Promoter-specific binding of Rap1 revealed by genome-wide maps of protein–DNA association

Jason D. Lieb[1], Xiaole Liu[2], David Botstein[3] & Patrick O. Brown[1]

Published online: 16 July 2001, DOI: 10.1038/ng569

We determined the distribution of repressor-activator protein 1 (Rap1) and the accessory silencing proteins Sir2, Sir3 and Sir4 *in vivo* on the entire yeast genome, at a resolution of 2 kb. Rap1 is central to the cellular economy during rapid growth, targeting 294 loci, about 5% of yeast genes, and participating in the activation of 37% of all RNA polymerase II initiation events in exponentially growing cells. Although the DNA sequence recognized by Rap1 is found in both coding and intergenic sequences, the binding of Rap1 to the genome was highly specific to intergenic regions with the potential to act as promoters. This global phenomenon, which may be a general characteristic of sequence-specific transcriptional factors, indicates the existence of a genome-wide molecular mechanism for marking promoter regions.

Introduction

Transcription factors control the expression of specific genes by acting selectively at genomic loci that are often predicted imperfectly by the factor's *in vitro* DNA-binding properties. Understanding where regulatory proteins bind to the genome *in vivo*, what determines where they bind, and the role of site-specific binding to genomic DNA in determining their regulatory specificity is critical to understanding the mechanism and logic of transcriptional regulation.

Rap1 is a well-studied DNA-binding protein that performs a diverse set of tasks in *Saccharomyces cerevisiae*. In its most thoroughly understood function, Rap1 binds to $[C_{(1-3)}A]_n$ repeats at chromosome ends, where it regulates telomere length by recruiting the Rap1-interacting factor proteins Rif1 and Rif2, and represses telomeric transcription through its interaction with Sir2, Sir3 and Sir4 (refs. 1,2). The Rap–Sir complex also accomplishes a function essential for mating by repressing transcription at the silent mating-type loci *HML* and *HMR* (mating-type cassettes, left and right)[3]. Rap1 binds to DNA through two Myb-type helix–turn–helix motifs[4]. The Sir proteins do not bind DNA directly but rely on interactions with each other, Rap1, histones H3 and H4 and other proteins to associate with specific genomic loci[1,5,6].

In addition to their joint role at telomeres and silent mating-type loci, Rap1 and the Sir proteins have independent functions. For example, Sir2 is an nicotinamide adenine dinucleotide-dependent histone deacetylase required for ribosomal DNA (rDNA) silencing and yeast longevity, functions that do not directly involve Rap1, Sir3 or Sir4 (ref. 7). Likewise, the Sir proteins are not required for Rap1's essential function as a sequence-specific transcriptional regulator of many essential *S. cerevisiae* genes, including genes encoding ribosomal proteins and glycolytic enzymes[8]. At some promoters, Rap1 does not act directly as a transcriptional activator but instead acts as a factor that allows binding by other regulatory proteins[9]. A human Rap1 homolog was shown to be localized to chromosome ends and involved in telomere length regulation, but its function in transcriptional regulation is not known[10].

The function of Rap1 in regulating the transcriptional program of *S. cerevisiae* is still unclear, mainly because Rap1 mutations that abolish DNA binding are lethal[11] and overexpression of Rap1 is toxic[12]. These impediments make it difficult to use genetic screens or genome-wide surveys of changes in transcript levels[13] to identify loci that are subject to *RAP1* regulation. To obtain a global picture Rap1's role in the regulatory logic of the yeast cell, and to investigate its genome-wide DNA-binding specificity *in vivo*, we mapped the binding sites of Rap1 and the silencing proteins Sir2, Sir3 and Sir4 across the entire yeast genome in wildtype cells.

Results

Determining sites at which Rap1 interacts with the yeast genome

To crosslink proteins at their sites of interaction with DNA, we added formaldehyde to a culture of wildtype yeast growing exponentially in rich media. We sonicated extracts from these yeast to shear chromatin to an average size of 1 kb, and then used them in immunoprecipitation (IP) reactions with polyclonal antibodies raised against Rap1 or Sir protein peptides. We purified, amplified and fluorescently labeled the DNA fragments enriched in each IP and hybridized these to whole-genome yeast DNA microarrays containing 12,943 unique spotted DNA segments representing every open reading frame (ORF) and intergenic region. The results allowed us to identify which segments of the genome were enriched in the IP, and thereby to construct a genome-wide map of *in vivo* protein–DNA interactions[14,15].

To determine the specific genomic targets of Rap1, we repeated IPs independently six times, each in parallel with a control IP using antibodies specific to the hemagglutinin epitope tag, which is not present in wildtype strains, or a mock IP without antibody. For each IP, we assigned a percentile rank to each arrayed genomic

[1]*Department of Biochemistry,* [2]*Stanford Medical Informatics and* [3]*Department of Genetics, Stanford University, Stanford, California 94305-5428, USA. Correspondence should be addressed to P.O.B. (e-mail: pbrown@cmgm.stanford.edu).*

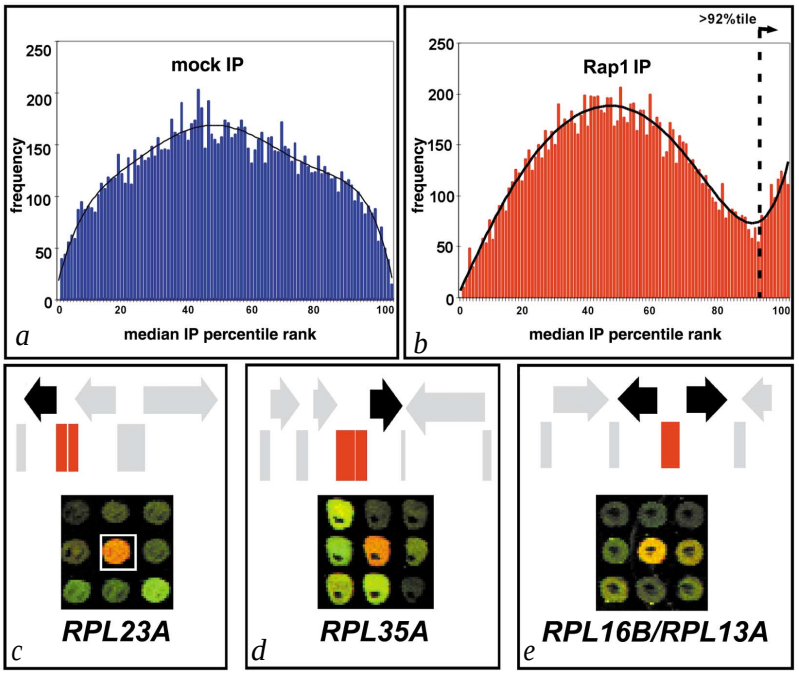

Fig. 1 Determination of Rap1 targets. We repeated Rap1 IPs independently six times, each in parallel with a mock IP without Rap1 antibodies. For each IP, we assigned each arrayed genomic DNA segment a percentile rank. The rank was based on degree of enrichment detected for corresponding fragments in the IP, relative to genomic DNA. We then used the six percentile rank values to calculate the median percentile rank value for each spot. ***a***, The unimodal distribution of median rank values in mock IPs indicates that very few fragments are strongly selected in control experiments (bin size =1 percentile). ***b***, The bimodal distribution of median rank values in Rap1 IP experiments results from consistent enrichment of particular fragments in Rap1 IPs, which is not the case for the control IPs. We defined Rap1 targets as DNA segments whose median percentile rank of enrichment falls to the right of the trough of the bimodal distribution. For Rap1 IPs, this point corresponds to the 92nd percentile, defining 6.3% of arrayed spots as Rap1 targets. ***c***–***e***, Translation of array data into the map in Fig. 2. Each panel shows a small section of a microarray. The center spot in each panel represents the promoter for an RPG and the fluorescence indicates that it is enriched in the Rap1 IP. In the schematic map, arrows indicate ORFs (black, RPGs; gray, others) and their direction of transcription; bars represent intergenic regions (red, Rap1 targets; gray, others).

DNA segment based on the relative enrichment of the corresponding sequences in the IP. We calculated the median of the percentile rank values for each genomic segment for the six control IPs (Fig. 1*a*) and the six Rap1 IPs (Fig. 1*b*), respectively. High median ranks resulted when particular genomic DNA fragments were enriched consistently in the IP experiments. The bimodal distribution of medians in the Rap1 IPs resulted from a subclass of fragments enriched consistently in Rap1 IPs, which was not found with the control IPs. We classified DNA segments whose median percentile rank of enrichment fell to the right of the trough of the bimodal distribution as Rap1 targets (Fig. 1*b*–*e*). We excluded from the list of targets DNA segments with equally high median ranks in control experiments, or segments for which the measurement of enrichment was technically inadequate in more than three of the six Rap1 IPs.

In the Rap1 IPs, 727 arrayed DNA segments, which represented 365 intergenic regions, 338 ORFs and 24 other genomic features, were selected by these criteria (Web Table A). These 727 segments were grouped into 294 contiguous clusters that we called 'foci' (Fig. 2). Foci arose because a single site of protein–DNA interaction would allow immunoprecipitation of randomly sheared genomic DNA fragments (0.5–2 kb) that hybridized not only to the DNA segment representing the actual binding site but also to its genomic neighbors. In fact, for 85% (618 of 727) of the arrayed genomic segments that detected Rap1 IP-enriched DNA, an array element representing an adjacent genomic segment also detected enrichment.

The extent of Rap1 and Sir protein binding to telomeres and silent mating loci

Rap1 and the Sir proteins are bound to telomeres in yeast[1,16–19]. Consistent with these findings, 149 arrayed segments detected enrichment of telomeric DNA in the Rap1 IPs, defining foci representing 30 of the 32 chromosome ends. Only telomeres at the right end of chromosome three (3R) and the left end of chromosome 4 (4L) lacked detectable binding (Fig. 2). Because of potential cross-hybridization among homologous telomeric regions, however, we could unambiguously ascribe Rap1 binding to only 17 telomeres on 14 different chromosomes (Methods). We obtained similar results, including the lack of detectable association with 3R and 4L, for Sir2, Sir3 and Sir4 (Web Figs. A–C).

Previous chromatin IP experiments found that Rap1 and the Sir proteins extend 2–4 kb inward from the right end of chromosome 6, consistent with the 2- to 4-kb inward spread of telomere position effect at that telomere[18,19]. However, other studies have indicated that the extent of binding at each chromosome end may be variable; for example, Sir3 binding has been reported to extend inward about 10 kb on chromosome 5R[19]. We found that each of the proteins remained associated with chromatin for variable distances ranging from 600 bp to 19 kb inward from the chromosome end. Despite the variability in binding distance, the extent of each subunit's binding was highly correlated with that of other Rap–Sir complex members at each end, as would be expected for proteins that physically interact (Web Fig. D). Our results are consistent with the general transcriptional derepression observed 6–8 kb from the chromosome ends in *rap1*, *sir2*, *sir3* or *sir4* mutants[13]. However, transcriptional silencing at native telomeres is variegated and discontinuous[20], and the precise relationship between transcriptional silencing and occupancy by the Rap–Sir complex has not been established.

The Rap–Sir complex also performs a function essential for mating by silencing transcription at the mating-type loci *HML* and *HMR*[1]. Consistent with this function and prior studies, we found that Rap1 and Sir3 were localized to a 7.1-kb stretch of DNA at *HML* bounded on the left by *YCL069W* and on the right by *YCL064C*, whereas Sir2 and Sir4 IPs selected a nearly continuous set of fragments from telomere 3L to *YCL064C*, a distance of 15 kb (Web Fig. E). The association of all four proteins with *HMR* was less extensive and was restricted to a 3.5-kb region defined by previously characterized heterochromatic boundary elements that were discovered through restriction endonuclease accessibility and transgene-silencing assays[21].

Rap1 binds to the promoters of 362 ORFs, including 122 of the 137 ribosomal protein genes

Mutations in *RAP1* that abolish DNA binding are lethal[11] and overexpression of Rap1 is toxic[12], frustrating attempts to identify loci that are subject to Rap1 regulation through genetics or

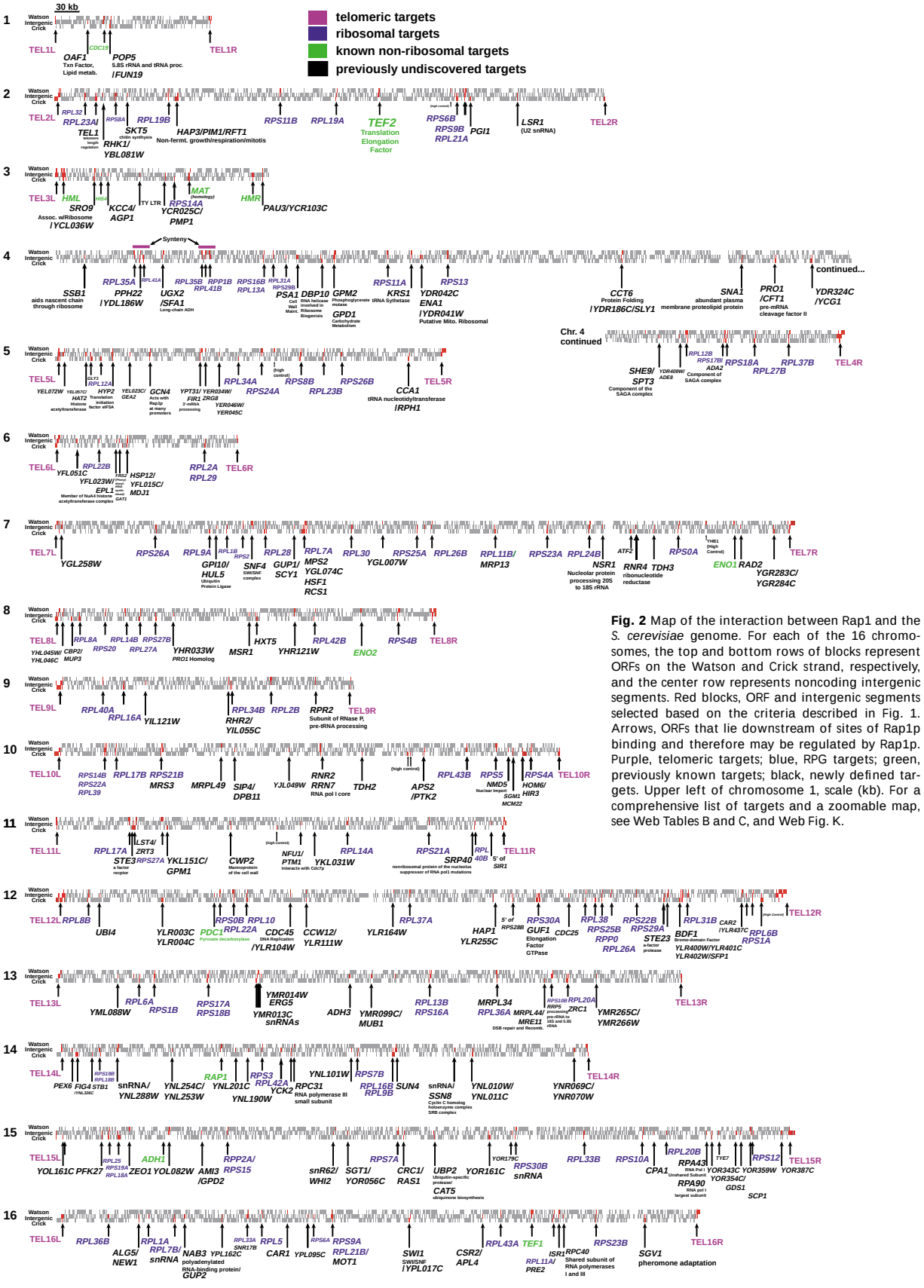

Fig. 2 Map of the interaction between Rap1 and the *S. cerevisiae* genome. For each of the 16 chromosomes, the top and bottom rows of blocks represent ORFs on the Watson and Crick strand, respectively, and the center row represents noncoding intergenic segments. Red blocks, ORF and intergenic segments selected based on the criteria described in Fig. 1. Arrows, ORFs that lie downstream of sites of Rap1p binding and therefore may be regulated by Rap1p. Purple, telomeric targets; blue, RPG targets; green, previously known targets; black, newly defined targets. Upper left of chromosome 1, scale (kb). For a comprehensive list of targets and a zoomable map, see Web Tables B and C, and Web Fig. K.

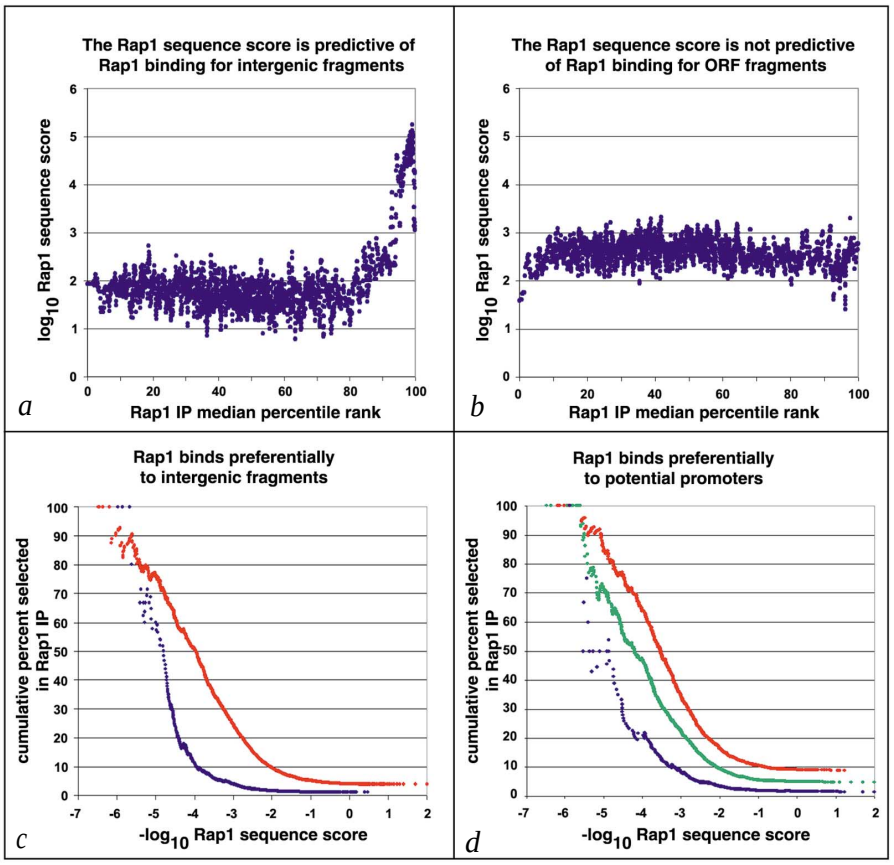

Fig. 3 Rap1 binds preferentially to potential promoters ***a***, The $\log_{10}$ of the Rap1 sequence score (moving median; window size, 20) plotted as a function of the median Rap1 IP percentile rank of all intergenic segments. The Rap1 sequence score is strongly predictive of Rap1 binding for intergenic segments. ***b***, Same as *a*, but for ORF segments. ***c***, The cumulative percentage of non-telomeric ORFs (blue) and non-telomeric intergenic segments (red) that are Rap1 targets, plotted as a function of the negative $\log_{10}$ of their sequence scores. Thus, segments containing the best Rap1p binding sequences are to the left. For any given sequence score, a higher proportion of intergenic fragments with at least that score is selected in the Rap1 IPs, compared with the proportion of ORFs selected by IP. All non-telomeric segments are plotted. ***d***, Same as *c*, except that the three classes of intergenic fragments, potential non-promoters (blue), single-promoters (green) and double-promoters (red), are plotted. For any given sequence score, a higher proportion of potential double- and single-promoter fragments with at least that score is selected in the Rap1 IPs, compared with the proportion of non-promoter fragments selected by IP. All non-telomeric intergenic segments are plotted.

genome-wide surveys of transcript level changes[13]. Mapping of the Rap1-binding foci provides a way to identify genes that Rap1 may regulate. In general, we considered genes transcriptional targets if Rap1 bound the intergenic region immediately upstream of the corresponding ORF; 362 ORFs satisfied this criterion (Fig. 2; Web Tables B and C).

The largest functional group of Rap1 targets is the ribosomal protein genes (RPGs). A typical RPG promoter includes one or two copies of the canonical Rap1-binding sequence at 250–400 bp upstream of the first codon, followed closely by one or two T-rich elements[22]. At least one strong Rap1-binding site is predicted in the promoters of 124 of 137 RPGs[22], but binding of Rap1 has been confirmed experimentally for relatively few specific ribosomal promoters[23]. Here, 122 RPG promoters were enriched in the Rap1 IPs (Fig. 2). Of the 124 RPG promoters predicted to have strong Rap1-binding sites, 118 were selected in our assay (exceptions were ribosomal protein (*RP*)*L15A*, *RPL38*, *RPL42B*, *RPS24B*, *RPS28A* and *RPS30B*). Conversely, 13 RPGs do not contain a previously predicted Rap1-binding site in their upstream region; nine of these were not enriched in Rap1 IPs (*RPS16A*, *RPS31*, *RPS25B* and *RPS29A* were enriched). Closer examination of these four RPG promoters showed that *RPS25B* and *RPS29A* contain non-consensus Rap1-binding sites[22]. Therefore, only the promoters for *RPS16A* and *RPS31* were selected unexpectedly in our experiments. Other general regulatory transcription factors (GRFs) may substitute for Rap1 at the 15 RPG promoters that did not appear to bind Rap1 *in vivo*. Eight of these promoters have a predicted Abf1 (ARS binding factor) binding site, and one has a predicted Reb1 (RNA polymerase I enhancer binding protein) binding site[22]. The transcriptional profiles of the 15 RPGs whose promoters were not Rap1 targets (*RPL3*, *RPL4A*, *RPL4B*, *RPL15A*, *RPL15B*, *RPL18B*, *RPL38*, *RPL42B*, *RPP1A*, *RPP2B*, *RPS22B*, *RPS24B*, *RPS28A*, *RPS28B* and *RPS30*) did not differ substantially from those of Rap1-regulated RPGs, consistent with published evidence that the binding site for one GRF can be exchanged with another[24], and that protein domains can be swapped among GRFs without a loss of function[25].

Rap1 targets form a continuous enzymatic pathway in glycolysis

Rap1 had previously been shown to bind to the promoters of 17 non-RPG ORFs[23], 14 of which were also identified by our experiments (exceptions were *YCR012W*, *YIL033C* and *YGL181W*). Among these known targets are five glycolysis genes: 3-phosphoglycerate kinase (*PGK1*), enolase I (*ENO1*), enolase II (*ENO2*), pyruvate kinase (*CDC19*) and pyruvate decarboxylase (*PDC1*)[23]. In addition to these previously identified targets, our results identified targets that complete a continuous enzymatic pathway in glycolysis from fructose-1,6-biphosphate to ethanol (Web Fig. F). These target genes encode enzymes that act downstream of phosphofructokinase and fructose-1,6-bisphosphatase, the key regulatory switch between glycolysis and glucogenesis. Newly identified Rap1 targets include the glyceraldehyde-3-phosphate dehydrogenase gene *TDH3* and the glycerol-3-phosphate dehydrogenase genes *GPD1* and *GPD2*. The phosphoglycerate mutase genes *GPM1* and *GPM2* link the previously known targets *PGK1*, *ENO1* and *ENO2*. The newly identified targets alcohol dehydrogenase genes *ADH1* and *ADH3* provide the link from acetaldehyde to ethanol.

Rap1 IPs reveal 185 newly identified binding targets

There were 185 ORFs downstream of newly identified Rap1-binding sites. In addition to RPGs, other genes involved in protein synthesis[23,26,27] were considerably over-represented among the

putative Rap1 targets, emphasizing the involvement of Rap1 in regulating the cell's capacity for protein synthesis (Web Fig. G). For example, our results indicate involvement of Rap1 in coordinating synthesis of ribosomal proteins and ribosomal RNA. Nine Rap1 targets are directly involved in rRNA production, including the 190-kD and 36-kD subunits of RNA polymerase I (*RPA190* and *RPA43*), an RNA polymerase I transcription factor (*RRN7*), a 40-kD subunit shared between polymerase I and polymerase III (*RPC40*), a polymerase III-specific subunit (*RPC31*) and three genes involved in rRNA processing, *RRP5*, *NSR1* (nuclear localization sequence binding protein) and *POP5* (processing of precursors). *SSB1* (stress-seventy subfamily B), which encodes a heat shock protein (HSP)70 chaperone associated with the ribosome–nascent chain complex, is coregulated with ribosomal genes, but the basis for that coregulation is unknown[28]. The identification of the promoter of *SSB1* as a Rap1 target provides a plausible hypothesis for how *SSB1* and ribosomal genes are coordinately regulated. Genes involved in sugar metabolism were also over-represented among Rap1 targets, whereas 'questionable' ORFs (Methods) and ORFs of unknown function were under-represented.

Rap1 binds to the promoters of the most heavily transcribed yeast genes

A striking common characteristic of genes that we identified as targets of Rap1 was a very high transcription rate during exponential growth (Fig. 4). The average transcription rate of the putative Rap1 targets is 45 mRNAs per hour, whereas the average for all transcripts is 7 mRNAs per hour[29]. Genes downstream of promoters bound by Rap1 account for an estimated 14,000 of the 38,000 transcripts produced each hour in rapidly growing cells[29], or 37% of all yeast mRNA transcripts, even though they account for only 5.4% of all yeast genes with measured transcription rates. Not all heavily transcribed genes are Rap1 targets and vice versa: 41% of transcripts above the 96th percentile of transcription rates are Rap1 targets, and 56% of Rap1 targets are above the 96th percentile of transcription rates. The high transcription rates observed for Rap1 targets remains after exclusion of RPG targets: 36% of the remaining targets are above the 96th percentile of transcription rates, and their average transcription rate is 13.2 mRNAs per hour, nearly twice the genome-wide average.

Rap1 imposes an RPG-like expression pattern on many of its target genes

We investigated whether Rap1 imposes a stereotypic program of expression on its target genes by examining their expression patterns in many diverse culture conditions. The Rap1 targets formed two main groups: 252 genes formed a cluster with a pattern of regulation similar to the RPGs and 113 genes were expressed in patterns not correlated with the RPGs (Web Fig. H). Even after exclusion of the RPGs, genes involved in protein synthesis and sugar metabolism were over-represented in the cluster with an RPG-like pattern of expression, whereas 51% of the genes in the 'uncorrelated' cluster were 'questionable' ORFs (Methods; Web Table D), ORFs of unknown function, or ORFs that were considered possible targets but shared an intergenic region with a known or likely target. This indicates that a substantial fraction of the genes in the 'uncorrelated' cluster may not be true transcriptional targets of Rap1. At some promoters, Rap1 acts as a factor that allows access to other regulatory proteins[9], and some of the genes in the 'uncorrelated' cluster may represent loci in this class. The identity of newly defined Rap1 targets and the regulation of their transcripts are consistent with the hypothesis that Rap1 directs a transcriptional program that promotes high growth rates in nutrient-rich environments by boosting the cells' capacity for protein synthesis and ATP production.

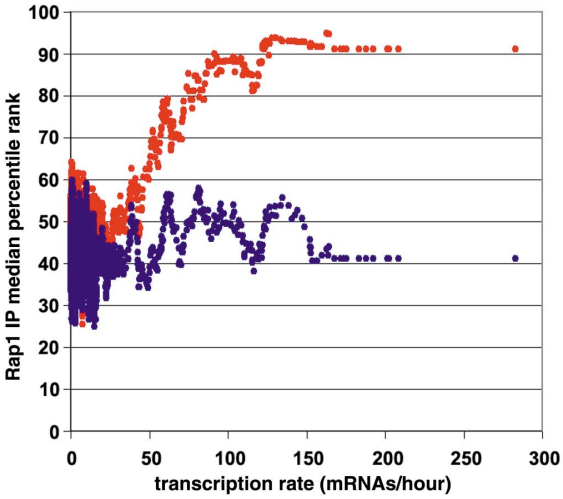

Fig. 4 Rap1 binds to the promoters of heavily transcribed genes. The median percentile rank of enrichment detected by the arrayed intergenic segment immediately upstream of each yeast ORF in Rap1 IPs (red, moving median; window size=20) or control IPs (blue), plotted as a function of the transcription rate (mRNAs/hour) of that ORF[29]. All ORFs are plotted.

The DNA sequence motif recognized by Rap1 can be inferred from the IP data alone

What determines the sites at which Rap1 binds to the genome? We used the computer program BioProspector to search for recurring motifs among Rap1 IP-enriched fragments[30]. The resulting motif matrix (Web Table E), represented by ACACCC RYACAYM, was nearly identical to the previously determined binding sequence for Rap1 (refs. 22,31–33). Therefore, we were able to localize the probable sites of interaction within the approximately 2-kb foci to just 13 bp. Although the *in vitro* properties of Rap1 binding have been established already, this matrix may more accurately reflect the *in vivo* specificity of Rap1, as it was derived from the entire spectrum of the natural targets of Rap1. This result also demonstrates that the *in vivo* binding sites of transcription factors with unknown DNA-binding specificities can be determined by this genomic mapping method.

We used the motif matrix derived from the IP data to assign a Rap1 motif 'sequence score' to every DNA segment represented on the array (Methods). Segments with one or more strong matches to the Rap1 motif matrix received high scores and segments with poor matches received low scores. Whereas the sequence score was strongly predictive of the binding of Rap1 for intergenic segments (Fig. 3*a*), we found no such relationship for ORFs (Fig. 3*b*). These data indicate that most of the 338 ORF sequences enriched in our IPs did not themselves bind Rap1. Indeed, one quarter of the selected ORFs were telomeric, and of the remainder, 90% were adjacent to a selected intergenic fragment. Therefore, these ORFs were selected presumably by virtue of the binding of Rap1 to the adjacent intergenic segment, leaving only 32 ORFs that seemed to be independently selected in the IPs (Methods). We classified 18 of these 32 ORFs as 'questionable' (Methods), and many of them lie upstream of known targets and may act as promoters. Furthermore, of the ORFs that have been experimentally confirmed to encode proteins, five lie upstream of ribosomal protein or mating genes, leaving only nine authentic ORFs that appear to bind Rap1: cytoplasmic chaperonin of the Cct ring complex (*CCT6*), threonine aldolase (*GLY1*), heat shock transcription factor (*HSF1*), alcohol acetyltransferase (*ATF2*), protein required for splicing of COB aI5 intron (*CBP2*), RNase P

ribonucleoprotein (*RPR2*), *MRPL49* (mitochondrial ribosomal protein, large subunit), ornithine aminotransferase (*CAR2*) and arginase (*CAR1*).

Preferential binding of Rap1 to potential promoters is not fully explained by the distribution of predicted Rap1-binding sequences

Although intergenic sequences generally contained more and better matches to predicted Rap1-binding sites than did ORFs, many ORFs did contain good matches. However, few of these ORFs were actually bound by Rap1 (Fig. 3*c*). For example, the 500 non-telomeric segments with the highest sequence scores (>95.5th percentile; Methods) included 322 intergenic sequences, 182 (57%) of which were Rap1 targets, whereas only 23 of the 163 ORF sequences in that group (14%) were Rap1 targets. Therefore, Rap1 binds intergenic sequences in preference to coding sequences, and the distribution of predicted Rap1-binding sequences does not explain this bias.

We next asked if, among the intergenic sequences, Rap1 bound preferentially to potential promoters. In *S. cerevisiae*, 75% of intergenic segments include putative promoters: 26% of intergenic segments are upstream of two divergently transcribed genes, 49% are upstream of one gene and downstream of another and 25% are presumed not to contain promoters because they are downstream of two convergently transcribed genes. Of the 278 non-telomeric intergenic fragments enriched specifically by Rap1 IP, 6% were non-promoters, 47% were upstream of one ORF and 46% were upstream of two divergently transcribed ORFs, almost twice the genomic frequency of this fragment class (26%). Of the 16 Rap1 IP-enriched non-promoter fragments, an adjacent upstream ORF and its promoter were selected in 13, and 3 were up- or downstream of an RPG. Therefore, each of these rare cases may be explained by co-selection with adjacent fragments that appeared to be authentic targets. The apparently exclusive binding of Rap1 to potential promoters indicates that, apart from telomeres, Rap1 acts primarily as a transcriptional regulator during rapid growth.

The specificity of Rap1 for binding to potential promoters rather than other intergenic regions cannot be accounted for only by the greater frequency of predicted Rap1-binding sites in promoter regions (Fig. 3*d*). For example, the 500 non-telomeric intergenic segments with the highest sequence scores (>91st percentile; Methods) included 188 potential double promoters, 57% (107 of 188) of which were Rap1 targets, and 258 potential single promoters, 39% (99 of 258) of which were Rap1 targets. However, only 9 of the 54 high-scoring non-promoter sequences (17%) in intergenic intervals were Rap1 targets.

A preference for binding to intergenic regions, and more particularly to potential promoters, may be a general property of proteins that act as promoters but recognize DNA motifs that are found throughout the genome[15]. Our data show that this preference operates *in vivo* on a genome-wide scale, indicating the existence of a molecular mechanism that serves to distinguish the potential binding sites that occur in coding sequences from those in intergenic regions.

Discussion

We have presented a detailed map of the sites at which four key regulatory proteins, Rap1, Sir2, Sir3 and Sir4, bind to the yeast genome at a resolution of approximately 2 kb. The binding distribution of these proteins tells us about the contributions these proteins make to the physiological and regulatory logic of the yeast cell. Rap1p is central to the cellular economy during rapid growth, targeting 294 loci, about 5% of yeast genes, and participating in the activation of more than one-third of all RNA polymerase II initiation events. The identities of 185 newly defined Rap1 targets and the regulation of their transcripts strongly suggests that Rap1 directs a transcriptional program that promotes rapid growth in nutrient-rich environments by boosting the cell's capacity for protein synthesis and ATP production.

We also investigated what information in the genome specifies the sites at which Rap1 binds to the genome in a living cell. Although the distribution of a specific sequence motif accounts for much of the specificity in Rap1 binding, the protein had a further striking specificity for binding to promoters in preference to coding regions and intergenic non-promoter segments. This preference did not appear to be accounted for by a local sequence motif, indicating the existence of a genome-wide mechanism that marks promoter regions in chromatin. What might be the basis for this distinction? Our attempts to find the missing specificity in sequence motifs surrounding the canonical Rap1-binding sites were unsuccessful. However, the context and *in vivo* properties of the Rap1-binding motif may provide additional specificity by determining promoter conformations favorable to cooperative binding between general and specific transcription factors[34]. Indeed, alignment methods like BioProspector produce simple models for how a sequence motif can specify Rap1 binding, and provide only a minimum estimate of the information present in local DNA sequences. Biases in the position and orientation of the Rap1-binding sequence with respect to the genes that Rap1 regulates indicate that long-range interactions may be important in the binding specificity of Rap1 *in vivo*. For example, Rap1-binding sites often occur in tandem between 250 and 450 bp upstream of RPGs[22] (Web Fig. I). Furthermore, among Rap1 targets whose downstream ORFs are regulated like RPGs (correlation, >0.7), 123 Rap1 motifs appear on the minus strand relative to the downstream ORF, whereas only 41 occur on the plus strand (Web Table F; Web Fig J). Differences between DNA at promoters and ORFs that are not readily recognized by simple motif search-and-alignment methods, like the ability to melt locally[35] or to phase nucleosomes, may further influence Rap1 binding.

The recruitment of complexes that modulate chromosomal architecture at promoters, such as SWI/SNF or RSC (Remodels the Structure of Chromatin) and histone acetylaces and deacetylaces, may also affect the binding specificity of Rap1[14,36]. The histone H4 acetyltransferase Esa1, a member of the Nucleosome Acetyltransferase of histone H4 (NuA4) protein complex, is recruited to many of the RPG promoters, and this recruitment is associated with binding of Rap1 (ref. 14). However, Esa1 is not likely be directly involved in specifying Rap1 targets, because binding of Rap1 seems to be a prerequisite for association of Esa1 with RPG promoters, and Esa1 occupancy at RPG promoters decreases in response to environmental changes, whereas Rap1 binding remains constant[14]. It remains possible that Esa1p itself, or the acetylation state of histones, acts to stabilize Rap1 binding at selected promoters.

The mechanism underlying the recruitment of Rap1 to promoters does not appear to depend on active transcription. Rap1 remains bound to nearly all of its targets during amino acid starvation[14], heat shock[14], and during stationary phase (J.D.L. and P.O.B., unpublished data), despite the fact that many of the downstream ORFs are transcriptionally silent in these conditions. Therefore, although it is possible that active transcription prevents the binding of Rap1 to certain loci, a lack of transcriptional activity does not hinder its recruitment. The persistence of the binding of Rap1 to the promoters of inactive genes strongly indicates that cofactors recruited by Rap1, or modifications made to Rap1 by cofactors, are major determinants of the transcriptional activity of genes downstream of Rap1 binding.

The global picture of the association of Rap1 and the Sir proteins with the genome provides new insights to the role that these proteins play in the regulatory logic of the yeast cell. By defining which sequences are functional Rap1p binding sites *in vivo*, and which apparently similar sequences are not, these data also provide a starting point for a systematic study to identify the currently unaccounted for sources of specificity in the binding and function of a transcription factor *in vivo*.

Methods

Online supplemental information. Full and detailed protocols for chromatin IP, DNA amplification, and microarray hybridization are available at: http://genome-www.stanford.edu/rap_sir/. The site also contains raw data, protocols, maps, annotated biochemical pathways, criteria for defining questionable ORFs, details about the motif-finding and score calculation, and a variety of other additional information not found in this text.

Strains. The following strains were used in these experiments: S288C; wild type (*MATα*), JRY4013; *MATα can1-100 his3-11 leu2-3,112 lys2Δ trp1-1 ura3-1*, JRY4563; JRY4013 but *sir2ΔTRP1*, JRY4578; JRY4013 but *sir4ΔHIS3*.

Array production and hybridization. We performed PCR of the individual segments, manufacture of DNA microarrays, and microarray hybridizations as previously described[15]. The whole-genome primer set and sequences can be obtained from Research Genetics (http://www.resgen.com). The reference hybridization probe for all experiments was a common pool of S288C (wildtype) genomic DNA that had been sonicated to the same length distribution as that of the IP samples. We then amplified and labeled the reference DNA using the same protocol used for the IP samples.

DNA sequence motif-finding software. The DNA sequence motif-finding software BioProspector adopts a Gibbs sampling algorithm to discover conserved DNA motifs, with the following improvements: the use of background Markov dependency to improve segment specificity, the sampling of alignments with two thresholds to allow variable motif copies in each sequence, the use of Monte Carlo simulation to test the statistical significance of a motif and the use of marginal distribution to find motifs with two blocks separated by a variable gap. Once a significant motif matrix Θ converges, it is used to calculate a sequence score for all arrayed amplicons of the yeast genome as follows: $\Sigma_{\text{all words } x \text{ width W in segment}}$ P(x generated from Θ)/P(x generated from background)

The 500 non-telomeric arrayed segments with the highest Rap1 sequence scores consisted of 322 intergenic segments, 163 ORFs, 5 transposon LTRs, 9 tRNA-encoding segments and 1 rDNA sequence.

Telomeric binding of the Rap1 and Sir proteins. We carried out Sir protein IPs in triplicate, in parallel with control IPs with Sir antibodies in Sir deletion strains. We defined binding to a chromosome end as 'unambiguous' if enrichment after IP was detected by at least one arrayed telomeric segment that contained less than 70% nucleotide sequence identity to any other arrayed DNA segment. The distances calculated from each chromosome end represent the innermost telomeric segment selected in our IP experiments. The measurements do not take into account the length of the arrayed PCR products or sonicated chromatin fragments, do not explicitly exclude arrayed DNA segments that contain homology to other loci and do not necessarily indicate continuous binding from the distal-most end of each chromosome to the proximal coordinate reported. More details are available at http://genome-www.stanford.edu/rap_sir/.

Internal distribution of the Sir proteins. The Sir proteins bound to fewer sites and had a simpler pattern of distribution than did Rap1 (Web Tables G–I). Most foci identified for the three Sir proteins were linked to the telomeres or to mating loci; only 12% of Sir2, 15% of Sir4 and 33% of Sir3 targets were not. Consistent with previous studies and with its specific function in longevity and silencing[7], we found that Sir2, but not Sir3, Sir4 or Rap1, was associated with the rDNA locus. We compared our chromatin maps with transcript measurements from Sir mutants, which show derepression at the telomeres and silent mating-type loci[13]. Transcript levels from nearly all non-telomeric genes targeted by the Sir proteins were unchanged (within ±1.5-fold) in their respective mutants. Therefore, the functional importance of the identified internal Sir protein targets remains to be determined.

'Questionable' ORFs. We assigned a single score to each ORF (Web Table I) intended to reflect the likelihood that it represented an authentic gene, based on the following criteria: codon usage, MIPS classification, overlap with another ORF, presence of a SAGE tag, YPD description, length of the predicted polypeptide and gene expression data. We subjectively weighted each criterion; they are listed above from 'heaviest' to 'lightest' weight. Using the score of 0 as a cutoff, we classified 874 ORFs as 'questionable'. Among ORFs with a score of <0, there were 10 named genes (1.1%): One encodes a structural RNA, one is now known to be noncoding, seven have no references or functions associated with them and only one has strong sequence similarity to a known gene.

Note: supplementary information is available on the Nature Genetics web site (http://genetics.nature.com/supplementary_info/).

Acknowledgments

We thank V. Iyer for initiating work on the intergenic yeast array, and other members of the Brown and Botstein labs for advice and technical support. We thank D. Brutlag and J. Liu for advice in developing BioProspector, J. Derisi for software that maps the IP data, and J. Rine for yeast strains. This work was supported by a grant from the National Human Genome Research Institute and by the Howard Hughes Medical Institute. P.O.B. is an associate investigator of the Howard Hughes Medical Institute, and the Helen Hay Whitney Foundation supports J.D.L.

Received 11 April; accepted 23 May 2001.

1. Moretti, P., Freeman, K., Coodly, L. & Shore, D. Evidence that a complex of Sir proteins interacts with the silencer and telomere-binding protein Rap1. *Genes Dev.* **8**, 2257–2269 (1994).
2. Shore, D. Telomerase and telomere-binding proteins: controlling the endgame. *Trends Biochem. Sci.* **22**, 233–235 (1997).
3. Aparicio, O.M., Billington, B.L. & Gottschling, D.E. Modifiers of position effect are shared between telomeric and silent mating-type loci in *S. cerevisiae*. *Cell* **66**, 1279–1287 (1991).
4. Konig, P., Giraldo, R., Chapman, L. & Rhodes, D. The crystal structure of the DNA-binding domain of yeast Rap1 in complex with telomeric DNA. *Cell* **85**, 125–136 (1996).
5. Cockell, M. *et al*. The carboxy termini of Sir4 and Rap1 affect Sir3 localization: evidence for a multicomponent complex required for yeast telomeric silencing. *J. Cell Biol.* **129**, 909–924 (1995).
6. Hecht, A., Laroche, T., Strahl-Bolsinger, S., Gasser, S.M. & Grunstein, M. Histone H3 and H4 N-termini interact with Sir3 and Sir4 proteins: a molecular model for the formation of heterochromatin in yeast. *Cell* **80**, 583–592 (1995).
7. Guarente, L. Sir2 links chromatin silencing, metabolism, and aging. *Genes Dev.* **14**, 1021–1026 (2000).
8. Warner, J.R. The economics of ribosome biosynthesis in yeast. *Trends Biochem. Sci.* **24**, 437–440 (1999).
9. Morse, R.H. Rap, Rap, open up! New wrinkles for Rap1 in yeast. *Trends Genet.* **16**, 51–53 (2000).
10. Li, B., Oestreich, S. & de Lange, T. Identification of human Rap1: implications for telomere evolution. *Cell* **101**, 471–483 (2000).
11. Graham, I.R., Haw, R.A., Spink, K.G., Halden, K.A. & Chambers, A. *In vivo* analysis of functional regions within yeast Rap1p. *Mol. Cell. Biol.* **19**, 7481–7490 (1999).
12. Freeman, K., Gwadz, M. & Shore, D. Molecular and genetic analysis of the toxic effect of Rap1 overexpression in yeast. *Genetics* **141**, 1253–1262 (1995).
13. Wyrick, J.J. *et al*. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418–421 (1999).
14. Reid, J.L., Iyer, V.R., Brown, P.O. & Struhl, K. Coordinate regulation of yeast ribosomal protein genes is associated with targeted recruitment of Esa1 histone acetylase. *Mol. Cell* **6**, 1297–1307 (2000).
15. Iyer, V.A. *et al*. Genomic binding distribution of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
16. Conrad, M.N., Wright, J.H., Wolf, A.J. & Zakian, V.A. Rap1 protein interacts with yeast telomeres in vivo: overproduction alters telomere structure and decreases chromosome stability. *Cell* **63**, 739–750 (1990).
17. Gotta, M. *et al*. The clustering of telomeres and colocalization with Rap1, Sir3, and Sir4 proteins in wild-type Saccharomyces cerevisiae. *J. Cell Biol.* **134**, 1349–1363 (1996).
18. 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).
19. Hecht, A., Strahl-Bolsinger, S. & Grunstein, M. Spreading of transcriptional repressor Sir3 from telomeric heterochromatin. *Nature* **383**, 92–96 (1996).
20. Pryde, F.E. & Louis, E.J. Limitations of silencing at native yeast telomeres. *EMBO J.* **18**, 2538–2550 (1999).
21. Donze, D., Adams, C.R., Rine, J. & Kamakaka, R.T. The boundaries of the silenced HMR domain in *Saccharomyces cerevisiae*. *Genes Dev.* **13**, 698–708 (1999).
22. Lascaris, R.F., Mager, W.H. & Planta, R.J. DNA-binding requirements of the yeast

protein Rap1p as selected in silico from ribosomal protein gene promoter sequences. *Bioinformatics* **15**, 267–277 (1999).
23. Costanzo, M.C. *et al*. The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73–76 (2000).
24. Remacle, J.E. & Holmberg, S. A REB1-binding site is required for GCN4-independent ILV1 basal level transcription and can be functionally replaced by an ABF1-binding site. *Mol. Cell. Biol.* **12**, 5516–5526 (1992).
25. Gonçalves, P.M. *et al*. C-terminal domains of general regulatory factors Abf1p and Rap1p in *Saccharomyces cerevisiae* display functional similarity. *Mol. Microbiol.* **19**, 535–543 (1996).
26. Cherry, J.M. *et al*. Saccharomyces Genome Database. 12/01/1999 edn Vol. 1999 (2000).
27. Mewes, H.W. *et al*. MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **27**, 44–48 (1999).
28. Lopez, N., Halladay, J., Walter, W. & Craig, E.A. SSB, encoding a ribosome-associated chaperone, is coordinately regulated with ribosomal protein genes. *J. Bacteriol.* **181**, 3136–3143 (1999).
29. Holstege, F.C. *et al*. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
30. Liu, X., Brutlag, D.L. & Liu, J.S. BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac. Symp. Biocomput.* 127–138 (2001).
31. Graham, I.R. & Chambers, A. Use of a selection technique to identify the diversity of binding sites for the yeast Rap1 transcription factor. *Nucleic Acids Res* **22**, 124–130 (1994).
32. Buchman, A.R., Kimmerly, W.J., Rine, J. & Kornberg, R.D. Two DNA-binding factors recognize specific sequences at silencers, upstream activating sequences, autonomously replicating sequences, and telomeres in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **8**, 210–225 (1988).
33. Idrissi, F.Z. & Piña, B. Functional divergence between the half-sites of the DNA-binding sequence for the yeast transcriptional regulator Rap1p. *Biochem. J.* **341**, 477–482 (1999).
34. Tjian, R. & Maniatis, T. Transcriptional activation: a complex puzzle with few easy pieces. *Cell* **77**, 5–8 (1994).
35. Leblanc, B.P., Benham, C.J. & Clark, D.J. An initiation element in the yeast CUP1 promoter is recognized by RNA polymerase II in the absence of TATA box-binding protein if the DNA is negatively supercoiled. *Proc. Natl. Acad. Sci. USA* **97**, 10745–10750 (2000).
36. Peterson, C.L. & Workman, J.L. Promoter targeting and chromatin remodeling by the SWI/SNF complex. *Curr. Opin. Genet. Dev.* **10**, 187–192 (2000).