SUPPORTING MATERIAL

Batenchuk *et al.* Chromosomal position effects are linked to Sir2-mediated variation in transcriptional burst size. *Biophys. J.* (2011).

SUPPORTING METHODS

1. Strain generation

Promoter-reporter expression cassettes were obtained by inserting a PCR-amplified yeast-enhanced green fluorescent protein, *yEGFP*, into a pRS406 vector backbone (Stratagene) using SacI and EcoRI. The endogenous P_{ACT1} and P_{ADH1} promoters were isolated from genomic DNA (strain BY4741) by PCR amplification of a 1kb region upstream of the *ACT1* or *ADH1* start codon, and inserted into the modified pRS406 vector using BamHI or XhoI and EcoRI. The resulting vector (**Fig. S5**) was designed such that reporter expression following chromosomal integration is shielded from distal upstream effects by an *URA3* expression cassette and a *CYC1* transcriptional terminator positioned upstream of the reporter cassette. Promoter sequences were confirmed by sequencing.

Genomic integration of the reporter-expression cassette was achieved by PCR-mediated gene replacement of the KanMX cassette used to generate the deletion library in the parental strain BY4741 (Open Biosystems). Successful integrations were confirmed by testing uracil prototrophic strains for the loss of kanamycin resistance by replica plating onto YPD plates containing 2 mg/ml G418 (Winsent). PCR confirmation of 15 randomly selected strains confirmed for loss of kanamycin resistance yielded an efficiency of 100%.

Loci selected for integration include all but one non-dubious euchromatic open reading frames (ORFs). The only ORF not analyzed is YCL073C, which encodes a protein of unconfirmed function, since the corresponding deletion mutant is not available.

2. Media and growth conditions

Strains were arrayed in a 96-well plate format and stored at -20 ^{o}C in YPD media supplemented with 15% w/vol glycerol. Prior to analysis, plates were thawed on ice for 1 hr and 20 µL aliquots were inoculated into 400 µL YPD supplemented with 42 mg/l adenine. Following ~17 hours of growth at 30^{o}C and continuous shaking (250 rpm), optical density at 600 nm (OD_{600}) was measured using a Victor3V 1420 Multilabel Counter (PerkinElmer). Each sample was subsequently diluted to an OD_{600} of ~0.006 and incubated for an additional 6 hours. Samples were supplemented with 5 mM of nicotinamide (Sigma) prior to incubation when applicable. This concentration was previously shown in Ref. [1] to inhibit Sir2 activity in vivo.

3. Flow cytometry and data processing

Flow cytometry was carried out on a 1024-channel Beckman Coulter FC 500 MPL System equipped with a custom 96-well plate loader. A 20 mW Argon laser provided excitation at 488 nm and fluorescence emission (FL1) was collected through a 525±15 nm band-pass filter. Typically, 30000 events were recorded for each sample. Three replicates were analyzed for most samples.

Analyzing only events within a narrow forward-scatter (FS) and side-scatter (SS) gate capturing a small fraction of the total cell population is often used to minimize extrinsic variability (see e.g., Refs. [2,3]). This can be problematic since a small gate capturing a low number of cells will have high uncertainty in the estimate of the mean FL1 signal and its variance. Additionally, since FS/SS values vary across different genetic backgrounds, cells captured by a narrow fixed gate will not consistently correspond to the most representative sampling of the population.

To circumvent these issues, we calculated, the mean FL1 signal and its variance for events with identical FS and SS values acquired within an elliptical auto-gate capturing 80% of all events measured on the same day for the same promoter. This methodology, results in up to ~10^6 estimates for each point on a 1024-by-1024 grid of FS and SS values. The mean overall signal and its variance were subsequently calculated as the weighted average of individual estimates. This greatly reduces the extrinsic component of the noise (**Fig. S6**). The impact of sample-specific effects (e.g., gene deletion effects or the strain generation errors) were mitigated by excluding samples with a low event count (<5000) or with less than 40% of events within the FS/SS autogate from further analysis. Therefore, data was not available for all loci in triplicate. For P_{ACT1} expression no data is available for the following ORFs: YCR003W, YCR020W-B, YCR028C and YCR053W. For P_{ADH1} expression no data was available for the following ORFs: YCL076W, YCL075W, YCL034W, YCR002C, YCR020W-B, YCR053W and YCR107W.

To correct for autofluoresence (AF), we treated AF and *yEGFP* fluorescence as independent random variables. With this assumption, population-average FL1 signal attributable to the average protein abundance $\langle p \rangle$ is given by $\langle p \rangle = \langle p_{\mathrm{FL1}} \rangle - \langle p_{\mathrm{AF}} \rangle$ where subscripts FL1 and AF refer to the raw FL1 signal for *yEGFP* expressing and non-expressing strains, respectively. The noise associated with variation in protein abundance among them is then given by the equation $\eta^2 = \left(\sigma^2_{\mathrm{FL1}} - \sigma^2_{\mathrm{AF}}\right)\langle p \rangle^2$ where σ^2 terms are the variances of the raw FL1 signal for *yEGFP* expressing and non-expressing strains. Noise estimates from samples with an average FL1 value less than three standard deviations above the mean AF were excluded from further analysis.

The normalized experimental data is provided as a text file.

4. Enrichment analyses

Enrichment analysis was performed in MATLAB. Loci within low expression (LE) and high noise (HN) regions were identified when the number of measurements within the 20$^{\mathrm{th}}$ (expression) or the 80$^{\mathrm{th}}$ percentile (noise) acquired across nearest-neighbour positions was greater than that expected under a random model (hypergeometric test, $p < 0.01$). The position of loci within identified regions are displayed in **Fig. S2a** (low expression) and **Fig. S2b** (high noise).

Regions de-enriched in polymerase occupancy or Sir2-mediated histone acetylation (H3K9, H3K14 and H4K16) were identified by analyzing a 12.5 kb window surrounding the native open reading frame. Loci were classified as de-enriched in polymerase activity or as enriched in Sir2 activity when the number of measurements within the 25$^{\mathrm{th}}$ percentile was greater than that expected under a random model (hypergeometric test, $p < 0.005$). The identified regions are displayed in **Fig. S2c** (low polymerase activity) and **Fig. S2d** (high Sir2 activity). Data for this analysis was generated by Liu et al.[4] In all cases, low polymerase or high-Sir2 regions are associated with LE and HN for both promoters (hypergeometric test, $p < 0.05$).To ensure robustness of this analysis, we systematically characterized the effect of varying the percentile cutoffs by ±5%, the size of the moving window by ±2.5kb and critical p values by a two-fold increase or decrease. Consistent fold enrichment was observed for both polymerase and Sir2 activity as illustrated in **Fig. S2e** and **Fig. S2f**, respectively.

The positions most affected by nicotinamide treatment were identified using a similar approach. In this case, however, a position-specific effect was determined by evaluating the deviation from the average effect of nicotinamide treatment. First, the average effect of nicotinamide was fitted to a linear model (**Fig. S3a**). Next, the deviation from this trend was determined by examining the residuals of this fit (**Figs. S3c and S3d**). The positions within the 80$^{\mathrm{th}}$ percentile of the residuals for average expression or the 20$^{\mathrm{th}}$ percentile of the residuals for expression noise were classified as those most affected by the treatment. These positions are also significantly enriched in high Sir2 activity (**Fig. S3e**).

5. Model discrimination

We consider the steady state dependency of noise η on average protein abundance $\langle p \rangle$ for the model in Fig. 2a. The model predicts that the average protein abundance is given by:

$$\langle p \rangle = \frac{k_M k_P}{\gamma_M \gamma_P}\left(1 - \frac{k_{off}}{k_{off} + k_{on}}\right), \qquad \text{Eq. (S1)}$$

where k_M is the rate of mRNA synthesis, k_P is the rate of protein synthesis per mRNA, k_{on} is the promoter activation rate, k_{off} is the promoter deactivation rate, and γ_M and γ_P are the rates of mRNA and protein decay, respectively. The maximal average protein abundance is given by $p_{\max} = k_M k_P / \gamma_M / \gamma_P$.

The intrinsic noise η is derived in Ref. [5] and given by:

$$\eta^2 = \frac{k_P}{\gamma_M + \gamma_P}\frac{1}{\langle p \rangle} + \frac{\gamma_M \gamma_P k_{off}\left(\gamma_M + \gamma_P + k_{off} + k_{on}\right)}{k_{on}\left(\gamma_M + \gamma_P\right)\left(\gamma_M + k_{off} + k_{on}\right)\left(\gamma_P + k_{off} + k_{on}\right)}. \qquad \text{Eq. (S2)}$$

In the following, we consider four different variants of the dependency in Eq. (S2): Modulation of k_M (Model A), k_{off} (Model B) or k_{on} (Model C), or the absence of transcriptional bursting (Model D)

For Model A, varying k_M only impacts the noise through variation of the average protein abundance. Correspondingly, the noise given by Eq. (S2) assumes that k_{off} and k_{on} are constant across chromosomal positions. This dependency can be written as:

$$\eta_B^2 = \frac{C_1}{\langle p \rangle} + \frac{\tau_0^2 \tau_2}{\tau_1\left(1 + \tau_0\right)}\frac{\tau_0 + \tau_1 + \tau_0\tau_2 + 1}{\left(\tau_1 + \tau_0\tau_2 + 1\right)\left(\tau_0 + \tau_1 + \tau_0\tau_2\right)} = \frac{C_1}{\langle p \rangle} + C_0. \qquad \text{Eq. (S3)}$$

where the three time-scale parameters are given by $\tau_0 = \gamma_P/\gamma_M$, $\tau_1 = k_{on}/\gamma_M$ and $\tau_2 = k_{off}/\gamma_P$, and C_0 is a constant defined by these parameters.

For Model B, the noise dependency can be obtained by first expressing k_{off} in terms of the average protein abundance in Eq. (S1), then inserting the resulting function into Eq. (S2). The result is given by:

$$\eta_C^2 = \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{\tau_1}{1+\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)\left(\frac{C_2}{\langle p \rangle} - 1\right)}{\left(\frac{\tau_1}{\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)\left(\tau_1\frac{C_2}{\langle p \rangle} + 1\right)} \approx \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)}{\left(\frac{\tau_1}{\tau_0}\frac{C_2}{\langle p \rangle} + 1\right)}, \qquad \text{Eq. (S4)}$$

where $C_2 = p_{\max}$ and the approximation is valid when $\tau_0 << 1$.

For Model C, a similar approach yields a noise dependency given by:

$$\eta_D^2 = \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)^2 \left(\frac{C_2}{\langle p \rangle}\left(\frac{\tau_0\tau_2}{1+\tau_0} + 1\right) - 1\right)}{\left(\frac{C_2(\tau_0\tau_2 + 1)}{\langle p \rangle} - 1\right)\left(\frac{C_2(\tau_2 + 1)}{\langle p \rangle} - 1\right)} \approx \frac{C_1}{\langle p \rangle} + \frac{\left(\frac{C_2}{\langle p \rangle} - 1\right)^2}{\left(\frac{C_2(\tau_2 + 1)}{\langle p \rangle} - 1\right)}. \quad \text{Eq. (S5)}$$

where $C_2 = p_{\max}$ and the approximation is valid when $\tau_0 << 1$.

For Model D, absence of transcriptional bursting is achieved when $k_{off} + k_{on} >> 1$. Imposing this condition yields a noise dependency given by:

$$\eta_A^2 = \frac{k_P}{\gamma_M + \gamma_P}\frac{1}{\langle p \rangle} = \frac{C_1}{\langle p \rangle}, \quad \text{Eq. (S6)}$$

where $C_1 = k_P / (\gamma_M + \gamma_P)$.

Experimental data were fitted to the noise dependencies predicted for the four different models using robust nonlinear least squares implemented using Matlab's "fit" function. We used the approximate dependencies for Models C and D since the assumption $\tau_0 << 1$ reduces the number of parameters with negligible impact on the goodness of fit (data not shown). Estimated parameter values are given in **Table S2**.

We discriminated among different models by examining the logarithmic error ratios (LER) defined for two models *X* and *Y* whereby LER = $\log(SSE_Y/SSE_X)$ and SSE is the sum squared error. For the chromosome-wide datasets, the *p*-value associated with the null hypothesis that two models perform equally well was calculated by cross-validation. Briefly, the loci were split randomly into two sets of equal size, one for training and one for validation. For each split, the four models were fitted to the training dataset and the SSE calculated using the validation dataset. A second SSE was calculated by swapping the training and validation loci, and the process repeated 50 times to obtain a distribution of LER and its standard deviation for each model. An LER that is consistently positive throughout this process indicates that model *Y* performs better than model *X*. The *p* values were calculated with Matlab's ztest function using the estimated mean LER and its standard deviation. The results are given in **Table S3**.

A slightly different approach was used to analyze the impact of Sir2 deletion. Because of the low sample size in these experiments, we estimated the mean LER and its standard deviation from five estimated SSE values. The estimated SSE values were in turn obtained by removing one of the five measured loci from the dataset. The mean LER, its standard deviation and the *p* value associated with the hypothesis that two models fit the data equally well are given in **Table S4**.

7. References

1. Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M. & Sinclair, D. A. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast Sir2 and human SIRT1. *J. Biol. Chem.* **277**, 45099-45107 (2002).
2. Newman, J. R. S. *et al*. Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. *Nature* **441**, 840-846 (2006).
3. Becskei, A., Kaufmann, B. B. & Van Oudenaarden, A. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nat. Genet.* **37**, 937-944 (2005).
4. Liu, C. L. *et al*. Single-nucleosome mapping of histone modifications in S. cerevisiae. *PLoS biology.* **3** (2005).
5. Raser, J. M. & O'Shea, E. K. Control of stochasticity in eukaryotic gene expression. *Science* **304**, 1811-1814 (2004).

SUPPORTING FIGURES

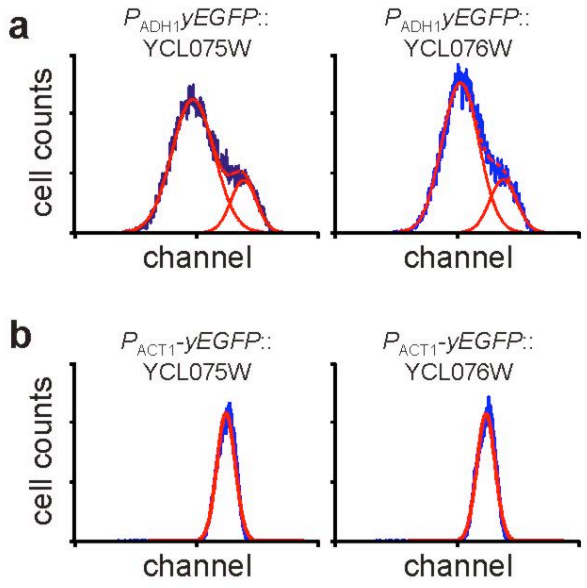

Figure S1: Promoter-dependent expression distributions

(**a**) Bimodal fluorescence histograms observed at positions near the left telomeric region for P_{ADH1} expression. Red curves illustrate the de-convolution of the histogram into two normal distributions. (**b**) Unimodal fluorescence histograms obtained at the same positions for P_{ACT1} expression.

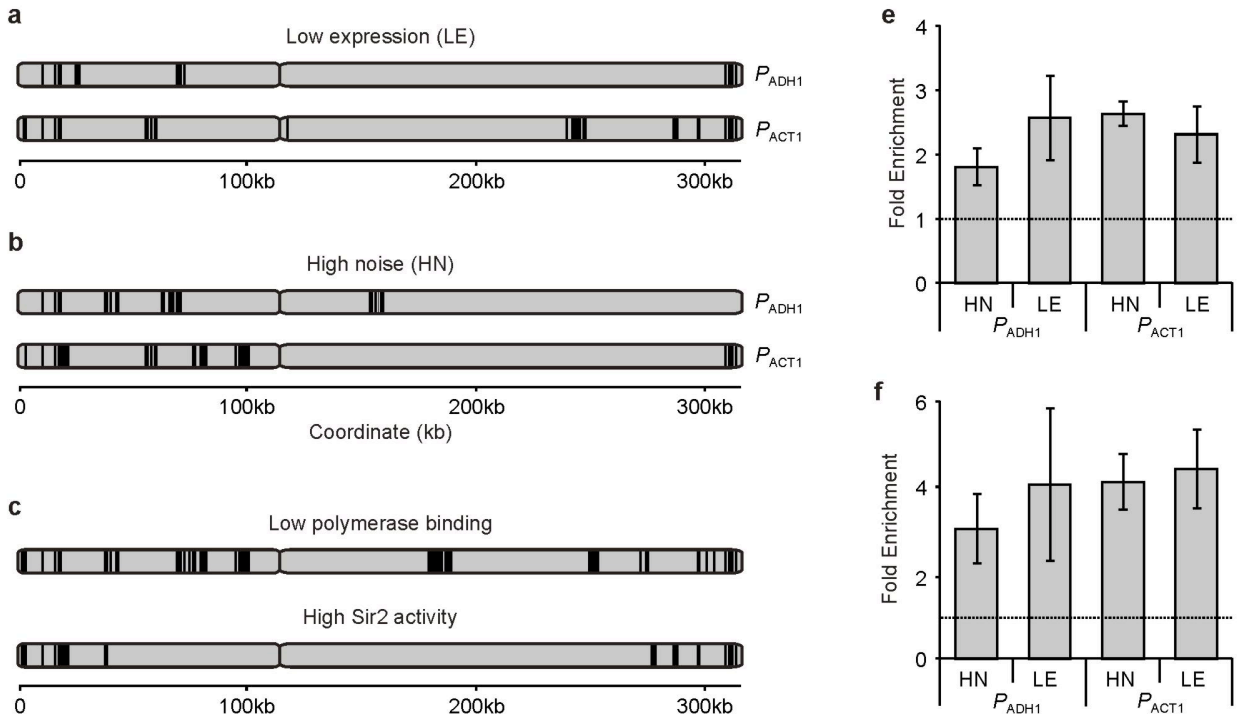

Figure S2: Analysis of low expression and high noise regions

(**a**) The loci within regions identified as low expression (LE) for the two promoters. (**b**) The loci within regions identified as high noise (HN) for the two promoters. (**c**) The loci identified as enriched in depletion of polymerase binding or depletion of Sir2-mediated histone acetylation (H3K9, H3K14 and H4K16). (**e**) Fold enrichment of LE and HN loci within regions of low polymerase binding. (**f**) Fold enrichment of LE and HN loci within regions of high Sir2 activity. Error-bars represent standard deviations calculated by systematic variation of the parameters used to classify loci. The details of the enrichment analysis are provided in Methods.

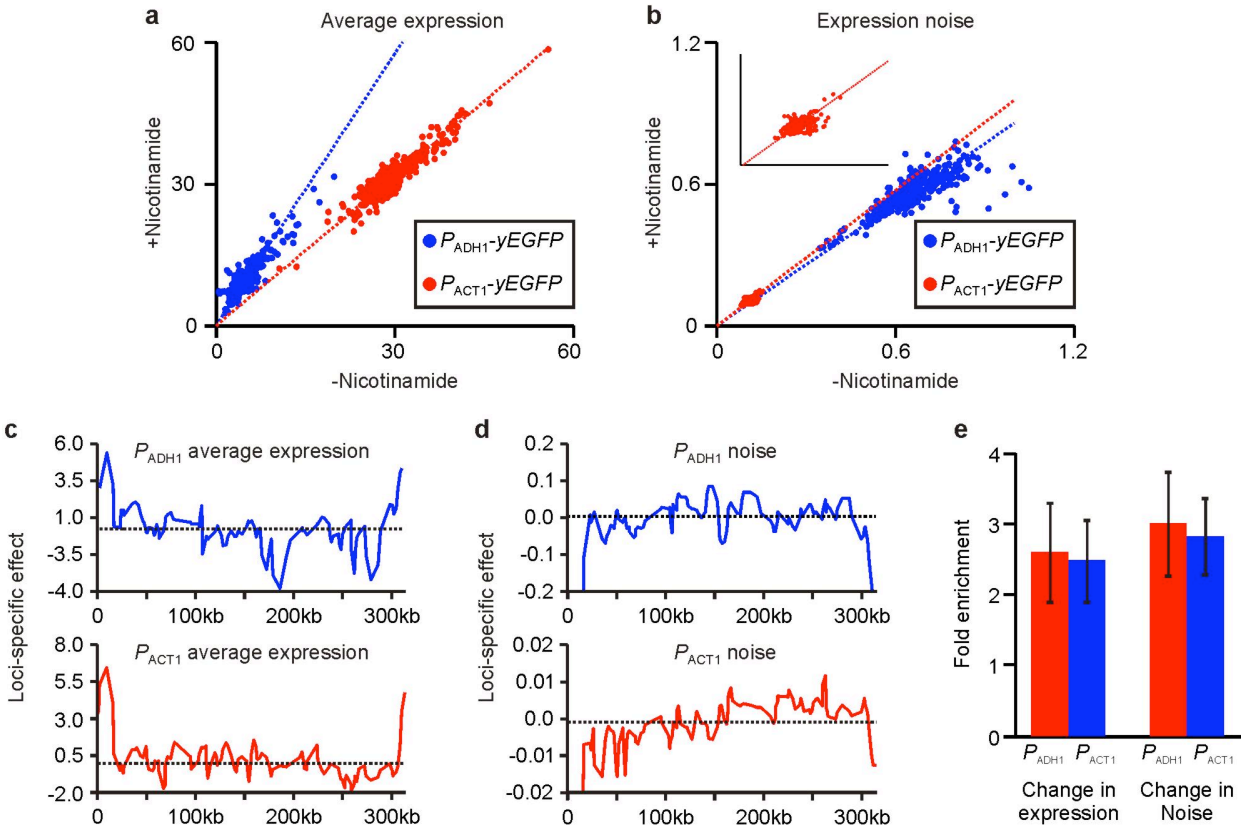

Figure S3: Effects of nicotinamide treatment

(**a**) Correlation between population-averaged expression in the presence and absence of 5mM nicotinamide. The slopes are 1.92 and 1.05 for P_{ADH1} and P_{ACT1} population-averaged expression, respectively. (**b**) Correlation between gene expression noise in the presence and absence of 5mM nicotinamide. The insert is an enlargement of the P_{ACT1} data. The slopes are 0.86 and 0.96 for P_{ADH1} and P_{ACT1} noise, respectively. (**c**) Position-specific effects of nicotinamide on average expression (see Methods). (**d**) Position-specific effects of nicotinamide on gene expression noise. (**e**) Enrichment of loci most affected by nicotinamide treatment within regions of high Sir2 activity. Error bars represent standard deviations calculated by systematic variation of the parameters used in the statistical analysis.

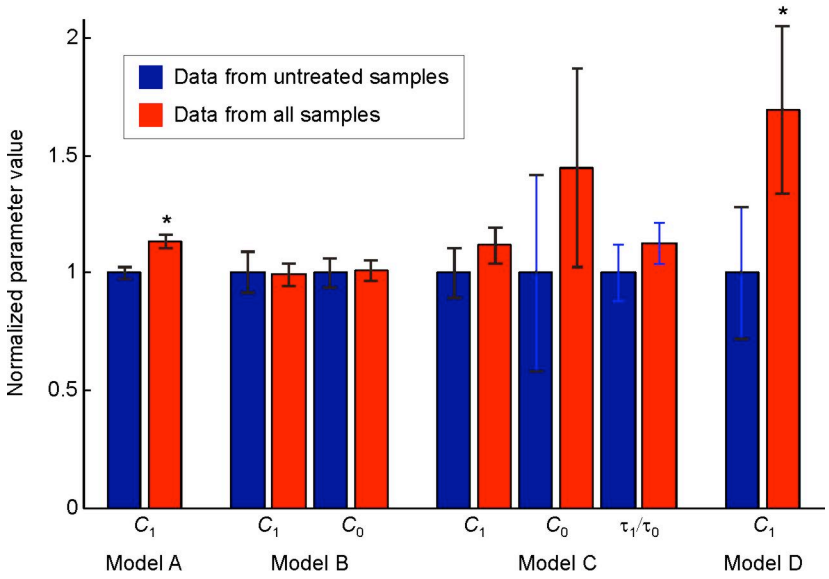

Figure S4: Impact of nicotinamide on fitted model parameters

The four models were fitted to data obtained from untreated P_{ADH1} samples only (blue) and all P_{ADH1} samples (red). The error-bars indicate 95% confidence intervals. Asterisk indicates a statistically significant change in the parameter value (non-overlapping 95% confidence intervals). The absence of a statistically significant difference in the fitted parameters indicates that the model obtained using untreated samples can account for the effect of nicotinamide treatment. All parameters were normalized to unity for the untreated data. Actual parameter values are given in Table S2. Only the parameter C_1 is shown for Model D (see Table S2 for details).

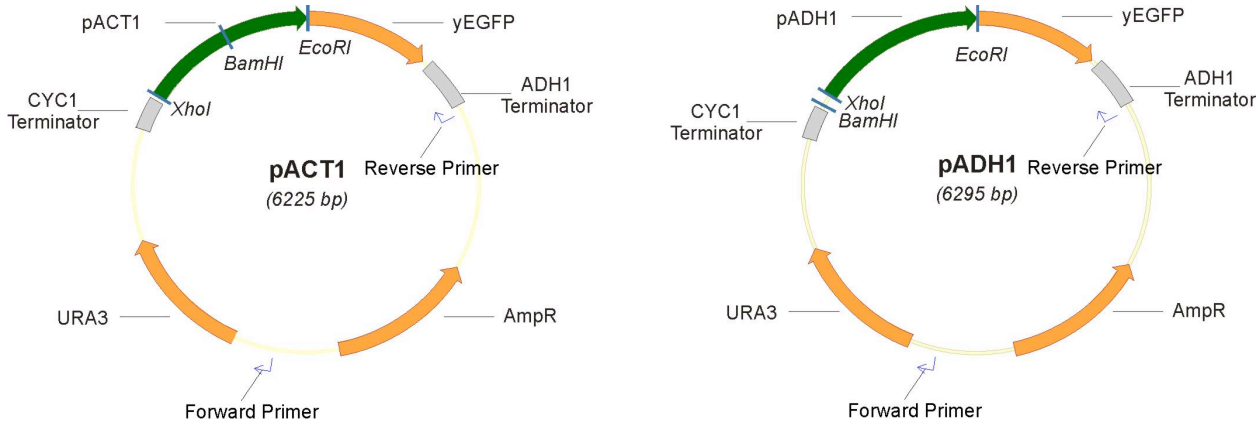

Figure S5: Maps of constructed plasmids

The Forward Primer and Reverse Primer indicate the positions used for genomic integration.

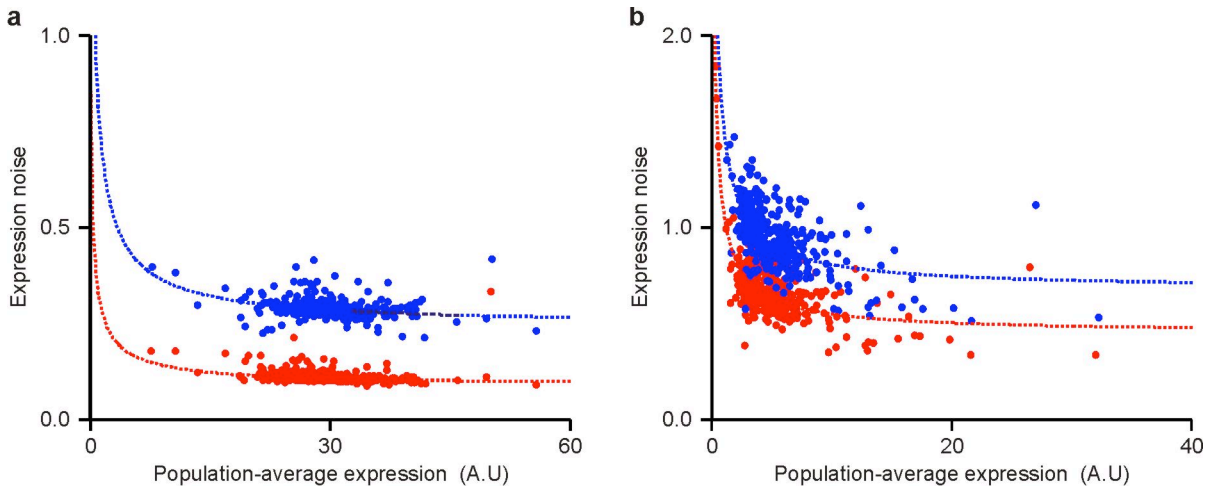

Figure S6: Extrinsic noise reduction

Reduction of extrinsic noise for (**a**) P_{ACT1} and (**b**) P_{ADH1} expression. Blue points are noise measurements without forward- and side-scatter gating. Red points are noise measurements within a fixed 80% autogate when cells are grouped by their forward- and side-scatter channel values. Estimates of the average expression and noise were obtained as the weighted average of the mean and the noise within these groups (see Methods). Broken curves are fits to Model B.

SUPPORTING TABLES

a **P_{ACT1}/P_{ADH1} correlations**

	r	*p* value
Average	0.29	1.50E-03
Noise	0.51	3.90E-09

b **P_{ACT1} noise correlations**

	r	*p* value
PolII	-0.48	1.50E-08
H3K14Ac	-0.52	4.10E-10
H3K9Ac	-0.6	1.20E-13
H4K16Ac	-0.63	3.80E-15

c **P_{ADH1} noise correlations**

	r	*p* value
PolII	-0.32	2.90E-04
H3K14Ac	-0.35	7.70E-05
H3K9Ac	-0.45	2.50E-07
H4K16Ac	-0.64	2.40E-15

Table S1: Correlation coefficients and *p* values

Correlation coefficients *r* and *p* values associated with the null hypothesis of no correlation for different experimental datasets. (**a**) Correlations between average expression and expression noise for the two promoters. (**b, c**) Anti correlations between RNA polymerase II binding and histone acetylation for noise in P_{ACT1} and P_{ADH1} expression respectively.

a P_{ADH1} data untreated samples

	C_1	CI	C_0 or C_2	CI	τ_1/τ_0	CI	τ_2	CI
Model C	1.70	±0.05	-	-	-	-	-	-
Model A	0.21	±0.02	0.95	±0.05	-	-	-	-
Model B	0.64	±0.07	46.11	±19.28	3.04	±0.36	-	-
Model D	1.70	±1.22	40.00	NA	-	-	3.9E+04	±2.0E+07

B P_{ADH1} data all samples

	C_1	CI	C_0 or C_2	CI	τ_1/τ_0	CI	τ_2	CI
Model C	1.93	±0.04	-	-	-	-	-	-
Model A	0.21	±0.01	0.94	±0.05	-	-	-	-
Model B	0.72	±0.05	66.73	±19.63	3.42	±0.27	-	-
Model D	2.88	±0.60	40.00	NA	-	-	4.7E+04	±3.8E+07

Table S2: Fitted model parameters for P_{ADH1} expression

CI is the confidence interval. The confidence interval for C_2 is not applicable (NA) since the parameter is fixed at the lower bound for the maximal average protein abundance. Note that the value of τ_2 is unconstrained.

a **P_{ACT1} data untreated samples**

	Slow vs fast kinetics			Size vs frequency		k_M vs k_{off}
	A vs D	B vs D	C vs D	A vs C	B vs C	A vs B
Mean	4.1E-02	-5.5E-03	-6.9E-04	4.2E-02	-4.8E-03	4.7E-02
Stdev	1.0E-01	7.0E-02	1.8E-03	1.0E-01	7.0E-02	6.2E-02
p-value	3.4E-01	5.3E-01	6.5E-01	3.4E-01	5.3E-01	2.2E-01

b **P_{ADH1} data untreated samples**

	Slow vs fast kinetics			Size vs frequency		k_M vs k_{off}
	A vs D	B vs D	C vs D	A vs C	B vs C	A vs B
Mean	5.5E-01	4.5E-01	-2.5E-05	5.5E-01	4.5E-01	9.4E-02
Stdev	1.4E-01	2.2E-01	8.8E-04	1.4E-01	2.2E-01	1.3E-01
p-value	5.8E-05	1.7E-02	5.1E-01	5.8E-05	1.8E-02	2.4E-01

c **P_{ADH1} data prediction of treated samples**

	Slow vs fast kinetics			Size vs frequency		k_M vs k_{off}
	A vs D	B vs D	C vs D	A vs C	B vs C	A vs B
Mean	1.7E+00	1.6E+00	5.0E-05	1.7E+00	1.6E+00	5.4E-02
Stdev	1.3E-01	1.2E-01	2.0E-05	1.3E-01	1.2E-01	8.4E-02
p-value	2.7E-37	3.7E-46	6.0E-03	2.7E-37	3.7E-46	2.6E-01

Table S3: Model discrimination using chromosome-wide data

The mean and standard deviation (stdev) of the log error ratio (LER) associated with different model comparisons (see Methods). Models A and B assume variation in burst size through k_M and k_{off}, respectively. Models C and D assume variation in burst frequency through k_{on} and absence of transcriptional bursting, respectively. The *p*-value is associated with the null hypothesis that the mean LER is greater than zero and obtained using a one-sided z-test. (**a, b**) Comparison of models fitted to untreated data for P_{ACT1} and P_{ADH1} expression, respectively. (**c**) Comparison of how well the different models predicts the effect of nicotinamide treatment.

	Slow vs fast kinetics			Size vs frequency		k_M vs k_{off}
	A vs D	B vs D	C vs D	A vs C	B vs C	A vs B
Mean	3.1E+00	2.0E+00	-6.1E-04	3.1E+00	2.0E+00	1.1E+00
SD	4.7E-01	2.0E-01	3.6E-05	4.7E-01	2.0E-01	5.6E-01
p-value	3.3E-11	1.3E-22	1.0E+00	3.3E-11	1.3E-22	2.7E-02

Table S4: Model discrimination using Sir2 deletion data

The mean and standard deviation (SD) of the log error ratio (LER) associated with different model comparisons (see Methods). Models A and B assume variation in burst size through k_M and k_{off}, respectively. Models C and D assume variation in burst frequency through k_{on} and absence of transcriptional bursting, respectively. The *p*-value is associated with the null hypothesis that the mean LER is greater than zero and obtained using a one-sided z-test.