was supplemented by manual inspection of the protein-level matches in the complete GenPept database to include genes with lower similarities if they occurred within co-linear regions of the genomes. The genome sequence was compared with that of MG1655 by the maximal exact match (MEM) alignment utility, (B.M., manuscript in preparation) an adaptation of MUMmer[30]. This program was based on suffix arrays rather than suffix trees, and exact rather than unique matches, coupled with a custom anchored-alignment algorithm that extends sequence homology into the regions separating contiguous co-linear exact matches. Inferences on biases in polymorphism patterns are based on χ^2 goodness-of-fit tests of a nested sequence of multinomial log–linear models. These predict symmetric elevated levels of A↔G, T↔C and G↔T polymorphisms, above a quasi-independent baseline generated from marginal frequencies in the co-occurrence matrix of synonymous third-codon differences. Further information may be found at our Website http://www.genome.wisc.edu/, including a Genome Browser displaying a comparative map of EDL933 and K-12.

Received 24 July; accepted 6 November 2000.

1. Riley, L. W. *et al.* Hemorrhagic colitis associated with a rare *Escherichia coli* serotype. *N. Engl. J. Med.* **308,** 681–685 (1983).
2. Karmali, M. A., Steele, B. T., Petric, M. & Lim, C. Sporadic cases of haemolytic–uraemic syndrome associated with faecal cytotoxin and cytotoxin-producing *Escherichia coli* in stools. *Lancet* **ii,** 619–620 (1983).
3. Mead, P. S. *et al.* Food-related illness and death in the United States. *Emerg. Infect. Dis.* **5,** 607–625 (1999).
4. Su, C. & Brandt, L. J. *Escherichia coli* O157:H7 infection in humans. *Ann. Intern. Med.* **123,** 698–714 (1995).
5. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277,** 1453–1474 (1997).
6. Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406,** 64–67 (2000).
7. Blattner, F. R. *et al.* Comparative genome sequencing of *E. coli* O157:H7 versus *E. coli* K 12. *Microb. Compar. Genom.* **2,** 174 (1997).
8. Hacker, J. *et al.* Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb. Pathog.* **8,** 213-225 (1990).
9. Perna, N. T. *et al.* Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66,** 3810–3817 (1998).
10. Plunkett, G. III, Rose, D. J., Durfee, T. J. & Blattner, F. R. Sequence of Shiga toxin 2 phage 933W from *Escherichia coli* O157:H7: Shiga toxin as a phage late-gene product. *J. Bacteriol.* **181,** 1767–1778 (1999).
11. Campbell, A. & Botstein, D. in *Lambda II* (eds Hendrix, R. W., Roberts, J. W., Stahl, F. W. & Weisberg, R. A.) 365–380 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983).
12. O'Brien, A. D. *et al.* Shiga-like toxin-converting phages from *Escherichia coli* strains that cause hemorrhagic colitis or infantile diarrhea. *Science* **226,** 694–696 (1984).
13. Walterspiel, J. N., Ashkenazi, S., Morrow, A. L. & Cleary, T. G. Effect of subinhibitory concentrations of antibiotics on extracellular Shiga-like toxin I. *Infection* **20,** 25–29 (1992).
14. Neely, M. N. & Friedman, D. I. Functional and genetic analysis of regulatory regions of coliphage H-19B: location of shiga-like toxin and lysis genes suggest a role for phage functions in toxin release. *Mol. Microbiol.* **28,** 1255–1267 (1998).
15. Grif, K., Dierich, M. P., Karch, H. & Allerberger, F. Strain-specific differences in the amount of Shiga toxin released from enterohemorrhagic *Escherichia coli* O157 following exposure to subinhibitory concentrations of antimicrobial agents. *Eur. J. Clin. Microbiol. Infect. Dis.* **17,** 761–766 (1998).
16. McDonough, M. A. & Butterton, J. R. Spontaneous tandem amplification and deletion of the shiga toxin operon in *Shigella dysenteriae* 1. *Mol. Microbiol.* **34,** 1058–1069 (1999).
17. Lawrence, J. G. & Ochman, H. Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* **44,** 383–397 (1997).
18. Francino, M. P., Chao, L., Riley, M. A. & Ochman, H. Asymmetries generated by transcription-coupled repair in enterobacterial genes. *Science* **272,** 107–109 (1996).
19. Blaisdell, J. O., Hatahet, Z. & Wallace, S. S. A novel role for *Escherichia coli* endonuclease VIII in prevention of spontaneous G→T transversions. *J. Bacteriol.* **181,** 6396–6402 (1999).
20. Le Page, F. *et al.* Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101,** 159–171 (2000).
21. Boyd, E. F., Li, J., Ochman, H. & Selander, R. K. Comparative genetics of the *inv–spa* invasion gene complex of *Salmonella enterica*. *J. Bacteriol.* **179,** 1985–1991 (1997).
22. Sharp, P. M., Kelleher, J. E., Daniel, A. S., Cowan, G. M. & Murray, N. E. Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc. Natl Acad. Sci. USA* **89,** 9836–9840 (1992).
23. Alm, R. A. & Trust, T. J. Analysis of the genetic diversity of *Helicobacter pylori*: the tale of two genomes. *J. Mol. Med.* **77,** 834–846 (1999).
24. Read, T. D. *et al.* Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28,** 1397–1406 (2000).
25. Ohnishi, M. *et al.* Chromosome of the enterohemorrhagic *Escherichia coli* O157:H7: comparative analysis with K-12 MG1655 revealed the acquisition of a large amount of foreign DNAs. *DNA Res.* **6,** 361–368 (1999).
26. Mahillon, J. *et al.* Subdivision of *Escherichia coli* K-12 genome for sequencing: manipulation and DNA sequence of transposable elements introducing unique restriction sites. *Gene* **223,** 47–54 (1998).
27. Lin, J. *et al.* Whole-genome shotgun optical mapping of *Deinococcus radiodurans*. *Science* **285,** 1558–1562 (1999).
28. Lukashin, A. V. & Borodovsky, M. GeneMark.hmm: new solutions for gene finding. *Nucleic Acids Res.* **26,** 1107–1115 (1998).
29. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215,** 403–410 (1990).
30. Delcher, A. L. *et al.* Alignment of whole genomes. *Nucleic Acids Res.* **27,** 2369–2376 (1999).

Acknowledgements

We thank T. Forsythe, M. Goeden, H. Kijenski, B. Leininger, J. McHugh, B. Peterson, G. Peyrot, D. Sands, P. Soni, E. Travanty and other members of the University of Wisconsin genomics team for their expert technical assistance. This work was funded by grants from the NIH (NIAID and NCHGR), the University of Wisconsin Graduate School and the RMHC to F.R.B., the NIH (NCHGR, NIAID) to D.C.S., HHMI/OTKA to G.P., an Alfred P. Sloan/DOE Fellowship to B.M., a CDC/APHL Fellowship to P.S.E., and an Alfred P. Sloan/NSF Fellowship to N.T.P. Sixteen University of Wisconsin undergraduates participated in this work and particular thanks are due to A. Byrnes for web-site development to complement this project, and to A. Darling for programming.

Correspondence and requests for materials should be addressed to N.T.P. (e-mail: perna@ahabs.wisc.edu). The GenBank accession number for the annotated sequence is AE00517H.

Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF

Vishwanath R. Iyer*†‡, Christine E. Horak§‡, Charles S. Scafe‖¶, David Botstein‖, Michael Snyder§ & Patrick O. Brown*#

** Department of Biochemistry and # Howard Hughes Medical Institute, Stanford University Medical Center, Stanford, California 94305, USA*
‖ Department of Genetics, Stanford University Medical Center, Stanford, California 94305, USA
§ Department of Molecular, Cellular, and Developmental Biology, Yale University, New Haven, Connecticut 06520, USA
† These authors contributed equally to this work

Proteins interact with genomic DNA to bring the genome to life; and these interactions also define many functional features of the genome. SBF and MBF are sequence-specific transcription factors that activate gene expression during the G1/S transition of the cell cycle in yeast[1,2]. SBF is a heterodimer of Swi4 and Swi6, and MBF is a heterodimer of Mbp1 and Swi6 (refs 1, 3). The related Swi4 and Mbp1 proteins are the DNA-binding components of the respective factors, and Swi6 may have a regulatory function[4,5]. A small number of SBF and MBF target genes have been identified[3,6–10]. Here we define the genomic binding sites of the SBF and MBF transcription factors *in vivo*, by using DNA microarrays. In addition to the previously characterized targets, we have identified about 200 new putative targets. Our results support the hypothesis that SBF activated genes are predominantly involved in budding, and in membrane and cell-wall biosynthesis, whereas DNA replication and repair are the dominant functions among MBF activated genes[6,11]. The functional specialization of these factors may provide a mechanism for independent regulation of distinct molecular processes that normally occur in synchrony during the mitotic cell cycle.

To identify the targets of SBF and MBF, we combined chromatin immunoprecipitation and microarray hybridization (Fig. 1). Proteins were crosslinked with formaldehyde to their target sites *in vivo*. DNA that was specifically crosslinked to either of the transcription factors was purified by immunoprecipitation using an antibody against either the native protein or an epitope tag that was fused to the protein. Polymerase chain reaction (PCR) analysis of immunoprecipitated DNA confirmed the specific association of Swi4, Swi6 and Mbp1 with several known target promoters, and other target promoters that are identified here (see Supplementary Information). After reversal of the crosslinks, immunoprecipitated DNA was amplified and fluorescently labelled with the Cy5 fluoro-

‡ Present address: Institute of Molecular and Cellular Biology, University of Texas at Austin, Austin, Texas 78712, USA.
¶ Present address: Applied Biosystems, Foster City, California 94404, USA.

phore. We also fluorescently labelled a separate control DNA sample with the Cy3 fluorophore. Genomic target loci were identified by comparative hybridization of the immunoprecipitated and control DNA probes to a DNA microarray. The ratio of the Cy5 to Cy3 fluorescence intensities measured at each DNA element in the microarray provided a measure of the extent of binding of the transcription factor to the corresponding genomic locus. As the transcription factors were expected to interact primarily with promoters, we constructed DNA microarrays that represented all the intergenic intervals, as well as all the predicted coding sequences in the yeast genome. We plotted fluorescence ratios on a representation of the yeast genome to generate a map of the distribution of these proteins on the entire genome (see Supplementary Information).

To identify the most probable targets of SBF and MBF, we ranked genomic loci according to their fluorescence ratios. We first determined the percentile rank for each array element in each immunoprecipitation experiment, and then the median percentile rank for each element across a set of experiments. The distribution of the median percentile ranks for the 11 Swi4 immunoprecipitation experiments was bimodal (Fig. 2a). Loci with median percentile ranks above 93.8 formed a group that was distinct from the bulk of the distribution. 163 loci (representing sequences upstream of 183 genes) had median percentile ranks above this threshold. The average ranks of these 163 loci were significantly higher than would be expected if the rankings in the individual immunoprecipitations were uncorrelated ($P = 8.0 \times 10^{-151}$). The genes associated with these loci therefore represent candidate SBF targets (Fig. 3). When a putative binding site fell between a pair of divergently transcribed genes, we considered both genes as potential transcriptional targets. We use 'target' to mean putative transcriptional target genes that are inferred from our immunoaffinity-microarray results.

To identify MBF targets, we considered loci with high ranks in immunoprecipitation experiments targeting Mbp1 and Swi6. The median percentile ranks from the Mbp1 plus Swi6 immunoprecipitation experiments did not show a bimodal distribution (Fig. 2b); however, loci with median percentile ranks above 94 were highly enriched for genes with the G1/S expression profile, characteristic of known MBF targets (Fig. 2c). On the basis of this analysis, and on the distribution of known MBF target genes in the ranked list, we chose a median percentile rank of 94 as a threshold for considering loci to be MBF targets. Eighty-seven loci (upstream of 98 genes) ranked above this threshold (Fig. 3). Forty-three loci ranked above the selection thresholds for both SBF and MBF, and thus seemed to represent targets of both factors (see Methods).

Three lines of evidence support the conclusion that this survey has identified *in vivo* targets of SBF and MBF. First, the putative *in vivo* binding sites we identified are highly enriched for sequences matching the defined consensus binding sites. Second, most of the genes downstream of the putative binding sites are periodically expressed during the cell cycle with a peak in G1/S, a pattern typical of the previously identified targets of these factors. Third, putative target genes with known functions are highly enriched for functions related to DNA replication, budding and the cell cycle. We have, however, probably misidentified some genes as putative targets of these factors, and failed to identify some targets.

The putative SBF and MBF targets that we identified were highly enriched for sequences matching the consensus binding sites for SBF and MBF[12]. Roughly 49% of the unique intergenic fragments that were bound by Swi4 contained a motif matching the defined SBF consensus binding site (CRCGAAA), whereas only about 10% of all intergenic elements contain this site ($P = 1.0 \times 10^{-30}$). Fifty-five per cent of the unique MBF intergenic targets contained the MBF consensus binding site (ACGCGN), whereas roughly twenty per cent of all intergenic fragments have such a site ($P = 1.0 \times 10^{-9}$). Compared with the genome overall, there was also a greater tendency for the SBF and MBF target promoters to have multiple binding sites for their respective factors. The consensus site for SBF is found with roughly equal frequency in open reading frames (ORFs) and intergenic regions in the yeast genome. However, when we used DNA microarrays that contained nearly all of the predicted ORFs, and the intergenic segments of the yeast genome, to map Swi4 targets, we observed that in virtually every case in which an ORF had a high percentile rank, its upstream intergenic fragment had a similarly high rank (data not shown). Thus, despite the frequent occurrence of consensus SBF binding sequences in coding sequences, SBF binds selectively to promoters and not to coding sequences.

We used the pattern-finding programs MEME[13] and Consensus[14] to identify sequence motifs that were enriched in the high ranking (percentile rank > 98 for SBF and > 97 for MBF) target intergenic fragments. Both of these programs identified the consensus binding sites for SBF and MBF (CGCGAAAA and ACGCGN, respectively), based solely on their overabundance in the respective target loci. This suggests that these sequence motifs are indeed the primary determinants of binding of the respective factors to promoters (data not shown; and Supplementary Information). However, not every promoter bound by SBF or MBF *in vivo* contained a recognizable consensus binding site. Moreover, most of the coding sequences and many of the promoters that contain the consensus sequences show no evidence of binding to these factors *in vivo*.

SBF binding to DNA is cell-cycle regulated, occurring maximally at G1 (refs 15, 16). We saw only a modest increase in fluorescence ratios measured for SBF target loci when cells were treated with α-factor, compared with the unsynchronized or benomyl-treated cultures, perhaps because the fraction of G1 cells in unsynchronized

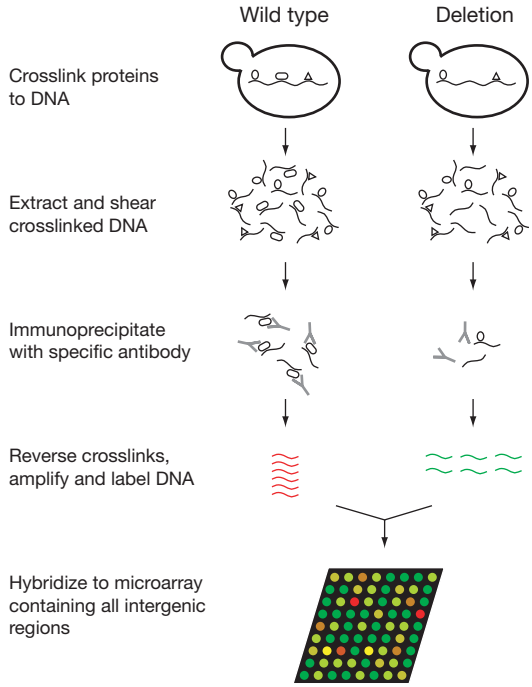

Figure 1 Strategy for analysing genome-wide protein–DNA interactions. The reference probe can either consist of DNA generated in parallel from a strain bearing a deletion of the gene encoding the protein of interest (as depicted), or of unfractionated genomic DNA amplified and labelled in the same manner. Alternatively, an epitope-tagged version of the protein of interest can be immunoprecipitated with an antibody directed against the epitope. The DNA microarray includes all of the intergenic regions or promoters from the genome. The Cy5/Cy3 fluorescence ratio for each locus reflects its enrichment by immunoprecipitation (IP) and therefore, in general, its relative occupancy by the cognate protein.

cultures or benomyl-treated cultures provided sufficient SBF-bound target sites for enrichment by immunoprecipitation.

In a comparison of the global gene expression patterns in asynchronous cultures of *swi4*Δ and wild-type cells, we found no significant differences in expression of most of the putative SBF targets, including such characterized targets as *CLN1*, *PCL1*, *PCL2* and *HO*, a result consistent with previous findings[11] (See Supplementary Information). Presumably, as the G1/S induction by SBF is superimposed on a basal level of SBF-independent transcription[17], the effect would be more readily detected in synchronized cells. However, the small set of genes whose transcript levels were significantly reduced in the *swi4*Δ cells included several SBF targets identified by chromatin immunoprecipitation, such as *CWP1*, *CWP2*, *CIS3*, *SVS1* and *SRL1*.

We examined the expression patterns of SBF and MBF target genes during the mitotic cell cycle[12], and sporulation[18], using published data (Fig. 4a). Whereas only 13% of all yeast genes are transcriptionally cell-cycle regulated[12], 66% of the SBF targets and 68% of the MBF targets, as defined by the immunoprecipitation results, were cell-cycle regulated ($P = 1.0 \times 10^{-44}$ and 1.0×10^{-22}, for SBF and MBF, respectively). Genes with maximal transcript levels during G1 and S phase showed the greatest enrichment, but there was also a less significant enrichment for genes whose transcripts peaked during other stages of the cell cycle. The putative target genes that were not cell-cycle regulated had slightly lower immunoprecipitation-selection percentile ranks and a lower frequency of consensus MBF or SBF sites than the cell-cycle regulated targets, suggesting that at least some of these genes represent false positives in our immunoprecipitation assay. This bias was more pronounced for the putative MBF targets. There was, however, a significant enrichment for the presence of SBF consensus sites in the promoters of SBF target genes that were not cell-cycle regulated. In the set of MBF target genes that were not cell-cycle regulated, there was a significant enrichment for genes with roles in DNA repair and replication, as described below.

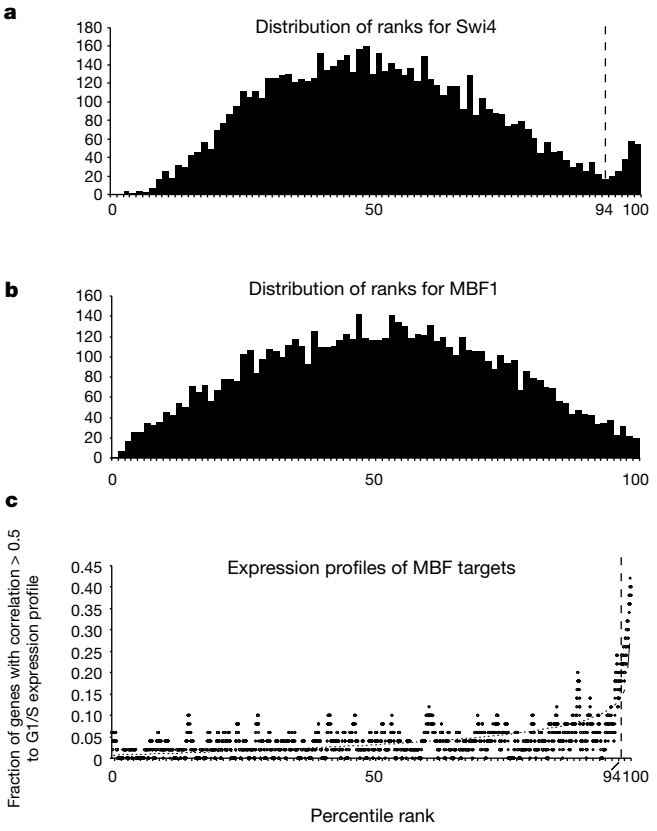

Figure 2 Defining criteria for identifying SBF and MBF targets. **a**, Distribution of median percentile ranks of Cy5/Cy3 fluorescence ratios from 11 independent microarray hybridizations analysing Swi4 binding. The threshold for defining SBF targets is indicated by a dotted line. **b**, Distribution of median ranks of Cy5/Cy3 fluorescence ratios from four independent microarray hybridizations analysing Swi6 and Mbp1 binding. **c**, Fraction of genes in a sliding 50-gene window whose expression profile has a correlation coefficient of ≥ 0.5 with that of a canonical G1/S regulated gene, as a function of the median percentile ranks of Cy5/Cy3 fluorescence ratios from the MBF hybridizations. The threshold for defining MBF targets is indicated by a dotted line.

The known functions of many of the putative targets that we have identified raised interesting questions about the regulation of the cell cycle (Fig. 3). The G1 cyclins *CLN1* and *CLN2*, as well as *PCL1* and *PCL2*, were among the SBF targets that we expected to find. Unexpectedly, the promoters of the G2/M-specific B-type cyclins *CLB1* and *CLB2* also seemed to be bound by Swi4 in our microarray analysis, and in a confirmatory PCR analysis (data not shown). The significance of this result is unclear; SBF may be involved in repressing the transcription of these genes during G1/S or, as the mitotic B-type cyclins are required to shut off SBF-induced transcription during G2 and M phase[15], this observation may point to a negative feedback loop in SBF regulation.

Why are two different transcription factors used to mediate nearly identical transcriptional programmes during the cell-division cycle in yeast? A possible answer is suggested by differences in the functions of the genes that they regulate. Many of the known transcriptional targets of SBF have roles in cell-wall biogenesis and budding[11]. Indeed, a remarkably large fraction of the putative SBF targets identified in this study encode proteins that are involved in various aspects of bud formation, and synthesis of membrane and cell-wall components, based on functional definitions in the MIPS (http://www.mips.biochem.mpg.de/proj/yeast/) and YPD (http://www.proteome.com) databases. Twenty-seven per cent of the known genes in the SBF target set fell in these functional categories, which is a significant enrichment ($P = 1.0 \times 10^{-5}$) over their representation in the genome as a whole (11%). Twenty-five per cent of the putative MBF target genes have known roles in DNA replication, recombination and repair (on the basis of MIPS and YPD definitions), which is a significant enrichment ($P = 1.0 \times 10^{-5}$) over their representation in the genome overall (6%)[2]. There was no enrichment of DNA metabolism-related genes among SBF targets, nor were cell-wall morphogenesis genes over-represented in the MBF targets.

The functional specialization of the putative targets of SBF and MBF was accompanied by differences in their overall expression programmes that suggest a possible regulatory logic to this division (Fig. 4b). The molecular programmes required for cellular growth and for the characteristic events at successive stages of the cell cycle occur with stereotyped synchronicity during simple mitotic proliferation. In different cell types, or during different physiological or developmental processes, however, the same molecular programmes may need to be orchestrated in different ways. In mitotically proliferating yeast, budding and DNA replication occur simultaneously; however, during the meiotic S phase, DNA replication occurs without budding. The converse is seen during mating (shmoo formation) and during pseudohyphal or invasive growth, when synthesis of membrane and cell-wall components occurs without concomitant DNA replication. Yeast cells must therefore have a mechanism to regulate genes devoted to DNA replication separately from those with major roles related to budding. As noted previously[19], we observed that many of the putative MBF targets were genes that are strongly induced during sporulation (Fig. 4; data from ref. 18). The product of the B-type cyclin *CLB6*, a prominent MBF target, is also required for initiation of S phase during meiosis[20], and several other MBF targets involved in DNA replication and repair have crucial roles in meiosis.

Column	Reference	IP strain	Antibody	Conditions
1	Genomic DNA	*SWI4*	Anti-HA	Unsynchronized
2, 3	Genomic DNA	*SWI4*	Anti-HA	α-factor
4	*SWI4*	*SWI4-3XHA*	Anti-HA	Unsynchronized
5	Genomic DNA	*SWI4-3XHA*	Anti-HA	Unsynchronized
6, 8	Genomic DNA	*SWI4-3XHA*	Anti-HA	α-factor
7	Genomic DNA	*SWI4-3XHA*	Anti-HA	Benomyl
9, 10	*swi4*Δ	*SWI4*	Anti-Swi4	Unsynchronized
11, 13	*swi4*Δ	*SWI4*	Anti-Swi4	α-factor
12, 14	*swi4*Δ	*SWI4*	Anti-Swi4	Benomyl
15	*swi6*Δ	*SWI6*	Anti-Swi6	Unsynchronized
16	Genomic DNA	*SWI6*	Anti-Swi6	Unsynchronized
17, 18	MBP1	*MBP1-3XMYC*	Anti-Myc	Unsynchronized

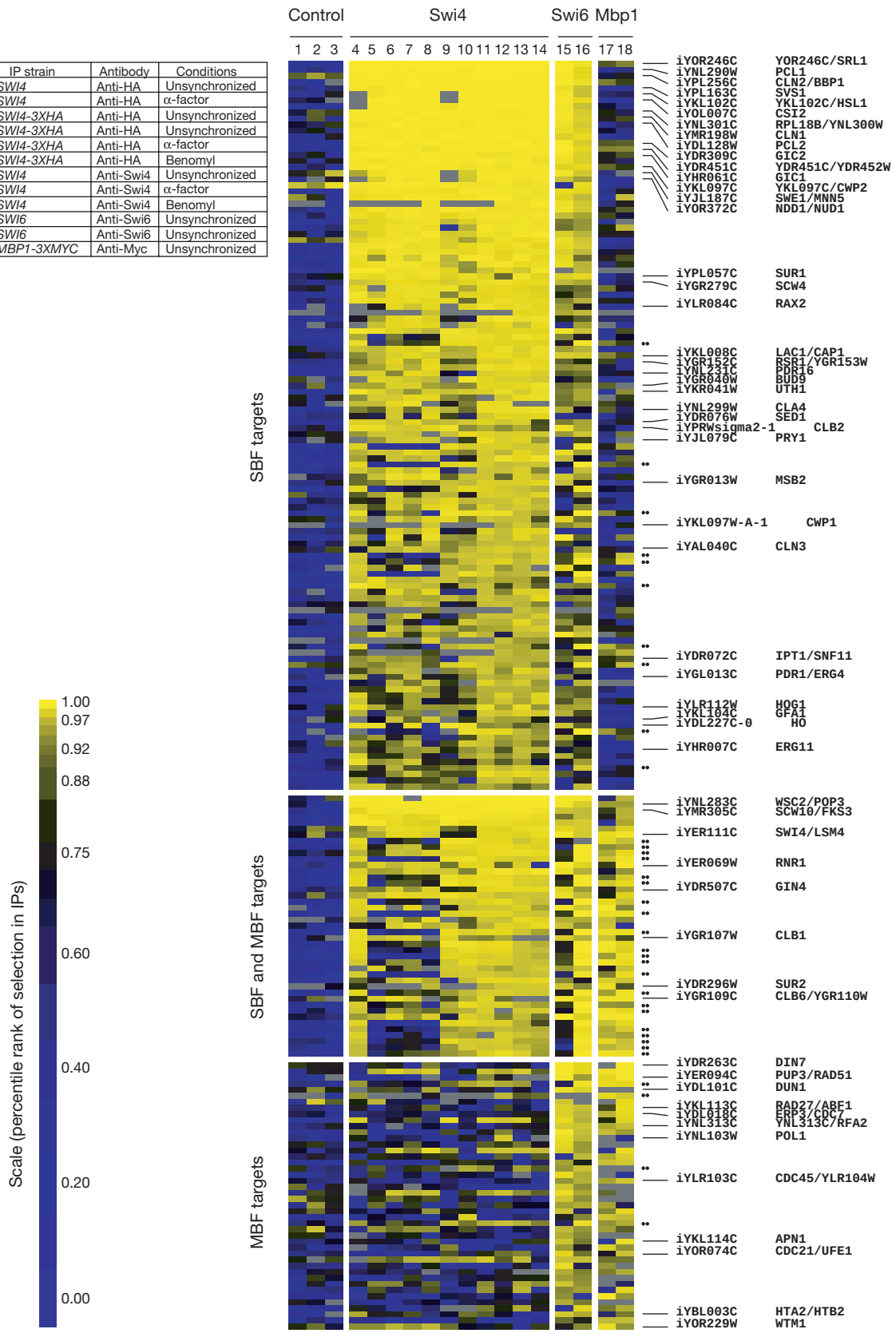

Figure 3 Genomic targets of SBF and MBF. Percentile ranks of intergenic fragments that meet selection thresholds are inicated (blue–yellow colour scale). Loci with ~70% overall nucleotide sequence identity to another yeast locus (potentially crosshybridizing) are indicated (closed circles). The combination of Cy3 and Cy5 labelled probes, the antibody used for IP (if used) and the culture conditions for each experiment are summarized (left panel). Experiments 9, 10, 17 and 18 involved independent crosslinking and IPs. DNA microarrays that included all yeast ORFs and other features, in addition to the intergenic fragments, were used for experiments 3, 8, 13 and 14.

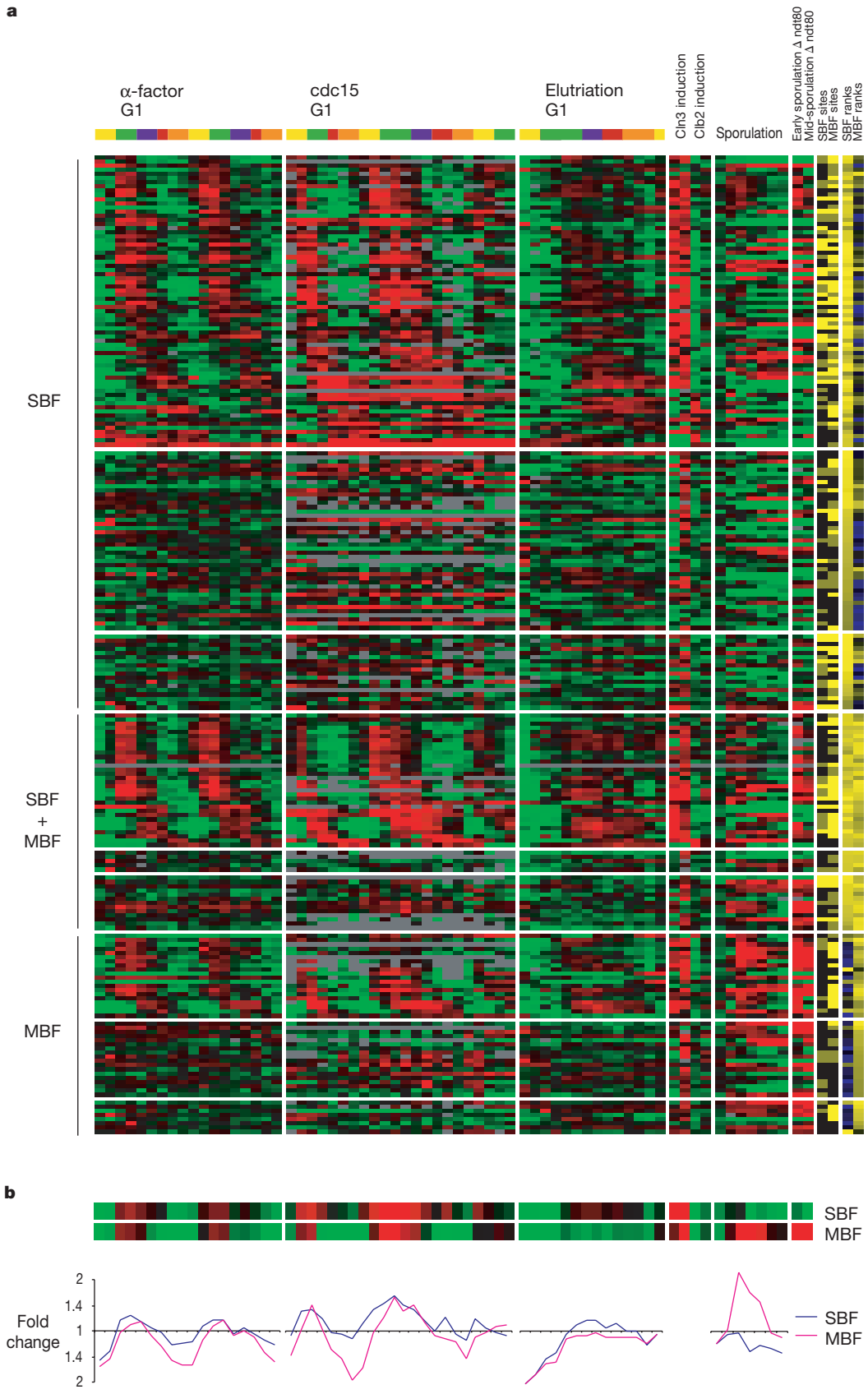

Figure 4 Expression profiles of SBF and MBF targets. **a**, Expression patterns of SBF and MBF targets are indicated (red–green colour scale). Cell-cycle data are from ref. 12 and sporulation data are from ref. 18. The stages of the cell cycle are: M/G1, yellow; G1, green; S, blue; S/G2l, red; and G2/M, orange. Yellow boxes indicate the presence of consensus binding sites in the intergenic sequences upstream of each ORF (right), and the median percentile rank in IPs of the upstream sequences is also indicated (blue–yellow colour scale), as in Fig. 3. For each set of targets, the top panel contains cell-cycle regulated genes, the bottom panel contains genes that are members of divergently transcribed pairs in which the other member was cell-cycle regulated, and the middle panel contains the remainder of the non-cell-cycle regulated genes. **b**, Average expression profiles of the cell-cycle regulated targets of SBF and MBF, computed by averaging the $\log_2$ (Cy5/Cy3) ratios. Note the specific induction of MBF targets during sporulation.

Several of the putative SBF targets are implicated in mating and pseudohyphal growth. *RSR1* and *WSC2* have a positive role in pseudohyphal growth; *CAP1*, *GIC1* and *GIC2* are involved in shmoo formation; *SRL1*, the highest-ranking target of SBF in our immunoprecipitations, is induced by *TEC1*, a positive regulator of pseudohyphal growth; and *CLN1* is required for pseudohyphal growth[21]. The expression of *TEC1* and *PHD1*, another positive regulator of pseudohyphal growth, was reduced in the *swi4* deletion strain (see Supplementary Information), which suggests a positive function for SBF in the regulation of these genes.

Our results support a model in which SBF is the principal controller of membrane and cell-wall formation[11], whereas MBF primarily controls DNA replication[6]. The need for both DNA replication and membrane and cell-wall biogenesis during G1/S in dividing cells accounts for the parallel functions of SBF and MBF in the mitotic cell cycle. It is possible that SBF has an independent and distinct role in mating and pseudohyphal growth, whereas MBF has a role in meiosis. In fission yeast, an Mbp1-like factor, p73pct1, is involved in regulating both mitosis and meiosis[22,23]. In budding yeast, it is thought that Swi6 is involved in meiosis, but to our knowledge, the role of Mbp1 in meiosis has not been examined directly[24]. Experiments that directly examine the function of SBF and MBF in mating and meiosis, respectively, may further clarify this system.

Our results indicate that although the sequences that define canonical binding sites for SBF or MBF are common in the yeast genome, only a fraction of these sequences are actually bound by the corresponding factors *in vivo*. Moreover, most of the *in vivo* binding sites are adjacent to genes that show a characteristic cell-cycle transcription pattern. The use of DNA microarrays, together with protein-affinity isolation methods, should be generally applicable to the investigation of molecular systems that involve physical interactions of proteins with the genome. □

Methods

DNA microarrays

We synthesized roughly 6,700 PCR primer pairs to amplify nearly every intergenic region in the yeast genome. Primers were designed to amplify genomic DNA immediately adjacent to the coding sequence of every ORF or non-ORF feature, up to the end of the neighbouring ORF or non-ORF feature. Intergenic elements were designated by prefixing the SGD (Saccharomyces Genome Database, http://genome-www.stanford.edu/saccharomyces) designated name of the ORF or non-ORF feature that was immediately to its left with an 'i'. The primers and their sequences can be obtained from Research Genetics (Huntsville, Alabama). We prepared DNA microarrays as described[25].

Crosslinking and immunoprecipitation

For chromatin immunoprecipitations, 5×10^8 cells were used with the conditions as described[26]. Briefly, cells were treated with formaldehyde to crosslink proteins and nucleic acids, and then lysed. Extracts were sonicated with a Branson 350 Sonifier with a microtip, at a power setting of 7 and a 60% duty cycle. Samples were pulsed eight times for 12 s to shear chromatin to a final average size of about 0.5–2 kilobases. We used antibodies at a 1:100 dilution for immunoprecipitation. Anti-haemmagglutinin antibodies and anti-Myc antibodies were from BabCo (Berkeley Antibody Company). Affinity-purified anti-Swi6 polyclonal antibody was a gift from K. Baetz and B. Andrews. The chromatin–antibody complexes were precipitated with Protein A Sepharose beads (Pierce) and washed twice with lysis buffer and twice with 1× TBS (150 mM NaCl, 20 mM Tris–HCl, pH 7.6) for 5 min each. Chromatin was eluted from the beads with 1% SDS, 1 × TE (50 mM Tris–HCl, 10 mM EDTA) and washed with 0.67% SDS, 1× TE. Crosslinks in the immunoprecipitated chromatin were reversed by heating at 65 °C for at least 10 h and the DNA was purified as described[27], and resuspended in 10 μl of 1× TE.

Microarray hybridizations and analysis

A detailed protocol for PCR amplification and fluorescent labelling of immunoprecipitated DNA is available at http://www.microarrays.org. Microarrays were hybridized at 65 °C for 6 h and washed as described previously, then scanned with a GenePix 4000A scanner (Axon Instruments). We quantified fluorescence intensities using GenePix Pro software (version 3.0), and uploaded data to a custom database for analysis. Data were filtered to exclude spots with obvious defects, or where the absolute signal intensity in both channels was below an empirically determined threshold. In considering loci whose percentile ranks were above the selected thresholds as putative targets, we excluded loci with similarly high ranks in control experiments (within 2× s.d. for the SBF targets, and within 1× s.d. for the MBF targets). The control experiments were immunoprecipitates performed using antibodies specific to the HA epitope tag, carried out with a strain lacking any of the HA fusion proteins (Fig. 3, columns 1, 2 and 3).

Received 30 August; accepted 1 December 2000.

1. Koch, C., Moll, T., Neuberg, M., Ahorn, H. & Nasmyth, K. A role for the transcription factors Mbp1 and Swi4 in progression from G1 to S phase. *Science* **261,** 1551–1557 (1993).
2. Koch, C. & Nasmyth, K. Cell cycle regulated transcription in yeast. *Curr. Opin. Cell Biol.* **6,** 451–459 (1994).
3. Andrews, B. J. & Herskowitz, I. The yeast *SWI4* protein contains a motif present in developmental regulators and is part of a complex involved in cell-cycle-dependent transcription. *Nature* **342,** 830–833 (1989).
4. Primig, M., Sockanathan, S., Auer, H. & Nasmyth, K. Anatomy of a transcription factor important for the start of the cell cycle in *Saccharomyces cerevisiae*. *Nature* **358,** 593–597 (1992).
5. Dirick, L., Moll, T., Auer, H. & Nasmyth, K. A central role for *SWI6* in modulating cell cycle Start-specific transcription in yeast. *Nature* **357,** 508–513 (1992).
6. Lowndes, N. F., Johnson, A. L. & Johnston, L. H. Coordination of expression of DNA synthesis genes in budding yeast by a cell-cycle regulated *trans* factor. *Nature* **350,** 247–250 (1991).
7. Ogas, J., Andrews, B. J. & Herskowitz, I. Transcriptional activation of *CLN1*, *CLN2*, and a putative new G1 cyclin (*HCS26*) by *SWI4*, a positive regulator of G1-specific transcription. *Cell* **66,** 1015–1026 (1991).
8. McIntosh, E. M. MCB elements and the regulation of DNA replication genes in yeast. *Curr. Genet.* **24,** 185–192 (1993).
9. Schwob, E. & Nasmyth, K. *CLB5* and *CLB6*, a new pair of B cyclins involved in DNA replication in *Saccharomyces cerevisiae*. *Genes Dev.* **7,** 1160–1175 (1993).
10. Measday, V., Moore, L., Ogas, J., Tyers, M. & Andrews, B. The *PCL2* (*ORFD*)-*PHO85* cyclin-dependent kinase complex: a cell cycle regulator in yeast. *Science* **266,** 1391–1395 (1994).
11. Igual, J. C., Johnson, A. L. & Johnston, L. H. Coordinated regulation of gene expression by the cell cycle transcription factor Swi4 and the protein kinase C MAP kinase pathway for yeast cell integrity. *EMBO J.* **15,** 5001–5013 (1996).
12. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9,** 3273–3297 (1998).
13. Bailey, T. L. & Elkan, C. in *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology* 28–36 (American Association for Artificial Intelligence, Menlo Park, California, 1994).
14. Stormo, G. D. & Hartzell, G. W. Identifying protein-binding sites from unaligned DNA fragments. *Proc. Natl Acad. Sci. USA* **86,** 1183–1187 (1989).
15. Koch, C., Schleiffer, A., Ammerer, G. & Nasmyth, K. Switching transcription on and off during the yeast cell cycle: Cln/Cdc28 kinases activate bound transcription factor SBF (Swi4/Swi6) at start, whereas Clb/Cdc28 kinases displace it from the promoter in G2. *Genes Dev.* **10,** 129–141 (1996).
16. Harrington, L. A. & Andrews, B. J. Binding to the yeast Swi4,6-dependent cell cycle box, CACGAAA, is cell cycle regulated *in vivo*. *Nucleic Acids Res.* **24,** 558–565 (1996).
17. Breeden, L. & Mikesell, G. Three independent forms of regulation affect expression of *HO*, *CLN1* and *CLN2* during the cell cycle of *Saccharomyces cerevisiae*. *Genetics* **138,** 1015–1024 (1994).
18. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282,** 699–705 (1998).
19. Johnston, L. H., Johnson, A. L. & Barker, D. G. The expression in meiosis of genes which are transcribed periodically in the mitotic cell cycle of budding yeast. *Exp. Cell Res.* **165,** 541–549 (1986).
20. Dirick, L., Goetsch, L., Ammerer, G. & Byers, B. Regulation of meiotic S phase by Ime2 and a Clb5,6-associated kinase in *Saccharomyces cerevisiae*. *Science* **281,** 1854–1857 (1998).
21. Madhani, H. D., Galitski, T., Lander, E. S. & Fink, G. R. Effectors of a developmental mitogen-activated protein kinase cascade revealed by expression signatures of signalling mutants. *Proc. Natl Acad. Sci. USA* **96,** 12530–12535 (1999).
22. Zhu, Y., Takeda, T., Nasmyth, K. & Jones, N. pct1$^+$, which encodes a new DNA-binding partner of p85cdc10, is required for meiosis in the fission yeast *Schizosaccharomyces pombe*. *Genes Dev.* **8,** 885–898 (1994).
23. Ayté, J., Leis, J. F. & DeCaprio, J. A. The fission yeast protein p73res2 is an essential component of the mitotic MBF complex and a master regulator of meiosis. *Mol. Cell. Biol.* **17,** 6246–6254 (1997).
24. Leem, S. H., Chung, C. N., Sunwoo, Y. & Araki, H. Meiotic role of *SWI6* in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **26,** 3154–3158 (1998).
25. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686 (1997).
26. Strahl-Bolsinger, S., Hecht, A., Luo, K. & Grunstein, M. *SIR2* and *SIR4* interactions differ in core and extended telomeric heterochromatin in yeast. *Genes Dev.* **11,** 83–93 (1997).
27. Orlando, V. & Paro, R. Mapping Polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin. *Cell* **75,** 1187–1198 (1993).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) and the Stanford University Web site (http://genome-www.stanford.edu/chromatinip/) or as paper copy at the London editorial office of *Nature*.

Acknowledgements

We thank J. DeRisi for mapping software; O. Aparacio for immunoprecipitation protocols; A. Khodursky for help with significance tests; and J. Lieb for comments on the manuscript and discussion. This work was supported by grants from the National Institutes of Health, and by the Howard Hughes Medical Institute. P.O.B. is an associate investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to P.O.B. (e-mail: pbrown@cmgm.stanford.edu) or M.S. (e-mail: michael.snyder@yale.edu).