1. Experimental Data Acquisition

The acquisition of data in this study was designed to enable integration of the global gene expression measured in samples from different time courses under different conditions. Eight time courses of cultures synchronized by the α-factor pheromone were sampled at >12 time points at equal time intervals. Four time courses, two of cultures of a Cdc6-shutoff strain at the depleted condition of $Cdc6^-$, and two of cultures of the parental strain at the condition of $Cdc6^+$, were sampled at 5min intervals. Genome-wide mRNA expression in these samples was measured relative to RNA extracted from asynchronous cultures of the parental strain at the condition of $Cdc6^+$. Four additional time courses, two of cultures of a Cdc45-shutoff strain at the inactivated condition of $Cdc45^-$, and two of cultures of the parental strain at the condition of $Cdc45^+$, were sampled at 6min intervals. Genome-wide mRNA expression in these samples was measured relative to asynchronous cultures of the parental strain at the condition of $Cdc45^+$ (Supplementary Tables I and II, and Datasets 1 and 2).

In both the $Cdc6^-$ and $Cdc45^-$ cells the initiation of DNA replication is prevented without delaying cell cycle progression. In the $Cdc6^-$ cells, depletion of the essential licensing factor Cdc6 prevents replication origin licensing by preventing the Mcm2-7 proteins from binding to origins during the cell cycle phase G1 (Piatti *et al*, 1995). The Mcm2-7 proteins are necessary for the formation and maintenance of the prereplicative complex. In the $Cdc45^-$ cells, inactivation of the essential initiation factor Cdc45 prevents origin firing and therefore also initiation of DNA replication at a step after Mcm2-7 loading. The Mcm2-7 proteins remain bound to origins even as the cells progress through S, S/G2 and G2/M (Tercero *et al*, 2000). The absence of replication in the $Cdc6^-$ and $Cdc45^-$ cells in this study was validated by flow cytometry measurement of the cellular DNA content in these samples. Nuclear division of these cells in the absence of DNA replication was validated by fluorescence microscopy measurement of the fraction of cells with divided chromatin in the $Cdc6^-$ and $Cdc6^+$ cultures (Supplementary Figures 1 and 2).

First, the experimental variation was designed to be orthogonal to the biological variation to enable computational detection and removal of experimental artifacts. Differences in sample batch, DNA microarray platform and protocols were implemented among and prevented within pairs of shutoff and control time courses. Samples were hybridized in batches of odd or even time points from each of the pairs of shutoff and control time courses.

Second, the environmental conditions, i.e., temperature and carbon source, of the synchronized Cdc6-shutoff strain at the depleted condition of $Cdc6^-$ differ from these of the Cdc45-shutoff strain at the inactivated condition of $Cdc45^-$. Therefore, the synchronized control and asynchronous reference $Cdc6^+$ and $Cdc45^+$ cultures were grown at the same environmental conditions as the $Cdc6^-$ and $Cdc45^-$ cultures, respectively, effectively canceling out mRNA expression variation that might be due to the variation in these environmental conditions (Supplementary information Section 1.2). Note also that mRNA expression variation between the $Cdc6^+/^-$ and $Cdc45^+/^-$ cells, e.g., due to the variation in the environmental conditions or the parental genetic background between these cells, is orthogonal to the biological variation between the $Cdc6^+$ and $Cdc6^-$ cells, and the $Cdc45^+$ and $Cdc45^-$ cells. Therefore, this variation was computationally detected and removed.

Third, the cell cycle period of the synchronized control $Cdc6^+$ culture differs from that of the $Cdc45^+$ culture. The time points, therefore, were selected to sample approximately similar cell cycle phases in both cultures, such that there is a one-to-one mapping among the time points of all eight time courses. This mapping was validated by the flow cytometry measurement of the cellular DNA content in these samples, and the immunoblot measurement of protein expression levels in the $Cdc6^-$ and $Cdc6^+$ cells. Of these time points, the first 12 approximately sample the exit from the pheromone-induced arrest and entry into G1 through S and S/G2 to the beginning of G2/M just before nuclear division in each time course. The time of nuclear division was approximately determined by the fluorescence microscopy measurement of the fraction of cells with divided chromatin in the $Cdc6^-$ and $Cdc6^+$ samples. Global gene expression measurements from the corresponding 12 samples from each time course were computationally integrated

Condition	Strain	Genotype	Parental Strain
$Cdc6^+$	W303-1a	*MAT***a** *ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100*	
$Cdc6^-$	YGP81	*MAT***a** *ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 CDC6::GAL-CDC6 (TRP1)*	W303-1a
$Cdc45^+$	YKL83	*MAT***a** *ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 UBR1::GAL-UBR1 (HIS3)*	W303-1a
$Cdc45^-$	YJT18	*MAT***a** *ade2-1 ura3-1 his3-11,15 trp1-1 leu2-3,112 can1-100 CDC45::cdc45-td (TRP1) UBR1::GAL-UBR1 (HIS3)*	YKL83

Supplementary Table I. The strains used in this study are based on W303-1a (Piatti *et al*, 1995; Tercero *et al*, 2000).

by using the mathematical framework of the higher-order singular value decomposition (HOSVD) (Supplementary information Sections 2 and 3).

1.1. Strains and Media

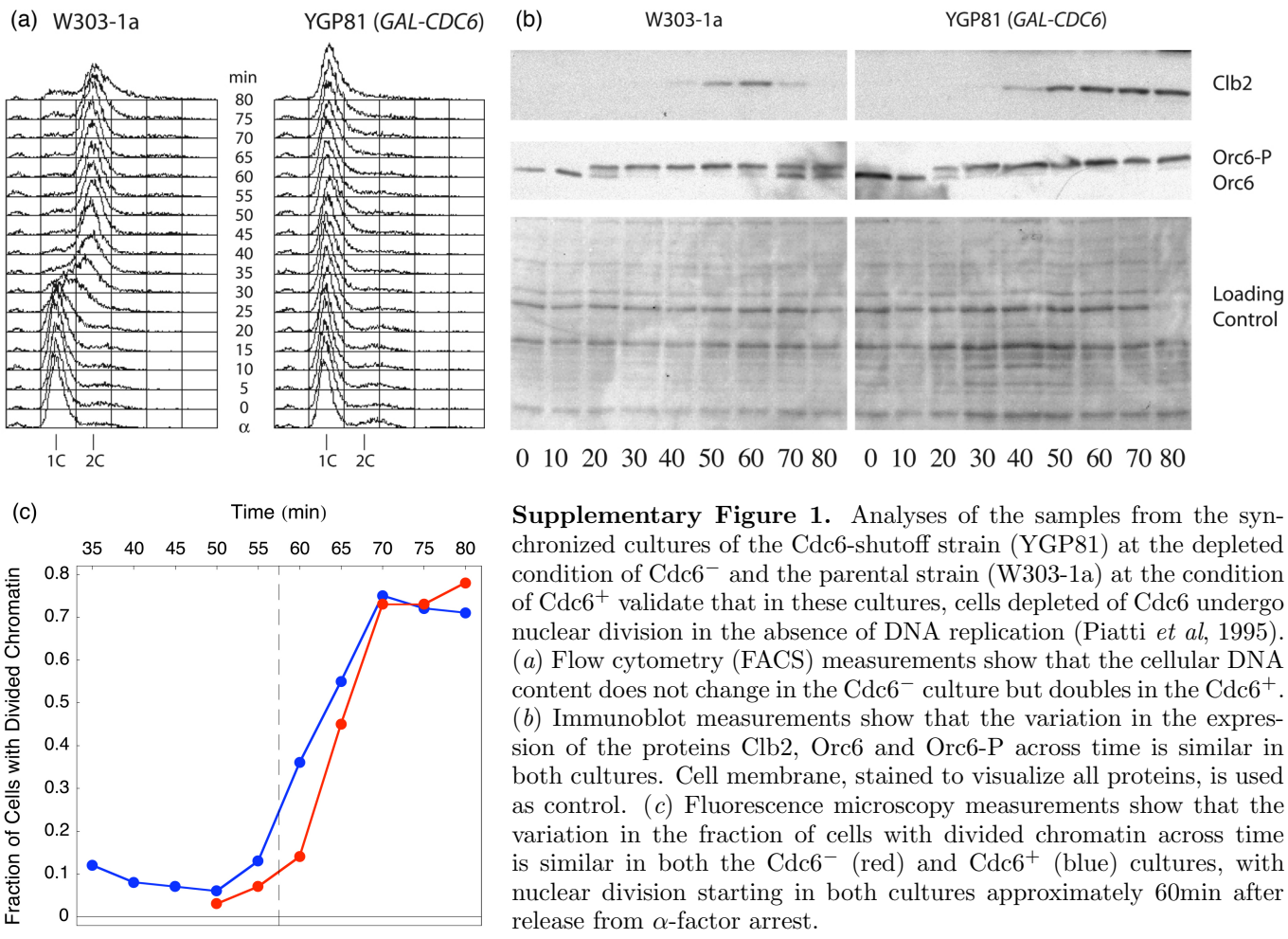
The strains used in this study are based on W303-1a (Supplementary Table I) as described (Piatti *et al*, 1995; Tercero *et al*, 2000).

Cells were grown in yeast extract, peptone and a carbon source of either glucose (YPD), raffinose (YPRaf), galactose (YPGal) or galactose and raffinose, each at a final concentration of 2% (YPRaf/Gal). Nocodazole and α-factor pheromone were also used, each at a final concentration of 5μg/ml.

1.2. Cell Synchrony and Conditions

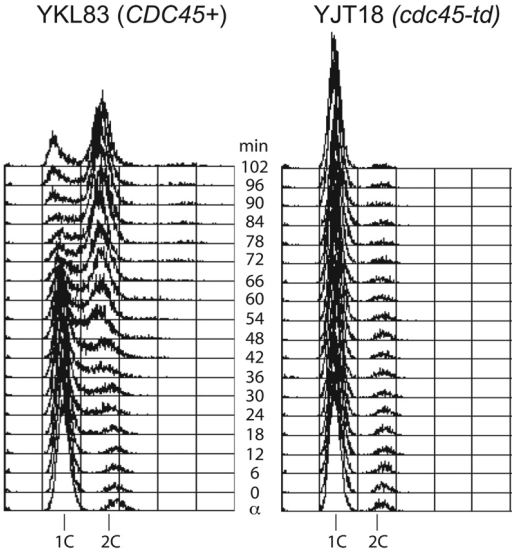
YGP81 has its only copy of the *CDC6* gene under the control of the glucose repressible GAL1-10 promoter. Cultures of both YGP81 and its parental strain (W303-1a) were grown to mid log-phase ($\sim 5 \times 10^6$ cells/ml) in YPGal (*CDC6* 'on') at 30°C. Nocodazole was added to arrest the cells in G2/M. Two hours later the cells were harvested and resuspended in YPD containing nocodazole to repress *CDC*6 transcription in the YGP81 cells (*CDC6* 'off', i.e., Cdc6$^-$) but not in the W303-1a cells (Cdc6$^+$). Cells were held for an additional hour to allow the unstable Cdc6 protein to be degraded, and then washed into YPD containing α-factor to synchronize them. After two hours the cells were washed once more and released into the cell cycle in YPD.

Samples were taken at the time of release and every 5min after for 80min. Cellular DNA content in these samples was measured by flow cytometry, i.e., fluorescent-activated cell sorting (FACS) (Supplementary Figure 1*a*). Expression of the proteins Clb2 and Orc6, as well as Orc6 phosphorylated by S-phase cyclin-depndent kinases (Orc6-P) was monitored by immunoblot analyses (Supplementary Figure 1*b*). The fraction of cells with divided chromatin was determined by fluorescence microscopy (Supplementary Figure 1*c*). The samples were also processed for RNA to be used in the DNA microarray hybridization as described (Gerke *et al*, 2006; Hu *et al*, 2007). Reference RNA was made from an asynchronous culture of the parental strain W303-1a grown to mid log-phase in the same environmental conditions as the synchronized cultures, i.e., in YPD at 30°C.



Supplementary Figure 1. Analyses of the samples from the synchronized cultures of the Cdc6-shutoff strain (YGP81) at the depleted condition of Cdc6$^-$ and the parental strain (W303-1a) at the condition of Cdc6$^+$ validate that in these cultures, cells depleted of Cdc6 undergo nuclear division in the absence of DNA replication (Piatti *et al*, 1995). (*a*) Flow cytometry (FACS) measurements show that the cellular DNA content does not change in the Cdc6$^-$ culture but doubles in the Cdc6$^+$. (*b*) Immunoblot measurements show that the variation in the expression of the proteins Clb2, Orc6 and Orc6-P across time is similar in both cultures. Cell membrane, stained to visualize all proteins, is used as control. (*c*) Fluorescence microscopy measurements show that the variation in the fraction of cells with divided chromatin across time is similar in both the Cdc6$^-$ (red) and Cdc6$^+$ (blue) cultures, with nuclear division starting in both cultures approximately 60min after release from α-factor arrest.

YJT18 has its only copy of *CDC45* as a heat inducible degron (*cdc45-td*) and a copy of *GAL-UBR1* to enhance inactivation of the Cdc45 protein. YJT18 and its parental strain YKL83 (just *GAL-UBR1*) were grown to mid log-phase in YPRaf at 24°C. Under these conditions *GAL-UBR1* is not induced and the Cdc45 protein is stable. The α-factor pheromone was added to synchronize the cells. Two hours later, once the cells were arrested, they were washed into YPRaf/Gal plus α-factor at 37°C and held for additional 45min to induce degradation of Cdc45 in YJT18 ($Cdc45^-$) but not in YKL83 ($Cdc45^+$). Cells were then washed and released into the cell cycle at 37°C in YPRaf/Gal and samples were taken at the time of release and every 6min after for 102min (Supplementary Figure 2). Reference RNA was made from an asynchronous culture of the parental strain YKL83 in environmental conditions similar to these of the synchronized cultures, i.e., grown in YPRaf at 24°C to mid log-phase and then harvested and resuspended in YPRaf/Gal at 37°C for two hours.



Supplementary Figure 2. Flow cytometry (FACS) measurements show that the cellular DNA content does not change in the culture of the Cdc45-shutoff strain (YJT18) at the inactivated condition of $Cdc45^-$, but doubles in the culture of its parental strain (YKL83) at the condition of $Cdc45^+$ as described (Tercero *et al*, 2000).

Supplementary Table II (on pp. 4 and 5). The DNA microarray experiments in this study. (*a*) Hybridization batches of the six even time points from each of the four pairs of shutoff and control time courses. (*b*) Hybridization batches of the six odd time points.

1.3. DNA Microarray Experiments

RNA isolation, construction of polylysine-coated DNA microarrays at the University of Texas (UT), sample labeling for and hybridization to the UT microarrays, and data acquisition (Supplementary information Dataset 1) were as described (Hu *et al*, 2007). The epoxy-coated DNA microarray platform of the Washington University (WU) Microarray Core and related experimental protocols (Supplementary information Dataset 2) were as described (Gerke *et al*, 2006).

Experimental variation was designed to be orthogonal to the biological variation. Differences in sample batch, DNA microarray platform and protocols were implemented among and prevented within pairs of shutoff and control time courses. Samples were hybridized in batches of odd or even time points from each of the four pairs of shutoff and control time courses (Supplementary Table II).

2. Mathematical Framework: HOSVD

The structure of the data in this study is of an order higher than that of a matrix. Each of the biological and experimental settings represents a degree of freedom in a cuboid, i.e., a third-order tensor of mRNA expression of genes × "x-settings," i.e., settings of the experimental variable x, e.g., time points × "y-settings," e.g., time courses. Unfolded into a matrix these degrees of freedom are lost and much of the information in the data tensor might also be lost.

We integrate these data by using a tensor higher-order singular value decomposition (HOSVD) (Supplementary information Section 3, and Mathematica Notebook). This tensor HOSVD (Supplementary information Section 2.1) was recently reformulated such that it separates the data tensor into a weighted sum of combinations, i.e., "subtensors," of three patterns each: one "eigenarray," i.e., a pattern of mRNA expression variation across the genes, one "x-eigengene," i.e., a pattern of expression across the x-settings, and one "y-eigengene," i.e., a pattern of expression across the y-settings. Each of these sets of mathematically orthogonal patterns, the eigenarrays, x-eigengenes and y-eigengenes, is computed by using the matrix singular value decomposition (SVD) (Supplementary information Section 2.2), as described (Golub and Van Loan, 1996; Alter *et al*, 2000; Nielsen *et al*, 2002; Alter and Golub, 2004; Alter, 2006; Li and Klevecz, 2006; Omberg *et al*, 2007).

(a)	Array ID	Sample Batch	Strain	Condition	Time	Time Point	Microarray Platform	Hybridization Batch	Synchronized Culture Label
1	SC18-083	1	W303-1a	$Cdc6^{+}$	5 min	2	UT	1	Cy5 (Red)
2	SC18-084	1	W303-1a	$Cdc6^{+}$	15 min	4	UT	1	Cy5 (Red)
3	SC18-085	1	W303-1a	$Cdc6^{+}$	25 min	6	UT	1	Cy5 (Red)
4	SC18-086	1	W303-1a	$Cdc6^{+}$	35 min	8	UT	1	Cy5 (Red)
5	SC18-087	1	W303-1a	$Cdc6^{+}$	45 min	10	UT	1	Cy5 (Red)
6	SC18-088	1	W303-1a	$Cdc6^{+}$	55 min	12	UT	1	Cy5 (Red)
7	SC18-091	1	YGP81	$Cdc6^{-}$	5 min	2	UT	1	Cy5 (Red)
8	SC18-093	1	YGP81	$Cdc6^{-}$	15 min	4	UT	1	Cy5 (Red)
9	SC18-092	1	YGP81	$Cdc6^{-}$	25 min	6	UT	1	Cy5 (Red)
10	SC18-094	1	YGP81	$Cdc6^{-}$	35 min	8	UT	1	Cy5 (Red)
11	SC18-095	1	YGP81	$Cdc6^{-}$	45 min	10	UT	1	Cy5 (Red)
12	SC18-096	1	YGP81	$Cdc6^{-}$	55 min	12	UT	1	Cy5 (Red)
13	STL-010	2	W303-1a	$Cdc6^{+}$	5 min	2	WU	2	Cy3 (Green)
14	STL-011	2	W303-1a	$Cdc6^{+}$	15 min	4	WU	2	Cy3 (Green)
15	STL-012	2	W303-1a	$Cdc6^{+}$	25 min	6	WU	2	Cy3 (Green)
16	STL-013	2	W303-1a	$Cdc6^{+}$	35 min	8	WU	2	Cy3 (Green)
17	STL-014	2	W303-1a	$Cdc6^{+}$	45 min	10	WU	2	Cy3 (Green)
18	STL-015	2	W303-1a	$Cdc6^{+}$	55 min	12	WU	2	Cy3 (Green)
19	STL-001	2	YGP81	$Cdc6^{-}$	5 min	2	WU	2	Cy3 (Green)
20	STL-002	2	YGP81	$Cdc6^{-}$	15 min	4	WU	2	Cy3 (Green)
21	STL-003	2	YGP81	$Cdc6^{-}$	25 min	6	WU	2	Cy3 (Green)
22	STL-004	2	YGP81	$Cdc6^{-}$	35 min	8	WU	2	Cy3 (Green)
23	STL-005	2	YGP81	$Cdc6^{-}$	45 min	10	WU	2	Cy3 (Green)
24	STL-006	2	YGP81	$Cdc6^{-}$	55 min	12	WU	2	Cy3 (Green)
25	STL-107	3	YKL83	$Cdc45^{+}$	6 min	2	WU	3	Cy5 (Red)
26	STL-108	3	YKL83	$Cdc45^{+}$	18 min	4	WU	3	Cy5 (Red)
27	STL-109	3	YKL83	$Cdc45^{+}$	30 min	6	WU	3	Cy5 (Red)
28	STL-110	3	YKL83	$Cdc45^{+}$	42 min	8	WU	3	Cy5 (Red)
29	STL-111	3	YKL83	$Cdc45^{+}$	54 min	10	WU	3	Cy5 (Red)
30	STL-112	3	YKL83	$Cdc45^{+}$	66 min	12	WU	3	Cy5 (Red)
31	STL-097	3	YJT18	$Cdc45^{-}$	6 min	2	WU	3	Cy5 (Red)
32	STL-098	3	YJT18	$Cdc45^{-}$	18 min	4	WU	3	Cy5 (Red)
33	STL-099	3	YJT18	$Cdc45^{-}$	30 min	6	WU	3	Cy5 (Red)
34	STL-100	3	YJT18	$Cdc45^{-}$	42 min	8	WU	3	Cy5 (Red)
35	STL-102	3	YJT18	$Cdc45^{-}$	54 min	10	WU	3	Cy5 (Red)
36	STL-103	3	YJT18	$Cdc45^{-}$	66 min	12	WU	3	Cy5 (Red)
37	STL-169	4	YKL83	$Cdc45^{+}$	6 min	2	WU	4	Cy3 (Green)
38	STL-170	4	YKL83	$Cdc45^{+}$	18 min	4	WU	4	Cy3 (Green)
39	STL-171	4	YKL83	$Cdc45^{+}$	30 min	6	WU	4	Cy3 (Green)
40	STL-172	4	YKL83	$Cdc45^{+}$	42 min	8	WU	4	Cy3 (Green)
41	STL-173	4	YKL83	$Cdc45^{+}$	54 min	10	WU	4	Cy3 (Green)
42	STL-174	4	YKL83	$Cdc45^{+}$	66 min	12	WU	4	Cy3 (Green)
43	STL-161	4	YJT18	$Cdc45^{-}$	6 min	2	WU	4	Cy3 (Green)
44	STL-162	4	YJT18	$Cdc45^{-}$	18 min	4	WU	4	Cy3 (Green)
45	STL-163y	4	YJT18	$Cdc45^{-}$	30 min	6	WU	4	Cy3 (Green)
46	STL-164y	4	YJT18	$Cdc45^{-}$	42 min	8	WU	4	Cy3 (Green)
47	STL-165	4	YJT18	$Cdc45^{-}$	54 min	10	WU	4	Cy3 (Green)
48	STL-166	4	YJT18	$Cdc45^{-}$	66 min	12	WU	4	Cy3 (Green)

Supplementary Table II*a*. The DNA microarray experiments in this study. Hybridization batches of the six even time points from each of the four pairs of shutoff and control time courses.

(b)	Array ID	Sample Batch	Strain	Condition	Time	Time Point	Microarray Platform	Hybridization Batch	Synchronized Culture Label
49	SC18-074	1	W303-1a	$Cdc6^+$	0 min	1	UT	5	Cy5 (Red)
50	SC18-075	1	W303-1a	$Cdc6^+$	10 min	3	UT	5	Cy5 (Red)
51	SC18-076	1	W303-1a	$Cdc6^+$	20 min	5	UT	5	Cy5 (Red)
52	SC18-077	1	W303-1a	$Cdc6^+$	30 min	7	UT	5	Cy5 (Red)
53	SC18-078	1	W303-1a	$Cdc6^+$	40 min	9	UT	5	Cy5 (Red)
54	SC18-079	1	W303-1a	$Cdc6^+$	50 min	11	UT	5	Cy5 (Red)
55	SC18-065	1	YGP81	$Cdc6^-$	0 min	1	UT	5	Cy5 (Red)
56	SC18-066	1	YGP81	$Cdc6^-$	10 min	3	UT	5	Cy5 (Red)
57	SC18-067	1	YGP81	$Cdc6^-$	20 min	5	UT	5	Cy5 (Red)
58	SC18-068	1	YGP81	$Cdc6^-$	30 min	7	UT	5	Cy5 (Red)
59	SC18-069	1	YGP81	$Cdc6^-$	40 min	9	UT	5	Cy5 (Red)
60	SC18-070	1	YGP81	$Cdc6^-$	50 min	11	UT	5	Cy5 (Red)
61	STL-019	2	W303-1a	$Cdc6^+$	0 min	1	WU	6	Cy3 (Green)
62	STL-020	2	W303-1a	$Cdc6^+$	10 min	3	WU	6	Cy3 (Green)
63	STL-021	2	W303-1a	$Cdc6^+$	20 min	5	WU	6	Cy3 (Green)
64	STL-022	2	W303-1a	$Cdc6^+$	30 min	7	WU	6	Cy3 (Green)
65	STL-023	2	W303-1a	$Cdc6^+$	40 min	9	WU	6	Cy3 (Green)
66	STL-024	2	W303-1a	$Cdc6^+$	50 min	11	WU	6	Cy3 (Green)
67	STL-041	2	YGP81	$Cdc6^-$	0 min	1	WU	6	Cy3 (Green)
68	STL-042	2	YGP81	$Cdc6^-$	10 min	3	WU	6	Cy3 (Green)
69	STL-043	2	YGP81	$Cdc6^-$	20 min	5	WU	6	Cy3 (Green)
70	STL-044	2	YGP81	$Cdc6^-$	30 min	7	WU	6	Cy3 (Green)
71	STL-045	2	YGP81	$Cdc6^-$	40 min	9	WU	6	Cy3 (Green)
72	STL-046	2	YGP81	$Cdc6^-$	50 min	11	WU	6	Cy3 (Green)
73	STL-087	3	YKL83	$Cdc45^+$	0 min	1	WU	7	Cy5 (Red)
74	STL-088	3	YKL83	$Cdc45^+$	12 min	3	WU	7	Cy5 (Red)
75	STL-089	3	YKL83	$Cdc45^+$	24 min	5	WU	7	Cy5 (Red)
76	STL-090	3	YKL83	$Cdc45^+$	36 min	7	WU	7	Cy5 (Red)
77	STL-091	3	YKL83	$Cdc45^+$	48 min	9	WU	7	Cy5 (Red)
78	STL-092	3	YKL83	$Cdc45^+$	60 min	11	WU	7	Cy5 (Red)
79	STL-077	3	YJT18	$Cdc45^-$	0 min	1	WU	7	Cy5 (Red)
80	STL-078	3	YJT18	$Cdc45^-$	12 min	3	WU	7	Cy5 (Red)
81	STL-079	3	YJT18	$Cdc45^-$	24 min	5	WU	7	Cy5 (Red)
82	STL-080	3	YJT18	$Cdc45^-$	36 min	7	WU	7	Cy5 (Red)
83	STL-081	3	YJT18	$Cdc45^-$	48 min	9	WU	7	Cy5 (Red)
84	STL-082	3	YJT18	$Cdc45^-$	60 min	11	WU	7	Cy5 (Red)
85	STL-151	4	YKL83	$Cdc45^+$	0 min	1	WU	8	Cy3 (Green)
86	STL-152	4	YKL83	$Cdc45^+$	12 min	3	WU	8	Cy3 (Green)
87	STL-153	4	YKL83	$Cdc45^+$	24 min	5	WU	8	Cy3 (Green)
88	STL-154	4	YKL83	$Cdc45^+$	36 min	7	WU	8	Cy3 (Green)
89	STL-155	4	YKL83	$Cdc45^+$	48 min	9	WU	8	Cy3 (Green)
90	STL-156	4	YKL83	$Cdc45^+$	60 min	11	WU	8	Cy3 (Green)
91	STL-139	4	YJT18	$Cdc45^-$	0 min	1	WU	8	Cy3 (Green)
92	STL-141	4	YJT18	$Cdc45^-$	12 min	3	WU	8	Cy3 (Green)
93	STL-142	4	YJT18	$Cdc45^-$	24 min	5	WU	8	Cy3 (Green)
94	STL-143	4	YJT18	$Cdc45^-$	36 min	7	WU	8	Cy3 (Green)
95	STL-145	4	YJT18	$Cdc45^-$	48 min	9	WU	8	Cy3 (Green)
96	STL-146	4	YJT18	$Cdc45^-$	60 min	11	WU	8	Cy3 (Green)

Supplementary Table II*b*. The DNA microarray experiments in this study. Hybridization batches of the six odd time points from each of the four pairs of shutoff and control time courses.

The SVD (Golub and Van Loan, 1996) separates a data matrix of, e.g., mRNA expression of genes × arrays into a weighted sum of combinations, i.e., "submatrices," of two patterns each: an eigenarray and its corresponding eigengene. The significance of each combination relative to all other combinations is defined in terms of the fraction of the overall mRNA expression information that this combination captures in the data matrix, and is proportional to its weight in this sum. The distribution of the overall expression information among the different combinations defines the "normalized entropy," i.e., information content of the data matrix. It was shown that the mathematical variables of the SVD, i.e., the significant and unique eigenarrays and eigengenes, can be interpreted in terms of the independent cellular states and the corresponding biological processes and experimental artifacts that compose the data matrix (Alter *et al*, 2000; Nielsen *et al*, 2002; Alter, 2006; Li and Klevecz, 2006).

The eigenarrays and eigengenes are in general unique, except in degenerate submatrix spaces, defined by eigenarrays and corresponding eigengenes that are of equal significance in the data matrix. Because the mathematical separation of a degenerate submatrix space into a weighted sum of combinations of eigenarrays and corresponding eigengenes is not unique, this separation may not be biologically interpreted. It was shown, however, that a unique orthogonal rotation in a degenerate submatrix space can be defined, that enables the interpretation of this subspace, by subjecting the rotated eigenarrays and corresponding eigengenes to unique constraints based on the experimental and biological settings (Alter *et al*, 2000). For example, such a unique rotation can be selected that effectively decouples patterns that correlate with the experimental variation from patterns that correlate with the biological variation.

The SVD can also be used to define a reconstruction of a data matrix in the subspace spanned by several selected eigenarrays and eigengenes, effectively filtering out all the remaining eigenarrays and eigengenes. This SVD reconstruction was shown to simulate observation of only the cellular states and biological processes that the selected eigenarrays and eigengenes represent. For example, this SVD reconstruction can be used to remove experimental artifacts from the data matrix (Alter *et al*, 2000; Nielsen *et al*, 2002; Li and Klevecz, 2006). Thus, it was shown that the mathematical operations of the SVD, e.g., classification, rotation or reconstruction in a subspace of selected eigenarrays and eigengenes, can be interpreted in terms of biological or experimental reality (Alter, 2006).

In analogy with the SVD, it was shown that the reformulation of the HOSVD enables its biological interpretation by mathematically defining (Omberg *et al*, 2007): (*i*) the relative significance of each combination of an eigenarray, an x-eigengene and a y-eigengene, in terms of the fraction of the overall mRNA expression information that this combination captures in the data tensor (Supplementary information Section 2.3.1); (*ii*) the normalized entropy, i.e., the information content of the data tensor in terms of the distribution of the overall expression information among the different combinations (Supplementary information Section 2.3.2); and (*iii*) a unique rotation of either the eigenarrays, the x-eigengenes or the y-eigengenes that span a degenerate subtensor space (Supplementary information Section 2.3.3).

In this study, we extend the analogy between the SVD and this HOSVD by using the reformulated HOSVD to mathematically define the reconstruction of a data tensor in the subspace of several selected eigenarrays, x-eigengenes and y-eigengenes, effectively filtering out all patterns of expression variation that are orthogonal to these selected patterns (Supplementary information Section 2.4).

2.1. HOSVD Definition

Let the data be structured as a third-order tensor, i.e., a cuboid, of mRNA expression of K-genes × L-x-settings × M-y-settings, where $K > LM$. Each element of the data tensor $\mathcal{T}$, i.e., $\mathcal{T}_{klm}$, is the expression measured for the kth gene under the lth x-setting and mth y-setting. Each column vector of $\mathcal{T}$, i.e., $\mathcal{T}_{:lm}$, lists the global gene expression measured under the lth x-setting and mth y-setting. The x-row vector $\mathcal{T}_{k:m}$ lists the expression measured for the kth gene under the mth y-setting across all x-settings. Similarly, the y-row vector $\mathcal{T}_{kl:}$ lists the expression measured for the kth gene under the lth x-setting across all y-settings.

The N-mode SVD, a tensor HOSVD, transforms the data tensor $\mathcal{T}$ to the reduced space of LM-eigenarrays × L-x-eigengenes × M-y-eigengenes by using the orthogonal transformation matrices U, V_x and V_y,

$$\mathcal{T} = \mathcal{R} \times_a U \times_b V_x \times_c V_y,$$
$$\mathcal{T}_{klm} = \sum_{a=1}^{LM} \sum_{b=1}^{L} \sum_{c=1}^{M} \mathcal{R}_{abc} U_{ka} V^T_{x,bl} V^T_{y,cm}, \qquad (1)$$

where $\times_a U$, $\times_b V_x$ and $\times_c V_y$ denote multiplications of the tensor $\mathcal{R}$ and the matrices U, V_x and V_y, which contract the first, second and third indices of $\mathcal{R}$ with the second indices of U, V_x and V_y or, equivalently, the first indices of U^T, V_x^T and V_y^T, respectively. In this space, the data are represented by the third-order "core tensor" $\mathcal{R}$, which in general is full.

The transformation matrix U defines the K-genes × LM-eigenarrays basis set, i.e., the set of LM orthonormal patterns of expression variation across the K genes. The vector in the ath column of U, i.e., $U_{:a}$, lists the global gene expression of the ath eigenarray. The transformation matrix V_x^T defines the L-x-eigengenes × L-x-settings, i.e., the set of L orthonormal patterns of gene expression variation across the L settings of the experimental variable x. The vector in the bth row of V_x^T, i.e., $V^T_{x,b:}$, lists the expression of the bth x-eigengene across all y-settings. Similarly, the transformation matrix V_y^T defines the M-y-eigengenes × M-y-settings. The vector in the cth row of V_y^T, i.e., $V^T_{y,c:}$, lists the expression of the cth y-eigengene across all x-settings.

The eigenarrays are orthonormal, i.e., normalized and orthogonal superpositions, i.e., weighted sums, of the global patterns of expression measured by the arrays. The x-eigengenes and the y-eigengenes are orthonormal superpositions of the gene expression patterns measured across the x-settings and y-settings, respectively.

This HOSVD holds for a tensor $\mathcal{T}$ of any order N. For a second-order tensor of N=2, i.e., a matrix, this HOSVD reduces to the matrix SVD (Golub and Van Loan, 1996).

2.2. HOSVD Computation by Using the SVD

The eigenarrays, which are listed in the transformation matrix U, are computed from the singular value decomposition (SVD) (Golub and Van Loan, 1996; Alter *et al*, 2000) of the matrix T_k, which is obtained by appending all column vectors of the data tensor along the K-genes axis,

$$T_k = (\mathcal{T}_{:11}, \ldots, \mathcal{T}_{:1M}, \ldots, \mathcal{T}_{:LM}) = UDV^T. \quad (2)$$

Note that U is independent of the order of the appended arrays. Similarly, the x-eigengenes and y-eigengenes, which are listed in the transformation matrices V_x and V_y, are computed from the SVD of the matrices T_l and T_m, which are obtained by appending all x-row vectors along the L-x-settings axis and all the y-row vectors along the M-y-settings axis, respectively,

$$T_l = (\mathcal{T}_{1:1}, \ldots, \mathcal{T}_{1:M}, \ldots, \mathcal{T}_{K:M}) = U_x D_x V_x^T, \quad (3)$$

$$T_m = (\mathcal{T}_{11:}, \ldots, \mathcal{T}_{1L:}, \ldots, \mathcal{T}_{KL:}) = U_y D_y V_y^T. \quad (4)$$

Following Equation (1), the core tensor $\mathcal{R}$ is computed by multiplying the data tensor $\mathcal{T}$ and the transformation matrices U, V_x and V_y,

$$\mathcal{R} = \mathcal{T} \times_k U^T \times_l V_x^T \times_m V_y^T. \quad (5)$$

The significance of each eigenarray, x-eigengene or y-eigengene is defined in terms of the fraction of the overall mRNA expression information that this pattern captures in the data matrix T_k, T_l or T_m, and is proportional to the corresponding singular value listed in D, D_x or D_y, respectively. These singular values are ordered in decreasing order, such that the patterns are ordered in U, V_x^T and V_y^T in decreasing order of their relative significance. Note that for a real data tensor the singular values are real and nonnegative.

It was shown that the mathematical variables of the SVD, i.e., the significant and unique eigenarrays and eigengenes, can be interpreted in terms of the independent cellular states and corresponding biological processes and experimental artifacts that compose the data matrix. The eigenarrays, x-eigengenes and y-eigengenes are in general unique, up to phase factors of ± 1 for a real data tensor, except in degenerate subspaces (Alter *et al*, 2000) . These subspaces are defined by equal singular values in either D, D_x or D_y, respectively. Eigenarrays, x-eigengenes or y-eigengenes that span degenerate subspaces may not be biologically interpreted.

It was shown, however, that a unique orthogonal rotation of these eigenarrays, x-eigengenes or y-eigengenes can be defined, that enables the interpretation of these patterns by subjecting the rotated patterns to unique constraints based on the experimental and biological settings (Alter *et al*, 2000). For example, the y-eigengenes $V_{y,c:}^T$ and $V_{y,m:}^T$, which satisfy $D_{y,cc} \approx D_{y,mm}$, span an approximately degenerate subspace (Supplementary Figure 5). Following the unique rotation R of these y-eigengenes, i.e., $RV_{y,c:}^T$ and $RV_{y,m:}^T$ (Supplementary Figure 6), the core tensor $\mathcal{R}$ is computed by multiplying the data tensor $\mathcal{T}$ and the transformation matrices U, V_x and the rotated RV_y,

$$\mathcal{R} = \mathcal{T} \times_k U^T \times_l V_x^T \times_m (RV_y^T),$$
$$RV_y^T = (V_{y,1:}^T, \ldots, RV_{y,c:}^T, \ldots, RV_{y,m:}^T, \ldots, V_{y,M:}^T). \quad (6)$$

Note that this core tensor $\mathcal{R}$ and the HOSVD of the data tensor $\mathcal{T}$ remain exact even if the subspace is only approximately degenerate.

2.3. HOSVD Reformulation

This HOSVD was recently reformulated in analogy with the SVD such that it separates the data tensor $\mathcal{T}$ into a superposition, i.e., a weighted sum, of $(LM)^2$ rank-1 subtensors,

$$\mathcal{T} = \sum_{a=1}^{LM} \sum_{b=1}^{L} \sum_{c=1}^{M} \mathcal{R}_{abc} \mathcal{S}(a,b,c),$$
$$\mathcal{S}(a,b,c) = U_{:a} \otimes V_{x,b:}^T \otimes V_{y,c:}^T, \quad (7)$$

where each subtensor $\mathcal{S}(a,b,c)$ is the outer product, denoted by $\otimes$, of three vectors, an eigenarray, an x-eigengene and a y-eigengene. The superposition coefficients, i.e., the weights, are the "higher-order singular values" tabulated in the core tensor $\mathcal{R}$. Note that for a real data tensor the higher-order singular values are real but not necessarily nonnegative.

The reformulated HOSVD separates the data tensor into a weighted sum of all possible combinations of three patterns of mRNA expression variation: one across the genes, one across the x-settings and one across the y-settings. This reformulation of the HOSVD enables its interpretation: It was shown that the significant and unique subtensors might be interpreted in terms of the cellular states, biological processess and experimental artifacts that compose the data tensor (Omberg *et al*, 2007).

2.3.1. *Relative significance of a subtensor.* The significance of a combination of patterns, i.e., a subtensor $\mathcal{S}(a,b,c)$, relative to all other combinations, is defined in terms of the "fraction" $\mathcal{P}_{abc}$ of the overall mRNA expression information that this combination captures in the data tensor, and is computed by using the higher-

order singular values tabulated in the core tensor $\mathcal{R}$,

$$0 \leq \mathcal{P}_{abc} \equiv \frac{\mathcal{R}^2_{abc}}{\sum_{a=1}^{LM}\sum_{b=1}^{L}\sum_{c=1}^{M}\mathcal{R}^2_{abc}} \leq 1. \quad (8)$$

Note that each fraction $\mathcal{P}_{abc}$ can be thought of as the probability that any given gene in the data tensor expresses both the corresponding bth x-eigengene and the cth y-eigengene.

2.3.2. *Normalized entropy of a data tensor.* The normalized "Shannon entropy," a measure of the information content or complexity of the data tensor, is defined in terms of the distribution of the overall mRNA expression information among the different subtensors, i.e., among the different combinations of patterns,

$$0 \leq d \equiv \frac{-1}{2\log(LM)}\sum_{a=1}^{LM}\sum_{b=1}^{L}\sum_{c=1}^{M}\mathcal{P}_{abc}\log(\mathcal{P}_{abc}) \leq 1. \quad (9)$$

A normalized entropy of $d = 0$ corresponds to an ordered and redundant data tensor in which all expression is captured by a single combination of one eigenarray, one x-eigengene and one y-eigengene. A normalized entropy of $d = 1$ corresponds to a disordered and random data tensor where all combinations are equally expressed.

2.3.3. *Approximately degenerate subtensor space rotation.* The space spanned by two or more subtensors is defined to be approximately degenerate if the higher-order singular values of these subtensors are approximately equal in magnitude and if $N-1=2$ of their $N=3$ indices are equal, such that they share all but one of the patterns of expression variation that compose them.

For example, the subtensors $\mathcal{S}(a,b,c)$ and $\mathcal{S}(k,b,c)$, that share the bth x-eigengene and the cth y-eigengene, and satisfy $|\mathcal{R}_{abc}| \approx |\mathcal{R}_{kbc}|$, span together an "approximately degenerate subtensor space," $\mathcal{R}_{abc}\mathcal{S}(a,b,c) + \mathcal{R}_{kbc}\mathcal{S}(k,b,c)$. The separation of this degenerate subtensor space into these two subtensors is not unique. These two mathematical subtensors, therefore, may not be biologically interpreted.

A single unique rank-1 subtensor $\mathcal{S}(a+k,b,c)$ is defined (Figure 1), that replaces the two subtensors in the superposition (Supplementary Figure 11), i.e., weighted sum of Equation (7),

$$\mathcal{R}_{a+k,b,c}\mathcal{S}(a+k,b,c) = \mathcal{R}_{abc}\mathcal{S}(a,b,c) + \mathcal{R}_{kbc}\mathcal{S}(k,b,c),$$
$$\mathcal{R}^2_{a+k,b,c} = \mathcal{R}^2_{a,b,c} + \mathcal{R}^2_{k,b,c}, \quad (10)$$

where $\mathcal{R}_{a+k,b,c}$ is the corresponding higher-order singular value of this subtensor. This subtensor is computed from the outer product of $U_{:,a+k}$, the normalized sum of the eigenarrays $U_{:a}$ and $U_{:k}$, weighted by the corresponding higher-order singular values $\mathcal{R}_{abc}$ and $\mathcal{R}_{kbc}$, and the shared x-eigengenes and y-eigengenes, $V^T_{x,b:}$ and $V^T_{y,c:}$,

$$\mathcal{S}(a+k,b,c) = U_{:,a+k} \otimes V^T_{x,b:} \otimes V^T_{y,c:},$$
$$U_{:,a+k} = \mathcal{R}^{-1}_{a+k,b,c}(\mathcal{R}_{abc}U_{:a} + \mathcal{R}_{kbc}U_{:k}). \quad (11)$$

Note that the HOSVD of the data tensor $\mathcal{T}$ remains exact even if the rotated subtensor space is only approximately degenerate. It was shown that this unique subtensor can be biologically interpreted (Omberg *et al*, 2007).

2.4. HOSVD Reconstruction

In analogy with the SVD, we use in this study the reformulated HOSVD to mathematically define the reconstruction of a data tensor in a subspace of several selected eigenarrays, x-eigengenes and y-eigengenes, effectively filtering out all patterns of expression variation that are orthogonal to these selected patterns. For example, HOSVD reconstruction in the subspace of the x-eigengenes and y-eigengenes that correlate with the biological variation across the x-settings and y-settings can be used to remove experimental artifacts from the data tensor. The transformation matrix U of Equation (2) is reconstructed by appending only the selected eigenarrays along the K-genes axis, and removing all remaining eigenarrays,

$$\begin{aligned} U &= (U_{:1}, \ldots, U_{:k}, \ldots, U_{:K}) \\ &\rightarrow (U_{:1}, \ldots, U_{:k}) \\ &= \tilde{U}. \end{aligned} \quad (12)$$

Similarly, the transformation matrices V_x and V_y of Equations (3) and (4) are reconstructed by appending only the selected x-eigengenes (Supplementary Figure 4) and y-eigengenes (Supplementary Figures 5 and 6) along the L-x-settings axis and the M-y-settings axis, and removing all remaining x-eigengenes and y-eigengenes, respectively,

$$\begin{aligned} V^T_x &= (V^T_{x,1:}, \ldots, V^T_{x,l:}, \ldots, V^T_{x,L:}) \\ &\rightarrow (V^T_{x,1:}, \ldots, V^T_{x,l:}) \\ &= \tilde{V}^T_x, \end{aligned} \quad (13)$$

$$\begin{aligned} V^T_y &= (V^T_{y,1:}, \ldots, V^T_{y,m:}, \ldots, V^T_{y,M:}) \\ &\rightarrow (V^T_{y,1:}, \ldots, V^T_{y,m:}) \\ &= \tilde{V}^T_y. \end{aligned} \quad (14)$$

The core tensor $\mathcal{R}$ is reconstructed by multiplying the data tensor $\mathcal{T}$ and the reconstructed transformation matrices $\tilde{U}$, $\tilde{V}_x$ and $\tilde{V}_y$,

$$\mathcal{R} \rightarrow \mathcal{T} \times_k \tilde{U}^T \times_l \tilde{V}^T_x \times_m \tilde{V}^T_y = \tilde{\mathcal{R}}. \quad (15)$$

The data tensor $\mathcal{T}$ is then reconstructed by multiplying the reconstructed core tensor $\tilde{\mathcal{R}}$ and transformation matrices $\tilde{U}$, $\tilde{V}_x$ and $\tilde{V}_y$,

$$\mathcal{T} \rightarrow \tilde{\mathcal{R}} \times_a \tilde{U} \times_b \tilde{V}_x \times_c \tilde{V}_y = \tilde{\mathcal{T}}. \quad (16)$$

Note that this reconstruction is mathematically equivalent to setting to zero the higher-order singular values in the core tensor $\mathcal{R}$, that correspond to the x-eigengenes and y-eigengenes which are to be removed or filtered out, and then computing the data tensor $\mathcal{T}$ following Equation (1). This HOSVD data tensor reconstruction, therefore, is analogous to the SVD reconstruction of a data matrix (Alter *et al*, 2000; Nielsen *et al*, 2002; Alter and Golub, 2004; Alter, 2006; Li and Klevecz, 2006), which was shown to give computationally similar results to those of the analysis of variance (ANOVA) following SVD detection of the experimental artifacts (Nielsen *et al*, 2002).

2.5. Subtensor Interpretation

A significant and unique subtensor is inferred to be representing an independent cellular state and the corresponding biological process or experimental artifact when a consistent biological or experimental theme is reflected in the interpretations of the patterns of the eigenarray, the x-eigengene and the y-eigengene that mathematically define the subtensor, taking into account the sign of the superposition coefficient of this subtensor, i.e., the sign of the corresponding higher-order singular value.

An eigenarray is parallel- and antiparallel-associated with the most likely parallel and antiparallel cellular states according to the annotations (Supplementary information Datasets 4 and 5) of the two groups of k genes each, with largest and smallest levels of expression in this eigenarray (Supplementary information Dataset 6) among all K genes, respectively. The P-value of a given association, i.e., $P(j;k,K,J)$, is calculated assuming hypergeometric probability distribution of the J annotations among the K genes, and of the subset of $j \subseteq J$ annotations among the subset of k genes, as described (Tavazoie *et al*, 1999),

$$P(j;k,K,J) = \binom{K}{k}^{-1} \sum_{i=j}^{k} \binom{J}{i} \binom{K-J}{k-i}. \qquad (17)$$

An x-eigengene or a y-eigengene is associated with a biological process or an experimental artifact when its pattern of expression variation across the x-settings or the y-settings, respectively, is interpretable.

2.6. HOSVD Data Classification

Inferring that subtensors represent independent cellular states and corresponding biological processes and experimental artifacts allows sorting the data by similarity in the expression of any chosen subset of subtensors.

For example, given three significant and unique subtensors that share the cth y-eigengene, two of which are $\mathcal{S}(a,b,c)$ and $\mathcal{S}(k,l,c)$, we plot the expression of each gene in the ath eigenarray along the $\theta = 0$-axis vs. that in the kth eigenarray along the $\theta = \pi/2$-axis, normalized by the overall expression of this gene in all three subtensors. In this plot, the distance of each gene from the origin is its amplitude of expression in the subtensor space spanned by the two subtensors relative to its amplitude of expression in the space spanned by the three subtensors. The "angular distance" of each gene from the 0-axis is its phase in the transition from the cellular state represented by the ath eigenarray to that represented by the kth eigenarray, $\theta_{:} = \arctan(U_{:k}/U_{:a})$. The genes are, therefore, sorted by their angular distances $\theta_{:}$.

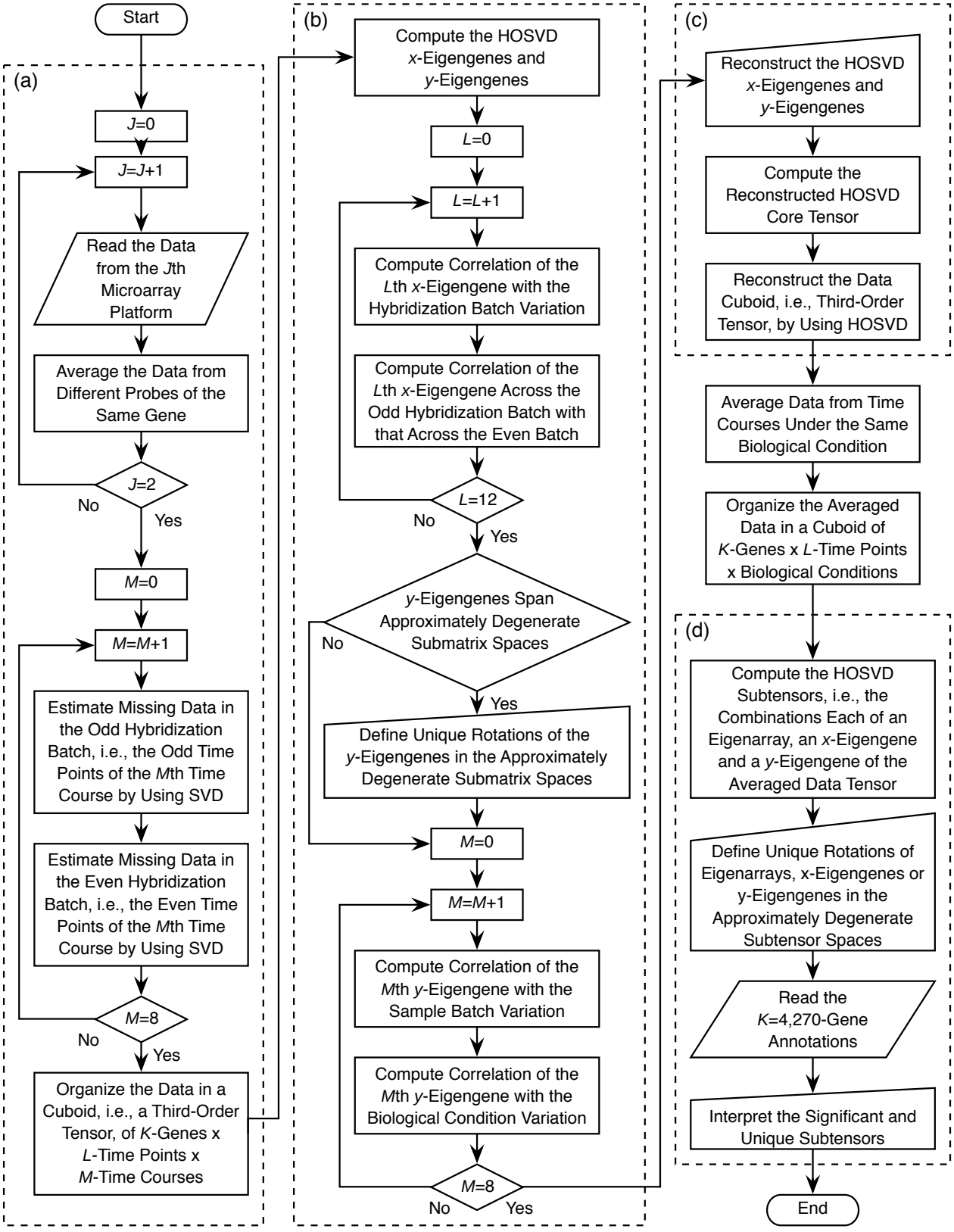
Similarly, we plot the expression of each x-setting in the bth x-eigengene along the $\theta = 0$-axis vs. that in the lth x-eigengene along the $\theta = \pi/2$-axis, normalized by the overall expression of this x-setting in all three subtensors. The distance of each x-setting from the origin is its amplitude of expression in the subspace spanned by the two subtensors relative to that by the three subtensors. The angular distance of each x-setting from the 0-axis is its phase in the progression from the biological process represented by the bth x-eigengene to that represented by the lth x-eigengene. The x-settings are sorted by $\theta_{:} = \arctan(V^T_{x,l:}/V^T_{x,b:})$ (Supplementary Figure 12).

Note that for visualization, the average of the expression of each gene across the x-settings is set to zero, such that gene expression is centered at its x-setting-invariant level (Box 1 and Figure 2, and Supplementary Figure 12).

3. Computational Data Integration

We organize the data in this study (Supplementary information Section 1) in a third-order cuboid, i.e., a tensor (Supplementary information Section 3.1), and computationally integrate it by using a tensor HOSVD (Supplementary information Section 2 and Figure 3, and Mathematica Notebook). This tensor HOSVD was recently reformulated and used in the integration of global gene expression from cell cycle time courses under different oxidative stress conditions. The picture that emerged identified conserved genes and cellular processes, some known from traditional assays and some previously unrecognized, as having significant roles in the time-dependent effects of these oxidative stress conditions on cell cycle progression (Omberg *et al*, 2007).

First, we use this HOSVD as described (Alter *et al*, 2000; Nielsen *et al*, 2002; Alter and Golub, 2004; Alter, 2006; Li and Klevecz, 2006; Omberg *et al*, 2007) to detect orthogonal patterns of expression variation across the time points and the time courses (Supplementary information Section 2.2). We show that this HOSVD enables decoupling of patterns that correlate with the experimental variation from patterns that correlate with the biological variation, because, in this study, the experimental variation was designed to be orthogonal to the biological variation (Supplementary information Section 3.2). Differences in sample batch, DNA microarray platform and protocols were implemented among and prevented within pairs of shutoff and control time courses. Expression patterns that vary among but are invariant within these pairs, therefore, correlate with the experimental



Supplementary Figure 3. Flowchart of the computational integration of the data. (*a*) Structure of the data cuboid (Supplementary information Section 3.1). (*b*) Detection of experimental artifacts (Supplementary information Section 3.2). (*c*) HOSVD data reconstruction and removal of experimental artifacts (Supplementary information Section 3.3). (*d*) Uncovering the cell cycle phase-dependent effects of Mcm2-7 origin binding (Supplementary information Section 3.4).

variation. Similarly, since the samples were hybridized in batches of odd or even time points from each of the pairs of shutoff and control time courses, patterns that are invariant across each hybridization batch, but show consistent variation between the batches, represent experimental artifacts. Patterns, that are invariant within the duplicated time courses and show consistent variation within the odd and even time points of the same time course, correlate with the biological variation, and are interpreted in terms of the cellular programs and biological processes that compose the data cuboid.

Second, we use this HOSVD to reconstruct the data cuboid in the subspace of gene expression patterns that correlate with the biological variation across the time points and the time courses, effectively filtering out, i.e., removing from the data cuboid, gene expression patterns that represent experimental artifacts (Supplementary information Section 2.4). We show that, in analogy with the SVD, this HOSVD enables filtering out experimental artifacts without eliminating genes or samples, since the patterns that correlate with experimental variation are decoupled from these that correlate with the biological variation (Supplementary information Section 3.3).

Third, after removing the experimental artifacts and averaging the duplicated time courses, we use this HOSVD as described (Supplementary information Section 2.3) to identify in the averaged data cuboid significant combinations of patterns of expression variation across the genes, the time points and the biological conditions of Mcm2-7 origin binding. We interpret these patterns in terms of the cell cycle phase-dependent effects of DNA replication and DNA replication origin activity on gene expression (Supplementary information Section 3.4).

3.1. Structure of the Data Cuboid

Integrating the data from the two different DNA microarray platforms, the $\log_2$ of the relative mRNA expression levels from different probes of the same gene in each DNA microarray are averaged. The 4270 genes that are selected have valid data in at least four time points in each of the six odd and six even samples of each time course, and in $\geq$92 of the 96 samples. These genes also have $\log_2$ of the relative mRNA expression $\geq$1, i.e., a twofold change in expression, in $\geq$12 of the 96 samples (Supplementary information Datasets 1 and 2). A relative expression level is defined valid if its signal to background ratio is >1 in both the synchronized culture and asynchronous reference.

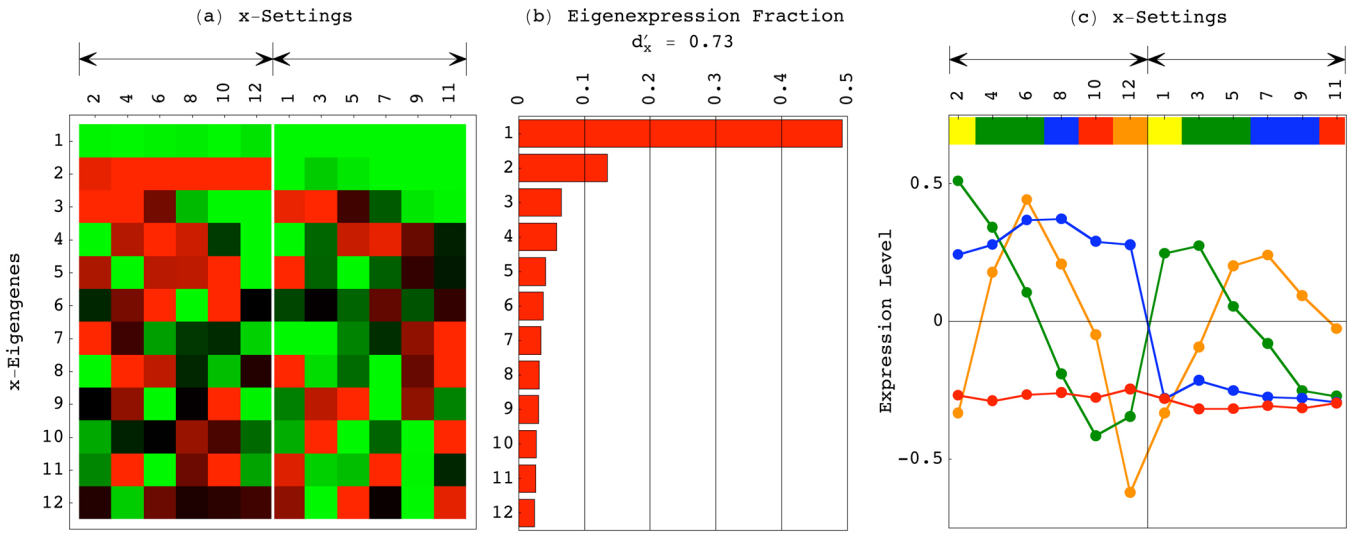
The structure of these data is of an order higher than that of a matrix. The biological and experimental settings, i.e., the time point or the cell cycle phase, and the time course or the strain, the environmental conditions, the sample batch and the DNA microarray platform and protocols, each represent a degree of freedom in a tensor. Unfolded into a matrix these degrees of freedom are lost. The data are, therefore, organized in a third-order tensor, i.e., a cuboid, tabulating the $\log_2$ of the relative mRNA expression of the 4270 genes across the 12 time points and across the eight time courses.

Of the 409,920 elements in the data cuboid, 2323 elements, i.e., <0.5%, are missing valid data. SVD (Golub and Van Loan, 1996; Alter *et al*, 2000) is used to estimate the missing data in the six odd and six even samples of each time course separately (Supplementary information Section 2.2) as described (Nielsen *et al*, 2002; Alter and Golub, 2004; Omberg *et al*, 2007). In each of these 16 sets of hybridization batches (Supplementary Table II), SVD of the expression patterns of the <4270 genes with no missing data uncovered six orthogonal patterns of gene expression, i.e., eigengenes. The most significant of these patterns, in terms of the fraction of the mRNA expression that it captures, is used to estimate the missing data in the remaining genes. For each of the 16 sets, the three most significant eigengenes and their corresponding fractions are almost identical to those computed for the 4270 genes after the missing data were estimated, with the corresponding correlations >0.99 (Supplementary information Mathematica Notebook). This suggests that the most significant eigengene, as computed for the genes with no missing data, is a valid pattern for estimation. This also indicates that this SVD estimation of missing data is robust to variations in the data selection cutoffs. After missing data estimation, the data from each sample are normalized to zero average and unit variance.

3.2. Detection of Experimental Artifacts

After missing data estimation, a tensor HOSVD of the data cuboid is used as described (Alter *et al*, 2000; Nielsen *et al*, 2002; Alter and Golub, 2004; Alter, 2006; Li and Klevecz, 2006; Omberg *et al*, 2007) to uncover orthogonal patterns of gene expression variation across the 12 time points and across the eight time courses, i.e., x-eigengenes and y-eigengenes. It was shown that significant and unique x-eigengenes and y-eigengenes, which are computed by using the SVD (Supplementary information Section 2.2), can be interpreted in terms of the independent biological processes and experimental artifacts that compose the data cuboid.

3.2.1. ***Hybridization batch artifacts.*** The first, third and fourth most significant x-eigengenes, which capture ~50%, 7% and 6% of the overall expression information in the data cuboid, respectively, describe variation that is consistent within the odd and even hybridization batches (Supplementary Figure 4). The second most significant x-eigengene describes invariant expression across each batch that varies consistently between the two batches. The correlation of this pattern with the experimental variation in hybridization batch is ~0.99. This pattern, therefore, is inferred to represent an hybridization batch artifact.



Supplementary Figure 4. The x-eigengenes, i.e., the HOSVD patterns of expression variation across the time points. (*a*) Raster display of the expression of the 12 x-eigengenes across the 12 time points, ordered by hybridization batch, with overexpression (red), no change in expression (black) and underexpression (green) around the time-invariant expression, that is represented by the first x-eigengene. The white line separates the even and the odd hybridization batches, indicated by the black arrows. (*b*) Bar chart of the fractions of mRNA expression that these orthogonal patterns capture in the data cuboid. The four most significant patterns capture ∼50%, 13%, 7% and 6% of the expression of the 4270 genes, respectively. (*c*) Line-joined graphs of the first (red), second (blue), third (green) and fourth (orange) expression patterns across the time points. The color bars indicate the cell cycle classifications of the time points in the averaged $Cdc6^+/45^+$ control (Supplementary Figure 7) as described (Spellman *et al*, 1998): M/G1 (yellow), G1 (green), S (blue), S/G2 (red) and G2/M (orange). The grid line separates the even and the odd hybridization batches, indicated by the black arrows. The first pattern is approximately time invariant. The second pattern correlates with the variation in hybridization batch. The third and fourth patterns describe oscillations consistent within the hybridization batches, that peak at M/G1 and G1/S and trough at S/G2 and G2/M, respectively.

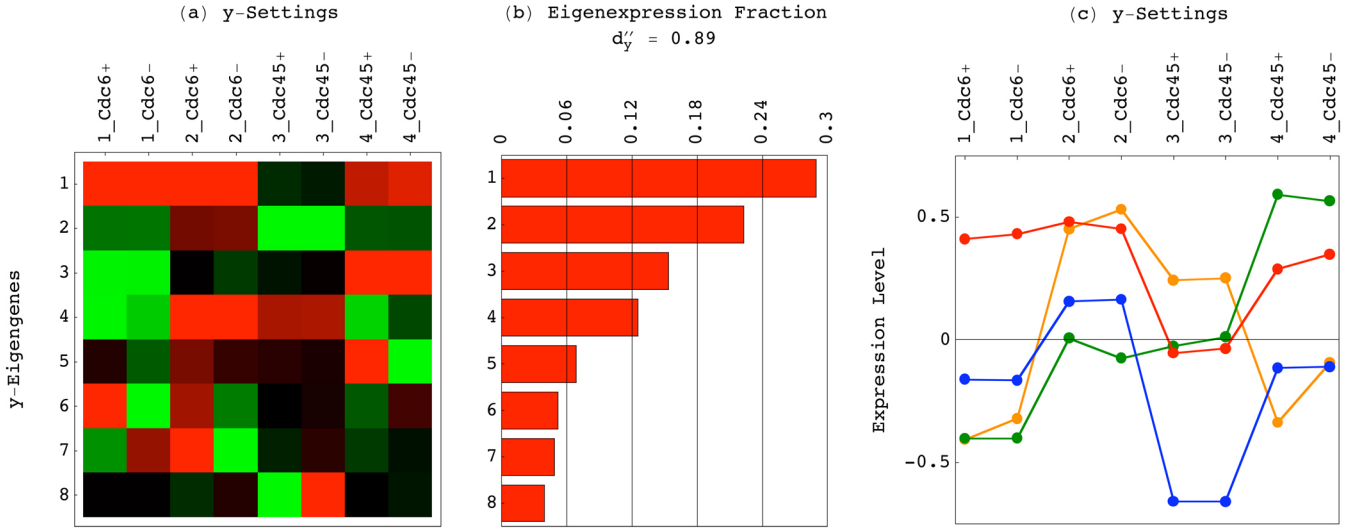
3.2.2. ***Sample batch and DNA microarray platform artifacts.*** The first and second most significant y-eigengenes capture each ∼25% of the expression of the 4270 genes and span an approximately degenerate submatrix space. The fifth through eighth y-eigengenes capture each ∼5% of the overall expression information, and span another approximately degenerate submatrix space (Supplementary Figure 5). These y-eigengenes are not unique and may not be biologically or experimentally interpreted.

We define, therefore, unique orthogonal rotations in these degenerate submatrix spaces (Supplementary information Section 2.2) as described (Alter *et al*, 2000). These rotations subject the rotated y-eigengenes to unique constraints based on the experimental variation in sample batch, DNA microarray platform and protocols, in order to decouple patterns that correlate with these experimental settings from patterns that correlate with the variation in the biological condition of Mcm2-7 origin binding across the time courses.

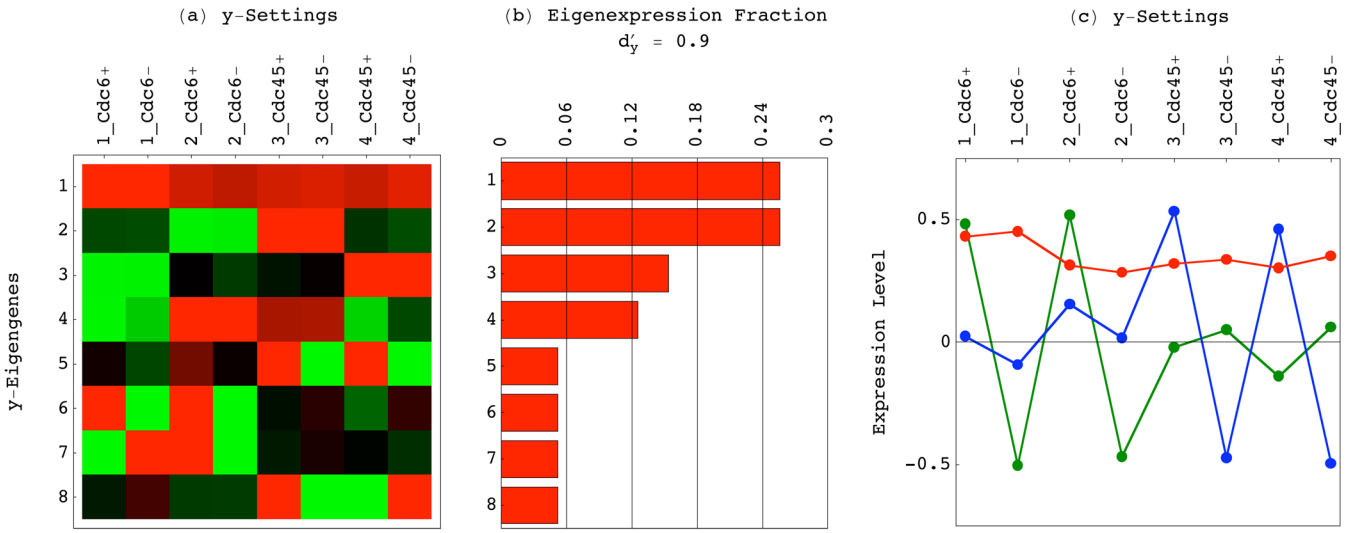
The first and second most significant y-eigengenes correlate with a time course-invariant pattern of expression as well as with a pattern of variation among and invariance within the four pairs of shutoff and control time courses (Supplementary Figure 5). Requiring the rotated second y-eigengene to be orthogonal to the time course-invariant pattern of expression, such that the rotated first y-eigengene is of maximal correlation with that pattern, gives a unique angle of rotation $\approx \pi/3$ in the two-dimensional submatrix space spanned by these two y-eigengenes,

$$V^T_{y,1:} \rightarrow RV^T_{y,1:} \approx -\sin(\pi/3)\, V^T_{y,1:} - \cos(\pi/3)\, V^T_{y,2:},$$
$$V^T_{y,2:} \rightarrow RV^T_{y,2:} \approx -\cos(\pi/3)\, V^T_{y,1:} + \sin(\pi/3)\, V^T_{y,2:}. \quad (18)$$

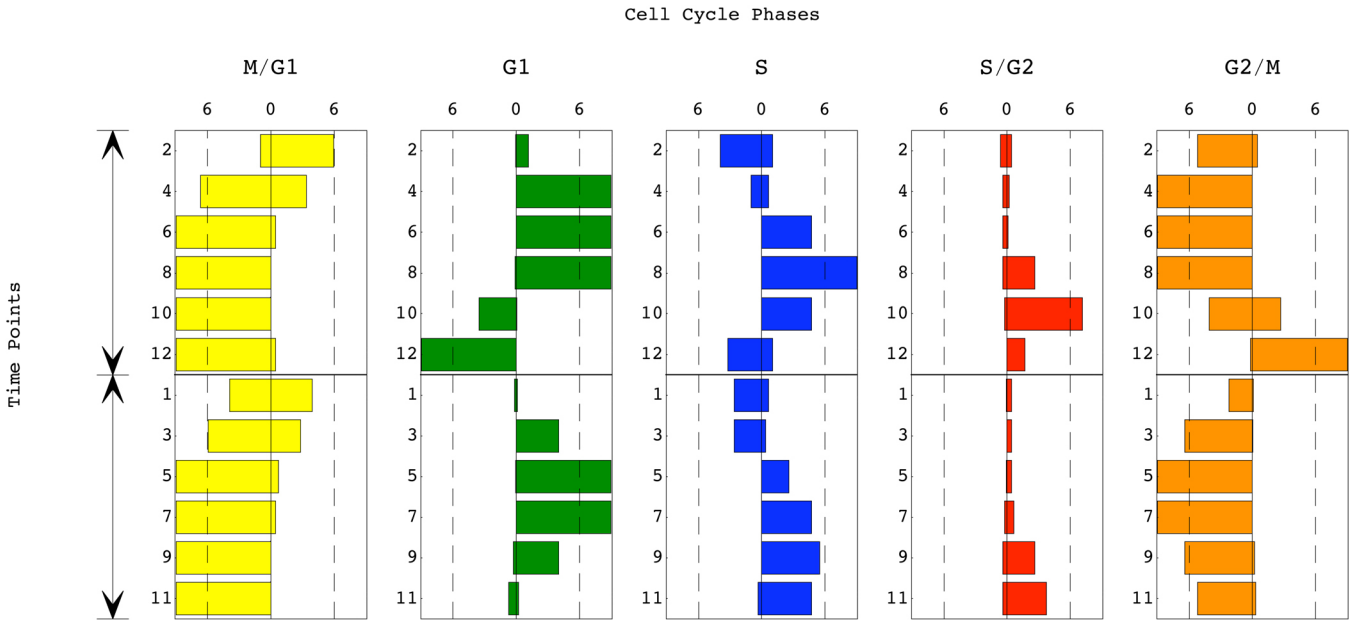
This unique rotation decouples the time course-invariant pattern, that corresponds to biological invariance among the time courses, and has ∼0.99 correlation with the rotated first y-eigengene, from the pattern of variation among and invariance within the four pairs of shutoff and control time courses, that corresponds to experimental variation among the time courses, and is described by the rotated second y-eigengene (Supplementary Figure 6).



Supplementary Figure 5. The y-eigengenes, i.e., the HOSVD patterns of expression variation across the time courses. (a) Raster display of the expression of the eight y-eigengenes across the eight time courses, ordered by sample batch (Supplementary Table II). (b) Bar chart of the fractions of mRNA expression that these orthogonal expression patterns capture. The first and second most significant patterns capture each ∼25% of the expression of the 4270 genes, and span an approximately degenerate submatrix space. The fifth through eighth patterns capture each ∼5% of the overall expression information, and span another approximately degenerate submatrix space. (c) Line-joined graphs of the first (red), second (blue), third (green) and fourth (orange) expression patterns across the time courses. These four patterns of variation among the four pairs of shutoff and control time courses correlate with a time course-invariant pattern of expression as well as with the variation in the experimental settings but not with the variation in the biological conditions of Mcm2-7 origin binding.



Supplementary Figure 6. The rotated y-eigengenes, i.e., the HOSVD patterns of expression variation across the time courses, after mathematical decoupling of the patterns that correlate with the experimental variation from the patterns that correlate with the biological variation. (a) Raster display of the expression of the eight rotated y-eigengenes across the eight time courses, ordered by sample batch (Supplementary Table II). (b) Bar chart of the approximate fractions of mRNA expression that these orthogonal patterns capture. (c) Line-joined graphs of the first (red), fifth (blue) and sixth (green) rotated expression patterns across the time courses. The first pattern correlates with a time course-invariant pattern of expression. The fifth pattern is of consistent variation between the two $Cdc45^+$ and the two $Cdc45^-$ time courses. The sixth pattern is of consistent variation between the two $Cdc6^+$ and the two $Cdc6^-$ time courses. These three patterns describe the variation in the biological condition of Mcm2-7 origin binding across the time courses.



Supplementary Figure 7. Probabilistic significance of the enrichment of the time points in the averaged Cdc6$^+$/45$^+$ control in overexpressed or underexpressed cell cycle-regulated genes. Following Equation (17), the P-value of each enrichment is calculated according to the cell cycle annotations of the genes (Supplementary information Dataset 4), assuming hypergeometric distribution of the J annotations among the K=4270 genes, and of the subset of $j \subseteq J$ annotations among the subset of k=200 genes with largest or smallest levels of expression at the corresponding time point, as described (Tavazoie *et al*, 1999). Bar chart of $-\log_{10}(P\text{-value})$ of the enrichment of each of the 12 time points in overexpressed (*right*) or underexpressed (*left*) M/G1 (yellow), G1 (green), S (blue), S/G2 (red) and G2/M (orange) genes. The grid line separates the even and the odd hybridization batches, indicated by the black arrows. These bar charts show that in the averaged control, the time points sample the exit from the pheromone-induced arrest and M/G1 (the first and second time points) and entry into G1 (the third through sixth time points) through S (the seventh through ninth time points) and S/G2 (10th and 11th time points) to the beginning of G2/M (12th time point) just before nuclear division. These are consistent with the flow cytometry, i.e., fluorescent-activated cell sorting (FACS) measurements of cell synchrony in the Cdc6$^+$ and Cdc45$^+$ cells (Supplementary Figures 1 and 2) as well as with previous global gene expression analyses of α-factor synchronized cultures (Spellman *et al*, 1998).

The sixth and seventh y-eigengenes describe expression variation among the Cdc6$^+$ and Cdc6$^-$ time courses, that is decoupled from that in the Cdc45$^+$ and Cdc45$^-$ time courses (Supplementary Figure 5). Requiring the rotated sixth y-eigengene to be of maximal correlation with a pattern of invariance within and consistent variation between the Cdc6$^+$ and Cdc6$^-$ time courses, such that the seventh rotated y-eigengene is almost orthogonal to that pattern, gives a unique angle of rotation $\approx \pi/3$ in the two-dimensional submatrix space spanned by these two y-eigengenes,

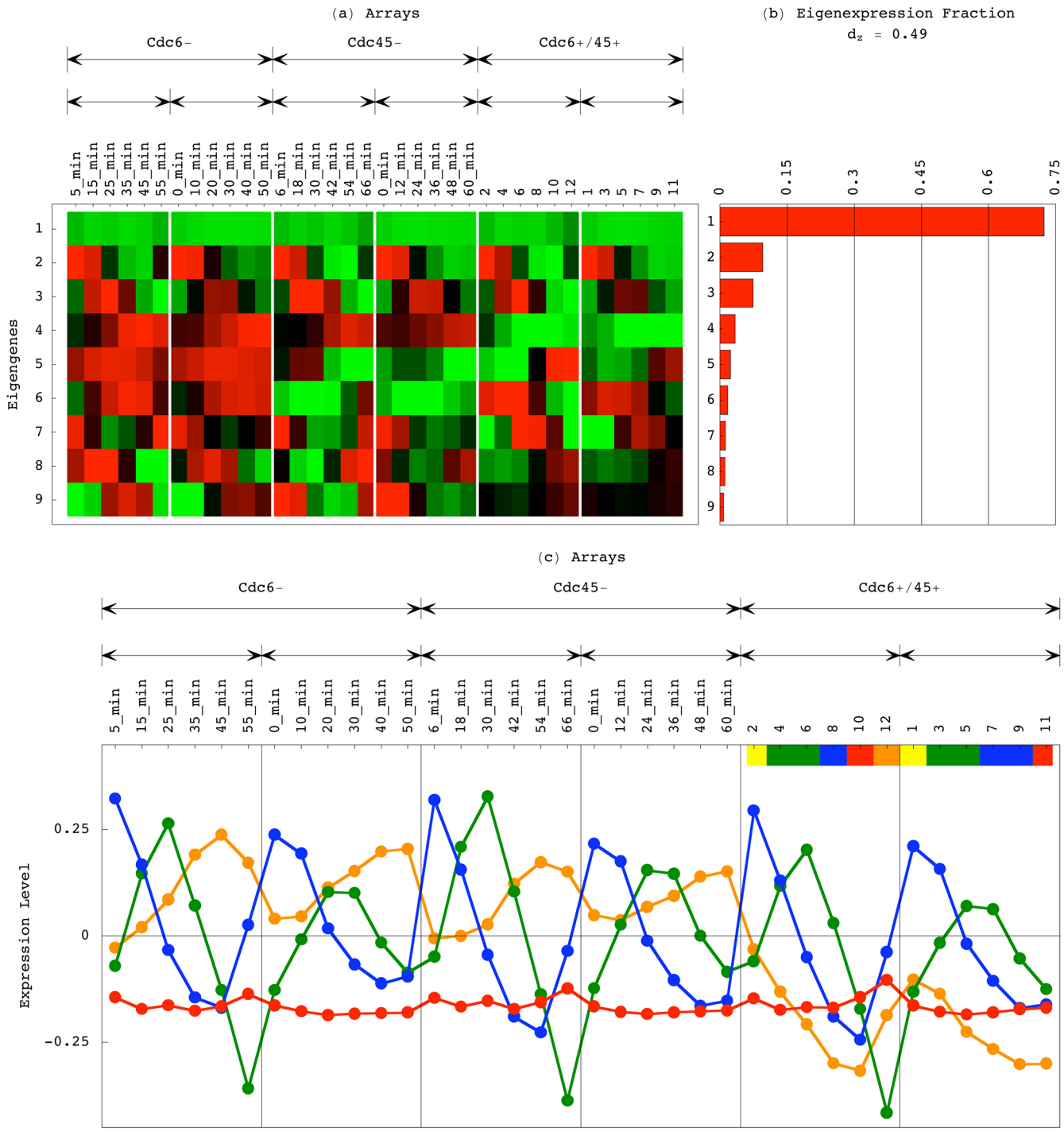
$$\begin{aligned} V^T_{y,6:} \rightarrow RV^T_{y,6:} &\approx \sin(\pi/3)\, V^T_{y,6:} + \cos(\pi/3)\, V^T_{y,7:}, \\ V^T_{y,7:} \rightarrow RV^T_{y,7:} &\approx -\cos(\pi/3)\, V^T_{y,6:} + \sin(\pi/3)\, V^T_{y,7:}. \end{aligned} \quad (19)$$

This unique rotation decouples the biological variation between the Cdc6$^+$ and Cdc6$^-$ time courses, that has $\sim$0.99 correlation with the rotated sixth y-eigengene, from the pattern of variation among and invariance within the two pairs of Cdc6-shutoff and control time courses, that corresponds to experimental variation among the time courses, and is described by the rotated seventh y-eigengene (Supplementary Figure 6).

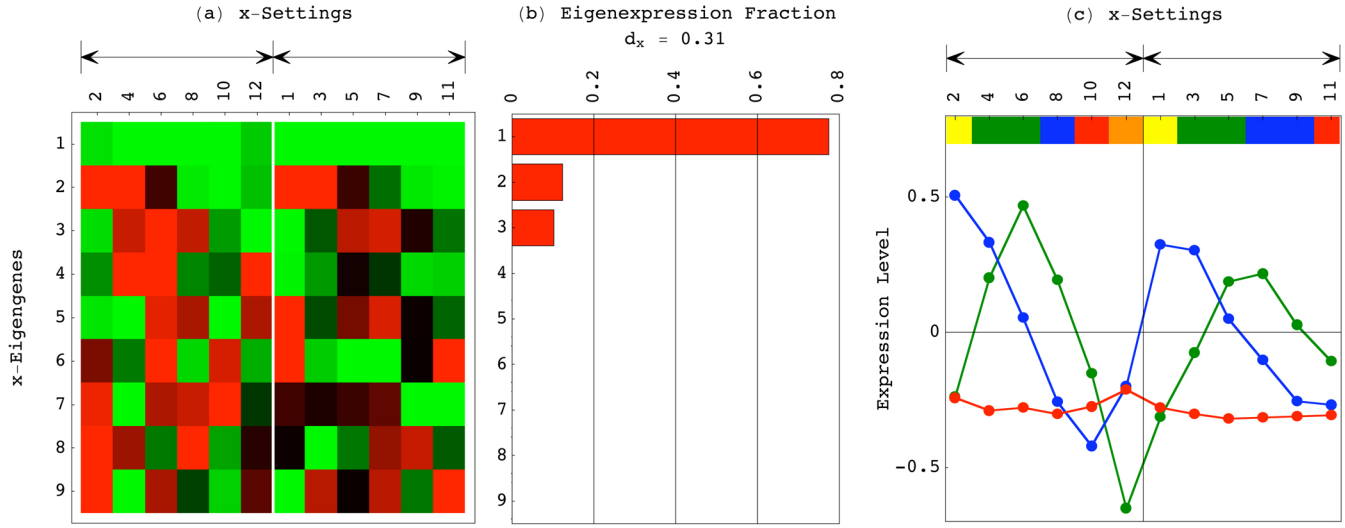
Similarly, requiring the rotated fifth y-eigengene to be of maximal correlation with a pattern of invariance within and consistent variation between the Cdc45$^+$ and Cdc45$^-$ time courses, such that the eighth rotated y-eigengene is almost orthogonal to that pattern, gives a unique angle of rotation $\approx \pi/4$ in the two-dimensional submatrix space spanned by these two y-eigengenes,

$$\begin{aligned} V^T_{y,5:} \rightarrow RV^T_{y,5:} &\approx -\sin(\pi/4)\, V^T_{y,5:} - \cos(\pi/4)\, V^T_{y,8:}, \\ V^T_{y,8:} \rightarrow RV^T_{y,8:} &\approx -\cos(\pi/4)\, V^T_{y,5:} + \sin(\pi/4)\, V^T_{y,8:}. \end{aligned} \quad (20)$$

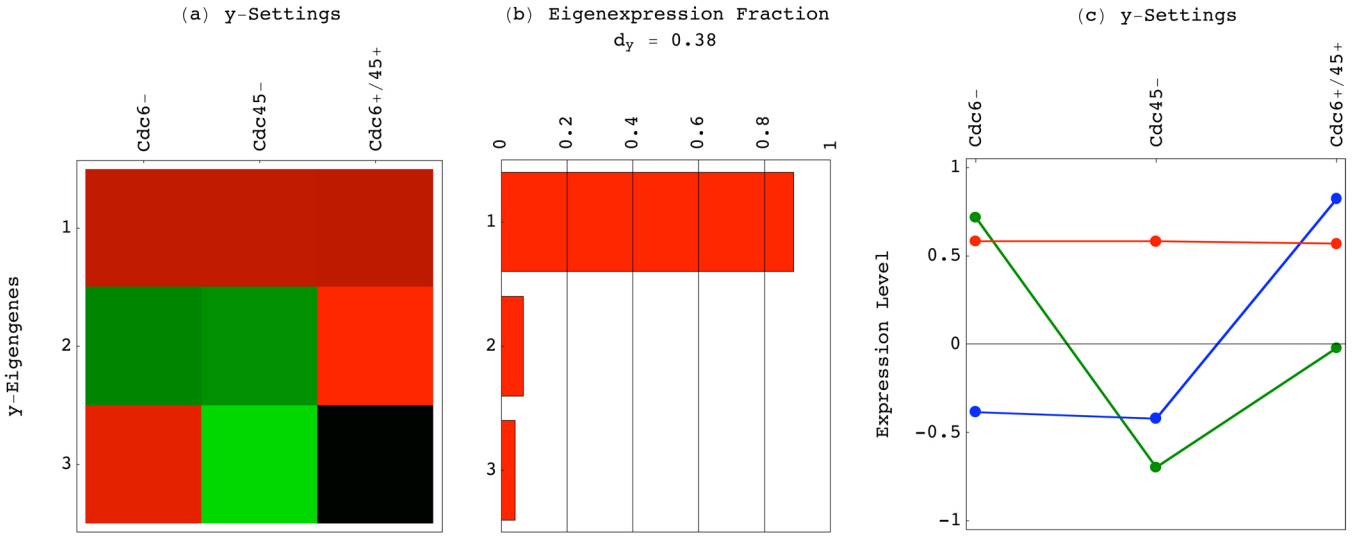
This unique rotation decouples the biological variation between the Cdc45$^+$ and Cdc45$^-$ time courses, that has $\sim$0.98 correlation with the rotated fifth y-eigengene, from the pattern of variation among and invariance within the two pairs of Cdc45-shutoff and control time courses, that corresponds to experimental variation among the time courses, and is described by the rotated eighth y-eigengene (Supplementary Figure 6).



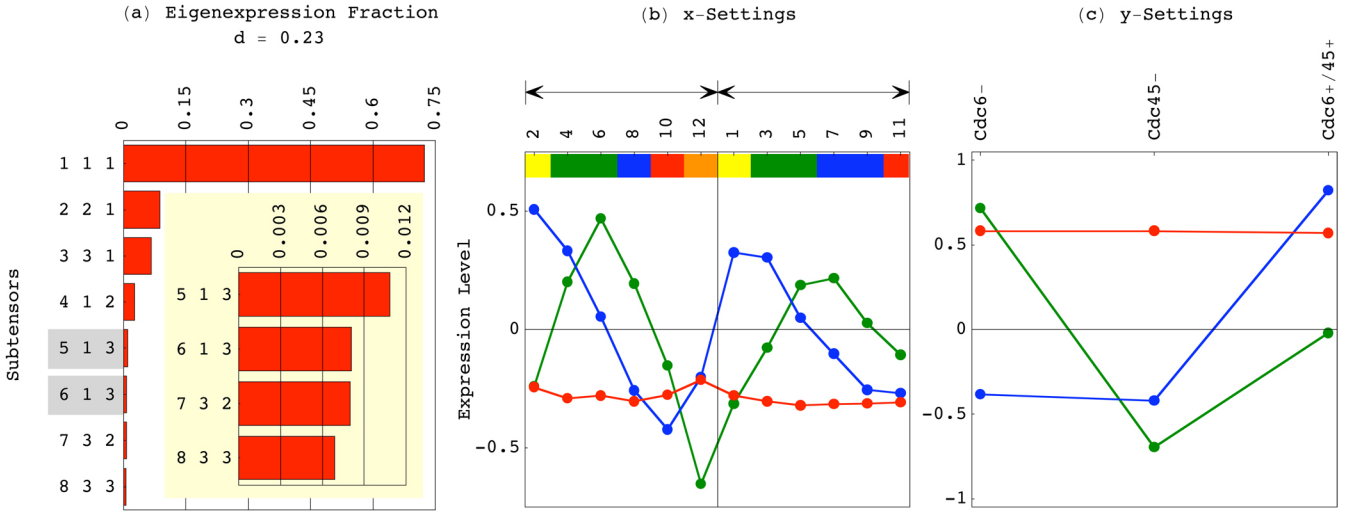
Supplementary Figure 8. The eigengenes V^T corresponding to the eigenarrays, i.e., the HOSVD patterns of expression variation across the 4270 genes, in the reconstructed and averaged data cuboid. (a) Raster display of the expression of the nine eigengenes across the 36 time points, ordered by condition and by hybridization batch, with overexpression (red), no change in expression (black) and underexpression (green) around the time-invariant expression, that is represented by the first eigengene. The white lines separate the conditions and the even and odd hybridization batches, also indicated by the black arrows. (*b*) Bar chart of the fractions of mRNA expression that these orthogonal patterns capture in the data cuboid. (*c*) Line-joined graphs of the first (red), second (blue), third (green) and fourth (orange) expression patterns across the conditions and time points. The color bars indicate the cell cycle classifications of the time points in the averaged Cdc6$^+$/45$^+$ control (Supplementary Figure 7) as described (Spellman *et al*, 1998): M/G1 (yellow), G1 (green), S (blue), S/G2 (red) and G2/M (orange). The grid lines separate the conditions and the even and odd hybridization batches, also indicated by the black arrows.



Supplementary Figure 9. The x-eigengenes, i.e., the HOSVD patterns of expression variation across the time points, in the reconstructed and averaged data cuboid. (a) Raster display of the expression of the 12 x-eigengenes across the 12 time points, ordered by hybridization batch, with overexpression (red), no change in expression (black) and underexpression (green) around the time-invariant expression, that is represented by the first x-eigengene. The white line separates the even and odd hybridization batches, indicated by the black arrows. (b) Bar chart of the fractions of mRNA expression that these orthogonal patterns capture in the data cuboid. The three most significant x-eigengenes, that were selected to reconstruct the data cuboid, capture >99% of the expression of the 4270 genes after the reconstruction. (c) Line-joined graphs of the first (red), second (blue) and third (green) expression patterns across the time points. The color bars indicate the cell cycle classifications of the time points in the averaged $Cdc6^+/45^+$ control (Supplementary Figure 7) as described (Spellman *et al*, 1998): M/G1 (yellow), G1 (green), S (blue), S/G2 (red) and G2/M (orange). The grid line separates the even and odd hybridization batches, indicated by the black arrows.



Supplementary Figure 10. The y-eigengenes, i.e., the HOSVD patterns of expression variation across the conditions, in the reconstructed and averaged data cuboid. (a) Raster display of the expression of the three y-eigengenes across the three conditions, with overexpression (red), no change in expression (black) and underexpression (green) around the condition-invariant expression, that is represented by the first y-eigengene. (b) Bar chart of the fractions of mRNA expression that these orthogonal patterns capture in the data cuboid. The first, second and third patterns capture ~89%, 7% and 4% of the overall expression information, respectively. (c) Line-joined graphs of the first (red), second (blue) and third (green) expression patterns across the conditions.



Supplementary Figure 11. Significant HOSVD subtensors before rotation in the approximately degenerate subtensor space spanned by $\mathcal{S}(5,1,3)$ and $\mathcal{S}(6,1,3)$. (*a*) Bar chart of the fractions of the eight most significant subtensors. The higher-order singular values corresponding to subtensors highlighted in gray, i.e., $\mathcal{S}(5,1,3)$ and $\mathcal{S}(6,1,3)$, are <0. The subtensor spaces $\mathcal{S}(5,1,3)$ and $\mathcal{S}(6,1,3)$ capture each $\sim$1% of the overall expression information in the data cuboid, and span an approximately degenerate subtensor space. The entropy of the data tensor is 0.23. (*b*) Line-joined graphs of the first (red), second (blue) and third (green) x-eigengenes, which define the expression variation across time in these subtensors. The color bars indicate the cell cycle classifications of the time points in the averaged Cdc6$^+$/45$^+$ control (Supplementary Figure 7). The grid line separates the even and odd hybridization batches, indicated by the black arrows. The first pattern is approximately time invariant. The second and third patterns describe oscillations consistent within the hybridization batches, that peak at M/G1 and G1/S and trough at S/G2 and G2/M, respectively. (*c*) Line-joined graphs of the first (red), second (blue) and third (green) y-eigengenes, which define the expression variation across the biological conditions. The first pattern is condition invariant. The second pattern correlates with underexpression in both conditions where DNA replication is prevented, i.e., in both the Cdc6$^-$ and Cdc45$^-$ cells, relative to the averaged Cdc6$^+$/45$^+$ control. The third pattern correlates with overexpression in the Cdc6$^-$ cells and underexpression in the Cdc45$^-$ cells relative to the control.

3.3. HOSVD Data Reconstruction

After the gene expression patterns that correlate with experimental variation are decoupled from these that correlate with the biological variation, we use this HOSVD to reconstruct the data cuboid in the subspace of the patterns that correlate with biological variation across the 12 time points and across the eight time courses, effectively filtering out patterns that represent experimental artifacts (Supplementary information Section 2.4).

Following Equations (13) and (14), the transformation matrices V_x and RV_y are reconstructed by appending only the selected x-eigengenes and y-eigengenes along the L-x-settings axis and the M-y-settings axis, and removing all remaining x-eigengenes and y-eigengenes, respectively,

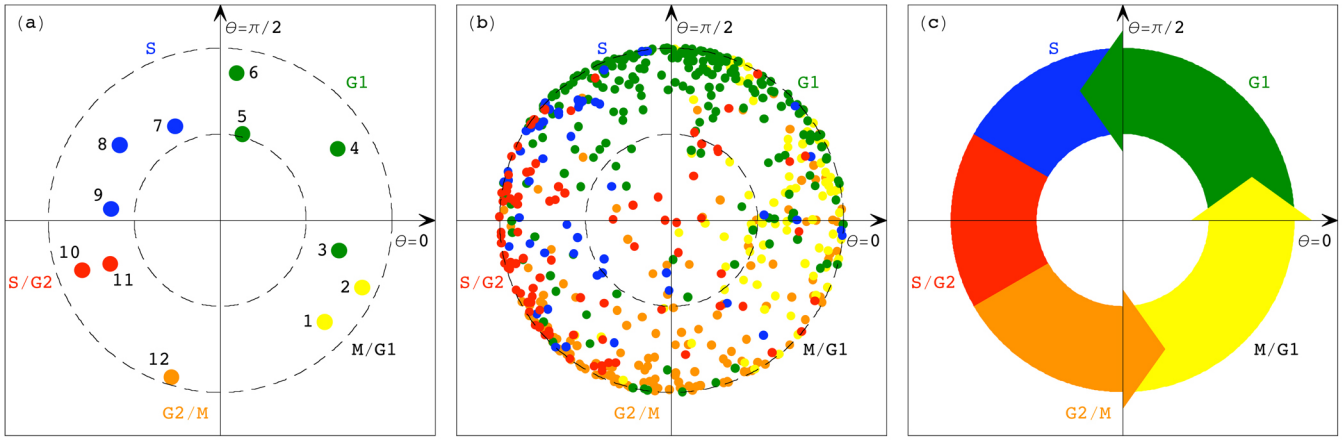
$$\begin{aligned} V_x^T &= (V_{x,1:}^T, \ldots, V_{x,12:}^T) \\ &\rightarrow (V_{x,1:}^T, V_{x,3:}^T, V_{x,4:}^T) = \tilde{V}_x^T, \qquad (21) \\ RV_y^T &= (RV_{y,1:}^T, \ldots, RV_{y,8:}^T) \\ &\rightarrow (RV_{y,1:}^T, RV_{y,5:}^T, RV_{y,6:}^T) = \tilde{V}_y^T. \qquad (22) \end{aligned}$$

Note that the rotated first, fifth and sixth y-eigengenes, that compose the reconstructed transformation matrix $\tilde{V}_y$ are orthogonal to gene expression variation between the Cdc6$^+$/$^-$ and Cdc45$^+$/$^-$ cells. Experimental artifacts that correlate with the variation in the environmental conditions or the parental genetic background between these cells are therefore also removed by the data cuboid reconstruction.

Following Equations (15) and (16), the reconstructed data tensor $\tilde{\mathcal{T}}$ is then computed by multiplying the data tensor $\mathcal{T}$ and the reconstructed transformation matrices $\tilde{V}_x$ and $\tilde{V}_y$,

$$\mathcal{T} \rightarrow (\mathcal{T} \times_l \tilde{V}_x^T \times_m \tilde{V}_y^T) \times_b \tilde{V}_x^T \times_c \tilde{V}_y^T = \tilde{\mathcal{T}}. \qquad (23)$$

Note that this HOSVD reconstruction of a data tensor (Supplementary information Section 2.4) is analogous to the previously described SVD reconstruction of a data matrix (Alter *et al*, 2000; Nielsen *et al*, 2002; Alter and Golub, 2004; Alter, 2006; Li and Klevecz, 2006), which was shown to give computationally similar results to those of the analysis of variance (ANOVA) following SVD detection of the experimental artifacts (Nielsen *et al*, 2002).



Supplementary Figure 12. The unperturbed cell cycle mRNA expression subspace spanned by the second and third HOSVD combinations, i.e., subtensors, $\mathcal{S}(2,2,1)$ and $\mathcal{S}(3,3,1)$ (Figure 1 and Table I). (*a*) The normalized expression of each time point in the second x-eigengene along the $\theta = 0$-axis vs. that in the third x-eigengene along the $\theta = \pi/2$-axis color-coded as described (Spellman *et al*, 1998) according to the cell cycle classification of the time point in the averaged Cdc6$^+$/45$^+$ control time course (Supplementary Figure 7). The distance of each time point from the origin is its amplitude of expression in the subspace spanned by these two x-eigengenes and therefore also by the two corresponding subtensors. The dashed half-unit and unit circles show that this subspace captures more than 25% of the global mRNA expression of each time point and close to 100% of the global expression of most of the time points, respectively. Sorting the time points according to their angular distances from the 0-axis is consistent with their order, which describes the progression through less than a period of the cell cycle, from M/G1 to G1 through S and S/G2 to the beginning of G2/M. This is consistent with the second and third x-eigengenes representing expression oscillations that peak at M/G1 and G1/S and trough at S/G2 and G2/M, respectively (Figure 1). (*b*) The normalized expression of each of the 576 cell cycle-regulated genes (Spellman *et al*, 1998) in the second eigenarray along the $\theta = 0$-axis vs. that in the third eigenarray along the $\theta = \pi/2$-axis, color-coded as described according to the cell cycle classification of the gene. The second and third eigenarrays which describe the global gene expression in the $\mathcal{S}(2,2,1)$ and $\mathcal{S}(3,3,1)$ HOSVD combinations are enriched with overexpressed M/G1 and G1/S genes and underexpressed S/G2 and G2/M genes, respectively (Table I). Of the 576 cell cycle-regulated genes, 528 have more than 25% of their expression in this subspace, outlined by the dashed half-unit line. (*c*) The HOSVD picture of the unperturbed *Saccharomyces cerevisiae* cell cycle, based on the consistent cell cycle themes reflected in the second and third eigenarrays and the second and third x-eigengenes that compose the second and third subtensors, respectively.

3.4. Uncovering the Cell Cycle Phase-Dependent Effects of Mcm2-7 Origin Binding.

After reconstruction of the data cuboid, the two Cdc6$^-$ time courses are averaged, and separately also the two Cdc45$^-$ and the four Cdc6$^+$ and Cdc45$^+$ time courses, and the data from each time point and condition are normalized to zero average and unit variance (Supplementary information Dataset 3). We find that the probabilistic significance of the enrichment of the time points in the averaged Cdc6$^+$/45$^+$ control in overexpressed or underexpressed cell cycle-regulated genes (Supplementary Figure 7) is consistent with the flow cytometry measurements of cell synchrony in the Cdc6$^+$ and Cdc45$^+$ cells (Supplementary Figures 1 and 2) as well as with previous analyses of α-factor synchronized cultures (Spellman *et al*, 1998).

The tensor HOSVD of the gene expression data cuboid of K=4270 genes across L=12 time points across M=3 conditions of Cdc6$^-$, Cdc45$^-$ and control (Supplementary information Dataset 3) is used (Box 1 and Supplementary information Section 3.3) as described (Omberg *et al*, 2007) to uncover combinations, i.e., subtensors, of three patterns of expression variation, one across the genes (Supplementary Figure 8), one across the time points (Supplementary Figure 9), and one across the conditions (Figure 10), and rotate the approximately degenerate subtensor spaces $\mathcal{S}(5,1,3)$ and $\mathcal{S}(6,1,3)$ (Supplementary Figure 11) by defining the unique subtensor $\mathcal{S}(5+6,1,3)$ according to Equations (10) and (11) (Figure 1 and Table I).

It was shown that significant and unique subtensors can be interpreted in terms of the cellular programs and biological processes that compose the data cuboid (Omberg *et al*, 2007). We find that the seven most significant among the unique subtensors uncovered in this gene expression data cuboid capture ~94% of the mRNA expression of the 4270 genes. We also find that the interpretation of these seven significant and unique subtensors is robust to modifications in the processing of the data, e.g., changes in the data selection cutoffs.

Supplementary information Mathematica Notebook. (*a*) A Mathematica 5.2 code file, executable by Mathematica 5.2 and readable by Mathematica Player, freely available at http://www.wolfram.com/products/player/. (*b*) A PDF format file, readable by Adobe Acrobat Reader.

Supplementary information Dataset 1. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, tabulating relative mRNA expression levels of 4771 probes of the UT DNA microarrays that correspond to the K=4270 genes across 24 samples.

Supplementary information Dataset 2. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, tabulating relative mRNA expression levels of 8540 probes of the WU DNA microarrays that correspond to the K=4270 genes across 72 samples.

Supplementary information Dataset 3. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, tabulating the averaged $\log_2$ of the relative mRNA expression of the K=4270 genes across the L=12 time points and across the M=3 conditions of Mcm2-7 origin binding. The genes are sorted by their angular distances between the second and third HOSVD combinations (Supplementary information Section 2.6 and Dataset 6), which represent the unperturbed cell cycle expression oscillations (Figure 2 and Supplementary Figure 12). The angular distance of each gene is also listed.

Supplementary information Dataset 4. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing descriptions and genomic coordinates (Nieduszynski *et al*, 2007) of the 325 confirmed ARSs in *Saccharomyces cerevisiae*.

Supplementary information Dataset 5. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, reproducing cell cycle annotations (Spellman *et al*, 1998), DNA damage responses (Jelinsky and Samson, 1999), descriptions and genomic coordinates (Cherry *et al*, 1997) of the 4270 *Saccharomyces cerevisiae* genes.

Supplementary information Dataset 6. A tab-delimited text format file, readable by both Mathematica and Microsoft Excel, tabulating the eigenarrays and superpositions of eigenarrays that define the global gene expression patterns of the seven significant and unique subtensors of the averaged data cuboid. The expression levels of the genes in the intersections of the fourth through seventh HOSVD combinations, as computed by using the corresponding eigenarrays, are also tabulated. The ten significant among the 1294 genes that are underexpressed in the fourth and overexpressed in the fifth and sixth combinations are enriched in histone genes. The 100 significant among the 1412 genes that are overexpressed in the fourth and underexpressed in the fifth and seventh combinations are enriched in genes with ARSs near their 3' ends.

[1] Alter O (2006) Discovery of principles of nature from mathematical modeling of DNA microarray data. *Proc Natl Acad Sci USA* **103**: 16063–16064

[2] Alter O, Brown PO, Botstein D (2000) Singular value decomposition for genome-wide expression data processing and modeling. *Proc Natl Acad Sci USA* **97**: 10101–10106

[3] Alter O, Golub GH (2004) Integrative analysis of genome-scale data by using pseudoinverse projection predicts novel correlation between DNA replication and RNA transcription. *Proc Natl Acad Sci USA* **101**: 16577–16582

[4] Alwine JC, Reed SI, Stark GR (1977) Characterization of the autoregulation of simian virus 40 gene A. *J Virol* **24**: 22–27

[5] Brewer BJ (1988) When polymerases collide: replication and the transcriptional organization of the *E. coli* chromosome. *Cell* **53**: 679–686

[6] Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, Adler C, Dunn B, Dwight S, Riles L, Mortimer RK, Botstein D (1997) Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**: 67–73

[7] Cowan K, Tegtmeyer P, Anthony DD (1973) Relationship of replication and transcription of Simian Virus 40 DNA. *Proc Natl Acad Sci USA* **70**: 1927–1930

[8] DePamphilis ML (1988) Transcriptional elements as components of eukaryotic origins of DNA replication. *Cell* **52**: 635–638

[9] Donato JJ, Chung SC, Tye BK (2006) Genome-wide hierarchy of replication origin usage in *Saccharomyces cerevisiae*. *PLoS Genet* **2**: e141

[10] Gerke JP, Chen CT, Cohen BA (2006) Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* **174**: 985–997

[11] Gilmartin GM (2005) Eukaryotic mRNA 3' processing: a common means to different ends. *Genes Dev* **19**: 2517–2521

[12] Golub GH, Van Loan CF (1996) *Matrix Computations*, 3rd edn. Baltimore, Maryland, USA: Johns Hopkins University Press

[13] Herendeen DR, Kassavetis GA, Barry J, Alberts BM,

Geiduschek EP (1989) Enhancement of bacteriophage T4 late transcription by components of the T4 DNA replication apparatus. *Science* **245**: 952–958

[14] Hu Z, Killion PJ, Iyer VR (2007) Genetic reconstruction of a functional transcriptional regulatory network. *Nat Genet* **39**: 683–687

[15] Jelinsky SA, Samson LD (1999) Global response of Saccharomyces cerevisiae to an alkylating agent. *Proc Natl Acad Sci USA* **96**: 1486–1491

[16] Klevecz RR, Li CM, Marcus I, Frankel PH (2008) Collective behavior in gene regulation: The cell is an oscillator, the cell cycle a developmental process. *FEBS J* **275**: 2372–2384

[17] Li CM, Klevecz RR (2006) A rapid genome-scale response of the transcriptional oscillator to perturbation reveals a period-doubling path to phenotypic change. *Proc Natl Acad Sci USA* **103**: 16254–16259

[18] Liu B, Alberts BM (1995) Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science* **267**: 1131–1137

[19] Lycan DE, Osley MA, Hereford LM (1987) Role of transcriptional and posttranscriptional regulation in expression of histone genes in *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**: 614–621

[20] Nieduszynski CA, Blow JJ, Donaldson AD (2005) The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. *Nucleic Acids Res* **33**: 2410–2420

[21] Nieduszynski CA, Hiraga S, Ak P, Benham CJ, Donaldson AD (2007) OriDB: a DNA replication origin database. *Nucleic Acids Res* **35**: D40–D46

[22] Nielsen TO, West RB, Linn SC, Alter O, Knowling MA, O'Connell JX, Ferro M, Sherlock G, Pollack JR, Brown PO, Botstein D, van de Rijn M (2002) Molecular characterisation of soft tissue tumours: a gene expression study. *Lancet* **359**: 1301–1307

[23] Omberg L, Golub GH, Alter O (2007) Tensor higher-order singular value decomposition for integrative analysis of DNA microarray data from different studies. *Proc Natl Acad Sci USA* **104**: 18371–18376

[24] Orlando DA, Lin CY, Bernard A, Wang JY, Socolar JE, Iversen ES, Hartemink AJ, Haase SB (2008) Global control of cell-cycle transcription by coupled CDK and network oscillators. *Nature* **453**: 944–947

[25] Piatti S, Lengauer C, Nasmyth K (1995) Cdc6 is an unstable protein whose *de novo* synthesis in G1 is important for the onset of S phase and for preventing a 'reductional' anaphase in the budding yeast *Saccharomyces cerevisiae*. *EMBO J* **14**: 3788–3799

[26] Proudfoot N (2004) New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol* **16**: 272–278

[27] Rosenthal LJ, Brown M (1977) The control of SV40 transcription during a lytic infection: late RNA synthesis in the presence of inhibitors of DNA replication. *Nucleic Acids Res* **4**: 551–565

[28] Rosonina E, Kaneko S, Manley JL (2006) Terminating the transcript: breaking up is hard to do. *Genes Dev* **20**: 1050–1056

[29] Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA (2001) Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**: 697–708

[30] Snyder M, Sapolsky RJ, Davis RW (1988) Transcription interferes with elements important for chromosome maintenance in *Saccharomyces cerevisiae*. *Mol Cell Biol* **8**: 2184–2194

[31] Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* **9**: 3273–3297

[32] Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM (1999) Systematic determination of genetic network architecture. *Nat Genet* **22**: 281–285

[33] Tercero JA, Labib K, Diffley JFX (2000) DNA synthesis at individual replication forks requires the essential initiation factor Cdc45p. *EMBO J* **19**: 2082–2093

[34] Thomas GP, Mathews MB (1980) DNA replication and the early to late transition in adenovirus infection. *Cell* **22**: 523–533

[35] Toth M, Doerfler W, Shenk T (1992) Adenovirus DNA replication facilitates binding of the MLTF/USF transcription factor to the viral major late promoter within infected cells. *Nucleic Acids Res* **20**: 5143–5148

[36] Vilette D, Ehrlich SD, Michel B (1995) Transcription-induced deletions in *Escherichia coli* plasmids. *Mol Microbiol* **17**: 493–504

[37] Weiner AM (2005) E Pluribus Unum: 3' end formation of polyadenylated mRNAs, histone mRNAs, and U snRNAs. *Mol Cell* **20**: 168–170

[38] Wellinger RE, Prado F, Aguilera A (2006) Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol Cell Biol* **26**: 3327–3334

[39] Wolffe AP, Brown DD (1986) DNA replication in vitro erases a *Xenopus* 5S RNA gene transcription complex. *Cell* **47**: 217–227

[40] Wyrick JJ, Aparicio JG, Chen T, Barnett JD, Jennings EG, Young RA, Bell SP, Aparicio OM (2001) Genome-wide distribution of ORC and MCM proteins in *S. cerevisiae*: high-resolution mapping of replication origins. *Science* **294**: 2357–2360

[41] Xu W, Aparicio JG, Aparicio OM, Tavaré S (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* **7**: 276