Supplementary Information for

Transcription of functionally related constitutive genes is not coordinated

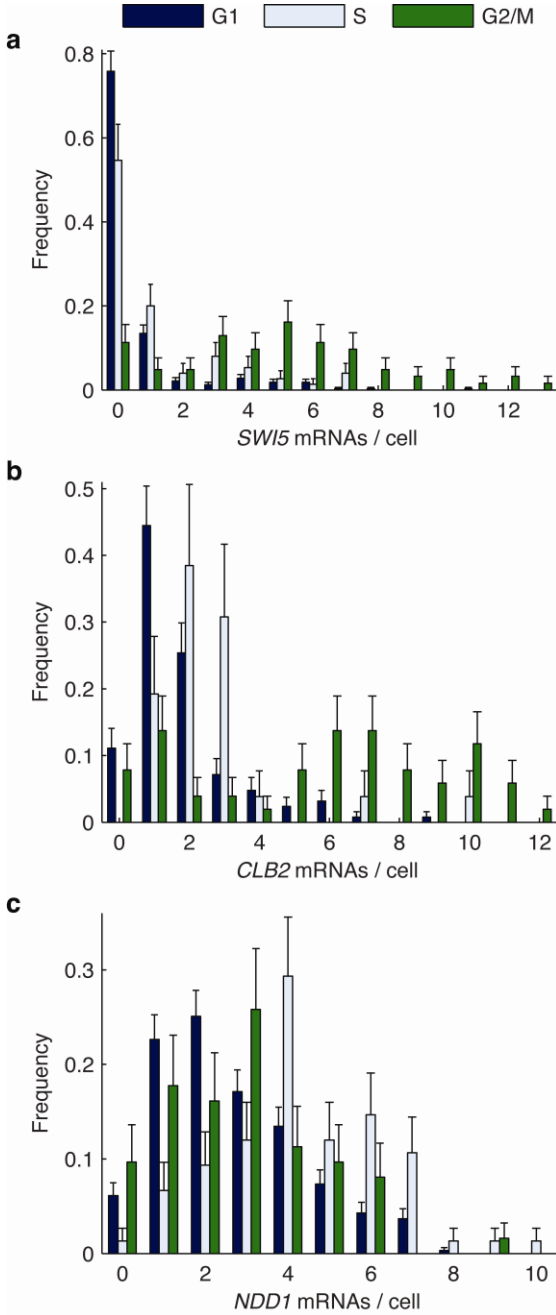*Saumil J. Gandhi, Daniel Zenklusen, Timothee Lionnet, and Robert H. Singer**

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine
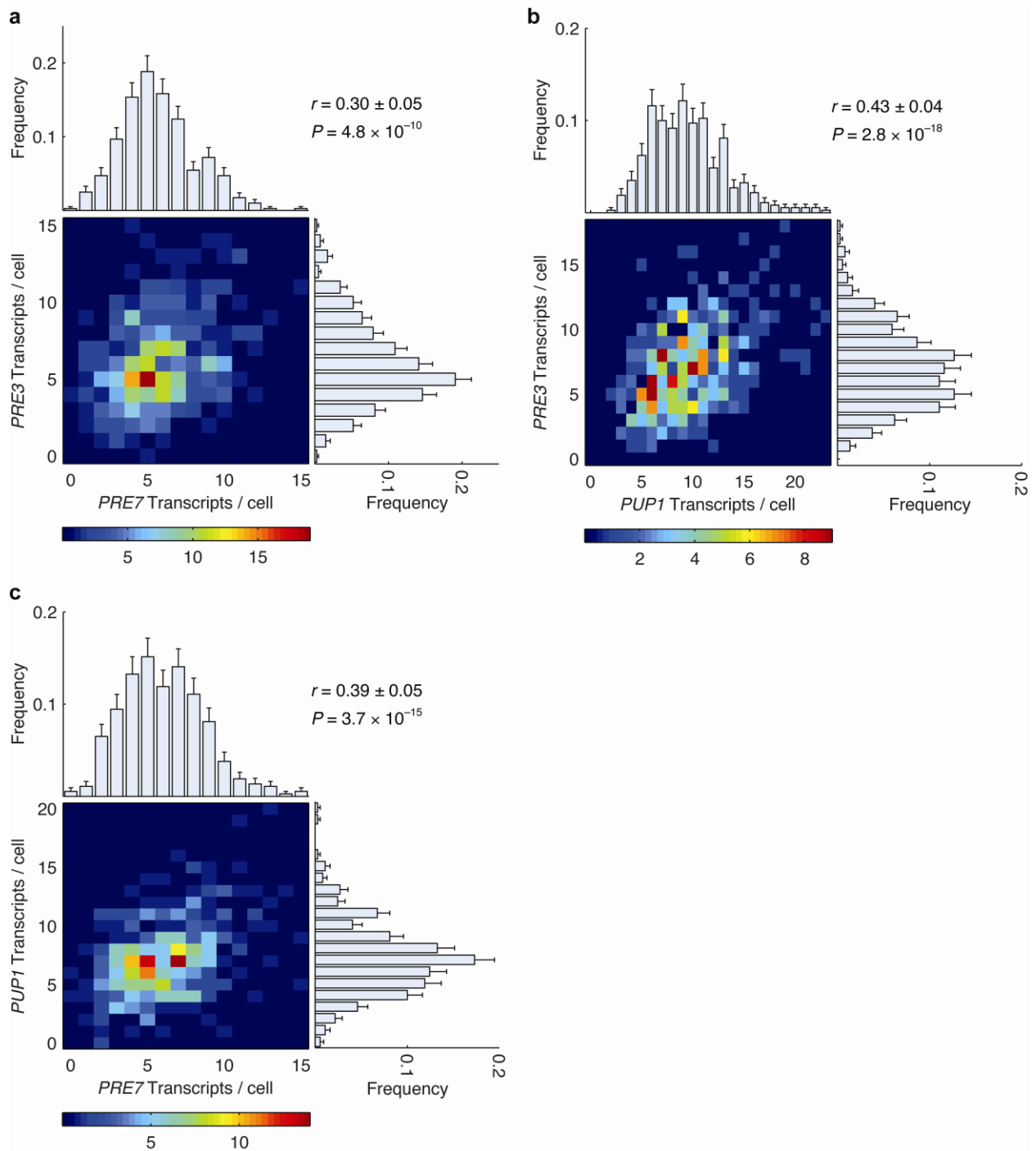
1300 Morris Park Avenue, Bronx, New York, 10461, USA

Correspondence: Robert H. Singer

robert.singer@einstein.yu.edu

SUPPLEMENTARY DATA

Supplementary Figure 1, related to Figure 3: Expression profiles of cell-cycle-stage-regulated genes. (**a-c**) *SWI5*, *CLB2*, and *NDD1* mRNA distributions for cells in G1 (red), S (blue), and G2 / M (green) phase. Error bars indicate s.e.m.

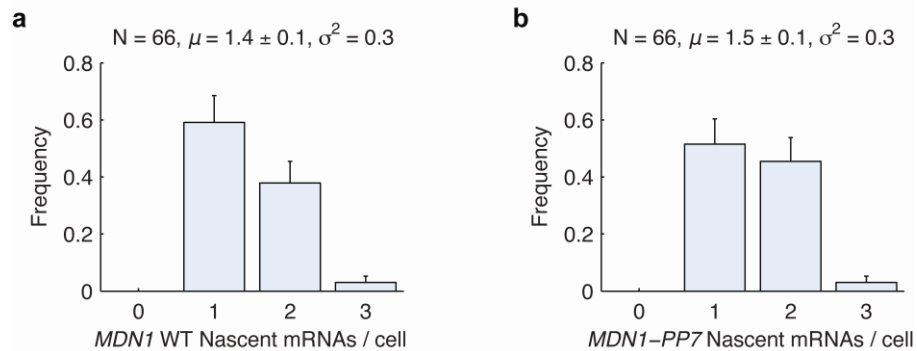**Supplementary Figure 2, related to Figure 5a: Pair-wise correlation coefficients for mRNAs of genes encoding β-subunits of the proteasome 20S core particle.** Error bars indicate s.e.m.
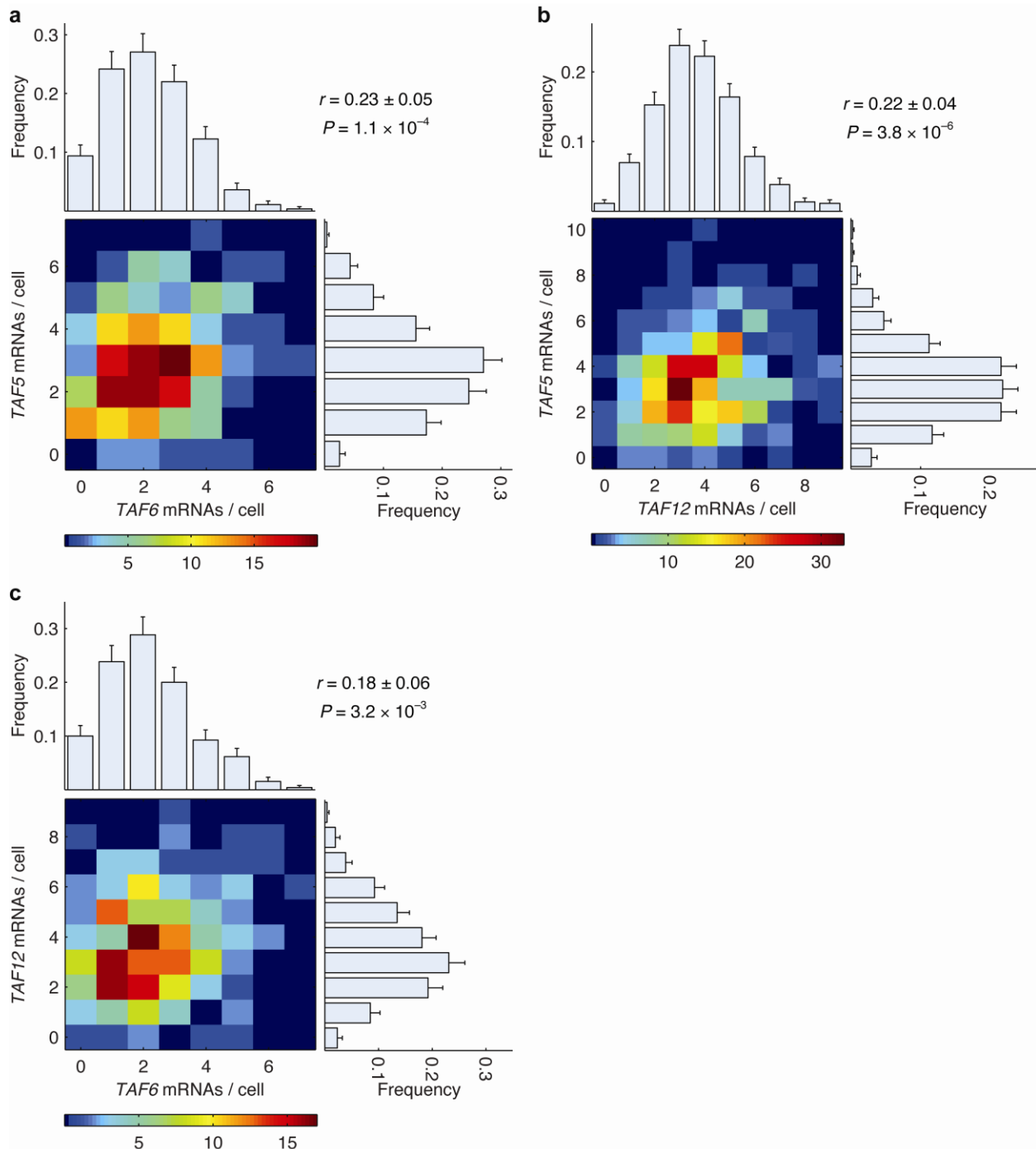
Supplementary Figure 3, related to Figure 5b: Pair-wise correlation coefficients for mRNAs of genes encoding TATA binding protein associated factors involved in transcription initiation. Error bars indicate s.e.m.

a

Frequency

$r = 0.43 \pm 0.05$
$P = 3.8 \times 10^{-15}$

RPB1 mRNAs / cell

RPB2 mRNAs / cell

Frequency

b

Frequency

$r = 0.26 \pm 0.05$
$P = 2.7 \times 10^{-7}$

RPB1 mRNAs / cell

RPB3 mRNAs / cell

Frequency

c

Frequency

$r = 0.23 \pm 0.06$
$P = 1.6 \times 10^{-4}$

RPB2 mRNAs / cell

RPB3 mRNAs / cell

Frequency
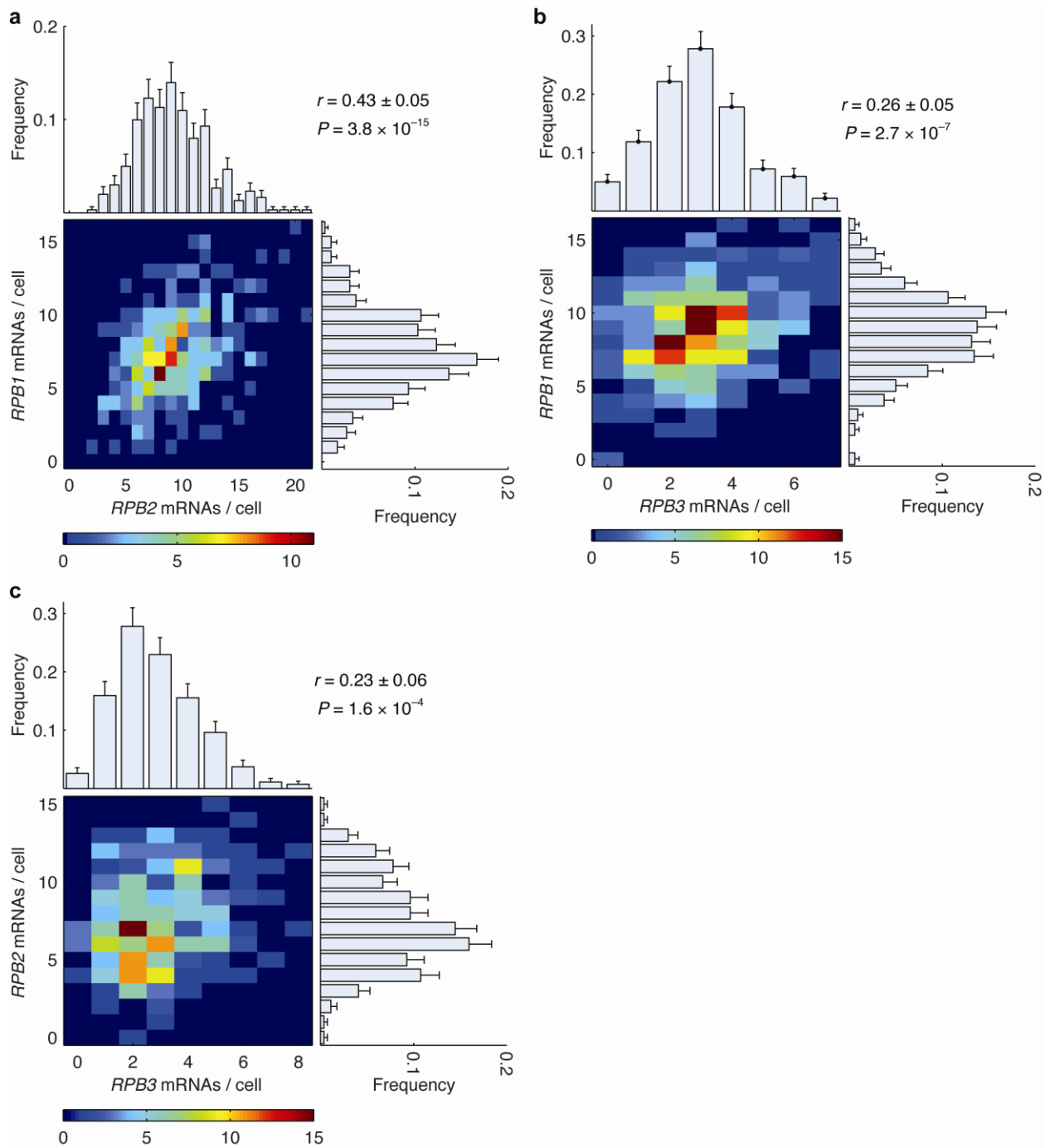
Supplementary Figure 4, related to Figure 5c: Pair-wise correlation coefficients for mRNAs of genes encoding subunits of RNA polymerase II. Error bars indicate s.e.m.

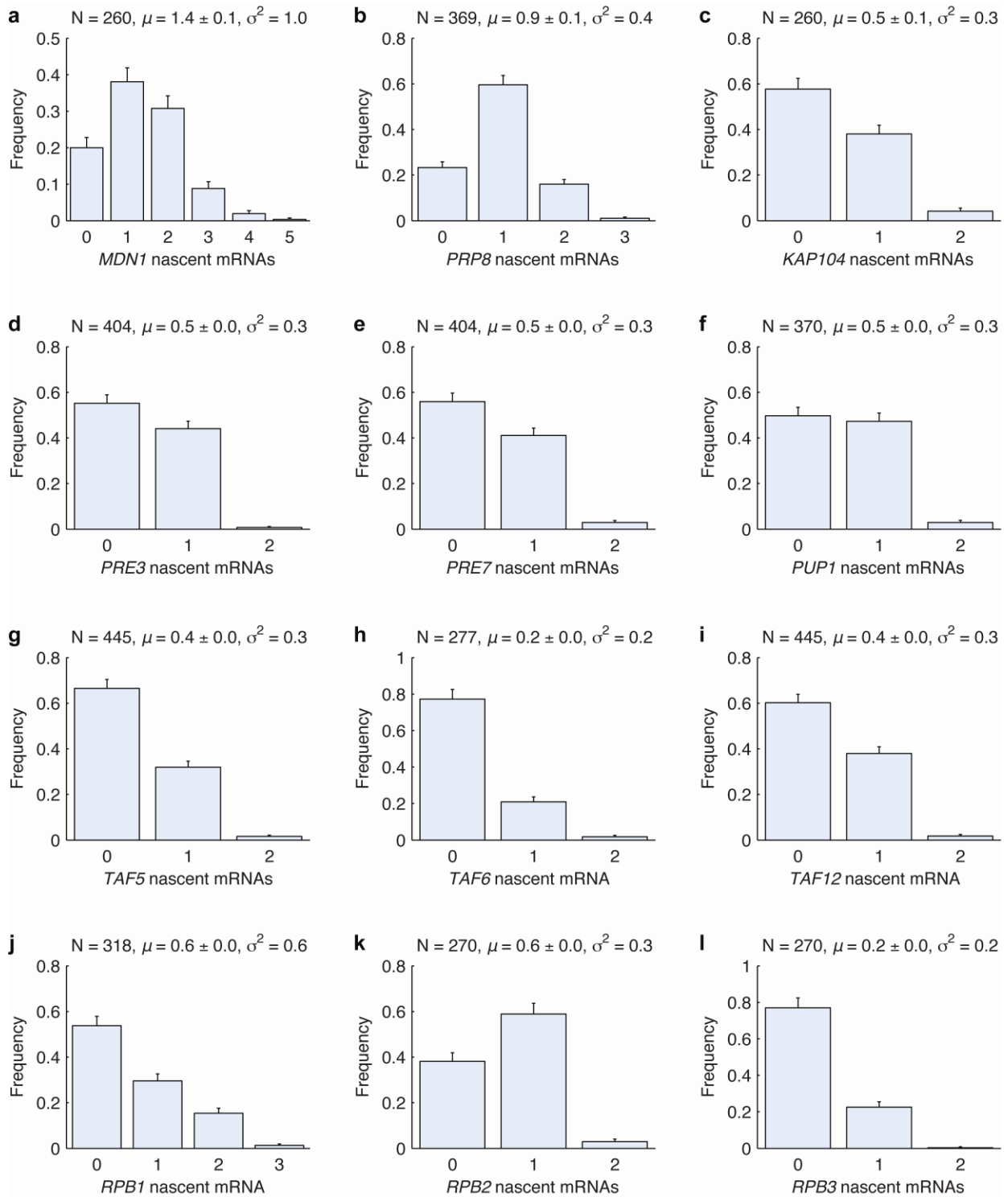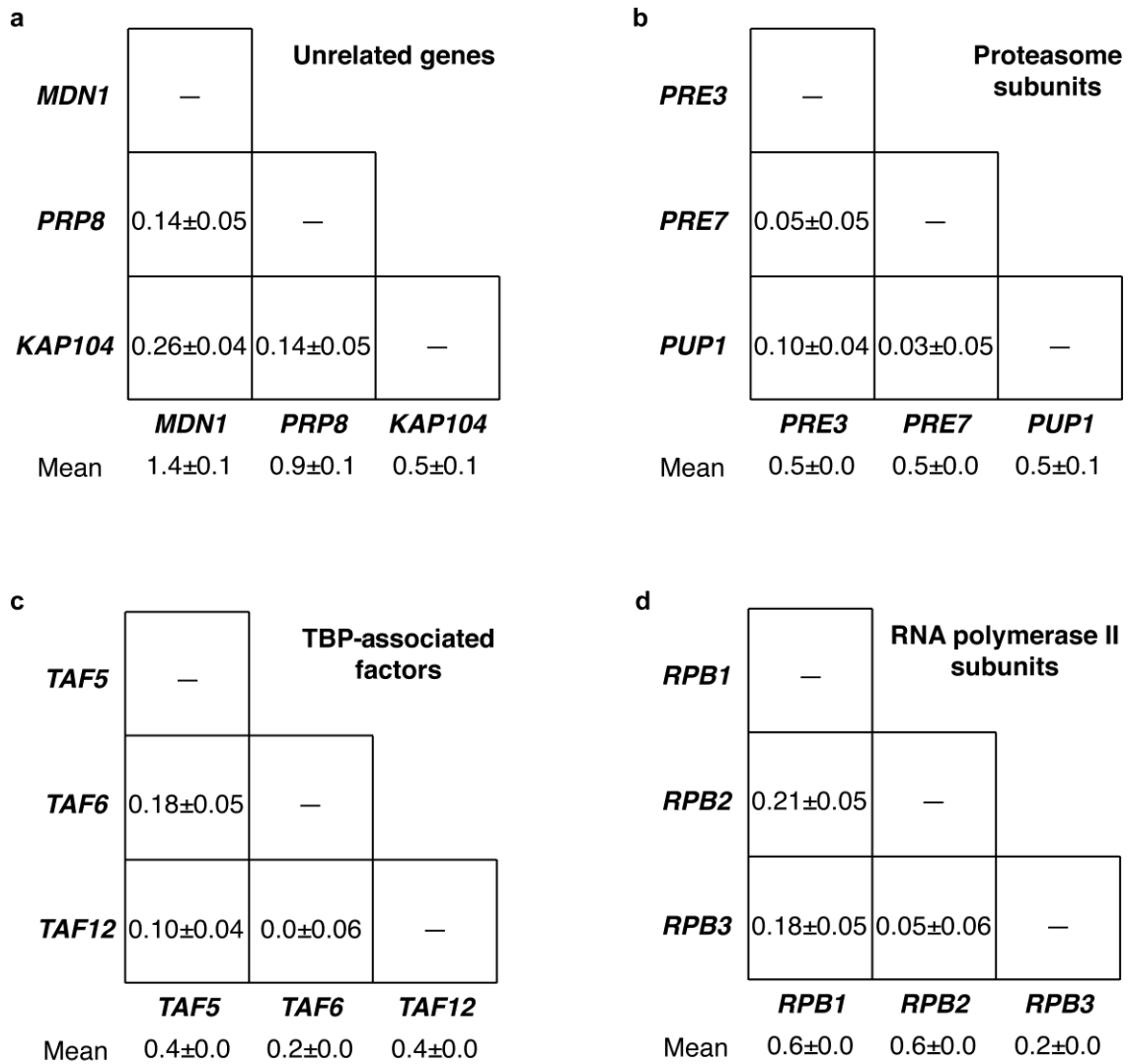**a** N = 66, $\mu = 1.4 \pm 0.1$, $\sigma^2 = 0.3$

b N = 66, $\mu = 1.5 \pm 0.1$, $\sigma^2 = 0.3$

MDN1 WT Nascent mRNAs / cell

MDN1–PP7 Nascent mRNAs / cell

Supplementary Figure 5, related to Figure 6: Distribution of nascent mRNAs per cell at the *MDN1* **transcription sites in the nucleus.** (**a**) The distribution of nascent mRNAs per cell at the wildtype (WT) *MDN1* allele. The number of nacent mRNAs were calculated by dividing the intensity of FISH signal at the transcription site in the nucleus by the intensity of individual mRNAs in the cytoplasm. (μ) and (σ^2) represent the mean and variance, respectively, for a population of N cells. (**b**) The distribution of nascent mRNAs at the *MDN1* allele with 24 PP7 hairpins inserted at the 3' end. The mean number of nascent mRNAs at the wildtype and PP7 alleles were equal, which demonstrates that both alleles are transcribed identically. Moreover, the nascent mRNAs of the two alleles were only weakly correlated ($r = 0.25 \pm 0.11$). Error bars indicate s.e.m.

a N = 260, μ = 1.4 ± 0.1, σ^2 = 1.0

b N = 369, μ = 0.9 ± 0.1, σ^2 = 0.4

c N = 260, μ = 0.5 ± 0.1, σ^2 = 0.3

d N = 404, μ = 0.5 ± 0.0, σ^2 = 0.3

e N = 404, μ = 0.5 ± 0.0, σ^2 = 0.3

f N = 370, μ = 0.5 ± 0.0, σ^2 = 0.3

g N = 445, μ = 0.4 ± 0.0, σ^2 = 0.3

h N = 277, μ = 0.2 ± 0.0, σ^2 = 0.2

i N = 445, μ = 0.4 ± 0.0, σ^2 = 0.3

j N = 318, μ = 0.6 ± 0.0, σ^2 = 0.6

k N = 270, μ = 0.6 ± 0.0, σ^2 = 0.3

**l** N = 270, μ = 0.2 ± 0.0, σ^2 = 0.2

Supplementary Figure 6, related to Figure 4 and Figure 5: Distribution of nascent mRNAs per cell at the transcription site in the nucleus.
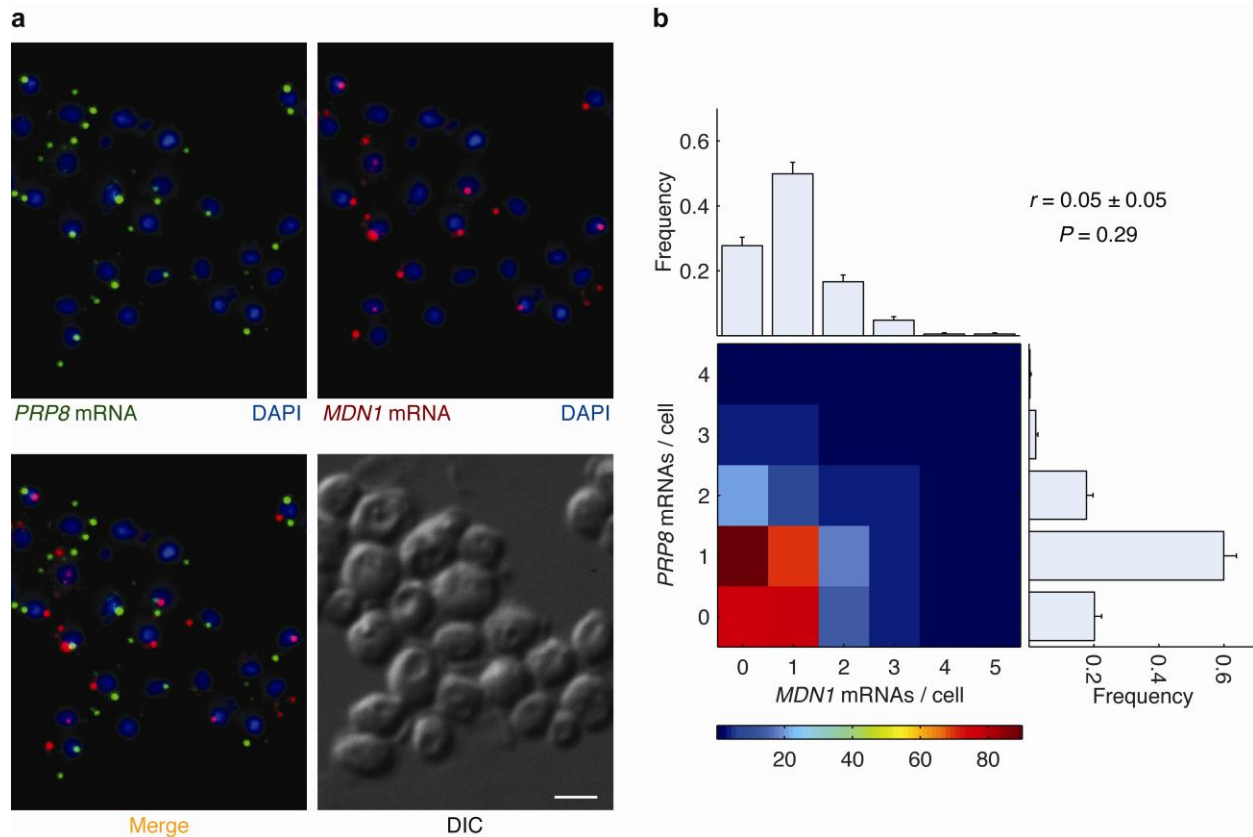
Supplementary Figure 6, related to Figure 4 and Figure 5: Distribution of nascent mRNAs per cell at the transcription site in the nucleus. (**a-l**) The number of nacent mRNAs for each gene were calculated by dividing the intensity of FISH signal at the transcription site in the nucleus by the intensity of individual mRNAs in the cytoplasm. (μ) and (σ^2) represent the mean and variance, respectively, for a population of N cells. Error bars indicate s.e.m.

a

MDN1	–		Unrelated genes
PRP8	0.14±0.05	–	
KAP104	0.26±0.04	0.14±0.05	–
	MDN1	*PRP8*	*KAP104*
Mean	1.4±0.1	0.9±0.1	0.5±0.1

b

PRE3	–		Proteasome subunits
PRE7	0.05±0.05	–	
PUP1	0.10±0.04	0.03±0.05	–
	PRE3	*PRE7*	*PUP1*
Mean	0.5±0.0	0.5±0.0	0.5±0.1

c

TAF5	–		TBP-associated factors
TAF6	0.18±0.05	–	
TAF12	0.10±0.04	0.0±0.06	–
	TAF5	*TAF6*	*TAF12*
Mean	0.4±0.0	0.2±0.0	0.4±0.0

d

|			
RPB1	–		RNA polymerase II subunits
RPB2	0.21±0.05	–	
RPB3	0.18±0.05	0.05±0.06	–
	RPB1	*RPB2*	*RPB3*
Mean	0.6±0.0	0.6±0.0	0.2±0.0

Supplementary Figure 7, related to Figure 4 and Figure 5: Correlation between nascent mRNAs at the transcription site in the nucleus. (**a**) Pair-wise correlation coefficients between the number of nacent mRNAs at the transcription site for three functionally unrelated genes. (**b**) Pair-wise correlation coefficients for nascent mRNAs of three genes encoding β-subunits of the proteasome 20S core particle. (**c**) Correlation coefficients for three genes encoding TATA binding protein associated factors involved in transcription initiation. (**d**) Correlation between three genes encoding subunits of RNA polymerase II. Error bars indicate s.e.m.
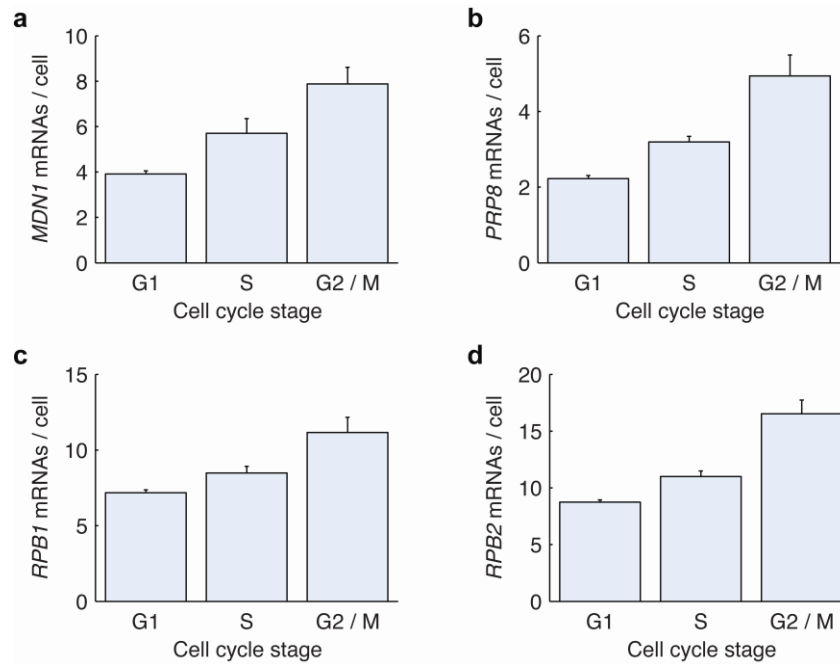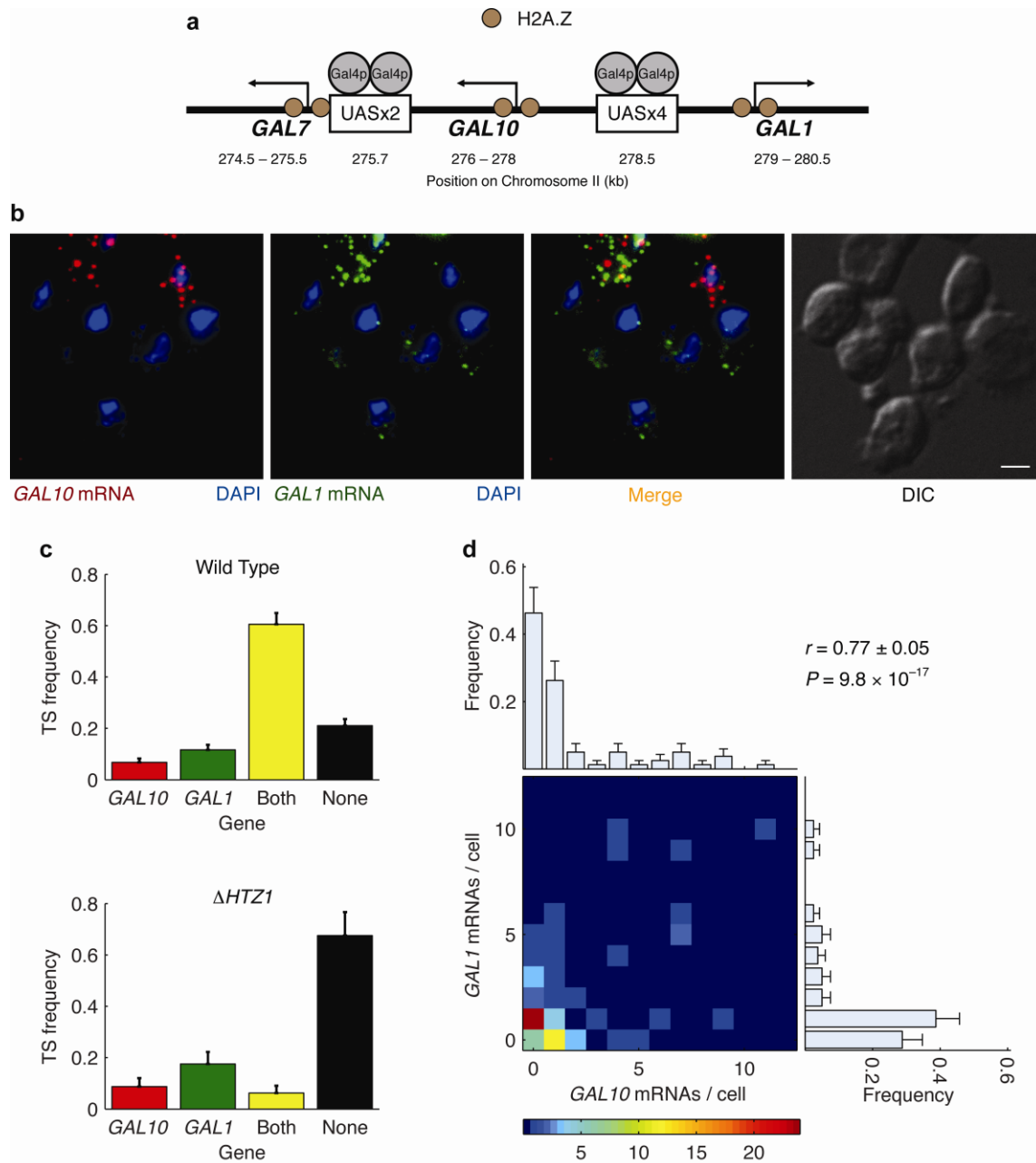
Supplementary Figure 8, related to Figure 7c-f: Mean mRNA abundance of constitutive genes increases over the cell cycle. (**a-d**) Experimentally measured average mRNA abundance across three different stages of the cell cycle. Differential interference contrast images were used to divide asynchronous cells into three different cell cycle stages based on morphology. Error bars indicate s.e.m.

Supplementary Figure 9, related to Figure 4 and Figure 7: Synchronizing effect of cell division disappears in cells with long cell cycles. (**a**) Representative FISH images of *PRP8* (green) and *MDN1* (red) mRNAs in cells grown in glucose-limiting media to achieve a doubling time of 14 hours. The scale bar in DIC image is 2 μm. (**b**) Heat map of number of *MDN1* and *PRP8* mRNAs in 397 cells shows that the correlation between these two genes is reduced to 0.05±0.05 from 0.26±0.05 in cells with a doubling time of 90 minutes. Marginal histograms: $\mu_{MDN1} = 1.0\pm0.05$, $\sigma_{MDN1} = 0.9$ (top); $\mu_{PRP8} = 1.0\pm0.04$, $\sigma_{PRP8} = 0.7$ (right). Error bars indicate s.e.m.
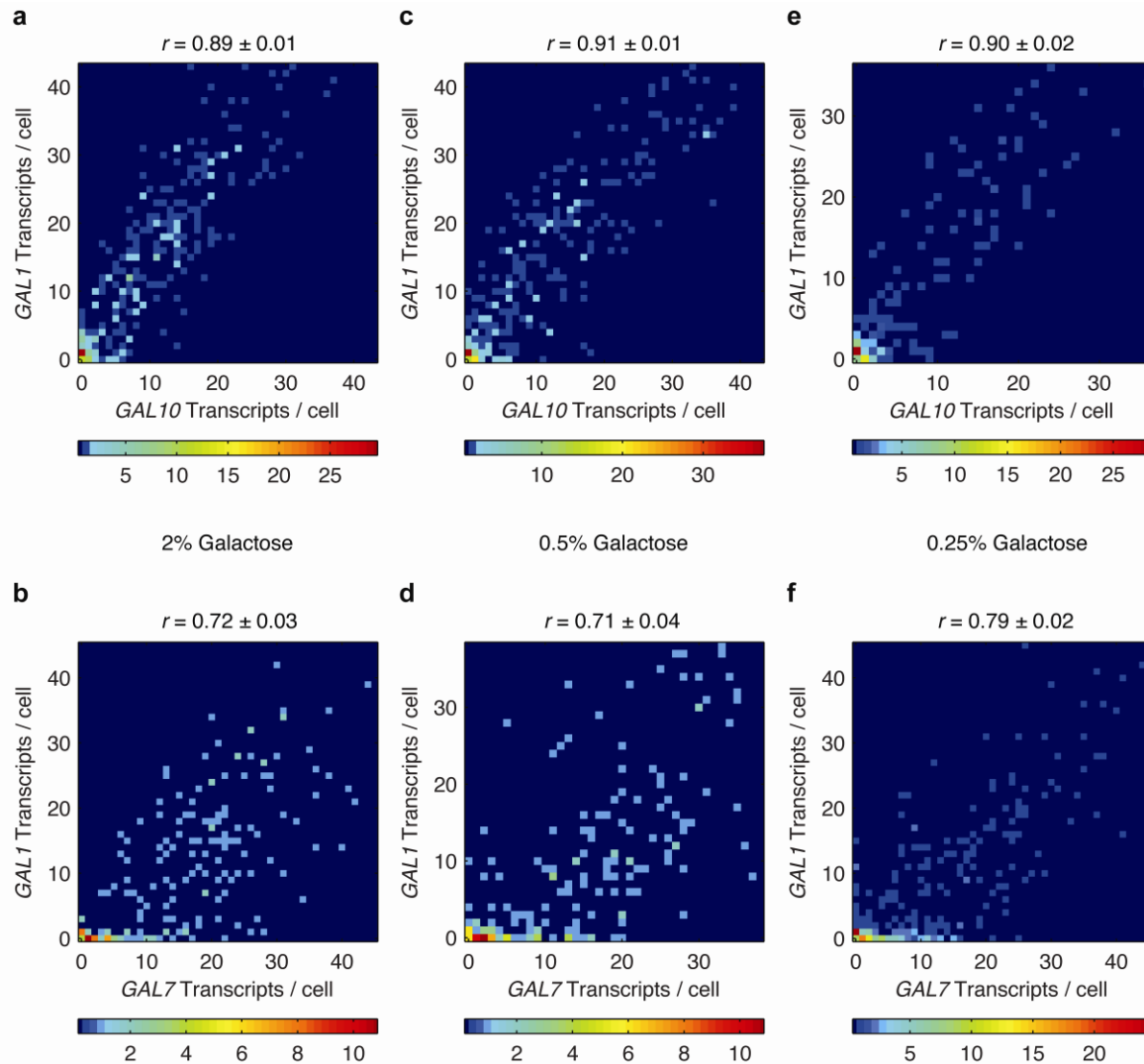
Supplementary Figure 10, related to Figure 2: Coordination in the expression of GAL genes is independent of the galactose concentration used for induction. All cells were induced for 15 minutes with varying concetrations of galactose, as indicated. Nascent or mature mRNAs were present in approximately 90% of cells upon induction with 2% galactose. About 75% of cells were induced with 0.5% galactose and 70% were induced with 0.25% galactose. (**a**) Heat map shows a high correlation coefficient (*r*) between the number of *GAL10* and *GAL1* mRNAs in cells induced with 2% galactose. (**b**) Correlation between *GAL7* and *GAL1* mRNAs in cells induced with 2% galactose. (**c, d**) Pair-wise correlations between GAL genes in cells induced with 0.5% galactose. (**e, f**) Pair-wise correlations upon induction with 0.25% galactose. Error bars indicate s.e.m.

Supplementary Figure 11, related to Figure 1 and Figure 2: Deletion of histone H2A variant H2A.Z reduces the correlation between *GAL1* and *GAL10*.

Supplementary Figure 11, related to Figure 1 and Figure 2: Deletion of histone H2A variant H2A.Z reduces the correlation between *GAL1* and *GAL10*. (**a**) Incorporation of H2A.Z makes the +1 and −1 promoter nucleosomes slightly unstable and is thought to promote gene activation by exposing the transcription start site. (**b**) Deletion of *HTZ1*, the gene encoding H2A.Z, led to a significant decrease in the expression of GAL genes. Only 40% of cells were induced after 30 minutes of induction with 2% galactose. *GAL10* mRNA (red) and *GAL1* mRNA (green) is shown along with DAPI (blue) and DIC image of cells. The scale bar is 2 μm. (**c**) Within the induced population, the percentage of cells actively transcribing both genes in the nucleus fell sharply compared to wild type cells (2% galactose, 15 minutes). (**d**) Heat map of number of *GAL10* and *GAL1* mRNAs per cell shows a reduced correlation between these two genes compared to wild type cells. The marginal histograms indicate that the mean mRNA abundance was also reduced significantly. $\mu_{GAL\ 10} = 2.0 \pm 0.4$, $\sigma^2_{GAL\ 10} = 12.1$ (top); $\mu_{GAL\ 1} = 3.0 \pm 0.4$, $\sigma^2_{GAL\ 1} = 30.5$ (right). Error bars indicate s.e.m.

| Gene	Mean mRNA abundance [1] (molecules/cell)	mRNA Half-life[4] (min)	Mean protein[19] abundance (molecules/cell)
MDN1	1.2	26	538
PRP8	1.2	38	468
KAP104	0.8	16	2130
PRE3	4.5	16	7250
PRE7	4	25	—
PUP1	3.5	32	11400
TAF5	1.3	18	14800 ± 203
TAF6	0.9	15	6510 ± 1290
TAF12	1.3	22	930 ± 45
RPB1	2.9	24	—
RPB2	3.9	26	18700
RPB3	3.3	18	10000 ± 668
GAL1	0.2	—	—
GAL7	0.1	18	—
GAL10	—	—	—
NDD1	0.4	30	799
SWI5	0.8	17	688
CLB2	1.1	13	339

Supplementary Table 1: Previously published mRNA abundances, mRNA half-lives, and protein abundances for all genes considered in this study.

SUPPLEMENTARY METHODS

Analytical solution for mRNA correlations

Model

The model presented here elaborates on previous work describing fluctuations of protein numbers in a dividing population of cells (Berg, 1978). In between two cell division events, mRNAs of a given gene are assumed to be transcribed randomly, independently of each other, at a rate of k_i messages per minute. Once an mRNA molecule has been generated, it decays with a uniform probability per unit time k_d :

$$\varnothing \; \underset{k_d}{\overset{k_i}{\rightleftarrows}} \; RNA$$
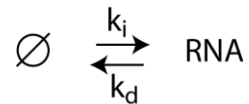
In between cell divisions, transcripts of each gene initiate and decay independently from the other gene. At cell division, the two mRNA populations get partitioned between daughter cells simultaneously, but independently.

In the following sections, we first extract the master equation that describes the time evolution of the number of mRNA of one gene between cell cycles. We then find the solution to this master equation, imposing as a boundary condition that the cells are in a steady state growth, i.e. the number of mRNAs in the cell is the same at each generation. From the master equation, we determine the time-dependent probability distribution for the number of mRNAs of that gene. We show that the probability distribution of the number of mRNAs follow a Poisson distribution at all times. The mean of that Poisson distribution varies over the cell cycle and is equal to the deterministic solution of the continuous differential equation system corresponding to our model.

We use the single gene mRNA probability distribution to compute the predicted values of the observables of the experiments. First, we compute the time-averaged probability distribution of the mRNA number for one gene. This is the formal equivalent of looking at unsynchronized cells that were fixed at a given time: each cell is at a different stage of the cell cycle. Therefore, the time-averaged distributions should reflect the experimentally observed distributions. We can then use the time averaged mRNA probability distribution to compute the mean number of transcripts presents in the cell. Finally, we can compute the correlation between two genes. For the sake of simplicity, we assume here that the two genes have equal appearance and decay rates.

16

Master equation

The probability that an mRNA molecule is transcribed during an infinitesimal time interval dt is equal to $k_i dt$. During the same time interval, an existing mRNA molecule will decay with a probability $k_d dt$. If there are m mRNA molecules in the cell, the probability that one decay event occurs during the interval dt is, therefore, equal to $mk_d dt$. Using these expressions, we can write the master equation that sets the time evolution of the probability $P(m,t)$ to have m transcripts in the cell at time t:

$$P(m, t + dt) = P(m,t) + k_d\, dt\, [(m+1)P(m+1,t) - mP(m,t)]$$
$$+ k_i dt\, [P(m-1,t) - P(m,t)] \qquad (1)$$

We introduce the generating function $G(z,t)$ as a mathematical tool to solve the master equation (1):

$$G(z,t) \overset{\text{def}}{=} \sum_{m=0}^{\infty} z^m P(m,t) \qquad (2)$$

We first determine the expression of $G(z,t)$. We can then compute the mRNA probability distributions $P(m,t)$ using the following property of the generating function:

$$P(m,t) = \frac{1}{m!} \lim_{z \to 0} \frac{\partial^m G}{\partial z^m} \qquad (3)$$

We multiply each side of (1) by z^m and sum over m to obtain:

$$\partial_t G = k_d(1-z)\partial_z G + k_i\,(z-1)G \qquad (4)$$

We perform the following variable change:

$$v = k_i(1-z)$$
$$w = k_i(1-z)e^{-k_d t}$$

The master equation then appears as a simpler form:

$$\partial_v G = -\frac{G}{k_d} \qquad (5)$$

We integrate the differential equation as follows:

$$G = A(w)e^{-v/k_d} \qquad (6)$$

where A is some function of the variable w. Therefore G has the form:

$$G(z,t) = A\{k_i(1-z)e^{-k_d t}\} \exp\left[-\frac{k_i}{k_d}(1-z)\right] \qquad (7)$$

The function A is unknown at this point, but we can determine its expression by imposing conditions corresponding to steady state growth.

Boundary conditions

We impose a steady state growth condition, meaning that the number of mRNAs in the cell is independent of the generation it belongs to. We also assume that mRNAs are equally partitioned between daughter cells at cell division.

Let's assume a cell has j mRNAs right before cell division. Then the probability to get m mRNAs in one of the daughter cells is given by the binomial partitioning probability:

$$\left(\frac{1}{2}\right)^j \binom{j}{m} \qquad (8)$$

Therefore, the probability of having m mRNAs in the cell at the beginning of the cell cycle can be expressed in terms of 1) the probabilities of having j mRNAs in the cell at the end of the cell cycle and 2) binomial partitioning (8):

$$P(m, t = 0) = \sum_{j=m}^{\infty} P(j, t = T_g) \left(\frac{1}{2}\right)^j \binom{j}{m} \qquad (9)$$

We introduce again the generating function $G(z, t) \overset{\text{def}}{=} \sum_{m=0}^{\infty} z^m P(m, t)$. The boundary condition can then be written in terms of G:

$$G(z, t = 0) = G\left(\frac{1 + z}{2}, t = T_g\right) \qquad (10)$$

Combining equations (7) and (10), we obtain the following condition on the unknown function A:

$$\forall x, \quad A(x) = A\left(\frac{x}{2} e^{-k_d T_g}\right) \exp\left(\frac{x}{2k_d}\right) \qquad (11)$$

The condition in (11) is satisfied by the following function:

$$A(x) = \exp\left(\frac{1}{2k_d} \frac{1}{1 - \frac{1}{2}\exp(-k_d T_g)} x\right) \qquad (12)$$

We write the expression of A from (12) into (7) and obtain the full expression of $G(z, t)$:

$$G(z, t) = \exp\left[-\frac{k_i}{k_d}(1 - z)\left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)}\right)\right] \qquad (13)$$

Time-dependent probability distribution

Using the property of the generating function (3), we calculate the probability distribution of the mRNA number $P(m,t)$ from the expression of generating function $G(z,t)$ in (13):

$$P(m,t) = \frac{1}{m!} \exp\left[-\frac{k_i}{k_d}\left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)}\right)\right] \left(\frac{k_i}{k_d}\left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)}\right)\right)^m \quad (14)$$

We recognize here the expression of a Poisson distribution

$$P(m,t) = \frac{\lambda(t)^m\, e^{-\lambda(t)}}{m!} \quad (15)$$

where the mean $\lambda(t)$ is dependent on time:

$$\lambda(t) = \frac{k_i}{k_d}\left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)}\right) \quad (16)$$

As expected, the time-dependent mean is equal to the deterministic solution of the continuous differential equation system based on our model. The stochastic calculation shows that the mRNAs remain Poisson distributed around this varying mean during the entire cell cycle, even right after cell division.

The expression of $\lambda(t)$ is the product of k_i/k_d by a correction factor. k_i/k_d is the ratio of the generation rate divided by the decay rate of the mRNA molecules, which would be the equilibrium value of the mean number of transcripts in the absence of cell cycle. The absence of a cell cycle is formally equivalent to the limiting case where the cell cycle duration is much longer than the other time scales involved, i.e. $k_d T_g \to \infty$. In this limit, the correction factor is equal to 1 when t reaches infinity. Therefore the mean number of mRNA when the system reaches equilibrium is equal to the expected value: $(t \to \infty) = k_i/k_d$. In the same limiting case $(k_d T_g \to \infty)$, we find that the mean number of mRNA at the beginning of the cell cycle is as expected half the number of mRNAs reached as equilibrium: $(t = 0) = k_i/2k_d$.

Time averaged probability distribution

We first compute the probability distribution of the number of mRNA molecules in the cell averaged over the cell cycle $\langle P(m,t)\rangle_t$:

$$\langle P(m,t)\rangle_t \overset{\text{def}}{=} \frac{1}{T_g}\int_{t=0}^{T_g} P(m,t)\, dt \quad (17)$$

$$\langle P(m,t)\rangle_t = \frac{1}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^m \, e^{-\lambda(t)}}{m!} \, dt \qquad (18)$$

$$= \frac{1}{T_g} \int_{t=0}^{T_g} \frac{1}{m!} \left(\frac{k_i}{k_d} \left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)} \right) \right)^m \exp\left(-\frac{k_i}{k_d} \left(1 - \frac{\frac{1}{2}\exp(-k_d t)}{1 - \frac{1}{2}\exp(-k_d T_g)} \right) \right) dt \quad (19)$$

The development of this expression is not very informative and yields integrals that cannot be solved analytically. However, this integral can be readily estimated numerically. This is the case for all of the following statistical quantities.

Mean number of transcripts

We can use the expression of the time averaged probability distribution (18) to compute the average number of transcripts μ present in the cell over the cell cycle. It is an important number: it is the value we have access to in the experiments, as we average together cells from various stages of the cell cycle. By definition,

$$\mu = \sum_{m=0}^{\infty} m \, \langle P(m,t)\rangle_t \, dt \qquad (20)$$

We insert the explicit expression of $\langle P(m,t)\rangle_t$ obtained in (15):

$$\mu = \sum_{m=0}^{\infty} \frac{m}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^m e^{-\lambda(t)}}{m!} \, dt \qquad (21)$$

Variance of the number of transcripts

The variance is defined as follows:

$$\sigma^2 = \frac{1}{T_g} \int_{t=0}^{T_g} \sum_{m=0}^{\infty} m^2 \, P(m,t) \, dt \; - \; \left(\frac{1}{T_g} \int_{t=0}^{T_g} \sum_{m=0}^{\infty} m \, P(m,t) \, dt \right)^2 \qquad (22)$$

Hence,

$$\sigma^2 = \sum_{m=0}^{\infty} \frac{m^2}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^m e^{-\lambda(t)}}{m!} \, dt - \left(\sum_{m=0}^{\infty} \frac{m}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^m e^{-\lambda(t)}}{m!} \, dt \right)^2 \qquad (23)$$

Joint probability distribution of the number of transcripts

We consider here two genes within the same cell, with mRNAs that are produced and degraded independently. We compute the probability $Q(m, n, t)$ to observe the m transcripts of gene 1 and n transcripts of gene 2 at time t. For simplicity, we assume that the mRNAs expressed by the two genes have the same kinetics (equal values of k_i and k_d).

Since the transcripts are independent, $Q(m, n, t) = P(m, t)P(n, t)$. The correlation in the mRNA abundance of the two genes comes only from the fact that they undergo cell division in synchrony.

Using the expression of $P(m, t)$ in (18), we can express $Q(m, n, t)$ as follows:

$$Q(m, n, t) = \frac{\lambda(t)^{m+n} \, e^{-2\lambda(t)}}{m! \, n!} \quad (24)$$

Hence,

$$\langle Q(m, n, t) \rangle_t = \frac{1}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^{m+n} \, e^{-2\lambda(t)}}{m! \, n!} \, dt \quad (25)$$

Covariance of transcripts numbers expressed by two genes

The covariance of the number of mRNAs measures how the numbers of transcripts from two different genes change together within the same cell. It is a useful intermediate in the calculation of the correlation coefficient.

By definition, the covariance is equal to:

$$Cov \stackrel{\text{def}}{=} \sum_{m,n=0}^{\infty} \frac{1}{T_g} \int_{t=0}^{T_g} m \, n \, Q(m, n, t) \, dt \; - \; \langle n \rangle^2 \quad (26)$$

$$Cov \stackrel{\text{def}}{=} \sum_{m,n=0}^{\infty} \frac{mn}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^{m+n} \, e^{-2\lambda(t)}}{m! \, n!} \, dt \; - \; \left(\sum_{m=0}^{\infty} \frac{m}{T_g} \int_{t=0}^{T_g} \frac{\lambda(t)^{m} e^{-\lambda(t)}}{m!} \, dt \right)^2 \quad (27)$$

Correlation Coefficient

In the case of two identical genes, the correlation coefficient r is equal to the ratio of the covariance (27) divided by the variance (23):

$$r = \frac{Cov}{\sigma^2} \quad (28)$$

The value of r can be determined by numerical approximations of the sums in the numerator and the denominator.

Numerical estimates

In order to compute the numerical values of r for a range of parameters, we wrote a program in Matlab 7.0.1 (TheMathworks). The inputs were the generation time (set invariably to $T_g = 90$ minutes), the mean number of transcripts μ and the experimental half life $t_{1/2}$. From the values of these macroscopic observables we first computed the microscopic rates as follows:

$$k_d = \frac{\log(2)}{t_{1/2}} - \frac{\log(2)}{T_g}$$

$$k_i = \frac{\mu \, k_d}{1 - \frac{1}{k_d T_g}\left(1 - \frac{1}{2 - \exp(-k_d T_g)}\right)}$$

Note that the formula linking the molecular k_d to the half life includes a correction term corresponding to mRNA dilution during cell division. Formulas (23) and (27) were then estimated the following way: integrals were evaluated using the adaptive Gauss-Kronrod quadrature algorithm, and the discrete sums over the number of transcripts were truncated at a value arbitrarily set to the largest value between 20 and $5k_i/k_d$.

FISH Probe Sequences

Ts indicate amino-allyl modified bases.

CLB2 182   TGC GGC **T**GT TGA TCT **T**GA TAC GC**T** TTC TCT AG**T** AAA TTG AGG AA**T** CCG AC
CLB2 321   GCA GC**T** CAG CTC TCC CC**T** CCT TTA TAG AAG **T**TA GCG CAC CAA A**T**T GCT GA
CLB2 427   TC**T** ACC GAG **T**TA TGG ACT ACC **T**CA GTG CTT GA**T** CCA ACG CCT TT**T** AGG GG
CLB2 984   CCC C**T**T CTT TGA TT**T** CAT CTT CCG **T**AC ATG CAC CGT CTG TCT CTG A**T**G CG
CLB2 1242   ACC CGC CGC **T**AT AGT GTA **T**TA GAT TTC CA**T** CCC ATT TAC C**T**T TAC CTA AC

GAL1 – 1   **T**AT GCT CGG GCA C**T**T TTC GGC CAA **T**GG TCT TGG **T**AA TTC CTT TGC GC**T** AG
GAL1 – 2   ACC AG**T** CCG ACA CAG AAG GA**T** CAA TTG TGA CA**T** AAG AAC CG**T** CCA ACG GC
GAL1 – 3   GAA GAC AA**T** CCA CTG CCA GT**T** GGT ACA TCA CCC **T**CA CAG AAG ACT **T**GC AG
GAL1 – 4   **T**AG CAT CTT TG**T** TAA CCG TTC GA**T** GCC GGA TTC AA**T** ATC GCC GT**T** CCA GG
GAL1 – 5   **T**CG TGC TCG ATC CTT CTT TTC CAG AAA G**T**A AAA CAA CAC CG**T** ACG TGG CAG C
GAL1 – 6   TG**T** GCG ACA TCG **T**CA ACA CTA AAG CCC **T**GT TTC TTA **T**TG GCG AGA GAC **T**C

GAL7 80   TG**T** AAG CAG CC**T** CCT GTT GAC C**T**A ACC AAG GTC **T**TT TAG CTC TGT G**T**G GAG
GAL7 244   CC**T** CAT TGG AA**T** CAT CT CT GTG G**T**A AAA TAG GT**T** GAT CGA GCC **T**AA CGG CAG C
GAL7 428   AGG C**T**T ATG ATT TTC **T**CT TGC TTC TC**T** GGA GAG ATC G**T**C AGT CAA TGC **T**TG CC
GAL7 541   CTT GCG AAA C**T**T CAC TAG GGA **T**GG ATT CTA AGC ACC AAG C**T**T GGC CAT G**T**G G
GAL7 765   GGC GAG G**T**C CTC CTT CAC CA**T** TTG GTT AAA **T**TG GCT AAT **T**GA GGC AAG C**T**T C
GAL7 878   ACT CAA TTC A**T**C ACC AGT CGC A**T**T CAA AGG AGC C**T**G ATG GAT ACC CA**T** TG

GAL10 152   CTC A**T**A GAA GGG AA**T** GTG ATG CTT GG**T** CAA GAC CTC **T**AA CCT GGC **T**AC AG
GAL10 778   CGA CAC AAA CC**T** TCA TTT TCA **T**TG TAG GCC **T**CT AGG TAT **T**GC AGG GC**T** GC
GAL10 925   GCC C**T**A TCT GGT **T**TA GCC GTC AAG **T**TC AAA ACA **T**CA CCT GCT C**T**T CTG CC
GAL10 1059   GCA TAT C**T**T CAG CGG AAA A**T**C TGG CCT CGA CAC CCC **T**TA ACT GG**T** AAC CA
GAL10 1173   CAA GAA CAA C**T**G ATT GTC CGT TCA CTT TCA GG**T** CAA CAA TGC **T**GG CGC CC
GAL10 1448   ACC **T**GG AAA TTC GG**T** GTC CTT CTC A**T**T ATC TAT CAG CAT GTA CTC GGC GG

KAP104 99   CTA GTT GCA ACA CAT AGT CTT CGG CGG GCT TCC ATG TCG ATG CCA TCT TT
KAP104 425   ATA CCG GCG GTA GCT CTA TTG TTC TGA AGA TCC TGT AGG GAG TAG TGT TG
KAP104 1421 CCT GCC TCT TTC TTT TTC ACA ATA CGG GGT GCA ATG GGC TTG ATG TCC TC
KAP104 2870 AGT CCT TCT CAG GCA CTA CTA TTG TCG GGT CTT GAT GTG ATT TGG CTT CC

MDN1 794   TT**T** GTC GTG GAT AG**T** GTG GAC CT**T** AGG GAC GA**T** AAC GCC ACA GAT **T**GA CG
MDN1 860   CTC CCG AGT **T**GA CGA AGA GAG GAA ACC G**T**T TTA TGA G**T**A GGG ACA AAG G**T**T
MDN1 1104   CTA TAA GTA CCC A**T**C TCC CTT CTT **T**GA CCG CGG **T**AG CGA GAA CAC CAG CTC
MDN1 1210   TT**T** GCA GCC TTT ACA GTC **T**CT CCT CTG GA**T** GGA ATG GT**T** AGT TCG CGC **T**T
MDN1 4350   CAC C**T**T TCT GCA AGA AGC A**T**A TAG CCA C**T**G GCA GCA AGC **T**GC TCA TA**T** CC
MDN1 4511   CTA AGA CGG ATA G**T**C GGC GCA TTC C**T**T TTG TCC AAG **T**AA CAG AGC CAA **T**G

NDD1 45   GTA TCT GCA **T**TA TTG TCC G**T**A GAG GGC TGT CT**T** GAG GAA GTT GCA GT**T** GCC

NDD1 325   CCT TGT GAC AGC AGG AGG **T**AC AAG TGA ACC **T**AG AGC TTG C**T**G TTG TTG TTG

NDD1 430   **T**GC TCT GGG GA**T** GCA ACC AAA AA**T** TCG CTG GG**T** AAA GAG TTG CCG AAC **T**G

NDD1 536   **T**GC TGA GCA CCG **T**TA GAA TCT GT**T** ACA AAT CGA AG**T** GGT GTC TT**T** GCC GGA G

NDD1 994   **T**CG TTG TCC **T**TA GTT AGT G**T**G CTA GTG CTA T**T**G ATG GTA TCC **T**GG GGG GT

NDD1 1251 CTT GGG TAG C**T**C AGG AAT TTG **T**AA GGC AGA TGT ACC TG**T** GGT TGT ACT A**T**G C

PRE3 39   CAC CAT CCT TAA ATG **T**CA CGG CCA TAA **T**TG AGG TAC C**T**A AAC TGA C**T**T CGC CC

PRE3 104   ATC TGT CAC ACG G**T**T AGC TAT G**T**A CGC ACC AGT GG**T** GGT ACG TGA A**T**C AG

PRE3 164   **T**GC GTG TCT GC**T** GCA GAA CCG GAC C**T**A CAA CAC CAA A**T**T TTG TCA TGT AC**T** C

PRE3 270   CGT AAC ATA ATT CTT **T**GA ACA CCG AGG CAG CAG **T**CT CTG TGG AGG GGG **T**A

PRE3 556   TCC ACA CCA GCA GC**T** GTC AAA ACA ACC A**T**T CTT ATA ACA CCA CCG GAA GA**T** CC

PRE7 3   GTT CAA TGG GTG **T**AT TTG ACG CC**T** CCG AAG AGT AT**T** CTG ATG CAA TAG **T**GG CC

PRE7 193   **T**CG CCG TCT GC**T** GCA AAT CCA **T**TC GCC GAC A**T**G ACT ATG TTA **T**CA CCA CA

PRE7 385   **T**CG AAC GAA **T**AG ACA GCG CCC **T**TA CCA TCT TCG **T**CA AGA CCC GCA A**T**G AT

PRE7 582   GTA GCA GAA G**T**G AAC GAG TC**T** CTC ACC AGT **T**TG ATG ACT TCT **T**CC ACG GAC

PRP8 520   GTG GTG GTG G**T**G GAA GTG C**T**A AGT CGC TGT CC**T** CTT CAA AAC CAG GAG G**T**g

PRP8 572   ATG GCA CCA TCG GA**T** TAT CCA GTT CT**T** CGA TTT CGT ATC CAG GCG G**T**G GT

PRP8 655   gaa G**T**C CAC AAT TTC **T**TC AGC GTT TAT **T**TC GAA GTT GCT **T**GG AGG AGG TGG **T**G

PRP8 904   ttG GAG GCC ATG **T**CA CTG TGA GTA **T**TA ATA ATC TTT CGC AGA TGT TCC GG**T** G

PRP8 2617 GAG **T**AG TGG TAT AG**T** GCC TCT TAA GAA AT**T** TAG CCA CAC **T**CT CCA CGC Agg **T**t

PRP8 2879 TTT TGA TTG GGG C**T**G GCA TAC CAG GAA CA**T** CCC ATG GAA **T**AT TAG CCT **T**C

PRP8 2968 GTG AGC TCC ACG C**T**T TAT ACG TTC **T**CG ATT GTA A**T**G AGC TGC AGA **T**AC CC

PRP8 3108 ccA **T**AA CAG AAA ATA **T**GG TCG TAG CC**T** CTT CGG GAG TAA **T**TT CGG GAC CGT **T**t

PUP1 51   CTA CAA TGG TGG **T**AC CCG TGG AAG **T**TG CCT TAG G**T**T GTG TGT GCG AG**T** TT

PUP1 385   **T**GC GCA TGA A**T**G GAG AAC AAA **T**GA GAT CCC G**T**A GGG TCT ACA CC**T** GCA AC

PUP1 489   **T**GG CTT CCT CC**T** TCG TAA GG**T** CTT GCT TCC AG**T** GTG ATT CCA ACA CAG CCA **T**T

PUP1 683   CTC CTT CAG CAC AGC AG**T** TGT ACC CC**T** GGG GAA TT**T** GTA GCT TTT C**T**G CT

RPB1 3   **T**GG ACC TCT TT**T** ACT GTA CGG AG**T** GGA GCA CTA GAA **T**AC TGT TGT CC**T** ACC

RPB1 122   CCT AGG GTC GTT **T**AG ACC ACC AA**T** TTT CGC TCT CG**T** CTG GGT TTC A**T**C CA

RPB1 449   ACC **T**CC CCT TGA **T**AC GAG CTG AG**T** AGG ATC ATC **T**TC AGA AGG GAC A**T**C TG

RPB1 569   TTC CTC CGT ACT **T**AA AAC TCT TAG **T**TC TGG TTC A**T**C CGC ATC CCC CGT GG

RPB1 730   **T**CC TCA CCT CT**T** TGA GAT TCA T**T**G AAG GAA ATG GA**T** GGA CGC ACC GG**T** GG

RPB1 1195 AAA **T**CT ATA CGG TC**T** CCG CTA TCA CGA A**T**G ACG TAT **T**TG GCA CCG GGG **T**G

RPB2 758   AAT AGT ACG AGC **T**GA ACT ACC CTC ACG ACC A**T**A AAG CTT GAC **T**TG AAG GG

RPB2 976   **T**AC CAA GGG CAG **T**AC CAC GAC GAC CAA **T**AA AGT CTA A**T**G CAG TTT CAC GA**T** C

RPB2 1353 GTT CAC CCC AGT **T**AC CAG TAG CCA AAG CGT **A**CT TAA GAC CCG A**T**G TTA TGG **T**TT

RPB2 1734 **T**TG CTG GGT TTC **T**GT GAA CAC CGT GCC AAA CAC CA**T** TGA CGA AGA CCC **T**T

RPB2 2134 CGT TTG GCA GGA **T**CA ACA TCG AGA **T**CG TTT TCC **T**CA TTT GCC TCT GCA GG

24

RPB2 2274   TAC CCA TCG CAG ATT GGT AAG TGT TAC GCG GAG ATT GAT TAT GGT CAG GG

RPB3 87   TGG CTA AAG TGG GAA TCT CTG CTA TCA TGA CTC TAC GGA GAG AAT TGG CC
RPB3 273   CGC TTT CGC CGA ACG CTT GCA ATG TTA ACA CCA CTG AAC ACT TAT CAC AG
RPB3 458   CTT GGC GTG CTC TTT GGC TAT ACC CTT TTT CGC AAC GCA TGT CAG TTT CA
RPB3 510   TGA GCT TAT TCC AAG GAT CAT ACT CGA ACT CTA TGG CTG CTG CTG GAC CC
RPB3 873   TAC CCA TTT GAG ATG CAT TGG AGT AAG GAT CTT GCT CAG CGC CCG TCA TC

SWI5 – 1   tAG CAG AGT TGG ACG TAG CTG AAC TAA CTC TAC TCT TTG GAG TGA AGG GCG
SWI5 – 2   tAG GAC CTG GTT TTC CAA GAA GTT CAA TAC CCA GGC CCA AAG GTT TCT CTG
SWI5 – 3   TTA GGG TCT AAA TCA TTG AAA ACG CCT TCC CGA TGT CGT GCT TGC TGG CTa
SWI5 – 4   TGG CGA AAC CAT AAT CCC CGT TCC ATT CCC ATC ATA ACT ACG CTC TCT TC
SWI5 – 5   ACT GTG GAC CAT CAG AGT TCT CTA AGG TAG CCT GTA GTT GTT GTT GCT TG
SWI5 – 6   ACT CGG AGA AAG TAG AAA CCT TGA CAG ATT TTT CAT CAG AAA GTG GTC CCA GCG G
SWI5 – 7   GAA GGG TCA CTG CCT CTA GCA TTT GAG TAA TAA GAG TTG GTT TGT AGG TCA A

TAF5 196   GCT CTC AGC ATG GCC TCT GTA CGG TGG TAT CCC TTT TTA TTC AGG TAC TC
TAF5 381   CTA GTC GCC CGC TAG ATA CTA TAC CAC CTT CTG CAT CTC TCT TGG ATG AC
TAF5 1633   ACT AGC GCG GTG TGA GTA TCC ATG GAC CAT AGC CTG ACG GTT TTA TCT TC
TAF5 1854   TTC GAC ACG TCT TAT CGC TGG ATC CCG TAA ACA CAT AGC AAC CGT TCG GA
TAF5 2029   CCA TGG CCA CGC ATT TGT TTC AAC CTT TTG CCG GTA CCA ATA TCC CAT AC
TAF5 2176   AAG TAT CCA ATG AAT GGC TCA TCT GGT TCT GCG CTT GGC TCA GTA GTG GC

TAF6 294   CTT CTT CCT CAT CCA GGT AGT AAA CCG ACT GTC CTC CAC TTG TGT TAA CC
TAF6 369   TGC TAG CCA ATG TGT AGT AAA AGT TGG TAG ACG AGG CAC TTG CGG TAA TGG C
TAF6 525   TCA CAG AAG CAC TTG CTG TCG TCG ACG TTA CAG GTG TTT GGA GGC TGT TA
TAF6 747   TGG AAC CAG TTG GTG CAG GCC ACT ATC AGT TCG TAG TGA AGT CAA AGC TGC T

TAF12 62   ACT GTC TCG GTT GCA GTA CCA TAT TTT GCT GAG CTC CTG TGA TCG CAT CAG
TAF12 220   CTT GCG CAG CCT GTT GTC TTC TCC TAT TCA GTG TAA GGG CAT CAT ATA CCT G
TAF12 547   GCA GTT CAA TTT TCT TTT CTT GTA GCT GCT GCT TGG GCC CTT CGT CCG TCT G
TAF12 744   GCT GCT GCT GCT GCT GTT GTT GTT GTT GTT GTT GTA CCC TTT GTT GCT GC
TAF12 800   TCT TTG GTT TTG ACC CTG GCG TTG CTG TTG CTG TTG CTG TTG CTG TTG CT