *RSE1*, encoding a splicing factor (Chen *et al.* 1998); *STO1*, encoding a large subunit of the nuclear mRNA cap-binding protein complex (Colot *et al.* 1996); *FYV6*, encoding a protein of unknown function (Wilson 2002); and *PAP2*, encoding a non-canonical poly(A) polymerase (Vanácová *et al.* 2005).

To further examine the connection between large cell size and deviation from a gamma distribution, we focused on the deletion strains whose median cell size values were in the top 5% (the criterion used by (Jorgensen *et al.* 2002), to define their large, *lge*, mutants). Not only were all 49 deletion strains whose distribution differed most significantly from a gamma pattern in this group, but these mutants were also some of the largest ones in the entire collection (Figure S4A,B). Deviations from gamma-distributed cell size values are strongly associated with a very large cell size ($P < 2.2E\text{-}16$; based on the Wilcoxon rank sum test with continuity correction, between the two groups shown in Figure S4A). The remaining mutants with large cell size (shown as 'Other' in Figure S4) were enriched for the gene ontology group "mitotic cell cycle" (GO:0000278; $P = 1.64E\text{-}07$), probably reflecting the expected increase in cell size due to a cell cycle block. Hence, it appears

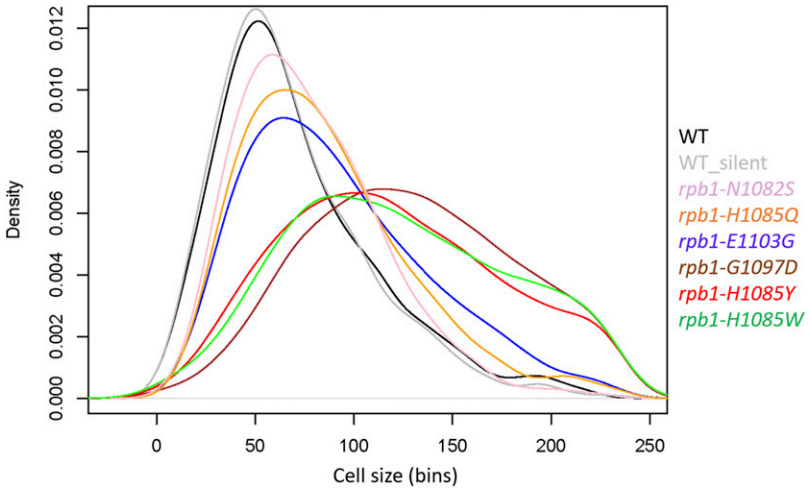

Figure 6 Density plots of cell size distributions from Pol II trigger-loop point mutants. The cell size of the indicated mutants was measured in standard YPD medium (1% yeast extract, 2% peptone, 2% dextrose), as described in Materials and Methods. Each strain was measured several times (see File S1; sheet 'figure6'), from which representative density plots are shown. On the y-axis are the density frequency values, while on the x-axis are the cell size bins, encompassing the cell size values shown in the corresponding spreadsheet associated with this plot (see File S1).

that while cell cycle blocks or presumed broad perturbation of gene expression can lead to a larger cell size, it is mostly presumed perturbations to global gene expression or associated processes that lead to deviations from gamma-distributed cell size values.

Next, we looked into the association between poor fitness and deviation from gamma-distributed cell size values. Of the 49 deletion strains whose distribution differed most significantly from a gamma pattern, 31 of the corresponding deletion mutants in a homozygous diploid background had also been reported to have reduced fitness compared to wild type cells in these culture conditions (Giaever *et al.* 2002). To answer if gamma-deviant mutants were also associated with an extreme reduction in fitness, we compared their fitness scores to the fitness scores of all other remaining 526 mutants with reduced fitness (Figure S4C). The 31 "Gamma_deviant" mutants had an overall significantly poorer fitness than the "Other" 526 mutants ((P = 6.163E-06; based on the Wilcoxon rank sum test with continuity correction). However, the difference in fitness was not as pronounced as was the difference in size (Figure S4C, *vs.* Figure S4A, respectively). Besides, more than a third (18 out of 49) of the mutants with cell size values that deviated the most from a gamma distribution pattern had a fitness level in these culture conditions that was indistinguishable from wild type, whereas all gamma-deviant mutants were also large size mutants. Hence, we conclude that although deviations from a gamma distribution of cell size values can be associated with poor fitness, the strength of that association is not nearly as great as that with large cell size.

Lastly, we also examined if mutants with extremely small mean cell size (*whi*; the 5% of mutants with the smallest median cell size, as defined by (Jorgensen *et al.* 2002)) are more likely to deviate from a gamma pattern of cell size distributions. Using the p-values of the Anderson-Darling test as a reference for gamma-deviant distributions, we found that there was no significant difference between *whi* mutants and strains that were not classified as size mutants by (Jorgensen *et al.* 2002) (P = 0.434, based on the Kruskal-Wallis one-way analysis of variance by ranks, followed by the post-hoc Nemenyi test). In contrast, the same analysis looking for deviations from gamma distributions indicated that large cell size mutants (*lge*) are different from strains that were not classified as size mutants by (Jorgensen *et al.* 2002) (P = 0.048, based on the Kruskal-Wallis one-way analysis of variance by ranks, followed by the post-hoc Nemenyi test). While there is a clear association between extreme large size and deviations from gamma distribution we documented (*e.g.*, see Figure S4), the size of mutants that are extremely small do not deviate from a gamma distribution. We illustrate the distribution of one of the most extreme *whi* mutants, *sfp1Δ* cells (Fig. S5), whose distribution fits a gamma distribution (P = 0.211, based on the Anderson-Darling test) as an example (Fig. S5). It is also worth noting that extreme variations in birth size (*sfp1Δ*: 11fL, WT: 22fL; *cln3Δ*: 34fL; calculated as described in (Truong *et al.* 2013; Soma *et al.* 2014)) or critical size (*sfp1Δ*: at 73% the size of WT (Jorgensen *et al.* 2004); *cln3Δ*: at >twofold the size of WT (Soma *et al.* 2014)) among these mutants are not necessarily associated with severe deviations from gamma distributions (Figure S5).

Validation of altered cell size distributions in selected homozygous diploid deletion mutants

In the outlier set of mutants we identified, we were intrigued by the preponderance of gene products connected to transcription. Hence, we decided to validate experimentally the size distributions of strains lacking *PAF1*, *CTK2*, *DEF1*, *BDF1*, *THO2*, *PAP2*, or *LDB7*. We used diploid strains carrying homozygous deletions of these genes, to minimize the effects of suppressors that may have been present in the haploid strains used by (Jorgensen *et al.* 2002). The cell size distributions for all these strains deviated from the gamma distribution of experiment-matched wild type cells (Figure 5; and File S1). These results strengthen the notion that perturbations in the control of gene expression may disrupt the distribution of cell sizes in a population.

Lastly, we also measured the DNA content of the strains lacking *PAF1*, *CTK2*, *DEF1*, *BDF1*, *THO2*, *PAP2*, or *LDB7*, to ask if their altered cell size distribution is associated with a particular cell cycle profile or ploidy abnormalities. From a genome-wide study, we had previously reported that loss of *PAF1*, *PAP2* or *LDB7* increased the percentage of cells that are in the G1 phase of the cell cycle (Hoose *et al.* 2012). Here, we confirmed this phenotype and found that an apparent G1 delay is also the case for cells lacking *CTK2* or *DEF1*, while the loss of *BDF1* or *THO2* does not lead to significant changes in the DNA content (Figure S6A). Hence, it appears that deviations from gamma distribution of cell size values are not obligately associated with a particular cell cycle profile, a conclusion reinforced by additional results we will describe later (see Figure S7).

Point mutations in the trigger loop of RNA polymerase II alter cell size

It has been proposed that global transcriptional output is tuned with cell size through some poorly characterized mechanism, perhaps by

increased RNA Pol II abundance or processivity, or altered mRNA stability in large cells (Zhurinsky *et al.* 2010; Marguerat and Bahler 2012; Mena *et al.* 2017; Padovan-Merhar *et al.* 2015). However, the cell size phenotypes of mutants that affect core RNA polymerase functions are not well-characterized, not least because only four (Rpb4,7,9,12) of the 12 subunits in the complex are non-essential in at least some genetic backgrounds (Myer and Young 1998; Giaever *et al.* 2002). Cells lacking any one of the non-essential RNA polymerase core subunits have been reported to be large (Jorgensen *et al.* 2002; Zhang *et al.* 2002), and usually display a G1 delay (Hoose *et al.* 2012).

To test the role of global transcription mechanism in cell size control, we examined a set of well-characterized point mutants carrying single amino acid substitutions in the largest Pol II subunit (Rpo21/Rpb1), which are either increased activity (biochemical and genetic "gain of function" GOF, E1103G, G1097D) or decreased activity (biochemical and genetic "loss of function" LOF: H1085Y, N1082S, H1085Q; genetic loss of function: H1085W)(Kaplan *et al.* 2012; Qiu *et al.* 2016; Kaplan *et al.* 2008; Braberg *et al.* 2013). We found that in all cases cell size was increased, correlating with the extent of catalytic rate alteration (Kaplan *et al.* 2008; Kaplan *et al.* 2012) and/or mutant growth rate defects (Kaplan *et al.* 2012; Malik *et al.* 2017; Qiu *et al.* 2016) (Figure 6). Interestingly, albeit the sizes of moderate mutants E1103G, N1082S, and H1085Q were moderately larger than wild type, the distribution pattern did not change (Figure 6). In contrast, the three severe alteration-of-function mutants (G1097D, H1085Y, H1085W) had a very large size, and their distribution deviated from the expected gamma distribution pattern (Figure 6). Pol II LOF and GOF mutants are distinguishable biochemically and genetically, though growth rates of these strains scale with the magnitudes of their biochemical defects and extent of their gene expression defects and genetic interactions. Similarly, Pol II mutant cell sizes, regardless of LOF or GOF status, correlate with their growth rates. We conclude that altering global transcription, severe or moderate, with gain- or loss-of-function Pol II mutations, increases cell size. Furthermore, severe alteration-of-function Pol II mutations abrogate the gamma distribution of cell size values.

Next, we measured the DNA content of these Pol II mutants. Similarly to the deletion mutants that we analyzed in Figure S6A, the severe alteration-of-function *rpb1* mutants (G1097D, H1085Y, H1085W) displayed a significant increase in the G1 DNA content (Figure S6B). Interestingly, both substitutions at position 1085 (Y or W) also displayed a cell cycle profile consistent with S-phase delay, because the peaks corresponding to un-replicated and replicated DNA were not separated (Figure S6B). The two moderate loss-of-function mutants (N1082S, H1085Q) had a modest increase of cells with G1 DNA content. The gain-of-function mutant (E1103G) did not display G1 or S-phase delay (Figure S6B). If anything, there was a slight increase of the G2/M DNA content in *rpb1-E1103G* cells, which along with their slightly larger cell size (Figure 6) and moderately slower proliferation rates (Kaplan *et al.* 2012; Malik *et al.* 2017), argues for a possible mitotic delay in this mutant. A potential mitotic delay could be consistent with Pol II increased activity mutants showing increased rates of minichromosome loss in chromosome segregation experiments (Braberg *et al.* 2013).

Altered cell cycle progression is not sufficient to alter the gamma distribution of cell size values

The altered cell cycle profiles of Pol II mutants and strains lacking genes involved in transcription (Figure S6) raised the question of whether abnormalities in cell cycle progression are the cause of the poor fits of cell size values to a gamma distribution in many of these mutants. To test this possibility, we examined the goodness-of-fit to a gamma distribution, using the Anderson-Darling associated p-values, for the following groups of deletion strains (Figure S7): Mutants displaying a 'High G1' DNA content, usually associated with a G1 delay (Hoose *et al.* 2012); mutants displaying a 'Low G1' DNA content, usually associated with a G2/M delay (Hoose *et al.* 2012); and mutants lacking genes of the DNA damage checkpoint biological process (Gene ontology group GO:0000077). These groups were compared to each other and all remaining strains analyzed by (Jorgensen *et al.* 2002). While some outliers in these groups had altered cell size distributions (*e.g.*, cells lacking the Chk1p checkpoint kinase in the GO:0000077 group), there was no statistically significant difference among these groups of mutants (based on the non-parametric Kruskal-Wallis test, $P = 0.4841$). Hence, cell cycle defects observed in several transcription mutants are not sufficient for explaining the significant deviations from the gamma distribution of cell size values in these strains. Instead, it is likely that a constellation of defects in gene expression or defects linked to transcriptional impact to the genome is the cause of cell size distribution derangement.

DISCUSSION

We discuss our results that cell size values in *S. cerevisiae* follow a gamma distribution and the role of global transcription in the control of cell size.

In biology, lognormal and gamma distributions have been proposed to describe tissue growth models (Mosimann and Campbell 1988). In both cases, the observed distributions are thought to arise from random fluctuations of many independent variables. Lognormal patterns reflect an aggregate multiplicative process generated from exponential patterns of growth (Koch 1966; Frank 2009). Despite random fluctuations, the growth of the overwhelming majority of cellular components is influenced proportionally, leading to lognormality (Koch 1966; Koch and Schaechter 1962). Similarly, gamma distributions represent the aggregate of many power-law and exponential processes (Frank 2009). With regards to cell size control, it is important to note that not only wild type cells but also size mutants, large or small, appear to maintain their size in a given environment. Such stationary size distributions, with their narrow range of coefficients of variation across populations with different mean size (Anderson *et al.* 1969), are accommodated by the properties of lognormal and gamma distributions (Hosoda *et al.* 2011; Dennis and Patil 1984; Kilian *et al.* 2005; Frank 2016).

Additionally, our data support the notion that global transcription mechanisms are necessary to balance expression with cell size (Figures 3-6), as postulated previously (Zhurinsky *et al.* 2010; Marguerat and Bahler 2012; Mena *et al.* 2017; Padovan-Merhar *et al.* 2015). The relationships between cell size and mRNA synthesis and decay rates are complex. It has been proposed for budding yeast that alterations to synthesis can correspondingly be buffered by changes in mRNA decay, and vice versa, enabling cells to maintain gene dosage in the face of global perturbations to gene expression (Sun *et al.* 2013; Sun *et al.* 2012; Haimovich *et al.* 2013). Cell size and genome replication are sources of potential perturbations to gene expression because an increase in cell volume will dilute concentrations of cellular factors unless compensated by global changes. Conversely, during replication, the dosage of the genome per cell doubles which can be countered by a concomitant increase in cell size. Prior work indicated that a subset of factors involved in mRNA turnover could also generate large cells. Here, we show that a number of additional mutants in genes with potential widespread roles in gene expression, including alterations to the Pol

II active site, lead to larger cells that can have altered distributions of sizes compared to wild type. A question that arises from this work is whether perturbed gene expression deregulates specific factors that control cell size homeostasis, or an increase in cell size is a consequence of globally defective gene expression, whereby alterations to global expression processes elicit buffering mechanisms that function through changes in cell volume. While an extreme size deviation is common among the gamma distribution-deviating strains, the actual distributions of these mutants appear to be of more than one class, suggesting either complex or distinct underlying mechanisms.

How could perturbations in global transcription alter the observed gamma distribution of cell size? In live cells, it appears that constitutive gene expression occurs stochastically in bursts at the single molecule level, with the burst magnitude and frequency leading to gamma-distributed, steady-state levels of the produced protein (Cai *et al.* 2006; Li and Xie 2011; Friedman *et al.* 2006). Changes in the burst magnitude and frequency of transcription events have been proposed to explain the scaling of mRNA counts with cell size (Padovan-Merhar *et al.* 2015). The reproductive property of the gamma distribution predicts that if the independent random variables themselves are gamma-distributed, then the aggregate of all these random variables will also be gamma-distributed (Johnson *et al.* 1994). Cell size is routinely viewed as a proxy for cell mass (Turner *et al.* 2012). Cell mass, in turn, is mostly determined by the accumulated macromolecules, especially proteins (Lange and Heijnen 2001). Hence, it is reasonable to speculate that global perturbations in the mechanics of gene expression that cause the steady-state levels of individual gene products to deviate from their gamma-distributed pattern could also perturb the aggregate gamma pattern of gene expression and its manifestation in cell size. Perturbation of global transcription through altered Pol II catalytic activity leads to changes in cell size, consistent with such a model. Conversely, a subset of specific gene expression perturbations may in aggregate lead to altered cell physiology that results in both extreme cell size and deviation from a gamma distribution.

We show here that some mutants are not just extremely large, their populations show distribution changes from wild type. Such changes in the distribution of sizes may occur for any number of reasons. Factors observed here may elicit cell size alterations directly through gene expression changes or indirectly through mRNA export defects or transcription-dependent DNA damage or recombination (TREX/THO complex members *hpr1*Δ, *tho2*Δ (Prado *et al.* 1997; Piruat and Aguilera 1998)), reduced ability to degrade stalled Pol II (*def1*Δ (Woudstra *et al.* 2002)), or widespread changes to transcription-dependent chromatin modifications or elongation control (*paf1*Δ (Van Oss *et al.* 2017)) or chromatin structure (RSC submodule components *ldb7*Δ, *htl1*Δ, *npl6*Δ (Wilson *et al.* 2006; Cairns *et al.* 1996)). Our DNA content data (see Figure S6) are not consistent with extreme chromosomal rearrangements in the mutants we examined. We also found that both haploid and diploid size distributions appear to fit better a gamma pattern (Table 1). However, we cannot exclude the possibility that possible aneuploidy or spontaneous diploidization in some of these mutants (especially RSC mutants (Sing *et al.* 2018), and *def1Δ* cells (Stepchenkova *et al.* 2018)) may contribute to deviations from gamma distributions. For Pol II mutants, sizes of cells correlate with several phenotypes: strain growth rates, biochemical and genetic defects, and the extent to which a specific mRNA's half-life was increased (Malik *et al.* 2017), raising intriguing questions about potential causative relationships between these phenotypes. Future experiments monitoring transcription events in single cells of different size could shed light in the relationship between global gene expression and size control.

ACKNOWLEDGMENTS

Methodology, C.D.K., M.P.; Formal Analysis, M.P.; Investigation, N.M. H.M.B., J.A., C.D.K., M.P.; Resources, C.D.K., M.P.; Data Curation, M.P.; Writing – Original Draft, M.P., Writing – Review and Editing, C.D.K., M.P.; Visualization, M.P.; Supervision, C.D.K., M.P., Funding Acquisition, C.D.K., M.P. We thank Marta Karas (Johns Hopkins University) for helpful suggestions for the R code we used. This work was supported by grants from the National Institutes of Health to M.P. (R01GM123139) and C.D.K. (R01GM097260), and the Welch Foundation to C.D.K. (A-1763).

LITERATURE CITED

Anderson, E. C., G. I. Bell, D. F. Petersen, and R. A. Tobey, 1969 Cell growth and division. IV. Determination of volume growth rate and division probability. Biophys. J. 9: 246–263. https://doi.org/10.1016/S0006-3495(69)86383-6

Anderson, T. W., and D. A. Darling, 1952 Asymptotic theory of certain" goodness of fit" criteria based on stochastic processes. Ann. Math. Stat. 23: 193–212. https://doi.org/10.1214/aoms/1177729437

Bailly, V., S. Lauder, S. Prakash, and L. Prakash, 1997 Yeast DNA repair proteins Rad6 and Rad18 form a heterodimer that has ubiquitin conjugating, DNA binding, and ATP hydrolytic activities. J. Biol. Chem. 272: 23360–23365. https://doi.org/10.1074/jbc.272.37.23360

Balakrishnan, R., J. Park, K. Karra, B. C. Hitz, G. Binkley *et al.*, 2012 YeastMine–an integrated data warehouse for Saccharomyces cerevisiae data as a multipurpose tool-kit. Database (Oxford) 2012: bar062. https://doi.org/10.1093/database/bar062

Björklund, M., M. Taipale, M. Varjosalo, J. Saharinen, J. Lahdenpera *et al.*, 2006 Identification of pathways regulating cell size and cell-cycle progression by RNAi. Nature 439: 1009–1013. https://doi.org/10.1038/nature04469

Bogomolnaya, L. M., R. Pathak, J. Guo, and M. Polymenis, 2006 Roles of the RAM signaling network in cell cycle progression in Saccharomyces cerevisiae. Curr. Genet. 49: 384–392. https://doi.org/10.1007/s00294-006-0069-y

Braberg, H., H. Jin, E. A. Moehle, Y. A. Chan, S. Wang *et al.*, 2013 From structure to systems: high-resolution, quantitative genetic analysis of RNA polymerase II. Cell 154: 775–788. https://doi.org/10.1016/j.cell.2013.07.033

Brachmann, C. B., A. Davies, G. J. Cost, E. Caputo, J. Li *et al.*, 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. https://doi.org/10.1002/(SICI)1097-0061(19980130)14:2<115::AID-YEA204>3.0.CO;2-2

Cai, L., N. Friedman, and X. S. Xie, 2006 Stochastic protein expression in individual cells at the single molecule level. Nature 440: 358–362. https://doi.org/10.1038/nature04599

Cairns, B. R., Y. Lorch, Y. Li, M. Zhang, L. Lacomis *et al.*, 1996 RSC, an essential, abundant chromatin-remodeling complex. Cell 87: 1249–1260. https://doi.org/10.1016/S0092-8674(00)81820-6

Chen, E. J., A. R. Frand, E. Chitouras, and C. A. Kaiser, 1998 A link between secretion and pre-mRNA processing defects in Saccharomyces cerevisiae and the identification of a novel splicing gene, RSE1. Mol. Cell. Biol. 18: 7139–7146. https://doi.org/10.1128/MCB.18.12.7139

Colot, H. V., F. Stutz, and M. Rosbash, 1996 The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. Genes Dev. 10: 1699–1708. https://doi.org/10.1101/gad.10.13.1699

Delignette-Muller, M. L., and C. Dutang, 2015 fitdistrplus: An R package for fitting distributions. J. Stat. Softw. 64: 1–34. https://doi.org/10.18637/jss.v064.i04

Dennis, B., and G. P. Patil, 1984 The gamma distribution and weighted multimodal gamma distributions as models of population abundance. Math. Biosci. 68: 187–212. https://doi.org/10.1016/0025-5564(84)90031-2

Frank, S. A., 2009 The common patterns of nature. J. Evol. Biol. 22: 1563–1585. https://doi.org/10.1111/j.1420-9101.2009.01775.x

Frank, S. A., 2016 The invariances of power law size distributions. F1000 Res. 5: 2074. https://doi.org/10.12688/f1000research.9452.1

Friedman, N., L. Cai, and X. S. Xie, 2006 Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys. Rev. Lett. 97: 168302. https://doi.org/10.1103/PhysRevLett.97.168302

Galitski, T., A. J. Saldanha, C. A. Styles, E. S. Lander, and G. R. Fink, 1999 Ploidy regulation of gene expression. Science 285: 251–254. https://doi.org/10.1126/science.285.5425.251

Giaever, G., A. M. Chu, L. Ni, C. Connelly, L. Riles *et al.*, 2002 Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387–391. https://doi.org/10.1038/nature00935

Ginzberg, M. B., R. Kafri, and M. Kirschner, 2015 Cell biology. On being the right (cell) size. Science 348: 1245075. https://doi.org/10.1126/science.1245075

González-Estrada, E., and J. Villaseñor, 2018 An R package for testing goodness of fit: goft. J. Stat. Comput. Simul. 88: 726–751. https://doi.org/10.1080/00949655.2017.1404604

Gregory, T. R., 2001 Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. Biol. Rev. Camb. Philos. Soc. 76: 65–101. https://doi.org/10.1017/S1464793100005595

Guo, J., B. A. Bryan, and M. Polymenis, 2004 Nutrient-specific effects in the coordination of cell growth with cell division in continuous cultures of Saccharomyces cerevisiae. Arch. Microbiol. 182: 326–330. https://doi.org/10.1007/s00203-004-0704-2

Gustafsson, C. M., L. C. Myers, Y. Li, M. J. Redd, M. Lui *et al.*, 1997 Identification of Rox3 as a component of mediator and RNA polymerase II holoenzyme. J. Biol. Chem. 272: 48–50. https://doi.org/10.1074/jbc.272.1.48

Haimovich, G., D. A. Medina, S. Z. Causse, M. Garber, G. Millan-Zambrano *et al.*, 2013 Gene expression is circular: factors for mRNA degradation also foster mRNA synthesis. Cell 153: 1000–1011. https://doi.org/10.1016/j.cell.2013.05.012

Hoose, S. A., J. A. Rawlings, M. M. Kelly, M. C. Leitch, Q. O. Ababneh *et al.*, 2012 A systematic analysis of cell cycle regulators in yeast reveals that most factors act independently of cell size to control initiation of division. PLoS Genet. 8: e1002590. https://doi.org/10.1371/journal.pgen.1002590

Hoose, S. A., J. T. Trinh, M. C. Leitch, M. M. Kelly, R. F. McCormick *et al.*, 2013 Saccharomyces cerevisiae deletion strains with complex DNA content profiles. FEMS Microbiol. Lett. 345: 72–76. https://doi.org/10.1111/1574-6968.12186

Hosoda, K., T. Matsuura, H. Suzuki, and T. Yomo, 2011 Origin of log-normal-like distributions with a common width in a growth and division process. Phys. Rev. E Stat. Nonlin. Soft Matter Phys. 83: 031118. https://doi.org/10.1103/PhysRevE.83.031118

Johnson, N. L., S. Kotz, and N. Balakrishnan, 1994 Gamma Distributions, pp. 337–414 in *Continuouns Univariate Distributions*. John Wiley & Sons, Inc., New York.

Jorgensen, P., J. L. Nishikawa, B. J. Breitkreutz, and M. Tyers, 2002 Systematic identification of pathways that couple cell growth and division in yeast. Science 297: 395–400. https://doi.org/10.1126/science.1070850

Jorgensen, P., I. Rupes, J. R. Sharom, L. Schneper, J. R. Broach *et al.*, 2004 A dynamic transcriptional network communicates growth potential to ribosome synthesis and critical cell size. Genes Dev. 18: 2491–2505. https://doi.org/10.1101/gad.1228804

Kaplan, C. D., H. Jin, I. L. Zhang, and A. Belyanin, 2012 Dissection of Pol II trigger loop function and Pol II activity-dependent control of start site selection in vivo. PLoS Genet. 8: e1002627. https://doi.org/10.1371/journal.pgen.1002627

Kaplan, C. D., K. M. Larsson, and R. D. Kornberg, 2008 The RNA polymerase II trigger loop functions in substrate selection and is directly targeted by alpha-amanitin. Mol. Cell 30: 547–556. https://doi.org/10.1016/j.molcel.2008.04.023

Kilian, H., H. Gruler, D. Bartkowiak, and D. Kaufmann, 2005 Stationary cell size distributions and mean protein chain length distributions of Archaea, Bacteria and Eukaryotes described with an increment model in terms of irreversible thermodynamics. Eur. Phys. J. E 17: 307–325. https://doi.org/10.1140/epje/i2004-10143-8

Koch, A. L., 1966 The logarithm in biology. 1. Mechanisms generating the log-normal distribution exactly. J. Theor. Biol. 12: 276–290. https://doi.org/10.1016/0022-5193(66)90119-6

Koch, A. L., and M. Schaechter, 1962 A model for statistics of the cell division process. J. Gen. Microbiol. 29: 435–454. https://doi.org/10.1099/00221287-29-3-435

Lange, H. C., and J. J. Heijnen, 2001 Statistical reconciliation of the elemental and molecular biomass composition of Saccharomyces cerevisiae. Biotechnol. Bioeng. 75: 334–344. https://doi.org/10.1002/bit.10054

Laughery, M. F., T. Hunter, A. Brown, J. Hoopes, T. Ostbye *et al.*, 2015 New vectors for simple and streamlined CRISPR-Cas9 genome editing in Saccharomyces cerevisiae. Yeast 32: 711–720. https://doi.org/10.1002/yea.3098

Li, G. W., and X. S. Xie, 2011 Central dogma at the single-molecule level in living cells. Nature 475: 308–315. https://doi.org/10.1038/nature10315

Liu, Y., G. Vidanes, Y. C. Lin, S. Mori, and W. Siede, 2000 Characterization of a Saccharomyces cerevisiae homologue of Schizosaccharomyces pombe Chk1 involved in DNA-damage-induced M-phase arrest. Mol. Gen. Genet. 262: 1132–1146. https://doi.org/10.1007/PL00008656

Malik, I., C. Qiu, T. Snavely, and C. D. Kaplan, 2017 Wide-ranging and unexpected consequences of altered Pol II catalytic activity in vivo. Nucleic Acids Res. 45: 4431–4451. https://doi.org/10.1093/nar/gkx037

Manukyan, A., J. Zhang, U. Thippeswamy, J. Yang, N. Zavala *et al.*, 2008 Ccr4 alters cell size in yeast by modulating the timing of CLN1 and CLN2 expression. Genetics 179: 345–357. https://doi.org/10.1534/genetics.108.086744

Marguerat, S., and J. Bahler, 2012 Coordinating genome expression with cell size. Trends Genet. 28: 560–565. https://doi.org/10.1016/j.tig.2012.07.003

Mena, A., D. A. Medina, J. Garcia-Martinez, V. Begley, A. Singh *et al.*, 2017 Asymmetric cell division requires specific mechanisms for adjusting global transcription. Nucleic Acids Res. 45: 12401–12412. https://doi.org/10.1093/nar/gkx974

Mosimann, J. E., and G. Campbell, 1988 Applications in Biology: Simple Growth Models, pp. 287–302 in *Lognormal distributions: theory and applications*, edited by E. Crow, L., Shimizu, K. Marcel, Dekker, Inc., New York.

Myer, V. E., and R. A. Young, 1998 RNA Polymerase II Holoenzymes and Subcomplexes. J. Biol. Chem. 273: 27757–27760. https://doi.org/10.1074/jbc.273.43.27757

Padovan-Merhar, O., G. P. Nair, A. G. Biaesch, A. Mayer, S. Scarfone *et al.*, 2015 Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. Mol. Cell 58: 339–352. https://doi.org/10.1016/j.molcel.2015.03.005

Piruat, J. I., and A. Aguilera, 1998 A novel yeast gene, THO2, is involved in RNA pol II transcription and provides new evidence for transcriptional elongation-associated recombination. EMBO J. 17: 4859–4872. https://doi.org/10.1093/emboj/17.16.4859

Prado, F., J. I. Piruat, and A. Aguilera, 1997 Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. EMBO J. 16: 2826–2835. https://doi.org/10.1093/emboj/16.10.2826

Pringle, J. R., and L. H. Hartwell, 1981 The *Saccharomyces cerevisiae* Cell Cycle, pp. 97–142 in *The Molecular and Cellular Biology of the Yeast Saccharomyces*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Qiu, C., O. C. Erinne, J. M. Dave, P. Cui, H. Jin *et al.*, 2016 High-Resolution Phenotypic Landscape of the RNA Polymerase II Trigger Loop. PLoS Genet. 12: e1006321. https://doi.org/10.1371/journal.pgen.1006321. Erratum-ibid. https://doi.org/10.1371/journal.pgen.1007158

Razali, N.M., and Y.B. Wah, 2011 Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. J. of stat. model.and ana. 2: 21–33.

Schwarz, G., 1978 Estimating the dimension of a model. Ann. Stat. 6: 461–464. https://doi.org/10.1214/aos/1176344136

Shapiro, S. S., and M. B. Wilk, 1965 An analysis of variance test for normality (complete samples). Biometrika 52: 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Si, F., D. Li, S. E. Cox, J. T. Sauls, O. Azizi *et al.*, 2017 Invariance of Initiation Mass and Predictability of Cell Size in Escherichia coli. Curr. Biol. 27: 1278–1287. https://doi.org/10.1016/j.cub.2017.03.022

Sing, T. L., M. P. Hung, S. Ohnuki, G. Suzuki, B. J. San Luis *et al.*, 2018 The budding yeast RSC complex maintains ploidy by promoting spindle pole body insertion. J. Cell Biol. 217: 2445–2462. https://doi.org/10.1083/jcb.201709009

Soma, S., K. Yang, M. I. Morales, and M. Polymenis, 2014 Multiple metabolic requirements for size homeostasis and initiation of division in Saccharomyces cerevisiae. Microb. Cell 1: 256–266. https://doi.org/10.15698/mic2014.08.160

Son, S., A. Tzur, Y. Weng, P. Jorgensen, J. Kim *et al.*, 2012 Direct observation of mammalian cell growth and size regulation. Nat. Methods 9: 910–912. https://doi.org/10.1038/nmeth.2133

Stepchenkova, E. I., A. A. Shiriaeva, and Y. I. Pavlov, 2018 Deletion of the DEF1 gene does not confer UV-immutability but frequently leads to self-diploidization in yeast Saccharomyces cerevisiae. DNA Repair (Amst.) 70: 49–54. https://doi.org/10.1016/j.dnarep.2018.08.026

Sun, M., B. Schwalb, N. Pirkl, K. C. Maier, A. Schenk *et al.*, 2013 Global analysis of eukaryotic mRNA degradation reveals Xrn1-dependent buffering of transcript levels. Mol. Cell 52: 52–62. https://doi.org/10.1016/j.molcel.2013.09.010

Sun, M., B. Schwalb, D. Schulz, N. Pirkl, S. Etzold *et al.*, 2012 Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. Genome Res. 22: 1350–1359. https://doi.org/10.1101/gr.130161.111

Truong, S. K., R. F. McCormick, and M. Polymenis, 2013 Genetic Determinants of Cell Size at Birth and Their Impact on Cell Cycle Progression in Saccharomyces cerevisiae. G3 (Bethesda) 3: 1525–1530. https://doi.org/10.1534/g3.113.007062

Turner, J. J., J. C. Ewald, and J. M. Skotheim, 2012 Cell size control in yeast. Curr. Biol. 22: R350–R359. https://doi.org/10.1016/j.cub.2012.02.041

Tzur, A., R. Kafri, V. S. LeBleu, G. Lahav, and M. W. Kirschner, 2009 Cell growth and size homeostasis in proliferating animal cells. Science 325: 167–171. https://doi.org/10.1126/science.1174294

Van Mullem, V., E. Landrieux, J. Vandenhaute, and P. Thuriaux, 2002 Rpa12p, a conserved RNA polymerase I subunit with two functional domains. Mol. Microbiol. 43: 1105–1113. https://doi.org/10.1046/j.1365-2958.2002.02824.x

Van Oss, S. B., C. E. Cucinotta, and K. M. Arndt, 2017 Emerging Insights into the Roles of the Paf1 Complex in Gene Regulation. Trends Biochem. Sci. 42: 788–798. https://doi.org/10.1016/j.tibs.2017.08.003

Vanácová, S., J. Wolf, G. Martin, D. Blank, S. Dettwiler *et al.*, 2005 A new yeast poly(A) polymerase complex involved in RNA quality control. PLoS Biol. 3: e189. https://doi.org/10.1371/journal.pbio.0030189

Vargas-Garcia, C. A., K. R. Ghusinga, and A. Singh, 2018 Cell size control and gene expression homeostasis in single-cells. Curr. Opin. Syst. Biol. 8: 109–116. https://doi.org/10.1016/j.coisb.2018.01.002

Warde-Farley, D., S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi *et al.*, 2010 The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res 38: W214–W220. https://doi.org/10.1093/nar/gkq537

Westfall, C. S., and P. A. Levin, 2017 Bacterial Cell Size: Multifactorial and Multifaceted. Annu. Rev. Microbiol. 71: 499–517. https://doi.org/10.1146/annurev-micro-090816-093803

Willis, L., and K. C. Huang, 2017 Sizing up the bacterial cell cycle. Nat. Rev. Microbiol. 15: 606–620. https://doi.org/10.1038/nrmicro.2017.79

Wilson, B., H. Erdjument-Bromage, P. Tempst, and B. R. Cairns, 2006 The RSC chromatin remodeling complex bears an essential fungal-specific protein module with broad functional roles. Genetics 172: 795–809. https://doi.org/10.1534/genetics.105.047589

Wilson, T. E., 2002 A genomics-based screen for yeast mutants with an altered recombination/end-joining repair ratio. Genetics 162: 677–688.

Woudstra, E. C., C. Gilbert, J. Fellows, L. Jansen, J. Brouwer *et al.*, 2002 A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. Nature 415: 929–933. https://doi.org/10.1038/415929a

Wu, C. Y., P. A. Rolfe, D. K. Gifford, and G. R. Fink, 2010 Control of transcription by cell size. PLoS Biol. 8: e1000523. https://doi.org/10.1371/journal.pbio.1000523

Zhang, J., C. Schneider, L. Ottmers, R. Rodriguez, A. Day *et al.*, 2002 Genomic scale mutant hunt identifies cell size homeostasis genes in S. cerevisiae. Curr. Biol. 12: 1992–2001. https://doi.org/10.1016/S0960-9822(02)01305-2

Zhurinsky, J., K. Leonhard, S. Watt, S. Marguerat, J. Bahler *et al.*, 2010 A coordinated global control over cellular transcription. Curr. Biol. 20: 2010–2015. https://doi.org/10.1016/j.cub.2010.10.002

Communicating editor: G. Brown