# Science
## AAAS

**Genetic Dissection of Transcriptional Regulation in Budding Yeast**

Rachel B. Brem, *et al.*
*Science* **296**, 752 (2002);
DOI: 10.1126/science.1069516

served similarity in fitness effects of interacting proteins was sufficient to explain the similarity in their rates of evolution. Path analysis based on the causal model shown in Fig. 3C indicated that the correlation between the fitness effects of interacting proteins contributes only slightly to the correlation between their evolutionary rates. Thus, similarity in fitness effects is not sufficient to explain the observed similarity in the evolutionary rates of interacting proteins.

We also considered two other alternatives to the coevolutionary hypothesis. First, interacting proteins might evolve at similar rates simply because they have similar numbers of interactors, and, as shown in Fig. 1, the number of interactors influences the rate of evolution. However, we found that proteins that interact do not have similar numbers of interactors ($r_{I_1 I_2} = 0.02$, $P = 0.26$). A second possibility is that interacting proteins evolve at similar rates because they exhibit structural homology and therefore have similar distributions of constrained sites. The most likely origin of structural homology between interacting proteins is duplication of the gene that encodes a homodimeric protein, followed by evolution of one copy of the gene. This process would result in homology not only between the structures, but also between the sequences, of interacting proteins. Hence, we have ensured that none of the interactions in our data set occur between proteins that exhibit detectable sequence similarity. Thus, to account for the similarity in evolutionary rates that we observe, structural similarity would have to be independent of sequence, which would be difficult to explain evolutionarily. In sum, having considered a number of alternative hypotheses, we conclude that the coevolution of interacting proteins may be largely responsible for the observed similarity in their rates of evolution.

Beyond describing the relation between a protein's interactions and its rate of evolution, the correlations presented here could find application in the rapid assessment of functional genomic data. Much as gene expression levels have recently been used to assess protein-protein interaction data sets (23), the correlation between protein interaction and evolutionary rate may allow one to use simple genomic sequence comparisons to statistically assess the quality of large interaction data sets. More generally, correlations between protein interaction, fitness effect, and evolutionary rate may provide a means by which multiple bioinformatic data sets can be quickly cross-referenced to assess the reliability of any single method or data set.

### References and Notes

1. A. C. Wilson, S. S. Carlson, T. J. White, *Annu. Rev. Biochem.* **46**, 573 (1977).
2. E. A. Winzeler *et al.*, *Science* **285**, 901 (1999).
3. S. A. Chervitz *et al.*, *Science* **282**, 2022 (1998).
4. A. E. Hirsh, H. B. Fraser, *Nature* **411**, 1046 (2001).
5. E. Zuckerkandl, *J. Mol. Evol.* **7**, 167 (1976).
6. R. E. Dickerson, *J. Mol. Evol.* **1**, 26 (1971).
7. B. Schwikowski, P. Uetz, S. Fields, *Nature Biotechnol.* **18**, 1257 (2000).
8. P. Uetz *et al.*, *Nature* **403**, 623 (2000).
9. T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
10. Our data set of yeast protein-protein interactions was compiled from (7), the "core data" of (9), and 625 interactions identified by the two-hybrid array method [described in (8)] listed at http://depts.washington.edu/sfields/yplm/data/new2h.html.
11. Putative orthologs were defined as reciprocal best hits (4, 24) with $P < 10^{-10}$ and >80% protein sequence alignment in reciprocal searches using Washington University Basic Local Alignment Search Tool for Proteins (WU-BLASTP) (3). Thus, if a query of all 19,099 nematode open reading frames (ORFs) with yeast ORF *i* yielded the set of hits {W} with $P < 10^{-10}$ and >80% alignment, and a query of all 6217 yeast ORFs with nematode ORF *j* yielded the set of hits {Y} with $P < 10^{-10}$ and >80% alignment, then the pair (ORF *i*, ORF *j*) was considered a putative ortholog only if *i* was the member of {Y} with the lowest $P$ value and *j* was the member of {W} with the lowest $P$ value.
12. To estimate the number of substitutions per amino acid site, we numerically solved the equation $q = [\ln(1 + 2K)]/2K$, where is the proportion of identical sites between aligned sequences. This method is described in (25) and tested further in (26).
13. Unless otherwise specified, the correlation coefficient $r_{X,Y}$ represents Spearman's rank correlation between variables $X$ and $Y$.
14. Y. Ho *et al.*, *Nature* **415**, 180 (2002).
15. A. C. Gavin *et al.*, *Nature* **415**, 141 (2002).
16. Supplementary fig. 1 is available on *Science* Online at www.sciencemag.org/cgi/content/full/296/5568/750/DC1.
17. V. Smith, K. N. Chou, D. Lashkari, D. Botstein, P. O. Brown, *Science* **274**, 2069 (1996).
18. H. Jeong, S. P. Mason, A. L. Barabasi, Z. N. Oltvai, *Nature* **411**, 41 (2001).
19. The $r_{IF}$ and $r_{FK}$ correlations were also significant when genetic footprint data (27) were used instead of parallel analysis data. When essential proteins are excluded from the analysis, which is expected to strengthen the correlation between fitness effect and evolutionary rate (4), $r_{FK} = -0.18$, $P = 1.3 \times 10^{-8}$, for 944 nonessential proteins in parallel analysis.
20. D. Kaplan, in *Advanced Quantitative Techniques in the Social Sciences*, J. de Leeuw, R. Berk, Eds. (Sage Publications, Thousand Oaks, CA, 2000), pp. 13–39.
21. J. D. Gibbons, in *Sage University Papers*, M. S. Lewis-Beck, Ed. (Sage Publications, Newbury Park, CA, 1993), pp. 3–29.
22. Examples of interprotein coevolution are described in (28, 29).
23. A. Grigoriev, *Nucleic Acids Res.* **29**, 3513 (2001).
24. M. C. Rivera, R. Jain, J. E. Moore, J. A. Lake, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 6239 (1998).
25. N. V. Grishin, *J. Mol. Evol.* **41**, 675 (1995).
26. D. Feng, R. Doolittle, *J. Mol. Evol.* **44**, 361 (1997).
27. V. Smith, D. Botstein, P. O. Brown, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6479 (1995).
28. P. D. Rawson, D. A. Brazeau, R. S. Burton, *Gene* **248**, 15 (2000).
29. K. K. Koretke, A. N. Lupas, P. V. Warren, M. Rosenberg, J. R. Brown, *Mol. Biol. Evol.* **17**, 1956 (2000).
30. B. Dunn provided genetic footprinting data. P. Uetz assisted with interaction data. J. Davis, M. Eisen, B. Kerr, D. Petrov, and C. Winter provided helpful discussion and comments on the manuscript.

# Genetic Dissection of Transcriptional Regulation in Budding Yeast

**Rachel B. Brem,\* Gaël Yvert,\* Rebecca Clinton, Leonid Kruglyak†**

To begin to understand the genetic architecture of natural variation in gene expression, we carried out genetic linkage analysis of genomewide expression patterns in a cross between a laboratory strain and a wild strain of *Saccharomyces cerevisiae*. Over 1500 genes were differentially expressed between the parent strains. Expression levels of 570 genes were linked to one or more different loci, with most expression levels showing complex inheritance patterns. The loci detected by linkage fell largely into two categories: cis-acting modulators of single genes and trans-acting modulators of many genes. We found eight such trans-acting loci, each affecting the expression of a group of 7 to 94 genes of related function.

Genetic linkage analysis has traditionally focused on mapping loci that affect one or a small number of organism-level phenotypes. DNA microarray technology (*1, 2*) makes it possible to apply such analysis to global patterns of gene expression, with the transcript abundance of each of thousands of genes treated as a quantitative phenotype (*3*). Although it has recently become clear that genetic variation has a strong effect on gene expression (*4–7*), little is known about the genetic basis of natural variation in expression levels (the number and type of loci involved, the effect of each locus, and the interaction between loci).

We carried out linkage analysis of global expression levels in a cross between two strains of the budding yeast *Saccharomyces cerevisiae*. The parents were haploid derivatives of a standard laboratory strain (BY) and a wild isolate from a California vineyard

Fred Hutchinson Cancer Research Center (FHCRC), 1100 Fairview Avenue North, D4-100, Seattle, WA 98109, USA and Howard Hughes Medical Institute.

\*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: leonid@fhcrc.org

(RM) (*8*). We first measured expression of 6215 genes in six independent cultures of each parent undergoing log-phase growth in a defined medium and found profound differences in expression (*9*). A total of 1528 genes showed differential expression at $P < 0.005$, whereas only 23 are expected by chance (*10*). At $P < 0.15$, we observed 3422 differences, compared to 724 expected by chance, suggesting that nearly half (2698 out of 6215) of all the genes in the genome are differentially expressed. Of the 1528 messages that were different at $P < 0.005$, 1165 differed by <twofold, 363 by >twofold, 147 by >fourfold, and 62 by >eightfold. Expression measurements in 40 haploid segregants from a cross between the two parents showed that parental differences in expression were highly heritable; the median proportion of the observed variation that is genetic was estimated to be 84% (*11*).

Genetic markers were identified with oligonucleotide microarrays using a method previously described by Winzeler *et al.* (*12, 13*). The resulting genetic map of 3312 markers covered >99% of the genome. Analysis of four segregants from one tetrad showed 2:2 marker segregation (Fig. 1), with 73 crossovers observed across the genome, which is close to the expected number of 86 (*14*). In the analysis of the cross, we found that mating type, kanamycin resistance, and auxotrophies to lysine, uracil, and leucine were correctly linked with LOD scores >9 to regions containing the genes *MAT*, *HO*, *LYS2*, *URA3*, and *LEU2*, respectively. We also carried out linkage analysis of flocculation (agglutination of cells in liquid culture). Although neither parent is flocculent, tetrad analysis showed 1 flocculent:3 nonflocculent segregation in the cross. We found linkage to a pair of loci, one containing the *FLO8* gene (chr V) and the other containing the *FLO1* gene (chr I) (*15*). *FLO1* encodes a cell wall protein responsible for agglutination of cells, and *FLO8* encodes a transcription factor that regulates *FLO1* expression (*16*). Sequencing of the corresponding BY and RM alleles showed that BY, but not RM, carried the S288c *flo8* null mutation (*17*) and that RM, but not BY, carried a short deletion in *FLO1*.

We next tested for linkage between markers and the abundance of each message (*18*). Expression levels of 570 messages showed linkage to at least one locus at $P < 5 \times 10^{-5}$ (53 are expected by chance). Two examples of segregation of gene expression with the genotype of a linked marker are shown in Fig. 2. Two hundred and five of the linkages remained significant at $P < 2 \times 10^{-6}$ (<1 is expected by chance). Message levels for all engineered auxotrophies linked to regions containing the respective genes.

Of the 1528 messages with parental differences at $P < 0.005$, 308 showed linkage to

at least one locus at $P < 5 \times 10^{-5}$. An additional 262 messages were not called different between the parents at $P < 0.005$, but showed linkage in the cross. This observation can be explained in three ways. First, a linkage may be a false positive; as noted above, we expect 53 false-positive linkages at $P < 5 \times 10^{-5}$. Second, a true difference in expression levels may exist between the parents, and be statistically significant in a comparison of 40 segregants separated by parental genotype, but not be statistically significant in a comparison of six replicates from each parent. Third, each parent may harbor several loci with alleles of opposite effect on message levels, reducing the parental difference relative to the range of the segregants. This phenomenon, called transgressive segrega-

tion, is common (*19, 20*) and is observed for a number of messages; an example is shown in Fig. 2B.

Conversely, levels of 1220 messages differed between the parents at $P < 0.005$ but did not show linkage to any locus; as noted above, we expect only 22 false-positive differences at $P < 0.005$. Simulations showed that if each of these differences in expression were caused by a single locus, we would expect to detect linkage for 97% of them (*21*). Simulations also showed that if multiple loci were involved and, for each differentially expressed message, the locus with the strongest effect accounted for more than a third of the difference, we would expect to detect linkage for >29% of these messages (*21*). Thus, detection of linkage for just 20% of the
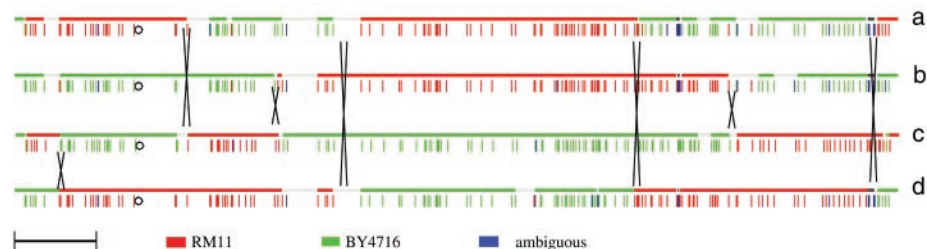


**Fig. 1.** Chromosome XII genotypes of four segregants (a, b, c, and d) isolated from a single tetrad. Each vertical bar represents one genetic marker, colored red, green, or blue when the genotype of the segregant at the marker is RM, BY, or ambiguous, respectively. The crosses indicate inferred positions of crossovers. The centromere position is shown by a black circle. The scale bar represents 100 kb.
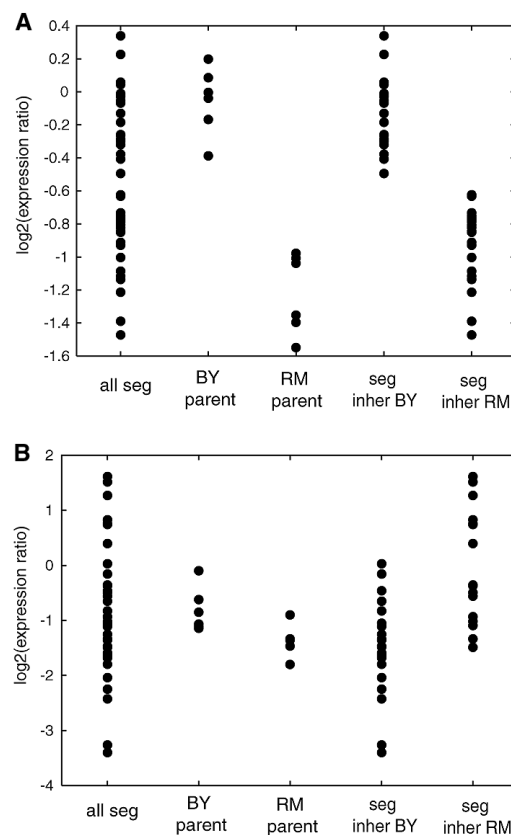


**Fig. 2.** Expression levels of parents and segregants for two genes that show linkage. In each panel, the first column shows expression levels for all 40 segregants, and the second and third columns show expression levels for six replicates of each parent. The fourth and fifth columns show expression levels for segregants that inherited the linked marker from BY and RM, respectively. (**A**) The gene is YLL007C, and the marker lies in YLL009C. (**B**) The gene is *XBP1* (YIL101C), and the marker lies in YIL060W. Note that, in this example, the effect of the locus is in the opposite direction from the difference between the parents, illustrating transgressive segregation.

differences in the real data (308 out of 1528) indicates that most messages are affected by multiple loci and that most loci account for less than a third of the total parental expression differences. We also used simulations to estimate the number of loci involved under a model in which each difference is caused by $n$ loci of equal effect (21). The results showed that we would expect to detect linkage for 82% of the differences if $n = 2$, 59% if $n = 3$, 49% if $n = 4$, and 39% if $n = 5$. The fact that in the real data set we detected linkage for only 20% of the differentially expressed messages implies that, under this model, $>5$ loci affect each message. In reality, both the number of loci contributing to expression differences and the distribution of their effects undoubtedly vary among genes, but our data are inconsistent with one or two major loci explaining the observed differences in expression for more than a small fraction of genes. Because the estimates above assume that all loci from one parent act in the same direction, existence of transgressive segregation implies even greater complexity.

To determine whether the loci found by linkage act in cis or in trans, we looked for messages whose levels were linked to markers within 10 kb of their own gene. Such "self-linkages" suggest that a polymorphism affecting a gene's expression lies within the gene itself or its regulatory region, rather than elsewhere in the genome. We found that 185 (32%) of the 570 messages that show linkage at $P < 5 \times 10^{-5}$ fell into this category. Because no message is expected to link to its own gene by chance (18), a more accurate estimate of 36% for the fraction of cis-acting loci is obtained by dividing the number of self-linkages by the expected number of true positives (517).

We next considered whether there are many trans-acting regulatory polymorphisms, each affecting one or a few messages, or a small number of such polymorphisms with effects on many messages. We divided the genome into 20-kb bins and counted the number of linkages to markers within each bin (Fig. 3). With a random distribution of linkages across the genome, no bin is expected to contain $>5$ linkages (22). In our data, 10 bins contained $>5$ linkages, ranging from 7 to 87. In two cases, two nearby bins contained $>5$ linkages and were combined into one group for future analyses. Over 40% of all linkages (231) fell into one of the eight groups (Table 1).

Groups 2 through 4 contain known members of the leucine biosynthesis, mating, and uracil biosynthesis pathways and link to *LEU2, MAT*, and *URA3*, respectively. Because Leu2 and Ura3 are biosynthetic enzymes rather than transcription factors, these linkages illustrate indirect transcriptional effects on other genes in the perturbed pathways. The other five loci represent natural polymorphisms between the parent strains with large transcriptional effects. The genes within each group appear to be functionally related based on annotated group members.

Group 5 contains 28 genes, 13 of whose products contain heme, regulate heme synthesis, or are involved in fatty acid or sterol metabolism. Five of them are known to be regulated by the heme-dependent transcriptional activator Hap1 (essential for anaerobic growth, which requires ergosterol metabolism), and the gene encoding Hap1 is in the linkage region for this group. The S288c *HAP1* allele was previously shown to carry a Ty1 insertion that reduces transcriptional activation of iso-1 cytochrome c (*CYC1*) by Hap1 (23). We amplified and sequenced the BY and RM *HAP1* alleles and found that BY, but not RM, carries this Ty1 insertion, consistent with our observation that *CYC1* and *CYC7* are underexpressed in all segregants inheriting the BY allele. These results strongly suggest that the other genes in group 5 are also regulated by Hap1. We searched for the known Hap1 binding site consensus sequence in the upstream regions of these genes and found sites containing at most one mismatch for 11 genes.

Group 1 contains 18 genes, 10 of which were found to be co-regulated in previous array experiments, with six expressed specifically in daughter cells during budding (24). The gene controlling this group may be *CST13*, which is in the region of linkage, shares the group's expression pattern, and is also expressed specifically in daughter cells during budding (24). Group 6 consists of 16 genes, all putative subtelomerically encoded helicases; these closely related genes may cross-hybridize on arrays. *SIR3*, a known transcriptional silencer active at telomeres, is within the linked region. Group 7 contains 94 genes, with 50 known to function in mitochondria of which 34 function as mitochondrial ribosomal proteins. This group shares 52 genes with a previously defined mitochondrial expression cluster (25). Several genes in the region of linkage function in mitochondria. Group 8 consists of 19 genes, 11 of which were previously shown to be expressed under acidic conditions in the presence of the transcription factors Msn2 and -4 (26), and 17 have at least one Msn2 and -4 binding site in their upstream region.

Unlike experiments that measure the correlations among transcript levels under different conditions, our approach allows causal connections to be made between modulator loci and the genes whose expression they directly or indirectly affect. In addition, studying naturally occurring alleles in the

**Fig. 3.** The number of linkages plotted against genome location. The genome is divided into 611 bins of 20 kb each, shown in chromosomal order from the start of chromosome I to the end of chromosome XVI. The dashed line is drawn at 5 linkages; no bin is expected to contain 5 linkages by chance (22). The regions with an unusually large number of linkages are marked 1 through 8 and correspond to the groups in Table 1.
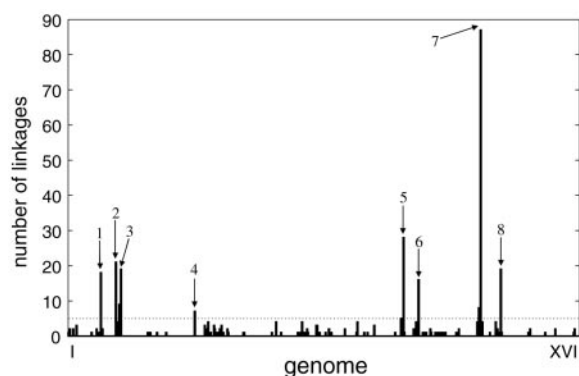
**Table 1.** Groups of messages linking to loci with widespread transcriptional effects. The location of the center of the linked bin is shown as chromosome:base pair. Lists of genes in each group are available as supplementary information (32).

| Group | Number of messages | Common function | Linkage bin | Putative regulator |
|---|---|---|---|---|
| 1 | 18 | Budding, daughter cell separation | II:550000 | CST13 |
| 2 | 21 | Leucine biosynthesis | III:90000 | LEU2 |
| 3 | 28 | Mating | III:190000 | MAT |
| 4 | 7 | Uracil biosynthesis | V:110000 | URA3 |
| 5 | 28 | Heme, fatty acid metabolism | XII:670000 | HAP1 |
| 6 | 16 | Subtelomerically encoded helicases | XII:1030000 | SIR3 |
| 7 | 94 | Mitochondrial | XIV:490000 | Unknown |
| 8 | 19 | Msn2/4-dependent induction in acid | XV:170000 | Unknown |

context of segregating variation allows the discovery of subtle effects obscured in strains with engineered knockouts.

We have found that regulatory genetic variation is characterized by a high rate of cis-acting alleles and a small number of trans-acting alleles with widespread transcriptional effects. Finally, genetic variation in physiological and behavioral quantitative phenotypes is known to be highly complex. Our results indicate that even in a single-cell organism grown in a controlled environment, variation in gene expression typically also has a polygenic basis.

### References and Notes

1. J. L. DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997).
2. L. Wodicka, H. Dong, M. Mittmann, M. H. Ho, D. J. Lockhart, *Nature Biotechnol.* **15**, 1359 (1997).
3. R. C. Jansen, J. P. Nap, *Trends Genet.* **17**, 388 (2001).
4. R. Sandberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 11038 (2000).
5. M. Primig *et al.*, *Nature Genet.* **26**, 415 (2000).
6. W. Jin *et al.*, *Nature Genet.* **29**, 389 (2001).
7. D. Cavalieri, J. P. Townsend, D. L. Hartl, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 12369 (2000).
8. Parent strains were BY4716, an S288c derivative (*MATa lys2Δ0*) (*27*), and RM11-1a (*MATa leu2Δ0 ura3Δ0 HO:kan*), a haploid derived from Bb32(3), a natural isolate collected by Robert Mortimer, as in (*28*). Auxotrophic deletions were used to select diploids during mating and as positive controls for linkage. Both strains were gifts from Lee Hartwell, FHCRC, Seattle, WA.
9. Cultures were grown to $1 \times 10^7$ cells/mL in synthetic C medium at 30°C and shaken at 175 rpm in a shaker bath. RNA was isolated with hot acid phenol and chloroform (*29*), followed by the RNA cleanup protocol for RNeasy columns (Qiagen, Valencia, CA). Samples were labeled and hybridized to cDNA microarrays, containing 6215 open reading frames, as described (*30*). Images were analyzed with GenePix software. Each two-color experiment involved one sample and one reference, with the same BY4716 RNA reference being used for all experiments. Two hybridizations were carried out for each sample, one with the sample labeled with Cy3 and the reference with Cy5, and one with the fluors reversed; for each gene, the two log ratios were averaged.
10. Differences in the expression between parent strains were assessed by comparing the six RM values to the six BY values with the Wilcoxon-Mann-Whitney test (*31*), and nominal *P* values are reported. To estimate the number of differences expected by chance at a given *P*-value threshold, we randomly assigned 6 of the 12 measurements to one group and 6 to the other, without regard for strain identity. We then compared the two groups, using the same statistical test we applied in comparing the two strains, and recorded how many genes had *P* values below the threshold. The randomization procedure was repeated 100 times, and the results were averaged.
11. Heritability was estimated as (segregant variance – parent variance)/segregant variance.
12. E. A. Winzeler *et al.*, *Science* **281**, 1194 (1998).
13. *GeneChip Yeast Genome S98* microarrays were purchased from Affymetrix (Santa Clara, CA). Genomic DNA was isolated using Qiagen columns, digested with DnaseI and end-labeled as described (*12*), and hybridized following Affymetrix instructions. We developed algorithms to select oligos as reliable genetic markers and to genotype them in the segregants. These are available as supplementary information (*32*).
14. J. M. Cherry *et al.*, *Nature* **387**, 67 (1997).
15. For each pair of markers with all flocculent segregants sharing one parental genotype at one marker and the opposite parental genotype at the other, a Chi-square test of independence was performed on a 2 × 4 contingency table of all 40 segregants according to their flocculation phenotype (+ or −) and genotype (BY/BY, BY/RM, RM/BY, or RM/RM) at the two markers. A single pair showed linkage with $P < 10^{-7}$.
16. O. Kobayashi, H. Yoshimoto, H. Sone, *Curr. Genet.* **36**, 256 (1999).
17. S. J. Kron, *Trends Microbiol.* **5**, 450 (1997).
18. We tested for linkage between a marker and a message by partitioning the segregants into two groups according to marker genotype and comparing the message levels between the groups with the Wilcoxon-Mann-Whitney test (*31*). The number of linkages expected by chance was estimated by empirical permutation tests (*33*).
19. M. C. deVicente, S. D. Tanksley, *Genetics* **134**, 585 (1993).
20. L. H. Rieseberg, M. A. Archer, R. K. Wayne, *Heredity* **83**, 363 (1999).
21. For each message differentially expressed between the parents at $P < 0.005$, we computed the expression difference, $x$, in units of standard deviation. We simulated a locus explaining a fraction $f$ of this difference by randomly assigning parent genotypes at the locus for 40 segregants and drawing the expression level of each segregant from one of two Gaussian distributions of unit variance, one with mean 0 and the other with mean $fx$ conditional on the genotype. We recorded the proportion $q$ of messages for which linkage was detected at $P < 5 \times 10^{-5}$ in the simulated data, averaged over 100 simulations. In the text, $q = 97\%$ for $f = 1$, and $q = 29\%$ for $f = 1/3$. For $n$ loci, each explaining a fraction $f = 1/n$ of the difference $x$, we computed the rate of detecting at least one locus as $1 - (1 - q)^n$.
22. The genome was divided into 611 bins of 20 kb each (the bins at ends of chromosomes were smaller). Excluding self-linkages, we found 385 linkages at $P < 5 \times 10^{-5}$. If these were randomly distributed across the genome, the number of linkages in any one bin would follow a Poisson distribution with a mean of 0.63. The probability that the bin with the most linkages would contain >5 is <0.03.
23. M. Gaisne, A. M. Becam, J. Verdiere, C. J. Herbert, *Curr. Genet.* **36**, 195 (1999).
24. A. Colman-Lerner, T. E. Chin, R. Brent, *Cell* **107**, 739 (2001).
25. T. R. Hughes *et al.*, *Cell* **102**, 109 (2000).
26. H. C. Causton *et al.*, *Mol. Biol. Cell* **12**, 323 (2001).
27. C. B. Brachmann *et al.*, *Yeast* **14**, 115 (1998).
28. R. K. Mortimer, P. Romano, G. Suzzi, M. Polsinelli, *Yeast* **10**, 1543 (1994).
29. The protocol for RNA isolation from yeast, developed by P. O. Brown's lab at Stanford University, can be found online at http://www.metazoa.com/viewprotocol.jsp?ProtocolId=2058&VersionNo=1.
30. T. G. Fazzio *et al.*, *Mol. Cell. Biol.* **21**, 6450 (2001).
31. B. W. Lindgren, *Statistical Theory* (Macmillan, New York, ed. 2, 1968).
32. Supplemental material is available at www.sciencemag.org/cgi/content/full/1069516/DC1.
33. G. A. Churchill, R. W. Doerge, *Genetics* **138**, 963 (1994).
34. We thank J. Delrow, M. Eberle, I. Furman, B. Garvik, J. Hartman, L. Iwasaki, C. Neal, J. Pincus, E. Ramos, C. Saunders, and P. Tata for technical assistance; and H. Coller, A. Colman-Lerner, E. Giniger, D. Gottschling, L. Hartwell, K. Markianos, J. Olson, and G. Smith for helpful discussions and comments on the manuscript. Supported by the Howard Hughes Medical Institute (of which L.K. is an associate investigator) and NIH postdoctoral fellowship GM64268 (R.B.). L.K. is a James S. McDonnell Centennial Fellow.

# Transcription Control by E1A and MAP Kinase Pathway via Sur2 Mediator Subunit

**Jennitte L. Stevens,[1] Greg T. Cantin,[1] Gang Wang,[1] Andrej Shevchenko,[2] Anna Shevchenko,[2] Arnold J. Berk[1,3]***

Sur2 is a metazoan Mediator subunit that interacts with the adenovirus E1A protein and functions in a mitogen-activated protein kinase pathway required for vulva development in *Caenorhabditis elegans*. We generated *sur2*$^{-/-}$ embryonic stem cells to analyze its function as a mammalian Mediator component. Our results show that Sur2 forms a subcomplex of the Mediator with two other subunits, TRAP/Med100 and 95. Knock-out of Sur2 prevents activation by E1A-CR3 and the mitogen-activated protein kinase–regulated ETS transcription factor Elk-1, but not by multiple other transcription factors. These results imply that specific activation domains stimulate transcription by binding to distinct Mediator subunits. Activation by E1A and Elk-1 requires recruitment of Mediator to a promoter by binding to its Sur2 subunit.

Regulation of transcription by RNA polymerase II (Pol II) is controlled by specific combinations of sequence-specific, DNA binding regulatory transcription factors (TFs) bound to a gene's promoter and enhancer regions. These TFs interact with several multiprotein complexes that remodel the chromatin context of the promoter, integrate signals from multiple TFs to control the frequency of transcription initiation, and regulate the efficiency of transcription elongation (1). One such

complex is the Mediator, an ~2-megadalton complex of 20 to 30 subunits that is believed to function as a molecular bridge by simultaneously interacting with both DNA-bound TFs and Pol II (1–4). The human Mediator subunit Sur2 was identified through its interaction with adenovirus E1A conserved region 3 (E1A-CR3), a potent activation domain that regulates early adenovirus genes (5). In *C. elegans*, mutations in *sur2* suppress an activated-*ras* multivulva phenotype and produce