Supporting Information

Guo et al. 10.1073/pnas.1120991110

SI Text 1: Analysis of 82 Genes Identified as Daughter-Specific

True Positives: 25 Genes. We categorize as true positives the 25 identified genes that are reported by Di Talia et al. (1) in their supplementary *Text S1* to be among 28 genes transcribed only in daughter cells, or particularly responsive to either Ace2 or Swi5: *AMN1*, *ASH1*, *BUD9*, *CTS1*, *CYK3*, *DSE1*, *DSE2*, *DSE3*, *DSE4*, *EGT2*, *GAT1*, *ISR1*, *NIS1* (mistyped in their supplementary text as *HIS1*, but evident from their figures as *NIS1*), *PCL9*, *PIR1*, *PRR1*, *PRY3*, *PST1*, *RME1*, *SCW11*, *SIC1*, *SUN4*, *YLR049C*, *YNL046W*, and *YPL158C*.

False Positives: 4 Genes. Colman-Lerner et al. (2) suggested that 19 genes were not daughter-specific in their Table 2. However, among these, 8 were subsequently confirmed by Di Talia et al. (1) to actually be daughter-specific: *BUD9*, *CYK3*, *PCL9*, *PST1*, *SIC1*, *YNL046W*, *NIS1*, and *RME1* (mistyped as *REM1*, but evident from their Figure 2a as *RME1*). Of the remaining 11 genes,

2 are not included on our microarrays: *YMR316C-A* and *YOR263C*;

4 are in the set we identified as daughter-specific: *CHS1*, *HO*, *PIR3*, and *TEC1*; thus, these are categorized as false positives (see comments below); and

5 are not in the set we identified as daughter-specific: *CDC6*, *FAA3*, *PCL2*, *YGR149W*, and *PIL1*; thus, these are categorized as true negatives.

True Negatives: 5 Genes. Refer to the description of false positives.

False Negatives: 3 Genes. We categorize as false negatives the 3 nonidentified genes that are reported by Di Talia et al. (1) in their supplementary *Text S1* to be among 28 genes transcribed only in daughter cells, or particularly responsive to either Ace2 or Swi5: *YLR414C*, *FTH1*, and *ESF2*.

Although this is not a proper quantitative estimate of the false discovery rate (FDR), from the above categorizations it suggests that the FDR is perhaps something in the ballpark of $4/29 = 0.138$. However, because many of the data from Colman-Lerner et al. (2) seem to have been overridden by more recent results (in particular, 8 of the 19 genes claimed not to be daughter-specific have subsequently been shown to actually be daughter-specific), this may be a high estimate of the true FDR.

Regarding the four false positives, we identified *HO*, which controls mating-type switching and is known to participate in mother/daughter differentiation (by being asymmetrically localized to mothers rather than to daughters), and *TEC1*, which plays a key role in regulating pseudohyphal growth and whose binding sites are suggestively enriched in our "late" cluster of daughter-specific genes, along with *STE12*, the key mating pheromone response transcription factor (TF). Taken together, these results suggest a linkage between mating-type/pheromone response pathways and how mothers and daughters differentiate. We also identified *CHS1*, a chitin synthase required to repair the septum after mother/daughter separation, which seems to be a Swi5 target rather than an Ace2 target; and *PIR3*, a cell wall protein. The presence of both *HO* and *CHS1* among our false positives suggests that sometimes a gene may be included in our list if it is mother- rather than daughter-specific, but is not present early in our time-course experiments. So false positives may include genes that are asymmetrically localized during mother/daughter differentiation to mothers, but do not appear until late in our time-course experiments.

SI Text 2: Detailed Algorithm for Selecting a Regularization Parameter γ

Input: Observed transcription profile g
Output: Regularization parameter $\hat{\gamma}$ and deconvolved transcription profile f
Parameters: Fit error boundaries $\phi_l = 1.05$, $\epsilon_l = 0.04$, $\phi_r = 1.40$, and $\epsilon_r = 0.32$

1 DECONVOLUTION (g, $\gamma \leftarrow 0$, ...) $\Rightarrow \varepsilon_0$ ▷determine ε_0 as the error of the best-fit estimator (no smoothing)
2 $\varepsilon_l \leftarrow \min(\varepsilon_0 \times \phi_l, \varepsilon_0 + \epsilon_l)$ ▷the left fit error boundary ε_l
3 BINARYSEARCH ($\varepsilon \leftarrow \varepsilon_l$, $\gamma \in [0.001, 0.01]$) $\Rightarrow \gamma_l$ ▷search for the left boundary of γ
4 $\varepsilon_r \leftarrow \max(\varepsilon_0 \times \phi_r, \varepsilon_0 + \varepsilon_r)$ ▷the right fit error boundary ε_r
5 BINARYSEARCH ($\varepsilon \leftarrow \varepsilon_r$, $\gamma \in [\gamma_l, 0.01]$) $\Rightarrow \gamma_r$ ▷search for the right boundary of γ
6 FINDELBOW ($\gamma \in [\gamma_l, \gamma_r]$) $\Rightarrow \hat{\gamma}$ ▷determine $\hat{\gamma}$ at the elbow of the L-curve
7 DECONVOLUTION (g, $\gamma \leftarrow \hat{\gamma}$, ...) $\Rightarrow f$ ▷deconvolve using $\hat{\gamma}$, determine f

1. Di Talia S, et al. (2009) Daughter-specific transcription factors regulate cell size control in budding yeast. *PLoS Biol* 7(10):e1000221.
2. Colman-Lerner A, Chin TE, Brent R (2001) Yeast Cbk1 and Mob2 activate daughter-specific genetic programs to induce asymmetric cell fates. *Cell* 107(6):739–750.

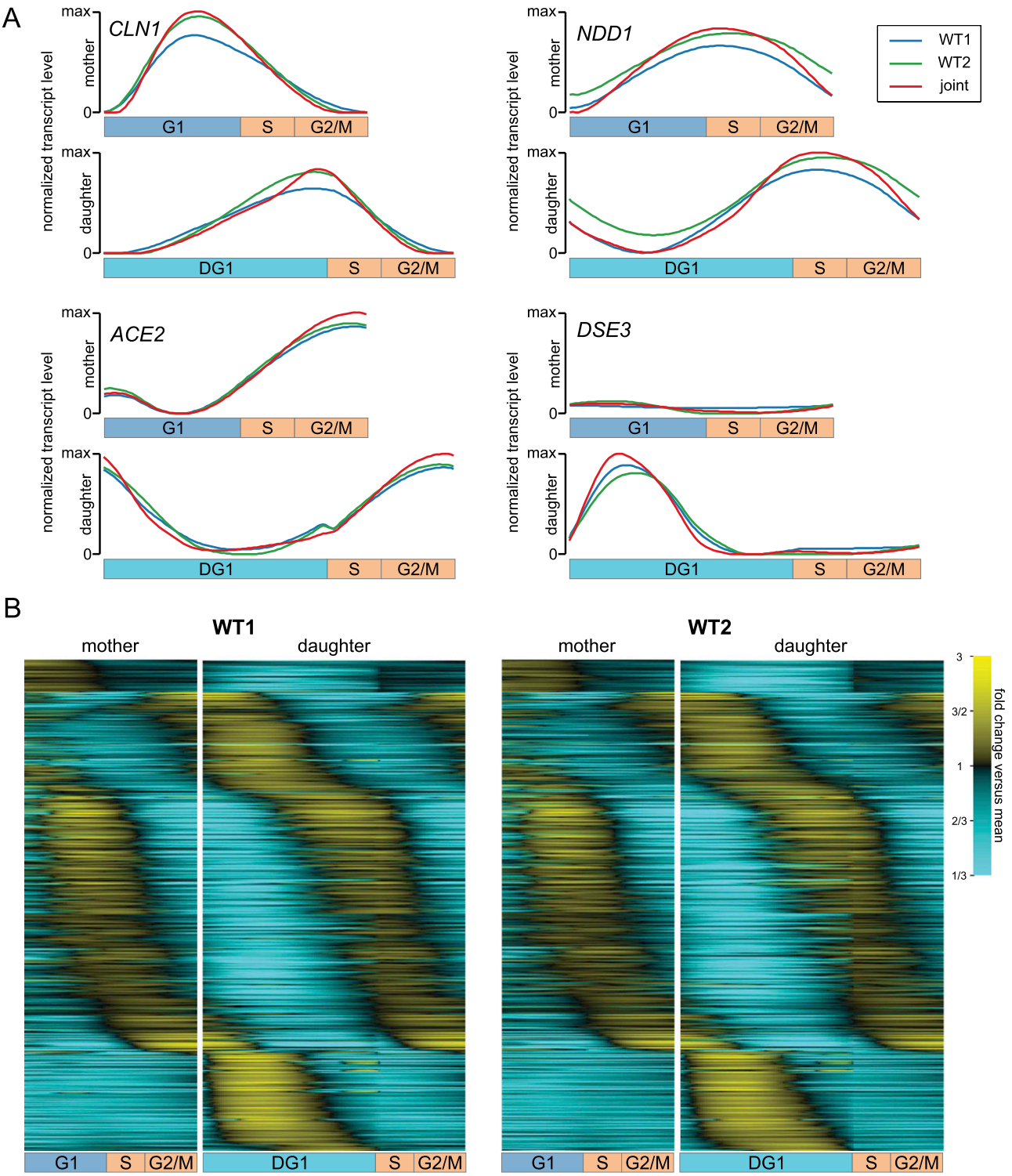

Fig. S1. Robustness of deconvolved profiles with respect to experimental variation across input data replicates. (*A*) Shown are the deconvolved profiles of *CLN1*, *NDD1*, *ACE2*, and *DSE3*, learned from WT1 data only (blue), from WT2 data only (green), and jointly from both replicates (red). (*B*) Heat maps depict transcript dynamics in the deconvolved transcription profiles of the 1,500 most strongly cell-cycle–regulated genes learned only from WT1 data and only from WT2 data, respectively. The similarity in the heat maps shows that the deconvolved profiles are essentially identical, regardless of the replicate from which they are learned.

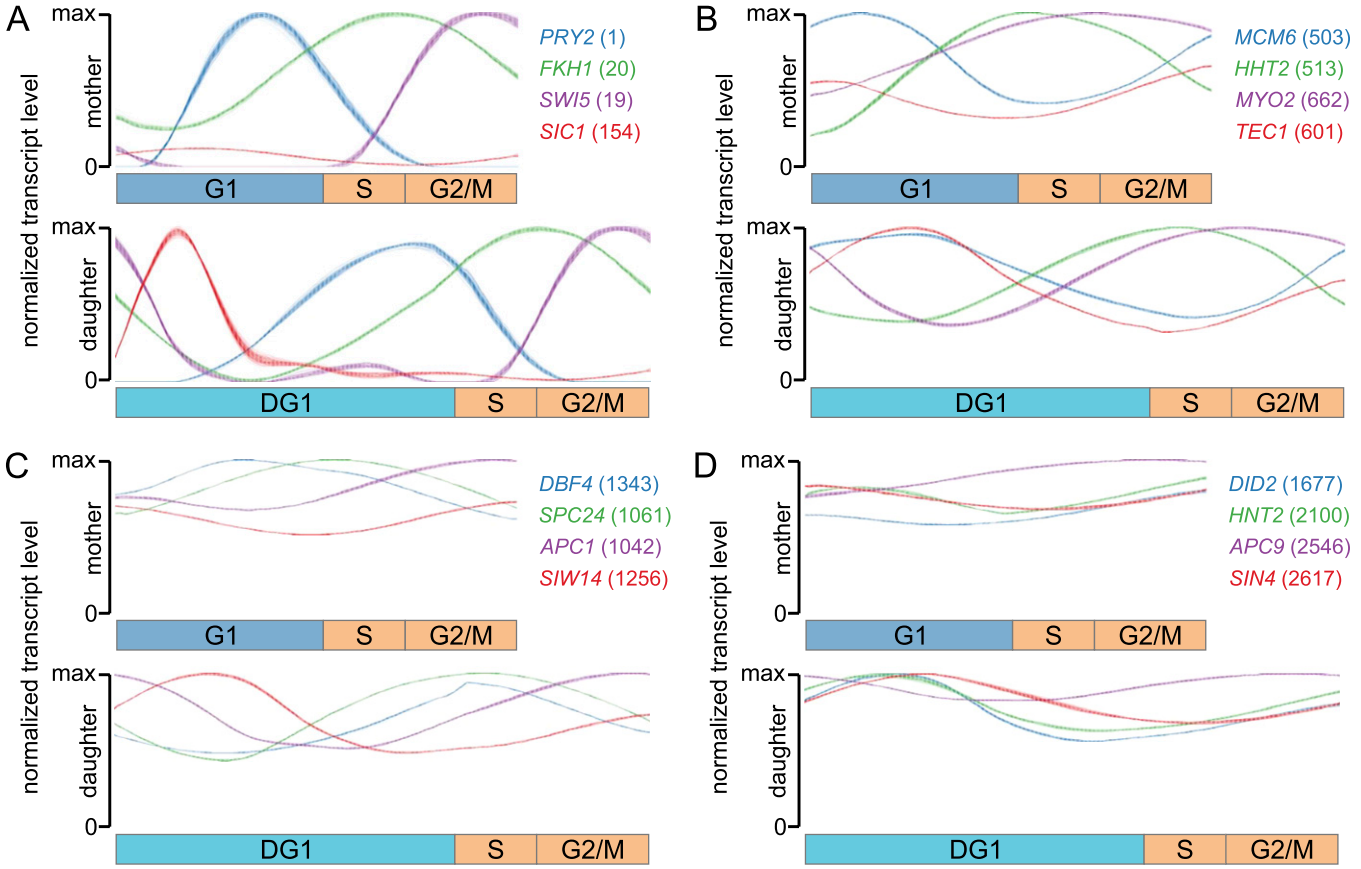

Fig. S2. Further examples of the robustness of deconvolved profiles with respect to uncertainty in CLOCCS (characterizing loss of cell-cycle synchrony) parameter estimates. Shown are 100 overlaid deconvolved transcription profiles for randomly selected genes with (*A*) high peak-to-trough ratio (PTR) scores (ranked in the top 500), (*B*) medium PTR scores (ranked 501–1,000), and (*C*) low PTR scores (ranked 1,001–1,500). In each case, we have chosen genes with transcription peaks in G1, S, G2/M, and daughter-specific G1 (DG1). (*D*) Shown are overlaid profiles for four genes with even lower PTR scores (ranked below 1,501). The 100 deconvolved transcription profiles for each gene were produced using 100 different CLOCCS parameterizations, each a random realization from the CLOCCS Markov chain. Numbers in parentheses indicate ranks of genes according to PTR score.

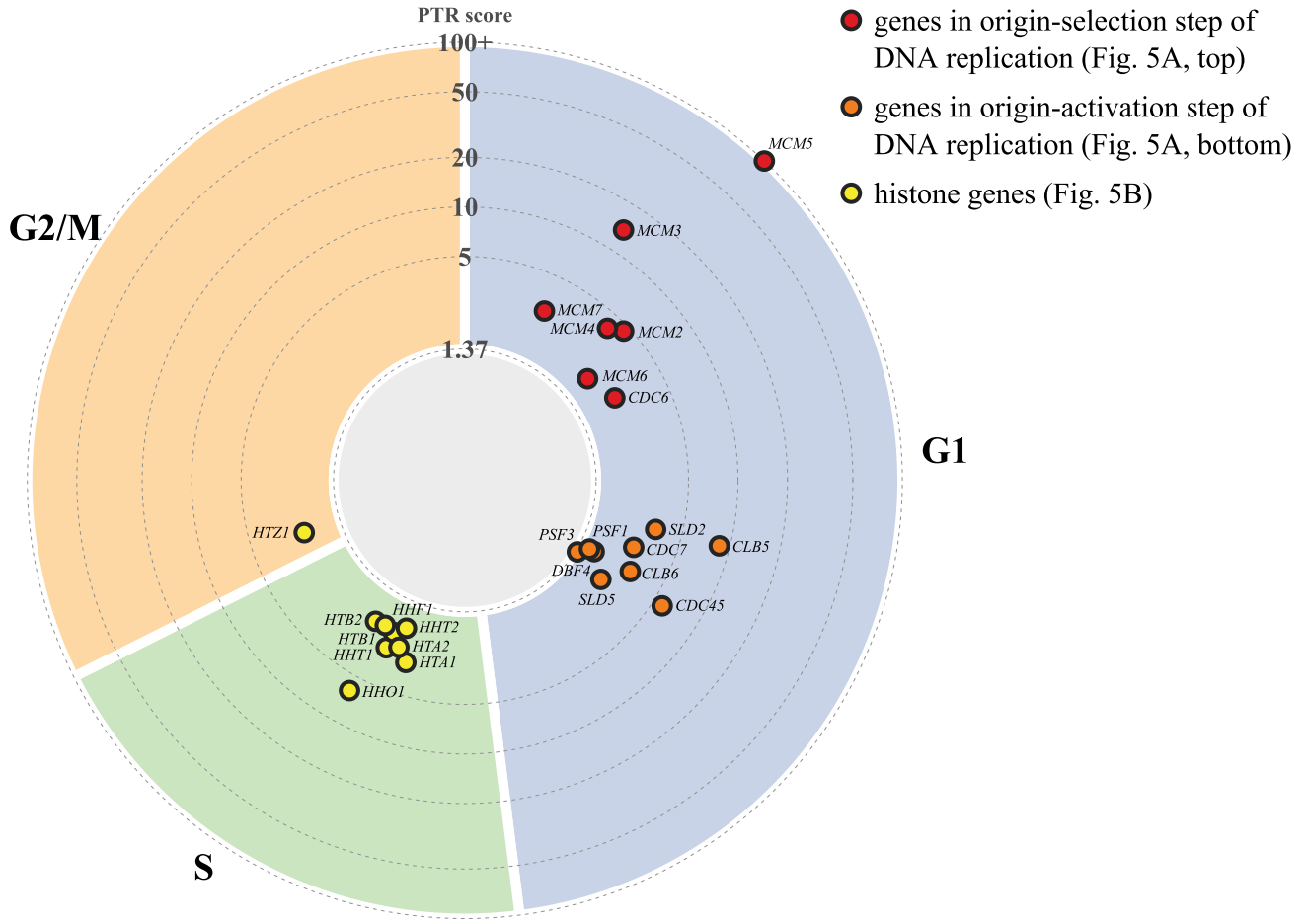

Fig. S3. Circular representation of peak timing of large subsets of the 1,500 most strongly cell-cycle–regulated genes. We provide a different representation of the peak timing of the genes shown earlier in Fig. 5, with the genes from each panel in Fig. 5 sharing a color here. In this representation, each gene is placed with an angle based on its peak timing and with a radius based on its (log) PTR score. Colored sectors demarcate the cell-cycle phases G1, S, and G2/M, and gray dashed circles indicate increasing PTR scores, starting from 1.37, the threshold corresponding to the 1,500 genes most strongly cell-cycle–regulated. This representation omits details of the profiles, but focuses on the peak timing and PTR score of each gene; the more compact focus allows more genes to be visualized at once with less clutter.

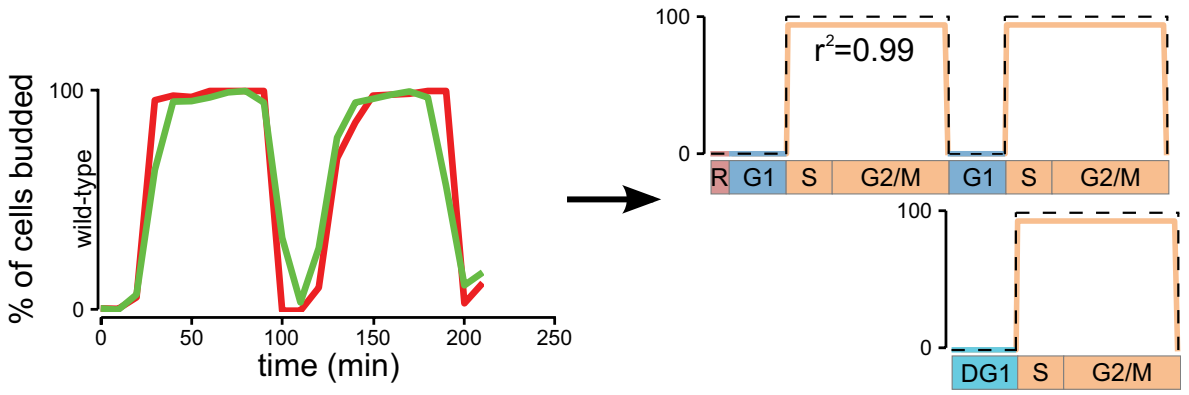

Fig. S4. Deconvolution of budding index measurements of Granovskaia et al. (1). (*Left*) Wild-type budding index measurements in red, along with the fit to those time series learned by our algorithm overlaid in green. (*Right*) Deconvolved budding profile. The true budding profile is shown as a dashed line for comparison (r^2 = 0.99). The cell-cycle parameters for CLOCCS are obtained from Mayhew et al. (2).

1. Granovskaia MV, et al. (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11(3):R24.
2. Mayhew MB, Robinson JW, Jung B, Haase SB, Hartemink AJ (2011) A generalized model for multi-marker analysis of cell cycle progression in synchrony experiments. *Bioinformatics* 27(13):i295–i303.

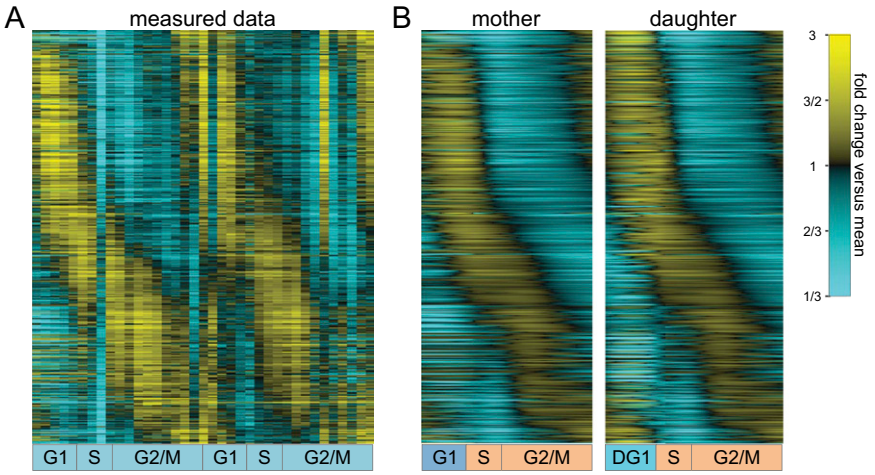

Fig. S5. Transcript dynamics of 598 periodic protein-coding mRNAs reported by Granovskaia et al. (1). (*A* and *B*) Heat maps depict the dynamics of transcripts in the measured data (*A*) and after deconvolution (*B*). Corresponding rows in the various heat maps represent the same gene.

1. Granovskaia MV, et al. (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11(3):R24.

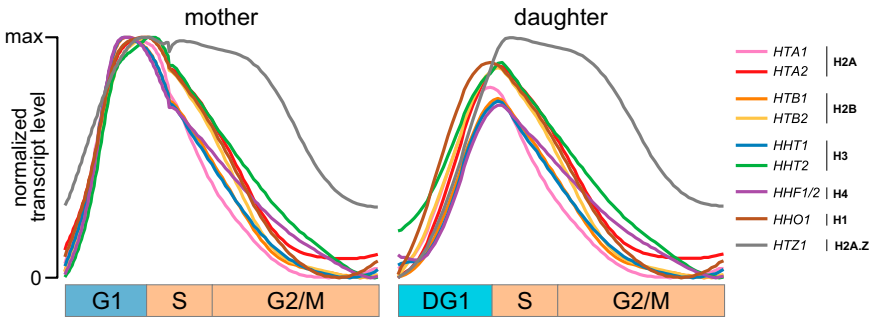

Fig. S6. Normalized deconvolved transcription profiles of histone genes from Granovskaia et al. (1). The deconvolved results are consistent with the observation of Fig. 5*B* that the H2A.Z histone variant (*HTZ1*) peaks uniquely later.

1. Granovskaia MV, et al. (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11(3):R24.

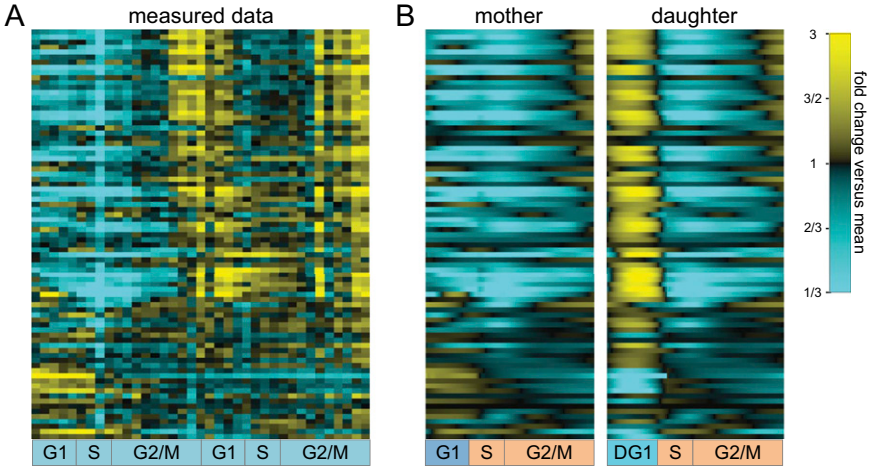

Fig. S7. Deconvolution of 82 identified daughter-specific genes, as shown in Fig. 6*B*, using the data of Granovskaia et al. (1). Most of these genes are transcribed primarily and almost entirely in the DG1 interval in the data of Granovskaia et al.

1. Granovskaia MV, et al. (2010) High-resolution transcription atlas of the mitotic cell cycle in budding yeast. *Genome Biol* 11(3):R24.

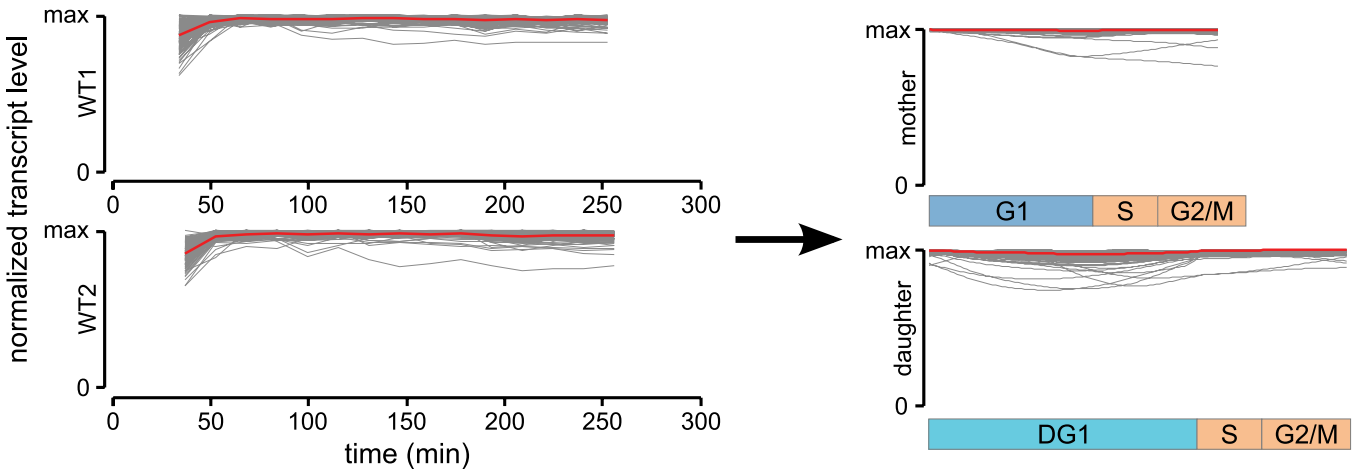

Fig. S8. Normalized transcription profiles of 129 ribosomal protein genes before and after deconvolution. The median transcription profile in each case is overlaid in red. The average of the 129 PTR scores decreased from 1.027 to 1.018 after deconvolution, suggesting that our deconvolution algorithm is effective at not sharpening noise.

A

	Peak ratio ≥ 2 (120 genes)	
One dominant peak (mother or daughter)	in mother G1	in daughter DG1
	1 gene (*PRY1*)	119 genes

B

		Peak ratio < 2 (1380 genes)			
Two dominant peaks (mother & daughter)		in mother G1 & in daughter DG1	in mother G1 & in daughter post-G1	in mother post-G1 & in daughter DG1	in mother post-G1 & in daughter post-G1
		364 genes	381 genes	345 genes	290 genes

		Compressed		
		High (corr≥0.9)	Medium (0.9>corr≥0.7)	Low (corr<0.7)
Delayed	High	13.7%	13.2%	17.0%
	Medium	8.2%	5.2%	8.5%
	Low	8.5%	7.1%	18.4%

delayed (30.2%)
compressed (30.5%)
mixed (20.9%)
uncorrelated (18.4%)

Fig. S9. Relationships of transcription profiles in G1 and DG1. First, as discussed in *Materials and Methods*, we separated our 1,500 most strongly cell-cycle–regulated genes into two groups: genes with one dominant peak and genes with two dominant peaks. (*A*) One-dominant-peak genes. In our cell-cycle branching process model, we allowed mother and daughter cells to transcribe genes differently during G1 and DG1, but assumed they share a common transcription program post-G1. Therefore, because these genes have only one dominant peak, it must occur either in mother G1 or in daughter DG1. Interestingly, we found only one gene (*PRY1*) in the first category, but 119 genes in the second. (*B*) Two-dominant-peak genes. We split the remaining 1,380 genes into four subgroups according to where their two dominant peaks occurred. For the 364 genes whose two dominant peaks are in mother G1 and in daughter DG1, we calculated two Pearson's correlation coefficients between the transcription profiles in G1 and DG1: one between the G1 profile and a compressed version of the DG1 profile (compressed) and the other between the G1 profile and the later segment of the DG1 profile (delayed). According to the strengths of these two correlation coefficients, we separated the 364 genes into nine groups and combined some of the groups into four categories, as shown.

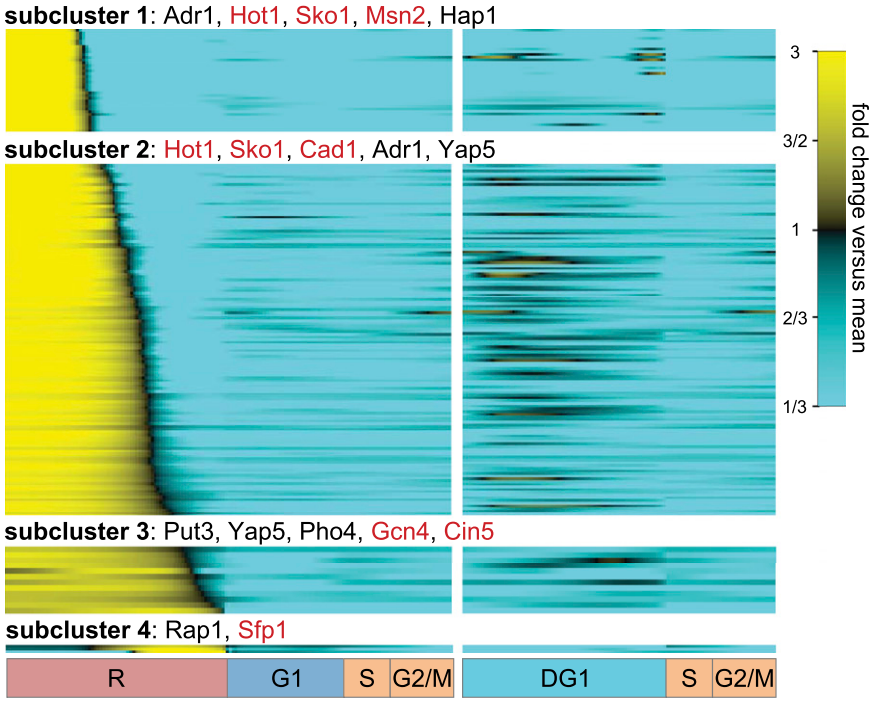

Fig. S10. Transcripts whose levels are elevated significantly under stress. In elutriation-based synchronization experiments, the initially collected cells—typically small cells early in G1—are released from synchrony after experiencing significant cold and osmotic stress. Thus, elevated transcript levels of a gene early in the time course could arise because the gene is necessary for early G1 events, or because the gene is part of a stress response, or both. If the former, we would expect to see high levels of transcription again later in the time course; if the latter, we would expect the high levels of transcription to be confined to the earliest samples of the time course. To identify genes whose high early transcription can be primarily attributed to stress response, we established two criteria: The integrated transcript level of a gene in the recovery (R) interval is at least half the total across all cell-cycle branches (R, G1 + post-G1, and DG1 + post-G1), and the peak transcript level in R is at least twice as high as that in mother (G1 + post-G1) or daughter (DG1 + post-G1) cells. We identified 184 genes satisfying these criteria, heat maps for which are shown. Gene Ontology (GO) (1) enrichment analysis reveals that the biological functions of many of these genes are relevant to the processes of vacuolar protein catabolic processes ($P < 4 \times 10^{-26}$), response to temperature stimuli ($P < 10^{-16}$), and response to abiotic stimuli ($P < 10^{-8}$), and similar processes, suggesting that these genes are likely indeed stress-response genes and, more specifically, responding to the cold temperatures during elutriation. On the basis of their deconvolved transcription profiles, we refined these genes into four subclusters according to the time at which the profiles first drop below their mean and looked for overrepresented transcription factors within the promoters of genes in each subcluster. Up to five overrepresented transcription factors for each subcluster are shown. Transcription factors that are involved in the regulation of genes during stress or amino acid starvation (e.g., Gcn4) are labeled in red. (As an aside, we used a simple PTR scoring scheme to identify cell-cycle–regulated genes. However, because PTR scores intentionally ignore the recovery interval to focus on the mother and daughter cell cycle, they do not take into account transcript levels during R, which for stress-response genes may be significantly elevated. In particular, 128 of the 184 genes listed here are also included in our set of 1,500 most strongly cell-cycle–regulated genes. As can be seen here, the transcript levels of these genes are often much higher in R than later in the cell cycle, indicating that their transcription is not exclusively regulated during the cell cycle, but also through varying environmental conditions and stress.)

1. Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25(1):25–29.

Table S1. Comparison of cell-cycle deconvolution methods

	Lu et al. (1)	Bar-Joseph et al. (2)	Qiu et al. (3)	Rowicka et al. (4)	Siegal-Gaskins et al. (5)	This study
Species	Eukaryote: *Saccharomyces cerevisiae*	Eukaryote: *S. cerevisiae*	Eukaryote: *S. cerevisiae*	Eukaryote: *S. cerevisiae*	Bacterium: *Caulobacter crescentus*	Eukaryote: *S. cerevisiae*
Data type	*1*) Basis experiments from synchronized cell-cycle experiments. *2*) Static transcription levels from populations of cells grown in a wide variety of conditions.	*1*) Budding index or FACS data. *2*) Synchronized cell-cycle time course.	Synchronized cell-cycle time course.	Synchronized cell-cycle time course (YMC).	Synchronized cell-cycle time course.	*1*) Budding index data and flow cytometry data. *2*) Two independent replicates of synchronized cell-cycle duration course.
Cell-cycle phases	G1, S, G2, M, M-to-G1	G1, S, G2/M	Not specified	G1(P: prereplicative), G1-to-S, S, G2, G2-to-M, M, M-to-G1.	SW, EPD, LPD, ST	G1, S, G2/M, DG1
Synchrony loss model	Static transcription. The fractions of cells in five cell-cycle phases were determined from the basis experiments.	Used budding index or FACS data to estimate the duration of each cell-cycle phase and cell growth variance in the population.	Used a mixture model to account for cells growing at slightly different rates.	Used transcription peaks of some characterized cell-cycle–regulated genes to determine cell-cycle phases and subphases.	Used a probabilistic model to estimate the total cycle time, SW-to-ST transition point, and cell-cycle distributions.	Used CLOCCS (characterizing loss of cell cycle synchrony) model to estimate cell-cycle lengths, initiation and growth variances, and distributions.
Synchrony loss factors recovered	Variation in cell-cycle rates.	Variation in cell-cycle rates.	Variation in cell-cycle rates.	Synchrony noise in initial population of cells.	*1*) Variation in cell-cycle rates. *2*) Variation in the physiological and developmental state of the cells; asymmetric cell-cycle division.	*1*) Synchrony noise in initial population of cells. *2*) Variation in cell-cycle rates. *3*) Asymmetric cell-cycle division. Newborn daughter cells require more time to reach a critical cell size.
Deconvolution model	Used transcription peaks of some characterized cell-cycle–regulated genes to determine cell-cycle stages. Used a system of weighted linear equations to fit the measured static transcription.	Used cubic splines to fit the time-series transcription data.	Used a polynomial model to fit the time-series transcription data.	Used a regularization-based approach on the maximum-entropy principle.	Converted the deconvolution problem to an optimization problem, using cross-validation to select an appropriate control parameter.	Used a wavelet-basis regularization approach to solve an ill-posed discrete inverse problem; avoided overfitting and oversmoothing by selecting a "good" regularization control parameter.
Deconvolution outputs	Transcript levels of a gene at each cell-cycle phase.	Refined transcription profiles of two cell cycles.	Refined transcription profiles.	Timing of the transcription peaks (and in some cases secondary transcription peaks).	"Single-cell" transcription profiles.	Single-cell transcription profiles, distinct for mother and daughter cells.

Table S1. Cont.

	Lu et al. (1)	Bar-Joseph et al. (2)	Qiu et al. (3)	Rowicka et al. (4)	Siegal-Gaskins et al. (5)	This study
No. genes identified (or used) as cell-cycle–regulated	From literature, authors picked 696 genes as cell-cycle dependent.	Inferred around 900 cell-cycle–regulated genes.	Not discussed.	Inferred 694 high-confidence cell-cycle–regulated genes, with an extended set of 1,129 genes.	Not discussed.	Ranked all genes, the top 1,500 of which show clear cell-cycle regulation.
Resolution in the deconvolved profiles	Static transcription; not applicable.	Not explicitly estimated.	Not explicitly estimated.	Resolution of transcription peaks around 2 min (~2% of cell cycle).	Not explicitly estimated.	Nominal temporal resolution $\leq$ 0.8 min (~1% of cell cycle). Effective temporal resolution of transcription peaks around 1 min (~1% of cell cycle).
Comments	Deconvolved static transcription levels, not transcription profiles.			Reported only the transcription peaks of genes instead of transcription profiles.	The proportion of cell types (SW, ST) at cell division, used to model the synchrony loss by asymmetric cell division, was determined by prior knowledge.	All cell-cycle parameters, including accurate distributions of mother and daughter cells, were estimated from the data. No prior knowledge needed.

1. Lu P, Nakorchevskiy A, Marcotte EM (2003) Expression deconvolution: A reinterpretation of DNA microarray data reveals dynamic changes in cell populations. *Proc Natl Acad Sci USA* 100(18):10370–10375.
2. Bar-Joseph Z, Farkash S, Gifford DK, Simon I, Rosenfeld R (2004) Deconvolving cell cycle expression data with complementary information. *Bioinformatics* 20(Suppl 1):i23–i30.
3. Qiu P, Wang ZJ, Liu KJ (2006) Polynomial model approach for resynchronization analysis of cell-cycle gene expression data. *Bioinformatics* 22(8):959–966.
4. Rowicka M, Kudlicki A, Tu BP, Otwinowski Z (2007) High-resolution timing of cell cycle-regulated gene expression. *Proc Natl Acad Sci USA* 104(43):16892–16897.
5. Siegal-Gaskins D, Ash JN, Crosson S (2009) Model-based deconvolution of cell cycle time-series data reveals gene expression details at high resolution. *PLoS Comput Biol* 5(8):e1000460.

Table S2. Cell-cycle parameters estimated by CLOCCS from budding index and flow cytometric measurements of DNA content

	Budding and flow		Flow only	
Cell-cycle parameters	WT1	WT2	WT1	WT2
From CLOCCS				
Length of R, min	94.387	101.904	94.279	101.954
Length of C, min	79.487	82.014	79.647	81.965
Length of DG1, min	44.318	37.436	44.326	37.425
Length of G1, fraction of C	0.153	0.165	—	—
Length of G1+S, fraction of C	0.349	0.391	0.349	0.391
Length of G2+M, fraction of C	0.651	0.609	0.651	0.609
Adjustment				
Length of postcytokinetic attachment, min	26	27	—	—
After adjustment				
Length of R, min	68.387	74.904	—	—
Length of C, min	79.487	82.014	—	—
Length of DG1, min	44.318	37.436	—	—
Length of G1, fraction of C	0.480	0.494	—	—
Length of G1+S, fraction of C	0.676	0.720	—	—
Length of G2+M, fraction of C	0.324	0.280	—	—

The parameters learned from budding index and flow cytometry were used for deconvolving all measured transcription profiles. The parameters learned only from flow cytometry were used for deconvolving budding index profiles (because deconvolving budding index profiles with the aid of parameters learned from those profiles would produce overly optimistic estimates of deconvolution performance).

Table S3. Full list of overrepresented transcription factors (TFs) of subclusters

Subcluster name or ID	No. genes	Overrepresented TFs with P value ≤ 0.005
In Fig. 6*B* subclusters		
Early	54	Ace2, Swi5, Sok2, Phd1, Ste12, Fkh2, Mcm1, Ash1, Fkh1, Skn7, Adr1, Tos8, Swi4, Mbp1, Dal81, Pho4, Yap5
Middle	8	Sok2, Ste12, Cin5, Yap6
Late	20	Mac1, Tec1, Put3, Mcm1, Ste12
In Fig. S10 subclusters		
1	38	Adr1, Hot1, Sko1, Msn2, Hap1, Sko2, Skn7, Pdr1, Cad1, Fkh2, Nrg1, Rtg3
2	131	Hot1, Sko1, Cad1, Adr1, Yap5, Msn2, Cin5, Sok2, Pdr1, Yap6, Ste12, Skn7
3	12	Put3, Yap5, Pho4, Gcn4, Cin5, Yap6
4	3	Rap1, Sfp1