2007

Calling cards for DNA-binding proteins

Haoyi Wang
Washington University School of Medicine in St. Louis

Mark Johnston
Washington University School of Medicine in St. Louis

Robi David Mitra
Washington University School of Medicine in St. Louis

Follow this and additional works at: http://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
Wang, Haoyi; Johnston, Mark; and Mitra, Robi David, 'Calling cards for DNA-binding proteins.' Genome Research. 17, 1202-1209. (2007).
hhttp://digitalcommons.wustl.edu/open_access_pubs/1984

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact engeszer@wustl.edu.
Calling cards for DNA-binding proteins

Haoyi Wang, Mark Johnston and Robi David Mitra

Genome Res. 2007 17: 1202-1209 originally published online July 10, 2007
Access the most recent version at doi:10.1101/gr.6510207
Methods

Calling cards for DNA-binding proteins

Haoyi Wang, Mark Johnston, and Robi David Mitra

Department of Genetics, Washington University, School of Medicine, Genome Sequencing Center, St. Louis, Missouri 63108, USA

Identifying genomic targets of transcription factors is fundamental for understanding transcriptional regulatory networks. Current technology enables identification of all targets of a single transcription factor, but there is no realistic way to achieve the converse: identification of all proteins that bind to a promoter of interest. We have developed a method that promises to fill this void. It employs the yeast retrotransposon Ty5, whose integrase interacts with the Sir4 protein. A DNA-binding protein fused to Sir4 directs insertion of Ty5 into the genome near where it binds; the Ty5 becomes a “calling card” the DNA-binding protein leaves behind in the genome. We constructed customized calling cards for seven transcription factors of yeast by including in each Ty5 a unique DNA sequence that serves as a “molecular bar code.” Ty5 transposition was induced in a population of yeast cells, each expressing a different transcription factor–Sir4 fusion and its matched, bar-coded Ty5, and the calling cards deposited into selected regions of the genome were identified, revealing the transcription factors that visited that region of the genome. In each region we analyzed, we found calling cards for only the proteins known to bind there: in the GAL1–10 promoter we found only calling cards for Gal4; in the HIS4 promoter we found only Gcn4 calling cards; in the PHO5 promoter we found only Pho4 and Pho2 calling cards. We discuss how Ty5 calling cards might be implemented for mapping all targets of all transcription factors in a single experiment.

[Supplemental material is available online at www.genome.org.]
To identify the regions of the genome into which Gal4DBD-Sir4 directed Ty5 insertion, we recovered the DNA immediately flanking the Ty5 and determined its nucleotide sequence. Genomic DNA from each His+ FOAr colony was cleaved with the restriction enzyme HincII, which cuts near the end of Ty5 and is ligated in dilute solution to favor their circularization and amplified by inverse PCR using primers complementary to the end of Ty5. The PCR products (of variable size) were labeled with Cy5 and used to probe a microarray of oligonucleotides that tile the yeast genome to identify regions of the genome flanking the Ty5 insertions (see Methods for details).

Seven regions known to be bound by Gal4 (GAL1-GAL10, GAL7, GAL2, GAL3, FUR4, GCY1, PCL10) (SCPD, http://rnlai.cshl.edu/SCPD; TRANSFAC, http://www.gene-regulation.com/pub/databases.html#transfac; Ren et al. 2000) are among the top 20 hybridization signals (see Methods for details of the analysis of the hybridization signals); two other known Gal4-regulated genes, MTH1 and GAL80, rank in the top 60 hybridization signals. (The data for all the genes that pass our significance criteria are provided in Supplemental Table 1).

Eight of the 13 promoters among the top 20 hybridization signals on the array that are not known to be Gal4 targets contain at least one Gal4-binding site (CCGGG,C,CCG) (Table 1). In an attempt to validate binding of Gal4 to these 13 promoters that are not known to be Gal4 targets, we immunoprecipitated Gal4 via the Myc epitope it carries and tested for coprecipitation of those regions of the genome. Three of the 13 promoters (SFL1-RUP1, YPL067C-YPL066W, YOR084W) were clearly enriched in the sample immunoprecipitated from cells with the Myc-tagged Gal4 compared to cells with an untagged Gal4 (Fig. 2A). Indeed, Gal4 regulates expression of these genes (Fig. 2B). Expression of the divergently transcribed genes flanking two of these promoters (SFL1-RUP1 and YPL067C-YPL066W) is induced by galactose via Gal4 (Fig. 2B, cf. lanes 3 and 1, and lanes 4 and 2); interestingly, Gal4 regulates expression of YOR084W in an unexpected way: It seems to repress its expression (cf. lanes 4 and 3, and lanes 6 and 5). Although 10 of the 13 potential Gal4 targets were not confirmed by the chromatin immunoprecipitation experiments, five of them have Gal4-binding sites and therefore could be Gal4 targets.

To estimate the sensitivity and specificity of the method, we turned to Gcn4, because it has a well-characterized DNA-binding specificity (Oliphant et al. 1989), many known targets in the genome (Natarajan et al. 2001; Pokholok et al. 2005), and many genes are known that are unlikely to be its target (Pokholok et al. 2005). In addition, Gcn4 was used to determine the specificity and the efficiency of the ChIP-chip method (Pokholok et al. 2005), enabling a direct comparison of the two methods. Using the same approach as for Gal4, we determined where in the genome Gcn4-Sir4 deposits Ty5. About 300 regions of the genome displayed significant hybridization to the array (see Methods for the criteria for significance). Twelve known Gcn4 targets are among the top 20 signals; the remaining eight all have perfect or recognizable Gcn4-binding sites (several of these genes are especially propitious Gcn4 targets because they encode enzymes involved in amino acid biosynthesis) (Table 1).

To estimate the specificity and sensitivity of this assay, we determined how many known Gcn4 target genes (defined by Pokholok et al. 2005) were not identified by our method (“false negatives”) and how many regions of the genome that are unlikely to be Gcn4 targets (also as defined by Pokholok et al. 2005) turned up in our assay (“false positives”). Fifty-one percent of the their genomic DNA was extracted and digested with three different restriction endonucleases with 4-bp recognition sequences that are present 300–1000 bp from the end of Ty5. The resulting fragments (containing Ty5 sequence on one end and the adjacent genomic sequence on the other end) were ligated in dilute solution to favor their circularization and amplified by inverse PCR using primers complementary to the end of Ty5. The PCR products (of variable size) were labeled with Cy5 and used to probe a microarray of oligonucleotides that tile the yeast genome to identify regions of the genome flanking the Ty5 insertions (see Methods for details).
known or likely Gcn4 target genes hybridized strongly enough to probes on the DNA microarray to pass our criteria for a positive signal. This false-negative frequency of 49% comes at a false-positive frequency of 2.5%. This is somewhat higher than the 25% false-negative frequency of the ChIP-chip method (at a false-positive frequency of 1%), which is perhaps not surprising given the results from ChiP-chip experiments (Fig. 4). In each of the three promoters we analyzed, we found calling cards for only those proteins known to bind to them (Fig. 4); in the \textit{GAL1-10} promoter, we only found Ty5 elements carrying the Gal4 bar code (Fig. 4A); in the \textit{HIS4} promoter, we found only Gcn4 bar codes (Fig. 4B) (Tice-Baldwin et al. 1989). In the \textit{PHOS} promoter, we found only bar codes corresponding to Pho4 and Pho2, and only when transposition was induced in cells starved for phosphate (Fig. 4C,D), as expected because Pho4 and Pho2 bind to DNA only when phosphate is scarce (Barbaric et al. 1996; Oshima et al. 1997).

This pilot experiment suggests that Ty5 can be used to identify proteins that bind to any region of the genome. The identity of the “bar codes” in these PCR products was determined by using them to probe a mini-array of the bar code sequences (Fig. 3, see Methods for details). The identity of the “bar codes” in these PCR products was determined by using them to probe a mini-array of the bar code sequences (Fig. 3, see Methods for details). The identity of the “bar codes” in these PCR products was determined by using them to probe a mini-array of the bar code sequences (Fig. 3, see Methods for details).

**Table 1. Top 20 targets of Gal4 and Gcn4**

<table>
<thead>
<tr>
<th>Target promoter</th>
<th>Known target?</th>
<th>Known by ChIP-chip?</th>
<th>Site present?</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{GAL1/GAL10}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{GAL7}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{GAL2}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{GAL3}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{FUR4}</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>\textit{PTR2/SRP40}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{GET3}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{SLF1/RUP1}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{YOH084W}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{CYC3}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{iYH0033W}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{YPL066W/O67C}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{YLR152C}</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>\textit{PUT1}</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>\textit{YCR061W}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{TS1}</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>\textit{GCY1/RYO1}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{NRM1}</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>\textit{GTO3}</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{KL10}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{GAL1}</td>
<td>No</td>
<td>No</td>
<td>Weak</td>
</tr>
<tr>
<td>\textit{ARG1}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{TRP1/SOK1}</td>
<td>No</td>
<td>No</td>
<td>Weak</td>
</tr>
<tr>
<td>\textit{AR63}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{CPA2/YMR1}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{LEU4/MET4}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{ILYS/YLR356W}</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{HIS5/PM5}</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{YLP086W-A}</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{IC12}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{ARG5,6}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{ASN1/NC4}</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{ARG4}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{LYS20}</td>
<td>No</td>
<td>Yes</td>
<td>Weak</td>
</tr>
<tr>
<td>\textit{CSH1}</td>
<td>No</td>
<td>No</td>
<td>Weak</td>
</tr>
<tr>
<td>\textit{SNO1/SN21}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{TEA1}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{IPT1/SNF1}</td>
<td>No</td>
<td>No</td>
<td>Weak</td>
</tr>
<tr>
<td>\textit{ARP1}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{HIS4}</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>\textit{PMP1}</td>
<td>No</td>
<td>No</td>
<td>Weak</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Known Gal4 targets as defined from three resources: TRANSFAC, SCPD, and Ren et al. 2000.
\textsuperscript{b}Known Gcn4 targets are as defined by Pokholok et al. 2005.
\textsuperscript{c}Known Gal4 targets as defined from three resources: TRANSFAC, SCPD, and Ren et al. 2000.
\textsuperscript{d}Known Gcn4 targets are as defined by Pokholok et al. 2005.
\textsuperscript{e}The consensus Gcn4 binding site is based on the weight matrix from TRANSFAC; a “weak” site is TGANTT.
reverse transcriptase produce no PCR products (data not shown). Total RNA was prepared, and RT-PCR was performed on the indicated targets. Control reactions lacking tin-associated proteins, because we used calling cards to identify cards can also be implemented for non-DNA-binding, chromatin-associated proteins, because we used calling cards to identify a known target of Mbt1, which is recruited to promoters of HXT genes by the Rgt1 transcriptional repressor (data not shown).

This method fills a gap in technology for characterizing DNA-binding proteins. Currently, we can identify the targets of any particular DNA-binding protein with the ChIP-chip technique, but to do the converse—identify the proteins that bind to a particular region of the genome—one would have to perform a ChIP-chip experiment on all DNA-binding proteins of an organism. Our calling card method promises to make this feasible.

Our method also provides an alternative to the ChIP-chip method for the genome-wide identification of targets of transcription factors, and can serve as an independent verification of the results obtained with the ChIP-chip method. Indeed, we were able to discover previously unidentified targets of Gal4, probably the best characterized transcription factor of yeast, perhaps because our method is very different from those that employ chromatin immunoprecipitation.

The calling card technology could be improved in several ways. Probably most important is to increase the number of transposition events sampled. For practical reasons we have been harvesting 3000–5000 independent transposition events in each experiment, but it should not be difficult to scale up the experiment and obtain more. This may be necessary because we did not find in the HIS4 promoter bar codes for Pho2 and Bas1, which are known to bind there (Tice-Baldwin et al. 1989). We identified two Pho2 bar codes among 18 that we analyzed by direct DNA sequencing in a preliminary experiment, suggesting that binding of these proteins would have been detected by hybridization to the microarray with a larger number of Ty5 transposition events. The number of transposition events could also be increased by improving the Ty5 transposition efficiency, which is relatively low compared with other Ty elements. This could also allow a shorter time of induction of transposition. Second, expression of the Ty5 calling card from the GAL1 promoter limits the conditions that can be tested. It would be better to use a different promoter, such as one that is activated by a gratuitous inducer such as tetracycline (Belli et al. 1998; Berens and Hillen 2003). Third, it has been speculated that the region of Sir4 that interacts with the Ty5 integrase also interacts with other proteins, which might interfere with the method in some cases. A clever solution to this potential problem—use of a heterologous pair of protein interaction domains on the DNA-binding protein and the integrase—was implemented by Zhu et al. (2003). That would also allow the method to be applied with a Sir4 strain, which would avoid the possibility of disruption of chromatin structure in certain regions of the genome. Fourth, fusing Sir4 to a DNA-binding protein could interfere with its ability to bind to DNA. This problem can be minimized by fusing Sir4 to each end of the protein (in different constructs). Finally, insertion of a calling card into a promoter could, in some cases, disrupt expression of the gene, which might prevent recovery of those cells. This problem can easily be solved by using a diploid strain.

We would like to reduce the false-positive and false-negative rates of our method. We empirically determined the significance cutoff using lists of genes that are likely or unlikely to be Gcn4 targets, as was done for the ChIP-chip method (Pokholok et al. 2005). This cutoff was applied to all experiments. We arbitrarily chose a significance cutoff that yielded 2.5% false positives, which results in a 49% false-negative rate. Similar performance (4% false positives and ~24% false negatives) was sufficient for application of the ChIP-chip method to genome-wide analysis of transcription factor targets in yeast (Harbison et al. 2004). Of course, the false-positive rate can be reduced by increasing the cutoff, but that comes at the expense of a higher false-negative rate. Advances in the experimental approach are likely to be necessary for significant improvement in the specificity and sensitivity of our method (Gabriel et al. 2006; Wheelan et al. 2006). One reason for this high false-positive rate might be the large number of cycles of the inverse PCR required to provide enough probe for hybridization to the DNA microarrays, which may result in over-amplification of some of the nonspecific insertions. Stochastic amplification of nonspecific insertions in the inverse PCR (“jackpotting”) could also contribute to the problem. Both problems should be ameliorated by performing the inverse PCR.
specific primers. Encoding a TF-Sir4 fusion and a plasmid carrying its matched Ty5 calling card. After transposition, the leaves behind when it visits a site in the genome. Each strain was cotransformed with a plasmid where both proteins bind and the adjacent genomic sequence, thereby revealing the false-positive rate will likely be improved if we can sample more transposition events. By coupling the calling card method to next-generation (massively parallel) sequencing technologies, it may be possible to identify genome-wide the binding locations of all yeast transcription factors in a single experiment. Induction of transpositions.

Methods

Strains and growth media

The sir4 deletion mutant yDV561 (MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, Gal+, psi2, Gal4-18-Myc) obtained from Dan Voytas (Zhu et al. 2003) was the host strain for Ty5 transposition. Chromatin immunoprecipitation was done from extracts of strain Z1319 (MATa, ade2-1, trp1-1, can1-100, leu2-3,112, his3-11,15, ura3, Gal+, psi2, Gal4-18-Myc) (Ren et al. 2000). Yeast strain BY4743 (MATa/MATA his3Δ1/His3Δ1 leu2Δ0/Leu2Δ0 ura3Δ0/Ura3Δ0 met15Δ0/Met15 LYS2/lys2Δ0) and homozygous gal4 deletion strain (Saccharomyces Genome Deletion Project, no. 31044) (MATa/MATa his3Δ1/His3Δ1 leu2Δ0/Leu2Δ0 ura3Δ0/Ura3Δ0 met15Δ0/Met15 LYS2/lys2Δ0) (Brachmann et al. 1998; Giaever et al. 2002) were used for reverse transcription PCR to measure gene expression. Yeast cells were grown in complete synthetic media with the addition of 2% glucose or galactose, unless described otherwise.

Construction of plasmids

To construct pBM5037 (Gal4DBD-Sir4-Myc), the region of SIR4 encoding amino acids 951–1200 was amplified in a PCR and fused to the Gal4 DNA-binding domain (amino acids 1–147 plus amino acids 877–881) in pOBD2 by “jackpotting” (Ma et al. 1987; Wach et al. 1994). Three copies of the Myc epitope were amplified using PCR and fused to the C terminus of Gal4DBD by gap repair. The entire ORF of each transcription factor was amplified in a PCR and used to replace Gal4DBD by homologous recombination. Gal4DBD-Sir4-Myc was linearized by cutting with XhoI (cuts once C-terminal to the Gal4DBD coding sequence) to serve as the recipient plasmid for gap repair to construct all the other TF-Sir4 fusions.

The plasmid pSZ293 with Ty5 expressed from the GAL1 promoter was obtained from Dan Voytas (Zhu et al. 2003). The XhoI–NotI fragment that includes GAL1::Ty5 was inserted between the XhoI and NotI sites of pRS316 (Sikorski and Hieter 1989) to generate pBM4735. Acal and FseI sites were engineered adjacent to the 3' long terminal repeat (LTR) to allow insertion of the 20-bp “bar codes.” The bar codes that identify each transcription factor were those developed for each gene in the Yeast Gene Knockout (YKO) collection (Yuan et al. 2005). Double-stranded oligonucleotides with the bar code sequences were inserted between the engineered Acal and FseI sites of the Ty5.

Induction of Ty5 transposition and inverse PCR

Since Ty5 is driven by the GAL1 promoter, transposition was induced by culturing cells in galactose medium for 2–3 d at room temperature. After induction, cells were plated on Glu – His medium to select for cells with transposition events. His+ cells were replica plated on – His, FOA-containing media to eliminate His+ colonies due to recombination of reverse-transcribed Ty5 with the transposon donor plasmid.

To map sites of Ty5 integration directed by Gal4-Sir4, 96 His+ FOA+ colonies were grown inYPD, and their genomic DNA was extracted and digested by HinP1I (1 µg in a 20 µL reaction). Five microliters of digested DNA was then ligated overnight at 15°C in 100 µL to encourage self-circularization. Five microliters of the ligated DNA was used as template for inverse PCR with

Figure 3. “Calling cards” for DNA-binding proteins. For each of seven transcription factors fused to Sir4 (Gal4, Gal80, Ste12, Bas1, Pho2, Gcn4, and Pho4), a unique 20-bp oligonucleotide was inserted into Ty5 to serve as a “molecular bar code,” thereby transforming Ty5 into a “calling card” that the TF leaves behind when it visits a site in the genome. Each strain was cotransformed with a plasmid encoding a TF-Sir4 fusion and a plasmid carrying its matched Ty5 calling card. After transposition, the calling cards deposited in the promoters of interest were recovered by a PCR with Ty5- and promoter-specific primers.

1206 Genome Research www.genome.org
sequences within Ty5. Six hundred nanograms of PCR products for each promoter were purified, labeled with Cy5, and hybridized to a mini-array of bar code oligonucleotides, using Genisphere’s Array 900DNA Cy3 and Cy5 labeling kits. Probes on the mini-array are 60-bp oligonucleotides consisting of three copies of the 20-bp bar code sequence. Each probe was printed in quadruplicate on the array. In addition, oligonucleotides of the LTR sequence were printed to serve as a positive control; three unrelated bar code oligonucleotides served as negative control.

Primers for PCR

OM6313: TAAAGCTCGGAATTCGACGTC
OM6188: ACAAGGAAAAACATAGACAGCC
OM6458: AGGGTTATGAGCCT
OM4960: CGTATGGAATTACGATCAGAC
OM6609: CTFTTGGGATTACCATACCAAC
OM6610: ATCGTATTACCTACG
OM6456: CCCATAACTGAAATAGCGATG
OM6606: AAGATCGGATGTCTATGCAG

DNA sequencing

The ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit was used for DNA sequencing. One hundred nanograms of PCR product or 1 µg of plasmid DNA was used as the template. The products of the reaction were separated and detected on an ABI 310 genetic analyzer.

Microarray analysis

We used two methods to identify the regions of the genome where calling cards were deposited due to the binding of the TF-Sir4 fusion protein. Each method requires a different type of hybridization control.

The Rosetta error model

We used the Rosetta error model to analyze the transcription factors Gal4 and Gcn4. In these experiments, our control was a sir4Δ strain containing a plasmid expressing Ty5 (pBM4735), but with no plasmid expressing a TF-Sir4 fusion. We induced transposition and performed inverse PCR as described above. We labeled the control reaction with the Cy3 (green) dye, the experimental reaction with the Cy5 (red) dye, pooled the reactions, hybridized them to the microarray, and imaged the slide. For each probe, we subtracted the intensity value observed in the control channel from the intensity value observed in the experimental channel. We then assigned each probe a $P$-value that gives the probability of the observed intensity difference, assuming no calling card was deposited at that location. As did Pokholok et al. (2005), we used the Rosetta error model to calculate this $P$-value. In this model, the difference in intensities between two technical replicates is assumed to be normally distributed, and the variance of this distribution increases with average probe intensity (see Supplemental Information for more details).

We chose our significance cutoff empirically by using the published test sets of positive and negative targets for Gcn4 (Pok-
holok et al. 2005). We selected a P-value threshold that mini-
mized the rate of false negatives at a false-positive rate of 2.5%. This
cutoff resulted in a false-negative rate of 55%. If a gene is
within 250 bp of a significant probe, then it is considered a target
of the transcription factor that is being analyzed.

The Maximum Likelihood Estimate of DNA Concentration
(MLEDC) method

The Rosetta error model works well when the distribution of
intensities in the control channel is similar to the distribution of
background intensities in the experimental channel. However,
we observed a significant increase in integration “hot spots”
when no TF-Sir4 fusion protein is present, rendering the Rosetta
error model inadequate. Therefore, we developed a second way to
analyze the calling card data. Using labeled genomic DNA as a
control, we estimated the concentration of DNA present at each
locus after recovery of calling cards and flanking genomic DNA
(see Supplemental Information). The maximum likelihood value
of DNA concentration is proportional to the average ratio of ex-
perimental to control intensities. We ranked the probes based on
their average ratio and empirically selected a cutoff as described
above. We selected a threshold that minimized the rate of false
negatives at a false-positive rate of 2.5%. This cutoff resulted in a
false-negative rate of 49%. Since this is slightly better than the
Rosetta error model, the data were analyzed using the MLEDC
method.

To understand better the nature of our false negatives, we
manually examined the intensities of these genes in the MLEDC
analysis—the majority of these features displayed little to no
fluorescence in the red channel, suggesting that these features
were categorized as negatives because no transposition event had
occurred in these samples, and not due to inaccurate assump-
tions in our error model. Data from probes covering telomere
regions were ignored (because Ty5 can insert into these regions of
the genome due to homologous recombination with Ty5 ele-
ments that reside there). HIS3 probes were also excluded because
HIS3 sequences from the Ty5 calling cards are present in the
inverse PCR product.

For the barcode array experiments, the raw intensity of each
probe on the array was normalized by dividing it by the raw
intensity of a probe containing LTR sequence. To eliminate the
random hopping background, we applied a stringent criteria: If
the probe gets a ratio >0.1 only in one experiment out of three
biological replicates, we count it as a random event and exclude
it from the data.

Chromatin IP

Chromatin immunoprecipitation was performed as previously
described (Aparicio 1999; Orlando 2000). Cultures were grown in
minimal medium with galactose. Bound proteins were cross-
linked to DNA in vivo by addition of formaldehyde, followed by
cell lysis and sonication to shear DNA. Individual transcription
factors were immunoprecipitated with antibody to their Myc epi-
tope tag, followed by reversal of the cross-links. DNA immuno-
precipitated from a Myc-tagged strain and from a control strain
with no Myc tag were used as template to amplify the promoter
of interest.

Reverse transcription PCR

Wild-type and Gal4 deletion strains were cultured in 50 mL of YP
medium with 2% glucose, 2% galactose plus 5% glycerol, or 5%
glycerol as carbon source. When the cultures reached an OD600 of
1.5, the cells were harvested and their RNA extracted. The same
amount of RNA from each sample was treated with DNase and
then reverse transcribed into cDNA using SuperScript II Reverse
Transcriptase from Invitrogen. The cDNA served as the template
in a PCR employing primers that amplify 200–300 bp of coding
sequence of the genes of interest. Twenty-five cycles were used
for each PCR. As a loading control, the ACT1 locus was amplified
by RT-PCR for each sample.

Acknowledgments

We thank Dan Voytas (Iowa State University) for generously pro-
viding reagents and advice, and Seth Crosby and Michael Heinz
(Washington University) for their expert assistance with the use
of DNA microarrays. We also thank Doug Chalker (Washington
University) for helpful suggestions, and Rick Young and Nancy
Hannett (Whitehead Institute, MIT) for providing strains ex-
pressing Myc-tagged Gal4 and Gcn4. This work was supported by
funds provided by the James S. McDonnell Foundation, and by
NIH grants R21RR023960 and S50HG003170-03.

References

Aparicio, O.M. 1999. Characterization of proteins bound to chromatin
by immunosuppression of yeast whole-cell extracts. In Current
Barbaric, S., Munsterkotter, M., Svaren, J., and Horz, W. 1996. The
homeodomain protein Pho2 and the basic-helix-loop-helix protein
Pho4 bind DNA cooperatively at the yeast PHO5 promoter. Nucleic
Acids Res. 24: 4479–4486.
activator/repressor dual system allows tight tetracycline-regulated
Constraints of resistance regulation in bacteria shape TetR for
Brachmann, C.B., Davies, A., Cost, G.J., Caputo, E., Li, J., Hieter, P., and
Boeke, J.D. 1998. Designer deletion strains derived from
Saccharomyces cerevisiae S288C: A useful set of strains and plasmids
for PCR-mediated gene disruption and other applications. Yeast
14: 115–132.
Gabriel, A., Dapprich, J., Kunkel, M., Gresham, D., Pratt, S.C., and
Genet. 2 e212. doi: 10.1371/journal.pgen.0020212.
Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S.,
Functional profiling of the Saccharomyces cerevisiae genome. Nature
Griffiths, A.D. and Tawfik, D.S. 2006. Miniaturising the laboratory in
Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J.,
Horak, C.E. and Snyder, M. 2002. ChiP-chip: A genomic approach for
to identify transcription factor binding sites. Methods Enzymol.
350: 469–483.
Ma, H., Kunes, S., Schatz, P.J., and Botstein, D. 1987. Plasmid
construction by homologous recombination in yeast. Gene
58: 201–216.
Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben,
Genome sequencing in microfabricated high-density picolitre
Natarajan, K., Meyer, M.R., Jackson, B.M., Slade, D., Roberts, C.,
shows that Gcn4p is a master regulator of gene expression during
specificity of DNA-binding proteins by selecting binding sites from
random-sequence oligonucleotides: Analysis of yeast GCN4 protein.


Received March 17, 2007; accepted in revised form June 11, 2007.