SI Text

Primer Sequences. p965 (5'-biotin-TAATGCAGCTGGCACGACAGG-3')

BKS8 (5'- ATGTATGTACAACACACATCGG-3')

pNot (5'- GGCCGCTCTAGCTGCATTAATG-3')

BKS7 (5'- TACCGAGCTCGAATTCGGAGG-3')

Recombinant TFIIS Proteins. Five-hundred-milliliter cultures of *E. coli* strain BL21(DE3) transformed with TFIIS expression plasmids were grown to an OD_{600} of 0.5 before being induced with 0.4 mM IPTG for 3 h at 37°C. Cells containing TFIIS or derivatives were pelleted, resuspended in 20 ml of Buffer A [20 mM Hepes (pH 7.5), 10 μM $ZnCl_2$, 50 mM NaCl, 10% glycerol, 5 mM imidazole, 1 mM PMSF, and 2 mM Benzamidine] containing 75 units Benzonase (Novagen) and sonicated. Sonicated lysates were centrifuged at 8,000 × *g* for 30 min at 4°C. Clarified lysates were passed over columns containing 2 ml of Talon-immobilized metal-affinity resin (Clontech) preequilibrated in Buffer A containing 500 mM NaCl (Buffer A 500). After washing the resin with 20 ml of Buffer A 500, TFIIS and truncation mutants were eluted with Buffer A 500 containing 10 mM to 200 mM imidazole. Fractions containing the desired proteins were combined and dialyzed against 20 mM Hepes (pH 7.9), 100 mM potassium acetate (KOAc), 10% glycerol, 10 mM 2-mercaptoethanol (B-ME), and 10 μM ZnCl and stored at -80°C. Proteins were estimated to be >95% pure by analysis on Coomassie-stained SDS/PAGE.

Immobilized Promoter Templates. For the experiments in Fig. 1 and 4, a 400-bp fragment containing 71 base pairs of the yeast HIS4 promoter centered around the TATA box, one Gal4 binding site, and upstream vector sequences was PCR amplified from pSH515 with biotinylated upstream primer p965 and downstream primer BKS8. For all

other experiments a 600-bp fragment containing additional sequence downstream of the transcription start site was PCR amplified from pSH515 with p965 and downstream primer pNot. The "no-promoter" template was created by PCR amplifying a 306-bp fragment that ends 14 bp downstream of the Gal4 site with p965, BKS7, and pSH515.

Immobilized Template Assays. Immobilized promoter assays were performed as described in refs. 1 and 2) and online at www.fhcrc.org/scince/basic/labs/hahn/. For Western blot analysis, clarified nuclear extracts (~3.6 mg/ml) were incubated with immobilized templates (5 μl, 1 pmol) in 100 μl of transcription buffer [20 mM Hepes (pH 7.6), 100 mM KOAc, 5 mM MgOAc, and 1 mM EDTA) with 250 ng of HaeIII-digested *E. coli* DNA for 40-60 min at room temperature. Gal4-AH, Gal4-VP16, Gal4-GCN4, rTFIIS, and rTFIIS truncation mutants were added during PIC formation where indicated. The templates were concentrated with a magnet and washed three times with 1 ml of transcription buffer containing 0.05% Nonidet P-40 and 2.5 mM DTT. Templates were resuspended in 10 μl of PstI buffer [100 mM NaCl, 50 mM Tris·HCl (pH 7.9), and 10 mM $MgCl_2$] with 10 units of PstI (Boehringer Mannheim) for 30 min at 37°C. The beads were concentrated with a magnet, and the supernatants were recovered. For Western blot analysis, proteins were electrophoresed on a 4–12% Bis-Tris gel (Invitrogen), transferred to Immobilon membranes (Millipore), and probed with antibodies to known PIC components. Proteins were detected by ECL (Amersham Pharmacia).

For quantitative MS analysis of PICs formed in the presence and absence of rTFIIS, reactions were scaled up 135 times compared with the reactions used to analyze PICs by Western blotting. After incubating nuclear extract prepared from *DST1* deletion strain Y4411 with templates for 60 min at room temperature, the templates were washed three times with transcription buffer containing 0.05% Nonidet P-40 and 2.5 mM DTT, followed by one wash with transcription buffer containing 0.003% Nonidet P-40. Templates were resuspended in 1.35 ml of PstI buffer with 405 units of PstI (Boehringer Mannheim) and incubated for 30 min at 37°C. The beads were concentrated with a magnet, and the supernatants were recovered.

Transcription Assays with Plasmid Templates. *In vitro* transcription reactions with plasmid templates were performed as described in refs. 1 and 3 and online at www.fhcrc.org/scince/basic/labs/hahn/. Briefly, ~120 μg of nuclear extract was incubated with 0.15 μg of the indicated transcription template in 25 μ ofl transcription buffer containing 2.5 mM DTT, 10 units of ribonuclease inhibitor (Promega), for 60 min (unless indicated otherwise) at room temperature. Gal4-AH, Gal4-VP16, rTFIIS, and rTFIIS derivatives were included in the incubation where indicated. Transcription was initiated by the addition of each nucleotide triphosphate (NTP) to 0.4 mM. Reactions were stopped after 2 min by the addition of 180 μl of stop solution (0.1 M sodium acetate, 10 mM EDTA, 0.5% SDS, 15 μg/ml tRNA). RNA was phenol/chloroform extracted, ethanol precipitated, and analyzed by primer extension or S1 nuclease protection. Products were quantified by PhosphorImager (Molecular Dynamics).

Transcription Assays with Immobilized Templates. *In vitro* transcription reactions with immobilized templates were performed as described in ref. 1 and online at www.fhcrc.org/scince/basic/labs/hahn/ with the following exceptions: 240 μg of nuclear extract and 2.5 μl (0.5 pmol) of HIS4-immobilized template were used per reaction. After a 60-min incubation, immobilized templates were concentrated with a magnet, and washed three times with 150 μl of transcription buffer containing 0.05% Nonidet P-40 and 2.5 mM DTT. After resuspension in 50 μl of transcription buffer containing 2.5 mM DTT and 20 units of ribonuclease inhibitor, transcription was initiated by the addition of 100 μM concentrations of each NTP. Reactions were stopped after 10 min by the addition of 360 μl of stop solution. The supernatant was removed from the beads, phenol/chloroform extracted, and ethanol precipitated. Products were mapped by primer extension as described above, except that actinomycin C1 was included (15 μg/ml) during the extension reaction.

Preparation of purified Pol II PICs for quantitative MS analysis. To analyze the composition and quantity of the isolated proteins, 2% of the samples were analyzed by silver-stained SDS/PAGE. The remaining samples were concentrated in Microcon 10

devices (Amicon), and buffer was exchanged by diluting the samples 10-fold with 20 mM Tris·HCl (pH 8.3), 50 mM NaCl, and 1 mM EDTA. After concentrating the samples to 30 μl, SDS was added to 0.3%, and the samples were heated at 100°C for 5 min. Proteins (~65 μg per sample) were reduced with 5 mM TCEP at 37°C for 30 min and then diluted with 250 μl of 20 mM Tris·HCl (pH 8.3), 1 mM EDTA, and 7.2 M urea. Isotopically heavy (+rIIS) or normal (-rIIS) cleavable ICAT reagent (Applied Biosystems) was added to 1.75 mM, and samples were incubated for 90 min at 37°C. Reactions were quenched by the addition of 10 mM DTT for 20 min at 37°C, samples were combined, and proteins were digested by the addition of 4 μg of endoproteinase Lys-C (Beohringer Mannheim) at 37°C for 3 h. SDS and urea concentrations were reduced to 0.01% and 1.2 M, respectively, by the addition of 20 mM Tris·HCl (pH 8.3) and 1 mM EDTA, and samples were digested with trypsin (sequencing grade modified, 1:25 wt/wt; Promega) overnight at 37°C. The sample was diluted with an equal volume of SCX Buffer A [5 mM KH_2PO_4 (pH 3)/25% CH_3CN], and the pH was adjusted to 3 with 10% trifluoroacetic acid (TFA). Peptides were fractionated by SCX HPLC (2.1 × 200 mm PolySULFOETHYL A (PolyLC) by running the following gradient: 0–15% SCX Buffer B [5 mM KH_2PO_4 (pH 3)/600 mM KCl/25% CH_3CN] over 30 min, 15–60% SCX Buffer B in 20 min, and 60–100% SCX Buffer B in 15 min at 0.2 ml/min. The 0.4-ml fractions were collected. ICAT labeled peptides from 21 fractions showing absorbance at 214 nm were pooled into seven fractions and purified over monomeric avidin cartridges (Applied Biosystems) as described (2). Samples were dried and resuspended in 5% CH_3CN and 0.1% TFA.

μLC-MS/MS Analysis. The configuration for capillary μLC-MS consists of a binary HPLC pump (HP1100; Agilent Technologies, Wilmington, DE), a microautosampler (Famos; Dionex LC Packings, San Francisco, CA), a precolumn (100 μm i.d. × 2 cm), and a microcapillary column (75 μm i.d. × 10 cm). Fused-silica capillary tubing with an integrated borosilicate frit (Integrafrit; New Objective, Cambridge, MA) was used for the precolumn. For the capillary column, one end of polyimide-coated fused-silica capillary (Polymicro Technologies, Phoenix, AZ) was manually pulled to a fine point with a torch.

The precolumn was packed with 5-μm, 200-Å Magic C18AQ resin (Michrom Bio-Resources), and the capillary column was packed with 5-μm, 100-Å Magic C18AQ. After equilibrating the system in RP Buffer A (0.1% formic acid) containing 10% CH_3CN, samples were loaded onto the precolumn at a flow rate of 12 μl/min for 5 min. Next, peptides were resolved by running 80-min gradients from 10–40% RP Buffer B (100% CH_3CN) at 0.3 μl/min and analyzed by μLC-MS/MS by using an LCQ DecaXP ion-trap mass spectrometer (ThermoFinnigan) as described (4).

Peptides were identified by searching MS/MS spectra against a yeast protein database (*Saccharomyces* Genome Database) with SEQUEST (5). Data analysis was performed by using a suite of software tools including XPRESS (6), INTERACT (6), PeptideProphet (7), and ProteinProphet (8). The data sets were filtered by using ProteinProphet probability of 0.9. Several lower scoring proteins were added to the list after manual validation of peptide assignments to MS/MS spectra. All peptide identifications and quantifications were confirmed by manual inspection of the data.

Computational Modeling of Quantitative Protein Expression Ratios. To assess the statistical significance of the data and to allow more transparent interpretation of protein groups with various degrees of enrichment in the affinity-purified samples compared with control, quantitative MS data were analyzed by using the model-based unsupervised clustering approach (for review see ref. 9). The underlying assumption of the method is that the observed quantitative MS data arises from several subpopulations (clusters) of proteins with distinguished distributions. Each ICAT protein abundance ratio r in the data set is assumed to be a realization from a Gaussian mixture distribution with the probability density function

$$f(r; M_K) = \sum_{k=1}^{K} \pi_k f_k(r \mid \mu_k, \sigma_k), \qquad [1]$$

where $\pi_k$is the probability that a protein belongs to the kth cluster ($\sum_k \pi_k = 1$, $k = 1 \ldots K$), and μ_k and σ_k are the mean and the standard deviation of each normal component, and M_K

is used here to represent all unknown parameters, including the number of clusters K. The posterior probability that protein i (i = 1..N) from the data set belongs to the cluster k given its ratio r_i is then given by

$$p(r_i \mid k) = \frac{\pi_k f_k(r_i \mid \mu_k, \sigma_k)}{\sum_{j=1}^{K} \pi_j f_j(r_i \mid \mu_j, \sigma_j)} \qquad [\mathbf{2}]$$

The mixture model given by Eq. 1 is fitted by using the expectation-maximization algorithm (10) to maximize the likelihood L of observing the data given the model M_K

$$\log L(M_K) = \sum_{i=1}^{N} \log f(r_i; M_K) \qquad [\mathbf{3}]$$

The EM algorithm computes the maximum likelihood estimate of L by iterating between the two steps: estimation of the cluster probabilities $p(r_i|k)$, Eq.**2**, given the current estimate of model parameters (E-step), and updating of the model parameters given the new cluster probabilities (M-step). The iterations continue until reaching convergence (no significant changes in model parameters and cluster probabilities between the subsequent iterations), thus obtaining the maximum likelihood $L(\hat{M}_K)$.

Because the EM algorithm, in general, is not guaranteed to find the global maximum, it is important to select a good set of the initial distribution parameters. To achieve this, before EM modeling, the data are clustered by using a k-means clustering algorithm with a fixed number of cluster components K. The mean ratios and the number of proteins in each cluster as determined by k-means clustering are used to initialize the EM algorithm.

The analysis is performed with an increasing number of mixture components K, starting with K = 1. The optimal number of components is then selected by using the Bayesian information criterion (BIC) (11):

$$BIC_K = -2\log L(\hat{M}_K) + \nu_K \log(N), \quad [4]$$

where ν_K is the number of independent parameters in the model M_K.

The method described above was first applied to the data shown in Fig. 1 by using the untransformed protein ICAT ratios. The model with $K = 5$ components resulted in the smallest BIC value (indicating the optimal number of components) and was selected for further analysis and biological interpretation, as discussed in the text. For each protein, five probabilities were computed representing the likelihood of that protein belonging to each of the five clusters. The list of proteins grouped into the five clusters is shown in SI Table 2.

The same analysis was performed by using the data shown in Fig. 4, except the data were log transformed before modeling. According to the BIC criterion, only two clusters were sufficient to model the data (stimulated and not stimulated proteins).

1. Ranish JA, Yudkovsky N, Hahn S (1999) *Genes Dev* 13:49-63.

2. Ranish JA, Yi EC, Leslie DM, Purvine SO, Goodlett DR, Eng J, Aebersold R (2003) *Nat Genet* 33:349-355.

3. Ranish JA, Hahn S (1991) *J Biol Chem* 266:19320-19327.

4. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R (1999) *Nat Biotechnol* 17:994-999.

5. Eng JK, McCormack AL, Yates JR (1994) *J Am Soc Mass Spectrom* 5:976-989.

6. Han DK, Eng J, Zhou H, Aebersold R (2001) *Nat Biotechnol* 19:946-951.

7. Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) *Anal Chem* 74:5383-5392.

8. Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003) *Anal Chem* 75:4646-4658.

9. McLachlan G, Peel D (2000) *Finite Mixture Models* (Wiley, New York).

10. Dempster AP, Laird NM, Rubin DB (1977) *J R Stat Soc Ser B* 39:1-38.

11. Schwartz G (1978) *Ann Stat* 6:461-464.