Measurement and modeling of transcriptional noise in the cell cycle regulatory network

David A Ball1, Neil R Adames1, Nadine Reischmann1, Debashis Barik2, Christopher T Franck3, John J Tyson1,2, and Jean Peccoud1,4,*

1Virginia Bioinformatics Institute; Virginia Tech; Blacksburg, VA USA; 2Department of Biological Sciences; Virginia Tech; Blacksburg, VA USA; 3Department of Statistics; Virginia Tech; Blacksburg, VA USA; 4ICTAS Center for Systems Biology of Engineered Tissues; Virginia Tech; Blacksburg, VA USA

Keywords: cell cycle, stochastic modeling, gene expression noise, Saccharomyces cerevisiae, single mRNA FISH

Abstract: Fifty years of genetic and molecular experiments have revealed a wealth of molecular interactions involved in the control of cell division. In light of the complexity of this control system, mathematical modeling has proved useful in analyzing biochemical hypotheses that can be tested experimentally. Stochastic modeling has been especially useful in understanding the intrinsic variability of cell cycle events, but stochastic modeling has been hampered by a lack of reliable data on the absolute numbers of mRNA molecules per cell for cell cycle control genes. To fill this void, we used fluorescence in situ hybridization (FISH) to collect single molecule mRNA data for 16 cell cycle regulators in budding yeast, Saccharomyces cerevisiae. From statistical distributions of single-cell mRNA counts, we are able to extract the periodicity, timing, and magnitude of transcript abundance during the cell cycle. We used these parameters to improve a stochastic model of the cell cycle to better reflect the variability of molecular and phenotypic data on cell cycle progression in budding yeast.

Introduction

The eukaryotic cell division cycle is a complex process in which cells grow, replicate their DNA, and then partition their replicated chromosomes and organelles between 2 new daughter cells. These events are coordinated by a family of cyclin-dependent protein kinases (CDKs), whose activities are controlled by a complex network of interacting proteins. The budding yeast cell cycle is regulated by a single, constitutively expressed CDK (Cdc28), whose activity and substrate specificity change in a consistent manner during each stage of the cell division cycle.1-5 The activity of Cdc28 in budding yeast is regulated by periodic changes in the availability of protein cofactors, called cyclins, each of which is produced and degraded during a specific phase of the cell cycle. Kinases, phosphatases, and CDK-inhibitory proteins (CKIs) modify the activities of these cyclin–CDK complexes. The abundances of the cyclins and many other regulatory proteins change during progression through the cell cycle because cyclin–CDK activity regulates their periodic synthesis and degradation.

Like any gene or protein regulatory network, the cell cycle is a noisy process; in genetically identical populations, there is significant cell-to-cell variation even in a uniform environment.6 The sources of non-genetic variation in gene expression are a combination of intrinsic noise (the probabilistic nature of transcription and translation events resulting from the low abundance of transcription factors, their target genes, and transcripts) and extrinsic noise (such as inequalities in the partitioning of cellular components during division).7 The abundance of a transcript or protein in a cell at any given time depends on the relative rates of synthesis and degradation and can be adequately described as a stochastic birth–death process with exponentially distributed waiting times.8

Mathematical modeling is now an integral component of cell cycle research. Computational biologists have developed models that are consistent with diverse sets of experimental data.8-17 Models have helped predict previously unknown interactions that could be tested experimentally.1,13,14,18-24 The first generation of models were “deterministic”, relying on nonlinear ordinary differential equations (ODEs) to capture the viability (and other phenotypic characteristics) of dozens of mutants of the cell cycle control system.25 However, other aspects of cell cycle progression, such as variability in interdivision time or cell size at division, cannot be explained by deterministic models.12 To
account for these characteristics, along with the partial viability of some mutants, requires “stochastic” models of cell cycle progression.\(^8,14\)

The first stochastic models of the cell cycle were derived from deterministic models by introducing extrinsic noise associated with partitioning of cellular components during cell division,\(^26\) by introducing Gaussian white noise into reaction rate equations,\(^27\) or by modifying Boolean deterministic models.\(^28-32\) However, experiments in yeast cells with increasing gene dosage suggest that most noise in the cell cycle is a consequence of small copy number per cell of some molecular regulators, especially in G₂ phase of the cell cycle.\(^14\) The early phenomenological models of noise in the cell cycle are not adequate to analyze the role of intrinsic noise in cell cycle fluctuations, which requires explicitly modeling molecular interactions at their source. Initially, stochastic models focused on protein activity levels and protein–protein interactions, since they are the products of gene expression and the agents of cellular activity.\(^26,27,29,31,33\) By combining time-lapse microscopy, microfluidics, and image processing, it has been possible to observe in single cells the trajectories through the cell cycle of regulator proteins tagged with fluorescent protein domains.\(^15,23,33,34\) However, cell cycle control also involves a complex system of transcriptional regulation. In fact, very little gene expression noise can be explained solely in terms of protein synthesis and degradation. A dominant proportion of intrinsic noise is the low abundance of mRNA.\(^11\)

Therefore, to develop better predictive stochastic models of cell cycle regulation, one must account quantitatively for the fluctuations of mRNA abundances in individual cells.

Accurate simulation of mRNA abundances, which are likely to be on the order of 5–10 molecules per yeast cell, requires the use of the Gillespie stochastic simulation algorithm (SSA), which is based on the assumption that the fluctuating reactions are pseudo-elementary, i.e., that they are adequately described by mass-action kinetics.\(^35\) Because most deterministic models of cell cycle regulation rely heavily on complex rate laws (Michaelis–Menten kinetics, Hill functions, zero-order ultrasensitivity), they are not easily converted into mass-action-type models, suitable for simulation by Gillespie SSA.\(^36\) To investigate the roles of intrinsic and extrinsic noise in cell cycle variability, Kar et al.\(^11\) performed this conversion on a simple deterministic model of the budding yeast cell cycle. Using Gillespie SSA, they showed that intrinsic noise dominates over extrinsic noise in determining the variability of cell size and cell cycle timing in budding yeast. However, to obtain the correct level of noise, they had to use mRNA abundances 4 times greater than the values reported on the basis of microarray studies.\(^11\) Also, they had to use mRNA half-lives (0.2–5 min) considerably shorter than the values reported on the basis of microarray studies.\(^11\)

A year later, Barik et al.\(^10\) presented a more realistic model of budding yeast growth and division, based on the idea that multisite phosphorylation of key cell cycle regulators is the primary source of nonlinearity in the control system. Even though the overall effect of multiple phosphorylations on protein behavior is nonlinear (i.e., switch-like), each phosphorylation event can be described by a mass-action rate law. Hence, the model can be simulated deterministically, by solving the nonlinear ODEs, or stochastically, by applying Gillespie SSA. The deterministic version of the model accurately reflected G₁/S size control and timing. The stochastic version accurately reproduced variability in cell cycle progression, provided that mRNA half-lives (1–5 min) are shorter than expected,\(^37\) and mRNA levels (5–10 molecules per cell) are larger than expected.\(^38\) The mRNA abundances in Barik’s model\(^10\) are in line with the values reported by Zenklusen and Singer\(^39\) for non-cell cycle-related genes, based on FISH measurements. Because FISH data on cell cycle-related genes was lacking, it was impossible for Barik et al.\(^10\) to calibrate their model accurately to the expected number and variability of regulatory mRNA molecules per cell.

In pursuit of more comprehensive and explanatory stochastic models of the cell cycle, we have used single mRNA FISH to measure the statistical distributions of transcripts for 16 cell cycle regulators in budding yeast.\(^39\) We found that cell cycle regulators known to be constitutively expressed showed transcript abundance patterns consistent with a Poisson distribution. mRNA distributions of periodically expressed genes were best modeled as two-component Poisson distributions, reflecting their regulated high and low states of transcription.\(^40\) We were also able to deduce the length of time in the cell cycle during which gene expression is elevated and the amplitude of the oscillations in gene expression.

We then used this data to refine the multisite phosphorylation model of Barik et al.\(^10\) To match our observed distributions, we modified the model to allow complete dissociation of Whi5 from SBF and of Net1 from Cdc14, and we adjusted transcription and translation rates to produce NET1 and WHI5 mRNA distributions that better matched our observations. Because the only periodically expressed transcript in this model refers to the G₁/S cyclins (Cln1, Cln2, Clb5, and Clb6, called “ClbS” in the model), we were able to reproduce the two-component Poisson distributions of CLN1 and CLN2 mRNAs in our data set, but not of other periodic transcripts.

Results

Statistical distributions of mRNA abundance

Individual transcripts from 16 cell cycle genes were counted in single cells using single molecule FISH.\(^39\) To this end, we used commercially available strains of *Saccharomyces cerevisiae* for which a GFP-coding sequence was inserted as an in-frame C-terminal fusion to a cell cycle control gene. A mixture of 5 fluorescently labeled oligonucleotides, each containing 5 internal fluorescent dye molecules, was hybridized to the GFP portion of these fused transcripts (Fig. 1A). Spots were visualized and counted with a spot-detection algorithm (Fig. 1B and C).\(^41\) Cells lacking any GFP fusion elements were used as a negative control and showed virtually no fluorescent spots (Table 1), indicating a high specificity of the probes to the GFP transcripts. As a positive control, *MDN1* transcripts were counted using probes against the coding region of *MDN1* in a *MDN1–GFP* strain, and again using the same GFP probe set used for the other strains. Data from these 2 probe sets exhibited similar mRNA distributions,
and MDN1 mRNA abundances were consistent with previous reports (Table 1).42

We then performed single mRNA FISH on asynchronous cultures of the 16 strains with GFP fused to a particular gene involved in cell cycle regulation. The average numbers of transcripts for each gene from one biological replicate are reported in Table 1. Similar results were obtained in 2 other biological replicates (Table S1). Our results, as well as those from other single mRNA FISH studies,42,43 consistently showed mRNA levels roughly 4-fold higher than large-scale transcriptome studies (Table S2). This discrepancy is probably due to the method of normalization used in transcriptome studies to infer average transcript abundance per cell (for discussion of this issue, see Supplementary Material). In fact, all transcriptome studies are in good agreement with our mRNA FISH data when they are normalized to 60 000 transcripts/cell (Table S2).

Test of transcriptional regulation
The genes studied here are involved in the M/G1 and G1/S transitions. Half of them are known to be constitutively expressed, and the other half are transcriptionally regulated (Table 1; Fig. S1). Furthermore, the periodic genes represent most of the known expression mechanisms for cell cycle genes, with the expression of at least one gene peaking at every stage of the cell cycle (Fig. S1).

We used maximum likelihood estimation to fit the mRNA data to a Poisson distribution. We found that BCK2, BUB2, CDC15, LTE1, MAD2, MCM1, and NET1 mRNAs all fit well to this distribution, confirming that these genes are constitutively transcribed (Fig. 1D; Table 1; Table S1; Fig. S2).44-48

Transcript data for 6 of the genes (CDC6, CDC20, CLB2, CLN1, CLN2, and SIC1) did not fit a Poisson distribution (Table 1; Table S1). Each of these distributions has a clearly visible shoulder or tail in the histogram (Fig. 1D; Fig. S2), consistent

Figure 1. Summary of single mRNA FISH method and mRNA distributions. (A) Schematic of how the FISH probes hybridize to target mRNAs. (B) Example image showing individual mRNA molecules. Image is a maximum intensity projection of a z-series with merged phase-contrast and fluorescence channels. (C) Detected single transcripts. The image in the left panel was processed as described in “Materials and Methods” to detect spots of single mRNAs. Scale bar = 5 µm. (D) Experimental distributions of mRNA for 16 cell cycle genes. Gray bars show the experimental histograms; red lines are the fit to a single component Poisson, and blue lines are the fit to a two-component Poisson. Data are from the biological replicate shown in Table 1. Similar results were obtained in 2 other biological replicates (Table S1).
with the hypothesis that these genes switch between 2 states of high and low transcription. Not surprisingly, all of these genes are annotated as periodically expressed (Table 1). We found that the distributions of mRNA abundance of these 6 genes are described well by two-component Poisson distributions, confirming that they have 2 expression states in the population (Fig. 1; Table 1; Table S1; Fig. S2).

Similar results were obtained in 3 biological replicates (Table 1; Table S1). There was generally good agreement between previous gene annotations and the outcomes of statistical tests to

| Gene                  | Periodica | FISH mean ± 95% CI  | G test P value | \( \phi^2 \) effect size | \( \beta \) ± 95% CI | \( \lambda_1 \) ± 95% CI | \( \lambda_2 \) ± 95% CI | G test P value | \( \phi^2 \) effect size |
|-----------------------|-----------|---------------------|----------------|--------------------------|------------------|-----------------|-----------------|------------------|--------------------------|
| BY4733 (WT no GFP)    | No        | 0.08 ± 0.02         |                |                          |                  |                 |                 |                  |                          |
| MDN1 (MDN1 probes)   | No        | 9.37 ± 0.23         | 0.02           | 0.04                     |                  |                 |                 |                  |                          |
| MDN1 (GFP probes)    | No        | 9.20 ± 0.27         | 4.47E-05       | 0.01                     |                  |                 |                 |                  |                          |
| BCK2                  | No        | 4.16 ± 0.21         | 0.78           | 0.01                     |                  |                 |                 |                  |                          |
| BUB2                  | No        | 3.04 ± 0.13         | 9.87E-04       | 0.04                     |                  |                 |                 |                  |                          |
| CDC15                 | No        | 3.24 ± 0.13         | 0.18           | 0.01                     |                  |                 |                 |                  |                          |
| LIE1                  | No        | 2.73 ± 0.09         | 0.00           | 0.10                     |                  |                 |                 |                  |                          |
| MAD2                  | No        | 2.50 ± 0.10         | 0.69           | 4.22E-03                 |                  |                 |                 |                  |                          |
| MCM1                  | No        | 6.06 ± 0.22         | 4.10E-03       | 0.05                     |                  |                 |                 |                  |                          |
| NET1                  | No        | 6.20 ± 0.25         | 0.60           | 0.02                     |                  |                 |                 |                  |                          |
| CDC6                  | Yes       | 4.07 ± 0.20         | 0.00           | 0.99                     | 0.67 ± 0.07      | 2.12 ± 0.26     | 8.08 ± 0.72     | 0.03             | 0.05                     |
| CDC20                 | Yes       | 4.16 ± 0.15         | 0.00           | 0.64                     | 0.63 ± 0.08      | 2.47 ± 0.29     | 7.08 ± 0.61     | 0.09             | 0.02                     |
| CLB2                  | Yes       | 3.01 ± 0.10         | 0.00           | 0.74                     | 0.65 ± 0.07      | 1.55 ± 0.20     | 5.69 ± 0.49     | 1.20E-10        | 0.07                     |
| CLN1                  | Yes       | 3.03 ± 0.23         | 9.99E-16       | 0.47                     | 0.85 ± 0.07      | 2.03 ± 0.28     | 8.56 ± 1.53     | 2.02E-03        | 0.09                     |
| CLN2                  | Yes       | 4.45 ± 0.13         | 0.00           | 4.49                     | 0.77 ± 0.03      | 2.09 ± 0.12     | 12.4 ± 0.55     | 5.92E-06        | 0.06                     |
| ESP1                  | Yes       | 2.89 ± 0.11         | 7.23E-07       | 0.05                     | 0.62 ± 0.26      | 2.15 ± 0.42     | 4.09 ± 0.75     | 0.11             | 0.01                     |
| SIC1                  | Yes       | 3.44 ± 0.12         | 0.00           | 0.62                     | 0.84 ± 0.03      | 2.47 ± 0.15     | 8.68 ± 0.79     | 4.46E-03        | 0.03                     |
| TEM1                  | Yes       | 2.83 ± 0.12         | 0.10           | 0.02                     | 0.53 ± 0.68      | 2.17 ± 0.89     | 3.57 ± 1.09     | 0.75             | 3.19E-03                 |
| WHIS                  | Yes       | 3.45 ± 0.13         | 1.11E-16       | 0.15                     | 0.52 ± 0.16      | 2.12 ± 0.41     | 4.90 ± 0.57     | 0.99             | 9.89E-04                 |

Expression of genes are ranked as cell cycle periodic by Cyclebase69 and by ref. 59; Transcripts/cell. The biological replicate from Figure 1 is shown. Similar results were obtained in other replicates (Table S1); Measured mRNA distribution fits to a Poisson distribution. G test of goodness of fit to a single Poisson distribution with \( \lambda = \text{FISH Mean} \). \( \phi^2 \) effect size. By convention, effect sizes of \( \phi^2 \leq 0.09 \) are considered small (green), \( 0.09 < \phi^2 < 0.25 \) are considered medium (orange), and \( \phi^2 \geq 0.25 \) are considered large (red). Null hypothesis for the G test P value is that the empirical data have a single Poisson distribution with mean \( \lambda \); Measured mRNA distribution fits to a two-component Poisson distribution. Same data set as for the single Poisson distribution. \( \beta \), fraction of the population with low transcript levels. \( \lambda_1 \), mean transcripts/cell for the fraction of the population with low transcript levels. \( \lambda_2 \), mean transcripts/cell for the fraction of the population with high transcript levels. G test P values and \( \phi^2 \) as for single-Poisson fit, but the null hypothesis is that the data have a two-component Poisson distribution with means \( \lambda_1 \) and \( \lambda_2 \).
differentiate constitutively expressed and transcriptionally regulated genes. In fact, there was an excellent correlation between whether a gene showed mRNA periodicity and whether the corresponding protein oscillates (Table S3). In this context, it is notable that although ESP1, TEM1, and WHI5 are ranked as periodic, their transcripts fit a single Poisson distribution in 2 of the 3 biological replicates, and neither Esp1-GFP nor Tem1-GFP display significant protein oscillations (Table S3).

Timing and amplitude of mRNA oscillations

We next asked if the proportion of highly expressing cells in the asynchronous population correlated with the proportion of the cell cycle in which the gene is expressed. We compared our FISH experiments to mean expression profiles in Cyclebase, derived from 6 different microarray experiments using synchronized cells (Fig. S1). For each periodic gene, we compared the fraction of cells with high mRNA abundance (Table 1) to the fraction of the cell cycle time in which gene expression is ≥50% of the peak level (full-width half-maximum; Fig. 2A–C). The good correlation between FISH and microarray data (Fig. 2C) indicates that the proportion of highly expressing cells does indeed correspond to the proportion of cell cycle time during which the gene is expressed.

The two-component Poisson analyses produce mean transcript numbers for the highly expressing population (λ). It seemed reasonable to expect that the ratio between the mean transcript numbers for the highly expressing subpopulation and for the population as a whole would correlate with the peak expression level of the gene divided by its mean expression level in microarray experiments. In fact, comparison of the 2 data sets shows a strong correlation (Fig. 2D). Therefore, distribution analysis of single mRNA FISH yielded accurate values for the magnitude of change in gene expression for regulated genes.

The amplitudes of CLB2 and SIC1 expression did not match in FISH and microarray data (the amplitudes are much higher in Cyclebase). Consequently, these outliers were excluded from the correlation analysis reported in Figure 2D. Nevertheless, when these transcripts were included, the correlation of expression between FISH and microarray data was still good (R² = 0.38, Figure 2).

**Figure 2.** Correlation between the timing and magnitude of gene expression from FISH experiments and Cyclebase microarray experiments. (A) CLB2 microarray data plotted on a linear scale with the minimum value set at 0 and the maximum set at 1. Used as an example of how we measure the length of time that the transcript level is greater than the half-maximum level (FWHM, full-width at half maximum). (B) CLB2 microarray data on a log log scale. Values are relative to the time-course mean. Used as an example of how we measure the peak:mean value and the total cell cycle time. (C) Correlation plot for each cell cycle gene showing fraction of the high expression population from FISH experiments (1 - β) compared with the fraction of cell cycle time in which transcript levels exceed FWHM level from Cyclebase. (D) Correlation plot for each periodic cell cycle gene showing the ratio of the mean transcript values of the high abundance population (λ) to the mean for the population from FISH experiments compared with the peak:mean values from Cyclebase. Data are from the biological replicate shown in Table 1. Similar results are obtained with all biological replicates. Values for CLB2 and SIC1 are not included (see text). Some genes have two peaks of expression in the Cyclebase data, producing two data points. R² values and P values from ANOVA statistical analysis of the linear regressions are shown. Null hypothesis for the P value is that there is no correlation between the FISH and Cyclebase data.
The mismatch for CLB2 and SIC1 may be attributable to the methods used to synchronize cells and to register the data in Cyclebase. In the microarray experiments, the abundance of CLB2 and SIC1 transcripts is suppressed in the first portion of the expression plots (Fig. S1). For CLB2, expression levels are much lower at the beginning of the time-trace than the minimum attained in the subsequent cell cycle, despite the fact that CLB2 levels should be high but decreasing in M phase. Similarly, SIC1 levels are artificially low at the start (the initial M phase), but this is clearly an artifact of how Cyclebase registers the experiments using α factor or a temperature-sensitive cdc28 allele (both of which arrest cells in G1). By contrast, experiments using a temperature-sensitive cdc15 allele, which arrests the cells in late M, did show an initial peak (Fig. S1). The lack of a starting peak for G1-arrested cells is likely due to the fact that the “initial M phase” of the time-registered experiments actually occurs at the end of the second cell cycle, by which time synchrony is reduced. The underestimated expression levels for the first part of the time-course data for CLB2 and SIC1 result in lower overall means and inflated relative peak values for these genes. The synchronization method is not a problem for the other periodic genes, as they are expressed earlier in the cell cycle than CLB2 and SIC1, and these genes correlated well with the Cyclebase data (Table 1; Fig. 2D).

To demonstrate that the low and high expressing populations in the FISH results correspond to cell cycle periodicity of expression, we performed a time course experiment using synchronized CLN2-GFP cells and monitored CLN2 mRNA levels relative to cell cycle progression, as determined by bud morphology and DNA content. As expected, CLN2 mRNA abundance was low in cells arrested in G1 by treatment with α-factor, and the single-cell mRNA levels fit a single-component Poisson distribution reasonably well (Fig. 3; Table 2). Within 30 min after release from G1 arrest, the proportion of cells expressing high levels of CLN2 mRNA increases, the mean transcript abundance of the high expressing population increases, and the mRNA abundances now fit two-component Poisson distributions (Fig. 3; Table 2). The shift in the cell population to higher CLN2 mRNA abundance corresponded to S-phase entry as shown.
by a corresponding shift of DNA content from 1N to 2N and by the appearance of small-budded cells (Fig. 3). A small population of high expressing cells persisted at later times, probably due to imperfect synchronization, since there were G₁ - and S-phase cells throughout the experimental time course (Fig. 3).

A second peak in CLN2 expression occurred at the 80 min time point as the cells entered a second cell cycle (Fig. 3; Table 2). In fact, CLN2 expression appeared to be more robust in the subsequent cycle despite a clear degradation in synchrony compared with the 30 min time point, suggesting that α-factor treatment might initially suppress the accumulation of CLN2 mRNA after cells are released.

Surprisingly, the proportion of cells that have no CLN2 mRNA was a minor portion of the population throughout the time course, and the mean CLN2 mRNA levels of the low-expressing population remained relatively constant (Fig. 3; Table 2). The persistence of low expressing cells might be due to imperfect synchronization of the population, since there was a significant population of G₁–S cells even 60 min after release (Fig. 3). However, the proportion of low expressing cells was much higher than the fraction of G₁–S cells (e.g., 75% low expressing cells but only ~35% G₁-S cells at 60 min; Fig. 3), indicating that there is measurable CLN2 basal transcription throughout the cell cycle.

Comparing mRNA and protein dynamics

Although we are examining only a small set of cell cycle-regulatory genes in this study, our results suggest that regulation of mRNA abundance is an excellent predictor for the regulation of protein abundance (Table S3). However, as revealed in multiple transcriptome-proteome studies, mRNA and protein steady-state levels are usually not well correlated (R² values are typically between 0.2–0.5). To better understand the influence of mRNA dynamics on protein dynamics, we decided to test if there were any correlations between the amplitudes of mRNA and protein oscillations and between mRNA and protein abundances.

Unfortunately, we were unable to compare mRNA and protein abundances in the same cells, because GFP fluorescence is not well preserved using the single mRNA FISH protocol. However, we previously described protein dynamics for the same 16 GFP-tagged strains,33 and we were able to compare the mRNA data from FISH to these protein data. We found that there was no significant correlation between the amplitudes of mRNA and protein oscillations (Fig. 4A, P = 0.2, null hypothesis is zero correlation). This poor correlation of mRNA to protein amplitudes probably reflects individual rate differences between synthesis and degradation of mRNA compared with protein. Furthermore, when we compared mean population levels of mRNA to protein (values are listed in Table S3), there was a mildly significant correlation reminiscent of transcriptome-proteome comparisons (Fig. 4B, black regression line, P = 0.03). However, we found very strong correlations between mRNA and protein abundances when we grouped the data by constitutive or periodic mRNA (Fig. 4B, blue regression line P = 1.8E-4 and red regression line P = 1.4E-3, respectively). We were able to use the protein population means for these correlations, because all of the proteins exhibit normal or Poisson distributions and do not consist of widely distinct populations expressing low or high levels of protein as there is for mRNA (Fig. S3).

Although separating the genes by their regulation showed good mRNA:protein correlations, the slope of the linear regression for the periodic genes was much higher than that of the constitutive genes (Fig. 4B, red regression line m = 354.4 and blue regression line m = 57.2, respectively). There is no reason to suppose that translation efficiency is so exquisitely sensitive to mRNA concentration differences of a few molecules per cell, since translation rates are largely governed by initiation and elongation kinetics.36,51 Instead, we postulated that the high slope of the correlation for periodic genes was due to the fact that most protein is produced in the high transcription state, and using

Table 2. Cell cycle-dependent CLN2 mRNA distributions in synchronized cells and derived gene expression parameters

| Cell cycle stage of CLN2 cells | FISH mean ± 95% CI* | Single Poisson* | Two-component Poisson* |
|-----------------------------|-----------------|---------------|------------------------|
|                             | G test P value | β ± 95% CI | λ₁ ± 95% CI | λ₂ ± 95% CI | G test P value | β ± 95% CI | λ₁ ± 95% CI | λ₂ ± 95% CI |
| asynchroneous from Table 1  | 4.45 ± 0.13  | 0.00 | 4.49 | 0.77 ± 0.03 | 2.09 ± 0.12 | 12.4 ± 0.55 | 5.92E-06 | 0.06 |
| asynchronous                 | 5.20 ± 0.11  | 0.00 | 18.08 | 0.70 ± 0.02 | 1.29 ± 0.08 | 14.46 ± 0.41 | 0.00 | 0.45 |
| 0 min α-factor release       | 2.94 ± 0.11  | 0.00 | 0.15 | 0.78 ± 0.14 | 2.25 ± 0.30 | 5.46 ± 1.09 | 0.03 | 0.02 |
| 15 min α-factor release      | 2.79 ± 0.11  | 0.00 | 2.78 | 0.83 ± 0.03 | 1.19 ± 0.09 | 10.86 ± 0.62 | 1.64E-07 | 0.08 |
| 30 min α-factor release      | 8.42 ± 0.19  | 0.00 | 23.77 | 0.64 ± 0.03 | 2.33 ± 0.15 | 19.10 ± 0.54 | 0.00 | 0.76 |
| 40 min α-factor release      | 6.61 ± 0.16  | 0.00 | 12.23 | 0.74 ± 0.03 | 2.54 ± 0.14 | 18.13 ± 0.66 | 0.00 | 0.57 |
| 60 min α-factor release      | 6.61 ± 0.19  | 0.00 | 11.51 | 0.75 ± 0.03 | 2.64 ± 0.16 | 18.66 ± 0.77 | 0.00 | 0.54 |
| 80 min α-factor release      | 12.59 ± 0.23 | 0.00 | 15.09 | 0.46 ± 0.03 | 3.14 ± 0.23 | 20.68 ± 0.47 | 0.00 | 0.92 |

*Transcripts/cell. Data are from cells shown in Figure 3; †Measured mRNA distribution fits to a Poisson distribution. G test of goodness of fit to a single Poisson distribution with λ = FISH Mean. β effect size. By convention, effect sizes of β ≤ 0.09 are considered small (green), 0.09 < β ≤ 0.25 are considered medium (orange), and β > 0.25 are considered large (red).49 Null hypothesis for the P value is that the empirical data have a single Poisson distribution with mean λ.49 Measured mRNA distribution fits to a two-component Poisson distribution. Same data set as for the single Poisson distribution. β, fraction of the population with low transcript levels. λ₁, mean transcripts/cell for the fraction of the population with low transcript levels. λ₂, mean transcripts/cell for the fraction of the population with high transcript levels. G test P values and β as for single-Poisson fit, but the null hypothesis is that the data have a two-component Poisson distribution with means λ₁ and λ₂.
lower population mRNA means overestimates the number of proteins/transcript. Therefore, we compared the protein abundance to the mRNA means of the high expressing population for the periodic genes (λ2 from Table S3). This results in a linear regression with the same slope as that for the constitutive genes and slightly fewer proteins/mRNA in the group of periodic genes compared with the constitutive genes (Fig. 4C).

All of the oscillating proteins in this study are known to be degraded via regulated ubiquitination and the 26S proteasome. Moreover, recent genomic, transcriptomic, and proteomic analyses have found that oscillating cell cycle proteins use sub-optimal codons to regulate translation in conjunction with cell cycle changes in tRNA availability and translation initiation rates. Therefore, the lower protein-to-mRNA ratio in periodic genes could be because the oscillating proteins have less efficient translation, enhanced degradation rates, or a combination of both.

These correlations also address the potential for the GFP tag to alter mRNA or protein dynamics. The good agreements of mRNA with protein levels suggest that GFP does not generally affect mRNA or protein synthesis or stability in this set of genes.

Including mRNA in a stochastic model of the cell cycle

We have previously described a multisite phosphorylation model of the budding yeast cell cycle, which accurately recapitulates experimental observations regarding natural variations in cell cycle timing and cell volume (see Supplementary Material for a brief description of the model). The model made certain assumptions about the mean levels of mRNA abundances of cell cycle regulators, because at the time there were no data on mRNA abundances in single cells for these genes. In light of our measurements of these mRNA distributions, we decided to revisit the model and assess its predicted mRNA distributions.

First of all, we compared our observed distribution of CLN2 mRNA (Fig. 1D) to our stochastic simulations of Mbs (mRNA for G1/S cyclin CLBS) in our original model (Fig. 9B in ref. 10). Our simulations predicted a much larger proportion of cells with high levels of CLN2 mRNA. To bring the model in line with our observations, we reasoned that the production of CLN2 mRNA has to be more abrupt, producing a transient peak in abundance to match the observed distribution of CLN2 mRNA. In order to produce a more abrupt burst of CLN2 mRNA, we allowed phosphorylation of terminal complex (CmpP2 = complex of Whi5P2 and SBF) by Cln3 and ClbS-dependent kinase, so that Whi5P2 is phosphorylated to Whi5P3 and SBF is released completely from the complex. We also modified the model to allow for phosphorylation of the RENT1 complex, leading to complete release of Cdc14. These simple changes resulted in steeper and more realistic peaks in SBF activity, Cdc14 release, and CLN2 mRNA (Mbs) abundance (compare Fig. 6 to Figs. 3 and 5 in ref. 10).

In addition, the distributions of 2 constitutive mRNAs in our model, WHI5 and NET1 transcripts, were shifted slightly to the right compared with our experimental observations. Therefore, to reproduce the observed distributions of WHI5 and NET1 transcripts, we reduced the mRNA production rates for these 2 genes and increased protein synthesis rates, such that total protein levels remained the same.

With these small adjustments, the model now captures quantitatively the observed distributions (Fig. 5) and quantities (Table 3) of CLN2, WHI5, and NET1 transcripts and the time-course of expression of the CLN2 gene (compare Fig. 6C to Fig. S1). All adjustments to the model are listed in Tables S3 and S4.
In addition to fitting the FISH data, simulations of the model with the new parameters corresponded well with the statistics of cell populations collected by Di Talia et al. Examining the timing of G1 phase, we see a size-dependent component (slope -0.55) for small birth size daughters and a size-independent component (slope -0.22) for larger daughters (Fig. S4); the corresponding values in Di Talia et al. are -0.7 and -0.3. In our simulations, the smallest quartile of mother cells show a weak, size-dependent component to the time spent in G1 (slope -0.18), but most mothers show no size dependence (slope -0.04; Fig. S4). Although Di Talia et al. obtained a single straight line with slope -0.1, the small amount of size control in mothers we observe in our new model is a marked improvement over our previous result (slope -0.36). Stochastic simulations with our new model produce cell cycle timing and cell volume distributions that correspond well with published observations (Table 4). Notice that the greatest variations occur in the timing of G1 when the number of molecules per cell is lower than in later phases of the cell cycle.

Having incorporated our mRNA observations in the model simulations, we wanted to determine how using the mRNA FISH data affected the outcome of stochastic simulations in comparison to using mRNA abundances reported by microarray studies. We ran simulations and collected statistics for populations in which each cell on average had 4-fold fewer numbers of mRNAs for all the genes included in the model. Although the average properties of cell cycle metrics remain more-or-less the same, the variability of these metrics increases significantly (Fig. 7). These results highlight the significance of mRNA levels in determining the amplitude of intrinsic noise and the importance of obtaining absolute mRNA quantities for stochastic modeling of gene expression.

Discussion

Recent attempts to build realistic, accurate, stochastic models of the CDK control system in budding yeast have been hampered by the lack of reliable data on the distribution of transcript numbers for cell cycle control genes in individual yeast cells. Computational models indicated that variability in cell cycle progression is extremely sensitive to the average number of gene transcripts per cell and to the expected half-lives of these mRNA molecules. Data available at the time, from yeast transcriptome studies, suggested that the average number of transcripts per cell was ~4-fold and that mRNA half-lives were ~20 min, and more recent studies have supported these values. Using these numbers, one can easily estimate that fluctuations in protein abundances per cell would be expected to be ± 50–100%, which is likely inconsistent with reliable behavior of any molecular control system.

Later measurements of mRNA transcripts for a handful of non-cell cycle control genes in single yeast cells by fluorescence in situ hybridization (FISH) suggested that the whole-transcriptome measurements might underestimate average mRNA abundances by ~4-fold, and a subsequent microarray study normalized according to this observation gave quantities of transcripts per cell in line with the mRNA FISH measurements. Furthermore, 20 min half-lives for mRNAs involved in cell cycle control seem to be inconsistent with how fast gene expression is turned on and off during normal cell cycle progression and in response to regulated promoters. In general, transcripts encoding regulatory proteins have shorter mean half-lives than transcripts coding for central metabolic proteins. More specifically, recent transcriptome analyses indicate that the transcripts for cell cycle genes compose a large proportion of the least stable transcripts. In fact, the half-lives for the 16 cell cycle genes included in the current study fall in the range of 5–10 min when measured by Dynamic Transcriptome Analysis (Table S1).

To help resolve uncertainties in mRNA quantities, we set out to measure the absolute quantities and distributions of mRNA transcripts of 16 cell cycle control genes in single yeast cells by FISH. Our data confirmed that the whole-transcriptome estimates are too low by ~4-fold. In addition, we were able to extract information about the dynamics of gene expression from a “snapshot” of the mRNA distribution in an asynchronous culture of yeast cells. Constitutive genes fit a single Poisson distribution, while periodic genes generally fit a two-component Poisson distribution. The low P values for some constitutive genes using the Poisson model (e.g., MCM1) and also some periodic genes for the two-component Poisson model (e.g., CLN1) suggest that these distributions may not fit the observed data exactly. However, the effect sizes (δ in Table 1; Table S1), Q–Q plots (Fig. S2), and density overlays (Fig. 1) suggest that the one- and two-component Poisson models generally fit the constitutive and periodic data well. Nevertheless, our observations that the mRNA distributions for some genes do not exactly fit single or two-component Poisson distributions might be a consequence of using an oversimplified two-state model for transcription.

Comparisons between mRNA and protein also yielded some interesting observations. We found that transcript periodicity corresponds to protein oscillations, and that constitutively transcribed genes usually produce constitutive protein levels. These might appear to be obvious correlations, but it is theoretically possible that a constitutively transcribed gene could produce protein oscillations via post-transcriptional mechanisms, or that these same mechanisms could also smooth out mRNA periodicity to yield constitutive protein levels. These 2 situations could arise in response to rapid developmental/environmental changes or as evolutionary intermediates as a gene is utilized for a new function, but we are not likely to see discordant mRNA and protein regulation during exponential growth in such ancient regulatory networks as those controlling the cell cycle.

Table 3. mRNA quantities from stochastic simulations of the model compared with observed quantities

| Gene | Mean | CV |
|------|------|----|
|      | Model | Expt | Model | Expt |
| CLB5 | 6.2  | 4.5  | 1.19  | 1.14  |
| NET1 | 6.8  | 6.2  | 0.21  | 0.43  |
| WHI5 | 4.2  | 3.5  | 0.49  | 0.67  |

Data from the biological replicates shown in Table 1. Similar results are obtained with all replicates; coefficients of variation (CV = standard deviation/mean).
Although there was no significant correlation between the amplitudes of mRNA and protein oscillations, mRNA abundances agreed very well with protein abundances. For constitutively expressed genes and periodic genes, the mean abundances of mRNA and protein correlated well, but only when grouped separately. On the other hand, using the transcript abundance of the high-expressing population \( \lambda_2 \) produced a correlation for periodic genes that fit well with the constitutive genes. These observations demonstrate one of the caveats of transcriptome–proteome correlations—using population means produces a poor mRNA-protein correlation when combining constitutive and regulated genes in the analysis, which is why various methodologies, such as principal component analysis, must be used to identify groups of differently regulated genes.

Three genes \( ESP1, TEM1, \) and \( WHI5 \) that Cyclebase labels as periodic exhibit mRNA distributions that are marginally fit by a single Poisson distribution. Examining their expression profiles in Cyclebase (see Fig. S1), we find that the data are quite noisy, that the changes in transcript abundance are quite small (~2-fold ratio between \( \lambda_1 \) and \( \lambda_2 \)), and that the majority of cells are grouped in the higher abundance population \( \beta < 0.5 \), reflecting their broad peaks of expression (Table 1; Table S1). Furthermore, in a previous study, we similarly found the Esp1 and Tem1 proteins do not show significant oscillations. Although these genes may indeed be weakly periodic, modeling their expression as constitutive produced an excellent fit of our simulations to the available experimental data. In addition, when performing mRNA–protein comparisons, it made no significant difference if we grouped \( ESP1, TEM1, \) and \( WHI5 \) with constitutive genes or with periodic genes (data not shown).

Using these measurements of mRNA distributions of cell cycle control genes, we were able to significantly improve our published model of the budding yeast cell cycle. The observed distributions of \( CLN1 \) and \( CLN2 \) transcripts suggested a major change in how

### Table 4. Population statistics of stochastic simulations compared with experimental values

| Quantity* | Daughter cell | Mother cell |
|-----------|---------------|-------------|
|           | Mean (model) | Mean (expt) | CV (model) | CV (expt) | Mean (model) | Mean (expt) | CV (model) | CV (expt) |
| \( T_{div} \) | 118.0 | 112.0 | 0.18 | 0.22 | 91.3 | 87.0 | 0.20 | 0.14 |
| \( T_{G1} \) | 40.6 | 37.0 | 0.43 | 0.50 | 13.5 | 16.0 | 0.47 | 0.50 |
| \( T_{SG2M} \) | 77.5 | 76.0 | 0.21 | 0.20 | 77.8 | 72.0 | 0.23 | 0.17 |
| \( V_{birth} \) | 31.3 | 28.0 | 0.29 | 0.22 | 46.9 | 40.0 | 0.29 | 0.18 |
| \( V_{div} \) | 70.9 | 0.25 | 89.1 | 0.30 |

* \( T_{div} \), time from birth to division; \( T_{G1} \), time from birth to bud formation; \( T_{SG2M} \), time from bud formation to division; \( V_{birth} \), daughter cell volume at birth; \( V_{div} \), mother cell volume at division; CV, coefficients of variation (CV = standard deviation/mean).

---

![Figure 5](image1.png) # Figure 5. mRNA distributions from stochastic simulations of the model compared with observed distributions. (A–C) Comparison of experimentally observed mRNA abundance to simulated distributions for (A) \( CLN2 \), (B) \( NET1 \), and (C) \( WHI5 \). Experimental data are from biological replicates shown in Table 1. Similar results are obtained with all biological replicates.
we modeled the phosphorylation of Whi5:SBF complexes and of Net1:Cdc14 complexes. The observed distributions of other genes prompted minor changes in the rate constant values for gene transcription and translation. After these changes were made, not only was our model brought into alignment with our measured gene expression profiles, but also it was more successful in fitting the observed distributions of other cell cycle metrics. Although the mRNA half-lives used in this model (1–5 min) are shorter than measured values (5–10 min),54,56 the discrepancy might easily be accounted for by some degree of mRNA gestation and senescence, as explained in the supplement to Barik et al.10 A new version of the model (under development), will include periodic expression of CLB2, SIC1, CDC6, and CDC20, using the data collected here to determine the correct mRNA distributions in the model. Furthermore, mRNA half-lives in the new model (2–10 min) are more in line with the best recent measurements.54,56

Our results show the benefits of absolute quantification of molecules in single cells. Because our approach counts mRNA molecules, we avoid the pitfalls of converting inherently relative data to absolute quantities by an intrinsically uncertain normalization factor. Remarkably, in addition to the benefits of having single-cell mRNA abundance measurements for mathematical modeling of gene expression, this data also provided insights into the dynamics of gene expression during cell cycle progression. We were able to determine if genes were constitutively expressed or regulated, and for the regulated genes, we were able to deduce the fraction of cell cycle time in which these genes are expressed and the magnitude of change in expression—all from analysis of static mRNA distributions in unperturbed asynchronous cells.

Although our experimental design precluded us from correlating transcript numbers to cell cycle phase, asynchronous cells can be staged using morphological markers such as the presence of a bud, myosin and/or Whi5 localization, and microtubule structures.4,57 Including such markers in automated analyses and correlating transcript numbers using 2-color FISH42 or fluorescent protein labeling of mRNA58 could produce high-definition data sets that include quantitative temporal information on single-cell behavior, for hypothesis testing and stochastic modeling of gene expression networks.

Materials and Methods

Yeast strains and media

Each S. cerevisiae strain used in this study included GFP inserted as an in-frame C-terminal fusion with a gene involved in cell cycle control.59 All strains were purchased from Invitrogen. Strains were tested for correct integration of the GFP tag by PCR-amplification of genomic DNA and growth characteristics, and protein dynamics were normal, suggesting that the fusions were functional.35 Cells were grown to an OD_{600} of 0.4 in synthetic complete (SC) growth medium with 2% glucose at 30 °C prior to fixation and hybridization of probes.

Cell cycle analysis

For cell cycle synchronization experiments with the CLN2-GFP strain, cells were grown as described above and arrested in G_{1} by the addition of 20 µM α-factor. Cell cycle arrest was monitored by removing 500 µL of culture, sonicating the cells (Branson Sonifier with microprobe, 10 s at power setting 4), and observing cells under differential interference contrast (DIC) microscopy until >95% of cells are unbudded with a small mating projection (~60 min). To release the cells from arrest, cells were harvested by centrifugation, washed once in sterile water, and resuspended in warm SC medium.
At various time points before and after synchronization, we removed a 60 mL sample with 50 mL to be processed for FISH, as described below, and 10 mL to be processed for fluorescence-activated cell sorting (FACS) analysis of DNA content and for bud morphology analysis (budding index). The 10 mL sample was centrifuged and washed twice in sterile nuclease-free water. The cell pellet, consisting of approximately $2 \times 10^8$ cells, was resuspended in 1 mL of water and divided equally into 2 micro-centrifuge tubes. The cells were sonicated, and an equal volume of absolute ethanol was added to fix the cells, which were stored at 4 °C overnight (or longer). Fixed cells were washed twice in sterile nuclease-free water before processing for FACS or budding index.

Samples for budding index analysis were resuspended in 1 mL of sterile nuclease-free water and stained with 2 µg/mL final concentration of Hoechst 33342 (Invitrogen, http://products.invitrogen.com/ivgn/product/H1399) for 2 h. Cells were then briefly sonicated, washed once in sterile nuclease-free water, and wet-mounted on slides for observation on a fluorescence microscope as described below except that single focus images were acquired. After acquisition, image files in TIFF format were automatically converted to Omero files and uploaded to an image storage server using Omero Insight’s importer.60 We used composite images of the phase-contrast and Hoechst channels to score unbudded G1 cells, small-budded S-phase cells in which the bud is <1/3 the size of the mother, large-budded G2/M cells with a single aggregate of DNA (nucleus), and large-budded anaphase/telophase cells with an elongated nucleus in the mother-bud neck or with mother and bud both containing a single nucleus, respectively.

Samples for FACS analysis of DNA content were sonicated, pelleted, and resuspended in 1 mL of RNase A (Sigma-Aldrich) solution (50 mM Tris-CL, pH 8.0, 15 mM NaCl, 2 mg/mL RNase A), and incubated at 37 °C for 4 h.61 After incubation, proteinase K (New England Biolabs) was added directly to the solution to a final concentration of 200 µg/mL, and the tube was incubated at 37 °C for a further 1 h. Cells were then centrifuged, resuspended in 50 mM Tris-CL, pH 8.0, and stained for immediate FACS analysis or stored at 4 °C. DNA staining was accomplished by adding 100 µL of cells to 1 mL of SYTOX green (Invitrogen, http://products.invitrogen.com/ivgn/product/S7020) solution (1 µM SYTOX Green in 50 mM Tris-CL, pH 8.0). The stained cells were briefly sonicated, and FACS analysis was performed on a BD LSR II using the Argon laser.

![Figure 7. Low mRNA abundance increases cell cycle variability. In these 4 panels we compare the means and coefficients of variation (CV, standard deviation/mean) of 5 cell cycle metrics ($T_{\text{G1}}$, interdivision time; $T_{\text{SG2M}}$, duration of unbudded phase; $T_{\text{G2M}}$, duration of budded phase; $V_{\text{Birth}}$, size at birth; $V_{\text{Division}}$, size at division) for mother cells and daughter cells. Blue bars, experimental measurements (from ref. 14); green bars, model simulations as described in the text; red bars, model simulations with 4-fold reduction in the average numbers of transcripts of all genes in the model.](image)
(488 nm) and FITC filter (505 LP splitter and 525/50 BP filter). Data was collected and analyzed with BD FACSDiva 7.0.

**FISH probes**

Five, 50-mer oligonucleotides were designed to target the GFP moiety of mRNA transcribed from the GFP-tagged genes (Supplementary Methods; Table 1). The oligos were synthesized commercially (BioSearch Technologies), and each oligo contained 5 modified Ts that carried an amino group. A Dyelight 550 NHS Ester labeling kit (Thermo Scientific, http://www.thermo.com/products/chemicals/formaldehyde.aspx) for 45 min. After washing 3 times in Buffer B (1.2 M Sorbitol, 100 mM KHPO₄, pH 7.5), cells were resuspended in spheroplast buffer (1.2 M Sorbitol, 100 mM KHPO₄, pH 7.5), 20 mM Ribonucleoside-Vanadyl complex (New England Biolabs, https://www.neb.com/products/s1402-ribonucleoside-vana-dyl-complex), 20 mM β-mercaptoethanol, and 25 U lyticase (Sigma-Aldrich, http://www.sigmaaldrich.com/catalog/product/sigma/sigma12524), and incubated at 30 °C for 7 min to partially digest the cell walls. Cells were placed on poly-L-lysine (Sigma-Aldrich) coated coverslips and allowed to settle for 30 min at 4 °C. The coverslips were washed once with Buffer B to remove unattached cells and stored overnight in 70% ethanol. The ethanol was removed, and the coverslips were incubated twice in 2× saline-sodium citrate (SSC) for 5 min and once in 40% formamide/2× SSC. Coverslips were then incubated with a mixture of the 5 probes (0.4 ng each, 2 ng total) overnight at 37 °C. Unbound probe was removed by washing the coverslips twice with 40% formamide/2× SSC, once with 2× SSC/0.1% Triton, and once with 1× SSC. The coverslips were then washed in a Hoechst 33342 solution to stain nuclei and mounted on slides with Prolong Gold mounting medium (Life Technologies, http://products.invitrogen.com/ivgn/product/P36930).

**Imaging**

All images were collected on an Axio Observer Z1 (Zeiss, http://microscopy.ziess.com/microscopy/en_de/products/imaging-systems/cell-observer.html) microscope equipped with a CoolSNAP HQ² CCD camera (Photometrics, http://www.photometrics.com/products/ccdcams/coolsnap_hq2.php), a halogen lamp for bright field imaging, and a 120 W metal halide lamp for fluorescence excitation. The microscope is controlled by custom software developed in MATLAB that relies on the API of the open-source microscopy control software, µManager. The custom software performs image-processing tasks during acquisition and, among other features, allows for the automatic identification of suitable fields of view (FOVs); therefore, maximizing the number of cells that can be sampled. For each strain, 20 FOVs, containing ~1000 cells in total, were automatically selected from a user-defined area of the slide. A z-stack was collected at each FOV containing 31 focal planes separated by 0.2 μm, and 2 color channels: Dyelight 550 (Chroma filter set SP102v1, exposure time: 2 s/plane), and Hoechst 33342 (Chroma filter set 49 000, exposure time: 150 ms/plane). In addition, a single phase-contrast image was acquired at the central plane for automated cell identification (see below).

**Image processing**

Phase-contrast images were segmented using custom software derived from Yeast Tree 1.6.3. The application relies on the MATLAB Image Processing toolbox. First, the function “imfill” was used to flood-fill local minima not connected to the image border, which fills in the center of the groups of cells. As each group of cells has slightly different levels to which the flood-fill will rise, we then searched the image histogram for intensities greater than the calculated background, taken from the border pixels, and with a frequency greater than the minimum cell area, generally set to 200 pixels. To keep only large groups of connected pixels, erosion (built-in function “imerode”) was performed, removing the outermost pixels of a region and eliminating small groups of pixels.

The next step is to separate these groups into individual cells. This was done with another call to “imerode” to cut the small necks that appear between touching cells. Once the cells are cut, the remaining connected regions were labeled with a call to the built-in function “bwlabel”, which identifies the individual cells and assigns each with a unique label. To finish, the cells were returned to their original sizes with a dilation (built-in function “imdilate”), which adds pixels around the edges of each cell.

Prior to the identification of spots, the 3-dimensional fluorescence z-stacks were reduced to 2D images by use of a maximum z-projection. Single spots, corresponding to mRNA molecules, were then identified using the algorithm described. First, a local background subtraction, in which the mean of a 19 × 19 pixel neighborhood is subtracted from the central pixel, was used to highlight the bright punctate spots from the autofluorescence.

---

**Table 5. List of oligonucleotides used in this study**

| Probe ID | Sequence*          |
|----------|--------------------|
| GFP 70   | TCCGATATGT GCATCACCTTG CACCCCTCCT ACTGAGCAAGA AATTGGTGCCC |
| GFP 304  | GGGATACCA TTAACCTTIG ACTTCAACGC ATGTCTCTAG TTCCCGTACTC |
| GFP 555  | TGGAACAGTTATGTTGTCT GTAAAAAGA CAGGCCTATC GCCAATTGGA GTA |
| GFP 628  | GCAAGCTTGA CAAAACCAAG AAGACCACTG TGGTCTCTTC TTTCCGTGGG ATC |
| GFP 682  | TAGAAGGTGGC GGGCCTATT TGATATGTC ATCCATGCA ATGTGAATCCTC |

*Nucleotides in bold are modified with amino groups for conjugation to a fluorophore.
background of the cells. Then, pixels 5 standard deviations above the mean pixel value of the entire image were chosen as initial seeds for the spot detection routine. Gaussian weighting was then used to move each seed to the center of the 2D Gaussian intensity distribution. Spots with an integrated intensity of <350 are likely the result of non-specifically bound individual probes and were therefore removed. Finally, to avoid double counting a single mRNA, if the centers of 2 spots were within 2 pixels of each other, then the dimmer of the 2 spots was removed.

Microarray data analysis

Microarray-derived relative gene expression data sets for each cell cycle gene were downloaded from Cyclebase. These data sets represent 6 independent time-course microarray experiments performed using synchronized budding yeast cell cultures. Because these data sets are from cells synchronized at different points in the cell cycle, Cyclebase provides them in a form in which all the experiments are time-registered to start at the same point in the cell cycle—roughly, late M/G1. Since the time points were different for the various experiments, we used the MATLAB cubic spline interpolation function to obtain high-resolution (0.001 min intervals) data sets for each experiment. We then obtained the means for 16 common time points (12 min intervals—the same as Spellman et al. across the 6 data sets and performed a cubic spline interpolation (0.001 min intervals) of the mean expression levels to plot the mean expression profiles over the course of 2 cell cycles (Fig. S1). These mean values were used for comparisons to the single mRNA FISH data.

Statistical analysis

The transcription of constitutively expressed genes was assumed to be the result of a random birth and death process with exponentially distributed waiting times. To characterize variability between experimental replicates, the following analyses were conducted on each experimental replicate separately. We used maximum-likelihood estimation to compare the distributions of mRNAs in unsynchronized populations to a single component Poisson function, with probability density function (PDF) defined as:

\[ P(x) = \frac{\lambda^x e^{-\lambda}}{x!} \]

where \( \lambda \) is the mean of the Poisson distribution.

For regulated genes, we assumed that our population includes 2 groups: one group of cells transcribing the gene at the low rate, and the other transcribing at a high rate. We make the further assumption that transcription in both groups has reached steady state, and so using maximum-likelihood estimation, the distribution was compared with a two-component Poisson with PDF:

\[ P_2(x) = \beta \frac{\lambda_1^x e^{-\lambda_1}}{x!} + (1-\beta) \frac{\lambda_2^x e^{-\lambda_2}}{x!} \]

where \( \lambda_1 \) and \( \lambda_2 \) are the means of the mRNA abundance in the 2 transcription states, and \( \beta \) is the proportion of cells with the lower mean transcript abundance (\( \lambda_1 \)), 0 \( \leq \beta \leq 1 \).

To determine if the experimental distributions, \( O(x) \), were adequately described by these distribution functions, we used likelihood ratio goodness of fit hypothesis tests, \( \phi^2 \) measures of effect size, and quantile-quantile plots.

The goodness of fit test statistic is:

\[ G = 2 \sum O_i \ln \left( \frac{O_i}{E_i} \right) \]

where \( i = 1, \ldots, k \) indexes the number of observed outcomes, and \( O_i \) and \( E_i \) represent the number of observed and expected counts for each outcome. To satisfy the asymptotic assumptions of this test, outcomes which had few counts were binned together, such that \( E_i > 5 \) in all cases. Under the hypothesis that the observed data arise from the theoretical distribution, \( G \) has an asymptotic Chi-square distribution with \( k - p - 1 \) degrees of freedom, where \( k \) is the number of data points, and \( p \) is the number of parameters estimated from the data (\( p = 1 \) for the Poisson model and \( p = 3 \) for the two-component Poisson model). The \( P \) value for this hypothesis test is calculated as the upper tail of the appropriate Chi-square distribution.

The measure of effect size \( \phi^2 \) is:

\[ \phi^2 = \frac{1}{n} \sum \left( \frac{O_i - E_i}{E_i} \right)^2 \]

where \( n \) is the sample size for a given gene. This measure summarizes the average normalized squared deviation between the number of observations in a cell and the number expected under a postulated model. Effect size conventions proposed by Cohen (1988) were used to classify small, medium, and large effects.

Quantile–quantile (Q–Q) plots (Fig. S2) were created by plotting the observed quantiles of the data against the corresponding theoretical quantiles of the proposed models. If data truly arise from the proposed model, then the points on the Q–Q plot will adhere closely to the \( X = Y \) line. Since these data and the corresponding models take non-negative integer values, a small amount of random noise was added (also known as jittering) to enhance visibility in the Q–Q plots.

Modeling

The model presented here is nearly identical to that described in Barik et al. with some exceptions noted in the modeling section of the results and listed in Tables S3 and S4. Deterministic simulations were performed in MATLAB using the ode15s solver. In deterministic simulations, we triggered cell division whenever ClbM concentration level drops below 12.5 nM as the cell exits mitosis. At cell division, cell volume is distributed to mother and daughter cells in a 60:40 ratio, but protein concentrations are unchanged.

Stochastic simulations were performed using Gillespie SSA. In stochastic simulations (as in the deterministic case), cell division occurs whenever ClbM concentration drops below 12.5 nM. To avoid multiple divisions caused by ClbM concentration fluctuating back and forth across 12.5 nM, we insist that ClbM concentration must increase above 20 nM before a subsequent division can be triggered by falling ClbM concentration. We use a pseudo-steady-state approximation for activation of the CLBS gene, under the assumption that SBF binds to and dissociates from the CLBS gene very rapidly. At cell division, we partitioned cell volume and all molecular species to the mother and daughter cells in a 60:40 ratio, except for Cln3, which was apportioned 75% to the mother cell and 25% to the daughter cell. This rule was intended to account for the fact that CLN3 expression in...
newborn daughters is 1/3 that of newly divided mother cells due to repression of CLN3 by Ace2 and Ash1.\textsuperscript{64}

In stochastic simulations, we followed both mother and daughter cells to generate a complete pedigree of growing and dividing cells. For each cell, we recorded $T_{Gi} =$ time from birth to onset of DNA synthesis (when ClbS rises above 25 nM) and $T_{j} =$ time from birth to Start (when SBF rises above 15 nM).

**Disclosure of Potential Conflicts of Interest**

No potential conflicts of interest were disclosed.

**Acknowledgments**

We are grateful to Daniel Zenklusen for generously providing advice in setting up FISH experiments, to Patrick Kramer and Björn Schwalb for providing published cDTA data, to Adria Carbo for help with FACS analysis, and to Matt Lux for helpful discussions. This project was supported by grants R01-GM078989 and R01-GM095955 from the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author Contributions**

DAB designed and performed the experiments, collected the data, analyzed the data, and wrote the manuscript. NRA designed and performed the experiments, collected the data, analyzed the data, and wrote the manuscript. NR performed the experiments and collected the data. DB performed the modeling and simulations. CF analyzed the data. JT designed the experiments, analyzed the data, and wrote the manuscript. JP designed the experiments, analyzed the data, and wrote the manuscript.

**Supplemental Materials**

Supplemental materials may be found here: [www.landesbioscience.com/journals/cc/article/26257](http://www.landesbioscience.com/journals/cc/article/26257)
34. Charvin G, Oikonomou C, Siggia ED, Cross FR. Origin of irreversibility of cell cycle start in budding yeast. PLoS Biol 2010; 8:e1000284; PMID:20087409; http://dx.doi.org/10.1371/journal.pbiol.1000284

35. Gillespie DT. General Method for Numerically Simulating Stochastic Time Evolution of Coupled Chemical Reactions. J Comput Phys 1976; 22:403-34; http://dx.doi.org/10.1016/0021-9991(76)90041-3

36. Sabouri-Ghomi M, Ciliberto A, Kar S, Novak B, Tyson JJ. Antagonism and bistability in protein interaction networks. J Theor Biol 2008; 250:209-18; PMID:17950756; http://dx.doi.org/10.1016/j.jtbi.2007.09.001

37. Wang Y, Liu CL, Storey JD, Tibshirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. Proc Natl Acad Sci U S A 2002; 99:5860-5; PMID:11972065; http://dx.doi.org/10.1073/pnas.09253897

38. Holstege FC, Jennings EG, Wyrick JJ, Lee TI, Hengartner CJ, Green MR, Golub TR, Lander ES, Young RA. Dissecting the regulatory circuitry of a eukaryotic genome. Cell 1998; 95:717-28; PMID:9845373; http://dx.doi.org/10.1016/S0092-8674(01)81641-4

39. Zenklusen D, Singer RH. Analyzing mRNA expression using single mRNA resolution fluorescent in situ hybridization. Methods Enzymol 2010; 470:641-59; PMID:20946829; http://dx.doi.org/10.1016/S0076-6879(10)70026-4

40. Pecoud J, Ycart B. Markovian Modeling of Gene-Expression levels. Mol Biosyst 2009; 5:1512-26; PMID:20023718

41. Ciardelli L, Stamfeld I, Romano MC. Ribosome traffic on mRNAs maps to gene ontology: genome-wide quantification of translation initiation rates and polysome size regulation. PLoS Comput Biol 2013; 9:e1002866; PMID:23382661; http://dx.doi.org/10.1371/journal.pcbi.1002866

42. Frenkel-Morgenstern M, Danon T, Christian T, Igarashi T, Cohen L, Hou YM, Jensen IJ. Genes adopt non-optimal codon usage to generate cell cycle-dependent oscillations in protein levels. Mol Syst Biol 2012; 8:572; PMID:22338280; http://dx.doi.org/10.1038/msb.2012.3

43. Tyers M, Jorgensen P. Protocols and the cell cycle: with this RING I do thee destroy. Curr Opin Genet Dev 2000; 10:54-64; PMID:10679394; http://dx.doi.org/10.1016/S0959-437X(00)00409-0

44. Grigull J, Minaimich S, Pootoolal J, Robinson MD, Hughes TR. Genome-wide analysis of mRNA stability using transcription inhibitors and microarrays reveals posttranscriptional control of ribosome biogenesis factors. Mol Cell Biol 2004; 24:5534-47; PMID:15169913; http://dx.doi.org/10.1128/MCB.24.12.5534-5547.2004

45. Miller C, Schwall B, Maier K, Schulz D, Düncek S, Zachar B, Mayer A, Sydow J, Marcinowski L, Dölken L, et al. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. Mol Biol Syst 2011; 7:458; PMID:21206491; http://dx.doi.org/10.4161/mbs.210.11.112

46. Munchel SE, Shurlzberger RK, Takizawa N, Wei K. Dynamic profiling of mRNA turnover reveals genespecific and system-wide regulation of mRNA decay. Mol Biol Cell 2011; 22:2787-95; PMID:21680716; http://dx.doi.org/10.1091/mbc.E11-01-0028