

Transcription factor binding site identification in yeast: a comparison of high-density oligonucleotide and PCR-based microarray platforms

Anthony R. Borneman · Zhengdong D. Zhang ·
Joel Rozowsky · Michael R. Seringhaus ·
Mark Gerstein · Michael Snyder

Received: 23 March 2007 / Revised: 18 May 2007 / Accepted: 20 May 2007
© Springer-Verlag 2007

Abstract In recent years, techniques have been developed to map transcription factor binding sites using chromatin immunoprecipitation combined with DNA microarrays (chIP chip). Initially, polymerase chain reaction (PCR)-based DNA arrays were used for the chIP chip procedure, however, high-density oligonucleotide (HDO) arrays, which allow for the production of thousands more features per array, have emerged as a competing array platform. To compare the two platforms, data from chIP chip analysis performed for three factors (Tec1, Ste12, and Sok2) using both HDO and PCR arrays under identical experimental conditions were compared. HDO arrays provided increased reproducibility and sensitivity, detecting approximately three times more binding events than the PCR arrays while also showing increased accuracy. The increased resolution provided by the HDO arrays also allowed for the identification of multiple binding peaks in close proximity and of

novel binding events such as binding within ORFs. The HDO array platform provides a far more robust array system by all measures than PCR-based arrays, all of which is directly attributable to the large number of probes available.

Keywords Tiling array · Chromatin immunoprecipitation · Microarray · Yeast

Introduction

With the availability of complete genome sequences, a number of different techniques have emerged to assist in the understanding of the relationship between DNA sequence and function on a genomic scale. Many of these methods involve the use of DNA microarray technology, in which thousands of DNA probes can be interrogated simultaneously (Schena et al. 1995). One application for DNA microarrays, which is rapidly gaining in popularity, is the chromatin immunoprecipitation microarray (chIP chip) protocol. Here, the *in vivo* interactions between DNA binding proteins and genomic DNA are assayed through immunoprecipitation of specific protein–DNA complexes from whole-cell lysates. This powerful technique therefore enables the entire downstream binding network for a particular transcription factor to be determined without prior knowledge of the binding specificity of that protein (Ren et al. 2000; Iyer et al. 2001).

ChIP chip was developed in yeast where the original microarrays consisted of spotted polymerase chain reaction (PCR) products representing all of the intergenic regions of *Saccharomyces cerevisiae* (Ren et al. 2000; Iyer et al. 2001; Horak et al. 2002). Unbiased PCR-based arrays, which contain both intergenic and open reading frame (ORF)

Electronic supplementary material The online version of this article (doi:10.1007/s10142-007-0054-7) contains supplementary material, which is available to authorized users.

A. R. Borneman · M. Snyder (✉)
Department of Molecular,
Cellular and Developmental Biology, Yale University,
New Haven, CT 06511, USA
e-mail: michael.snyder@yale.edu

Z. D. Zhang · J. Rozowsky · M. R. Seringhaus · M. Gerstein ·
M. Snyder
Department of Molecular Biophysics and Biochemistry,
Yale University,
New Haven, CT 06511, USA

Present address:

A. R. Borneman
The Australian Wine Research Institute,
Urrbrae,
Adelaide, SA 5064, Australia

DNA probes, have since been applied to the study of transcription factor binding by chIP chip in many systems including yeast (Lieb et al. 2001; Hahn et al. 2004) and mammalian cell lines (Martone et al. 2003; Euskirchen et al. 2004). Although inexpensive, the use of PCR products limits the effective resolution available for mapping factor binding to the size of the PCR product. For the arrays designed for yeast, this is generally determined by the size of the entire intergenic region for those regions less than 1 kb, or for larger sequences, the intergenic space is divided up into smaller products (Iyer et al. 2001; Lieb et al. 2001).

Recently, high-capacity oligonucleotide arrays have been developed, which are comprised of large numbers (40,000 to more than 6,000,000) of short oligonucleotides (25mers to 60mers) (Lipshutz et al. 1999; Hughes et al. 2001; Nuwaysir et al. 2002). These arrays have been shown to be useful for a number of microarray applications including mRNA expression, comparative genome hybridization, and chIP chip, as the large numbers of probe sequences allow entire chromosomes or even whole genomes to be analyzed (reviewed in Bertone et al. 2005).

While a number of studies have utilized high-density oligonucleotide (HDO) arrays for chIP chip (Cawley et al. 2004; Boyer et al. 2005; Pokholok et al. 2005; Bieda et al. 2006; Squazzo et al. 2006; Xu et al. 2006; Hayashi et al. 2007; Isogai et al. 2007; Zeitlinger et al. 2007; Zheng et al. 2007), there has been no comprehensive comparison made between the HDO and PCR array platforms, aside from a cursory comparison made using Gcn4 (Pokholok et al. 2005). We have sought to address this shortcoming by directly comparing the performance of the two array types. Due to the widespread use of *S. cerevisiae* for chIP chip experiments, yeast was chosen as the model system for this comparison (Horak et al. 2002; Buck and Lieb 2004). The transcription factors Tec1, Ste12, and Sok2 were selected for study as there are chIP chip results available for these three factors from identical strains grown under identical conditions using both PCR and HDO arrays (Borneman et al. 2006; Borneman et al. 2007). While the results from the two array studies show a majority of consistent binding targets, the HDO arrays displayed higher sensitivity, reproducibility, and resolving power, which is directly attributable to the increased number of probes afforded by this array platform.

Materials and methods

Yeast strains and growth conditions

All yeast strains and growth conditions were performed as described previously (Borneman et al. 2006).

Array design

Spanning the yeast genome Watson strand, 50mer oligonucleotides were tiled at 60 bp spacing (10 bp between adjacent oligonucleotides; revision 8/27/2004 obtained from SGD, <http://www.yeastgenome.org>) for a total of 202,618 oligonucleotides. In addition, 50mer oligonucleotides were designed at 120 bp spacing (70 bp between adjacent oligonucleotides) to span the Crick strand, 101,303 oligonucleotides. The start position for oligonucleotides on the Crick strand were generated with a 30 bp offset relative to those on the Watson strand. Arrays were synthesized on microarray slides using maskless photolithography by Nimblegen LLC (Iceland; Nuwaysir et al. 2002). The array design has been deposited in the Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>) under the platform accession number GPL3541.

PCR arrays consisted of PCR products representing all the non-ORF regions of *S. cerevisiae*. One PCR product was used to each intergenic region of less than 1 kb, while for those regions greater than 1 kb, the region was dissected into multiple parts until each part was less than 1 kb in length (Iyer et al. 2001).

Chromatin immunoprecipitation, DNA labeling, and array hybridizations

Chromatin immunoprecipitations and DNA labeling were performed as described previously (Borneman et al. 2006). Hybridizations were carried out according to manufacturer's protocols (Nimblegen LLC). ChIP chip data has been deposited in GEO under the series accession number GSE4450.

Data acquisition and analysis

After scanning, the two files corresponding to each channel (in .pair file format) were uploaded to the Telescope pipeline for high-density tiling array data analysis for normalization and scoring (Telescope; <http://www.tilescope.gersteinlab.org>; Zhang et al. 2007). Telescope first performs intra- and inter-slide scaling using quantile normalization, the results of which are then integrated using a sliding window approach, with a window of size 400 bp in genomic space used to integrate neighboring probes from replicate arrays. For each window centered at the genomic coordinate of each oligonucleotide probe, both the pseudo-median signal as well as a *p* value measuring the likelihood that the region was bound by the transcription factor were calculated. The pseudo-median was represented by the median of pair-wise averages of the \log_2 ratio of test to reference signals for all oligonucleotide probes within the window, while the *p*

value was calculated using a Wilcoxon paired signed rank test comparing test signal against reference signal for all oligonucleotide probes in the window. Regions corresponding to putative transcription factor binding sites were then determined using a threshold for both the signal and p value, as well as requiring the region have MaxGap less than or equal to 60 bp (the maximum gap allowed between probes above threshold) and MinRun greater than or equal to 120 bp (the minimum length region with probes above threshold). This is a modified version of the scoring scheme used in Cawley et al. (2004).

PCR verification

PCR verification of the chIP chip results were performed by designing oligonucleotides to twelve HDO-specific and either two or three PCR-specific regions for each factor in addition to three unenriched controls (Supplemental Table 4 online). Quantitative PCR (qPCR) was performed using a Applied Biosystems 7000 with these primer pairs and DNA from both tagged and untagged strains. Fold enrichment of tagged versus untagged signal was calculated for each primer pair with the average level of enrichment observed with the three control genes subtracted from this value to correct for small differences in the level of input DNA. Each qPCR value presented represents the average of three independent measurements of fold enrichment for each intergenic region.

Results

HDO arrays provide a high-resolution, unbiased view of transcription factor binding across the genome

PCR-based arrays have an effective minimum resolution that is determined by the size of the PCR products spotted on the array. With the recent availability of HDO arrays, we sought to compare their performance in terms of accuracy, sensitivity, and resolution relative to the PCR-based arrays for the analysis of chIP chip experiments in *S. cerevisiae*. High-quality data on the binding networks of six transcription factors including Ste12, Tec1, and Sok2, grown under nitrogen limitation conditions in diploid strains, were already available from PCR-based arrays (Borneman et al. 2006). Likewise, data from chIP chip studies of Ste12 and Tec1 using HDO arrays, but otherwise performed in the same strains and under identical conditions to those used in the PCR-based chIP chip experiments, are also available (Borneman et al. 2007). To complete a three-way comparison set, we therefore performed chIP chip on Sok2 using HDO arrays, again using the same strain and growth

conditions used previously, such that the results for chIP chip from three factors across two different array platforms could be directly compared to evaluate the performance of each platform.

The HDO arrays used for the Ste12 and Tec1 chIP chip analysis made use of a maskless photolithography array platform (Nimblegen; Nuwaysir et al. 2002) with oligonucleotides designed to tile each chromosome of *S. cerevisiae* such that a 50 bp oligonucleotide (50mer) was present every 60 bp on the Watson strand, while a 50mer was designed every 120 bp on the Crick Strand, for a total of 303,921 oligonucleotides. For the Sok2 chIP chip, this same HDO array design was used, with DNA immunoprecipitated from three independent cultures with an independent immunoprecipitation from an untagged strain used as a control for each experiment. DNA isolated from each tagged and untagged pair was competitively hybridized to separate arrays, thereby providing results from three biological replicates. In addition to using the identical strains and growth conditions as the previous PCR and HDO studies, after the isolation of the bound DNA, the same dye labeling protocols were also applied to prepare the DNA for microarray hybridization.

After array scanning, the raw intensity ratios obtained from the Sok2 HDO arrays, in addition to the raw results of the Ste12 and Tec1 arrays (Borneman et al. 2007), were normalized (quantile normalization) and the mean enrichment signals from the three biological replicates for each factor were plotted across the entire genome (chromosome IX is presented as an example in Fig. 1a). The high density of probes present on the HDO tiling arrays resulted in the appearance of recognizable peaks of increased tagged versus untagged signal in the chIP chip data, thereby providing a robust means for detecting regions of protein binding. Binding was determined for each factor using an automated analysis pipeline that was specifically designed for the analysis of microarray data from HDO tiling arrays (Telescope; <http://www.telescope.gersteinlab.org>; Zhang et al. 2007). This program identifies regions of high-tagged versus untagged ratios using a sliding window (400 bp) with Wilcoxon paired signed rank tests used to calculate the probability for this enrichment. A probability value cut off ($p \leq 0.0001$), in addition to a signal enrichment cut off (pseudo-median ≥ 1.25), were used to maximize the number of binding events observed for each factor while limiting the number of false peak calls. A total of 346, 337, and 385 independent binding events were identified for Tec1, Ste12, and Sok2, respectively (For a full list, see Supplemental Table 1).

In addition to providing a robust means by which to score regions of protein binding, the high density of probes provided by the HDO platform allowed for the detection of binding events that are impossible to distinguish using

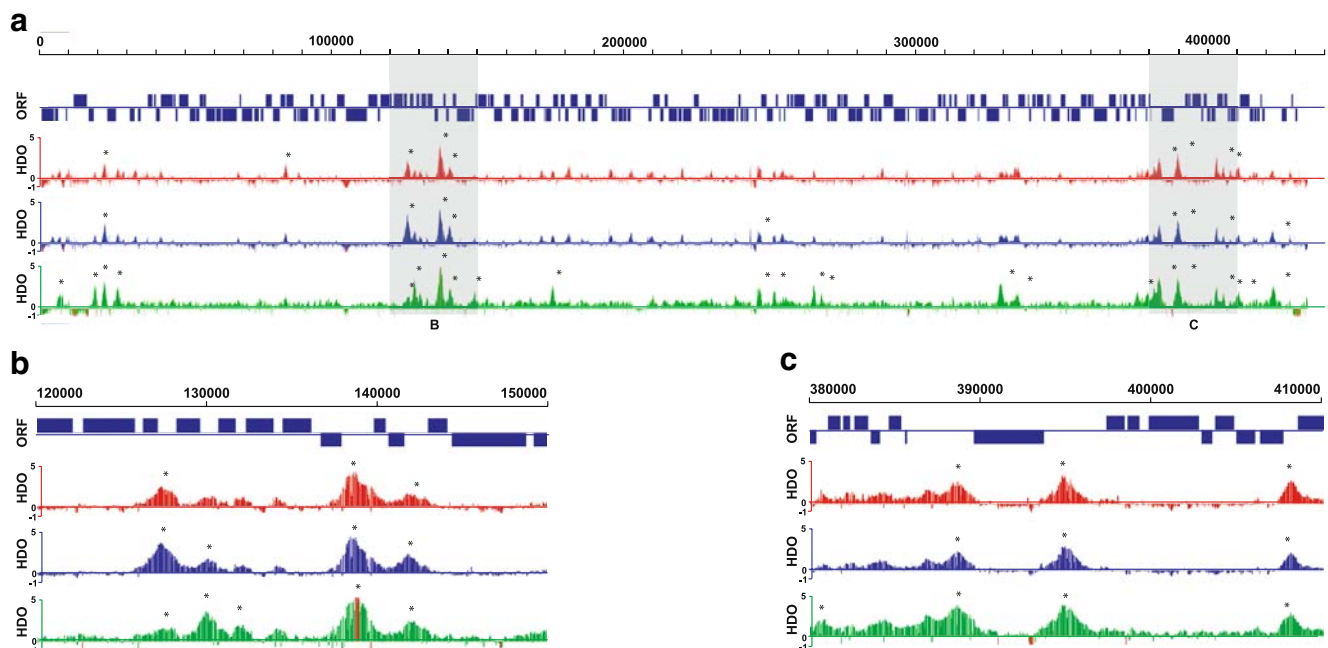


Fig. 1 High-resolution detection of protein–DNA interactions using HDO arrays. **a** A whole-chromosome view of the binding results for Tec1 (red), Ste12 (blue), and Sok2 (green) using HDO arrays across chromosome IX of *S. cerevisiae*. **b** An enlargement of the region

around YIL119W. **c** An enlargement of the region around YIR019C. The results are presented as the mean \log_2 ratio of tagged/untagged signals from three biological replicates with all peaks scored as being statistically enriched being indicated by asterisks

PCR-based arrays due to their lower resolution. These included the reproducible detection of multiple signal peaks (i.e., independent binding events) in single intergenic regions (Fig. 2a) and also of signal peaks occurring within coding regions (Fig. 2b), both of which take advantage of the increased resolution afforded by the HDO technology. Detection of coding region peaks was observed for a total of 17 ORFs (Supplemental Table 2) and has not been reported previously. For all of the ORFs listed, there is at least one line of evidence that they encode bona fide genes (SGD; <http://www.yeastgenome.org>) and thus are not simply dubious ORFs.

HDO arrays provide increased target reproducibility with fewer replicates

In addition to the resolution benefits that the HDO arrays were shown to offer, we wished to determine if these arrays provided more robust data than their PCR-based counterparts. Raw data was therefore independently processed and analyzed from the three replicate chIP experiments using Tec1, Ste12, and Sok2 and the HDO arrays. In addition, the raw chIP chip data from three biological replicates for Tec1, Ste12, and Sok2 published previously using the PCR arrays (Borneman et al. 2006) was reanalyzed, with each biological replicate also being scored independently. Replicate experiments from within each specific combination of transcription factor and platform were then compared to

observe the amount of reproducibility inherent in single array experiments (Fig. 3). Initially, Pearson correlation coefficient (r^2) values were calculated for each pair-wise combination of arrays within a factor and platform (Fig. 3a). The correlation values obtained did not show a significant difference between platforms, suggesting that on a probe-by-probe basis, neither platform was significantly more reproducible.

Correlation comparisons between entire array lists can often be overwhelmed by the large amount of data originating from features with insignificant signal strengths. Therefore, binding target lists were generated from each individual experiment [PCR, $SD \geq 4$ (Borneman et al. 2006); HDO, $p \leq 0.004$, pseudo-median signal ≥ 1.0 ; cut off values were relaxed due to the reduced statistical power of the single replicate experiments) thereby focusing only on the binding targets for the factors. Unlike the situation observed with Pearson correlations, comparison of target lists uncovered significant differences in reproducibility between the PCR and HDO platforms. Whereas the PCR-based arrays displayed an 11% overlap of targets in three out of three experiments, the majority (60%) of the targets from the HDO were shown to be present in all three replicate experiments. In addition, both array platforms were subsequently shown to display their best overlap between targets with the highest binding strengths. More than 95% of all of the overlapping hits from both platforms were located in the top 75% of the each list (Fig. 3b). This

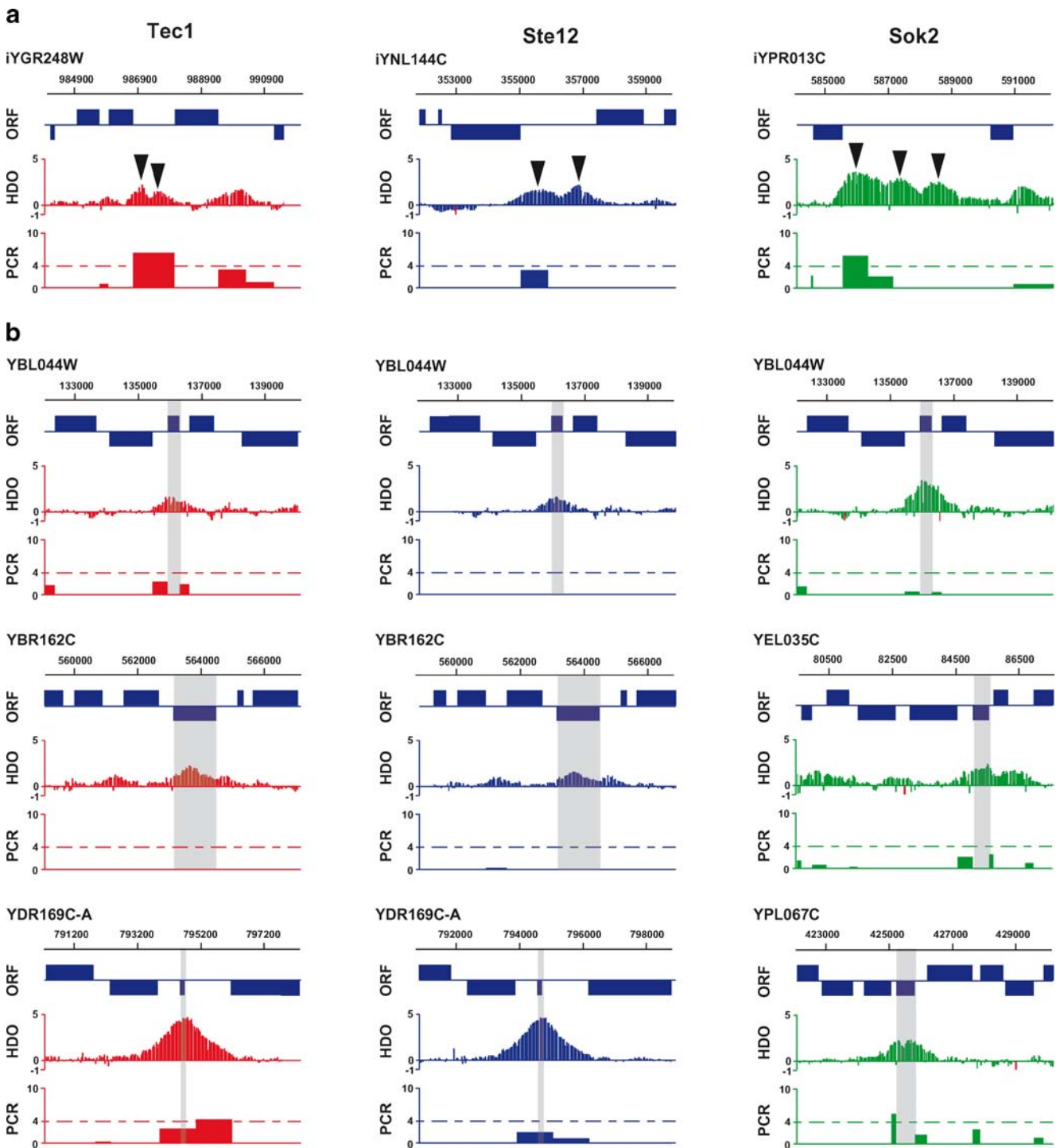


Fig. 2 HDO arrays allow for the detection of novel types of putative binding events. **a** Multiple signal peaks in a single intergenic region (black arrows). **b** Signal peaks within nondubious ORFs. The ORF

displaying the binding within its sequence is shaded (gray) to aid in its alignment with the chip signals presented below the ORF map

would suggest that for many lower scoring regions, the observed lack of reproducibility is due their proximity to the cut off, such that small variations in signal strength can result in identical regions from replicate experiments lying on opposing sides of the signal threshold.

These results clearly show that the HDO arrays provide greater reproducibility from fewer replicate experiments. This is most likely due to the large number of probes that are present in each unit of DNA length in the HDO platform. Unlike the PCR arrays, in which one PCR

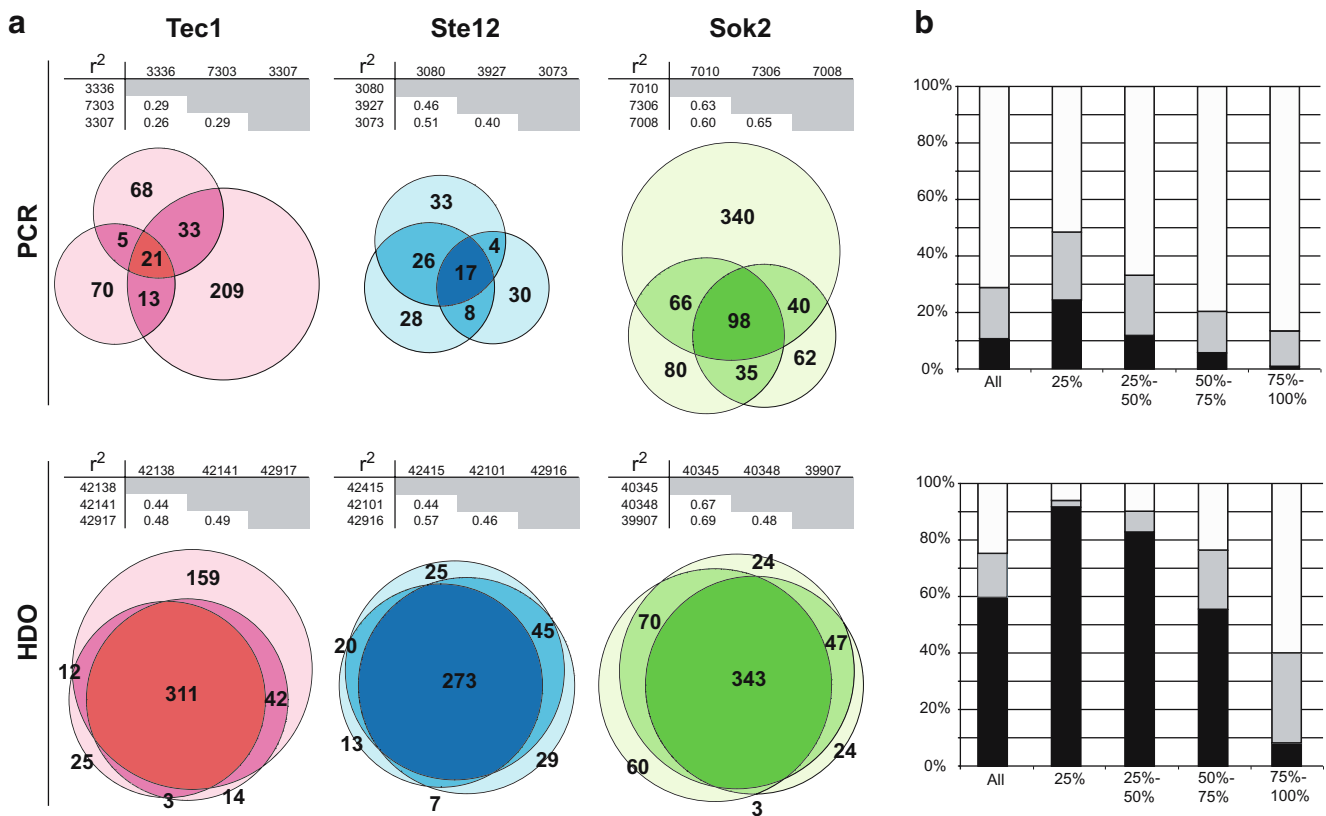


Fig. 3 The reproducibility of PCR and HDO arrays. **a** The results of three biological replicates for chIP chip experiments using either Tec1 (red), Ste12 (blue), or Sok2 (green), hybridized to either PCR or HDO arrays were analyzed independently and compared to determine the reproducibility of single-array experiments. Pearson correlation coefficients for each entire dataset are shown above Venn diagrams representing the overlap of hits scored for each individual experiment. **b** The proportion of targets present in either one out of three (white),

two out of three (gray), or three out of three (black) from the hit lists of the replicate experiments were determined as a function of their rank. Hits were called using either a standard deviation ($SD \geq 4$) cut off for the PCR-based arrays or using a p value ($p \leq 0.004$) for the HDO arrays. Each bin group was calculated from the rank order of the targets in descending order from the most significant hits (averaged across the three experiments) such that the 25% group represents the top 25% of hits and the 75–100% represents the bottom 25% of hits

product generally represents the entire region, the ability to calculate signal over a large number of probes in single experiment allows for accurate data to be extracted from a single HDO array. PCR arrays, on the other hand, require the statistical power provided by performing large numbers of replicates for robust data to be obtained, with some studies using up to ten independent biological replicate experiments for certain factors (Hahn et al. 2004).

PCR and HDO arrays identify similar but distinct downstream target genes

The most relevant method by which to assess the performance of the PCR and HDO arrays is to compare the targets identified by the two techniques. The bound intergenic regions determined by the tiling array analysis pipeline ($p \leq 0.0001$, pseudo-median signal ≥ 1.25) were therefore converted into gene lists by scoring ORFs which were 3' of and directly flanking the bound intergenic region

as being potential targets of that particular factor (see Supplemental Table 3 online). These lists were then compared to the binding lists determined for Tec1, Ste12, and Sok2 using PCR arrays (Borneman et al. 2006). Using five replicate experiments for each factor, the PCR arrays detected 143, 123, and 240 downstream targets for Tec1, Ste12, and Sok2, respectively, while the HDO arrays detected 430, 386, and 509 targets, respectively, from three replicate experiments. Calculating the union of these lists showed that, while the sizes were drastically different, the majority of the targets determined using the PCR arrays were also present in the HDO array lists (Fig. 4a), suggesting that the difference in the numbers of downstream targets may be due to sensitivity differences in the two platforms.

Interestingly, despite the large discrepancy in the sizes of the target lists between the PCR and HDO arrays, the combinatorial relationship between the transcription factors was still conserved (Fig. 4b). It had been shown previously

with the PCR-based arrays that Tec1 and Ste12 bound to a high proportion of common targets and that Sok2 overlapped with a smaller subset of these genes (Borneman et al. 2006). The HDO arrays also displayed this relationship, although there was a far greater level of common binding shown between all three factors, especially Tec1 and Ste12. Unlike the situation observed in the overlap between replicate array experiments, where much of the variation occurred in the bottom 25% of the hit lists, many of the differences between the factors, especially those observed for Sok2, were found in the top 50% of the lists, suggesting that these are real biological binding differences.

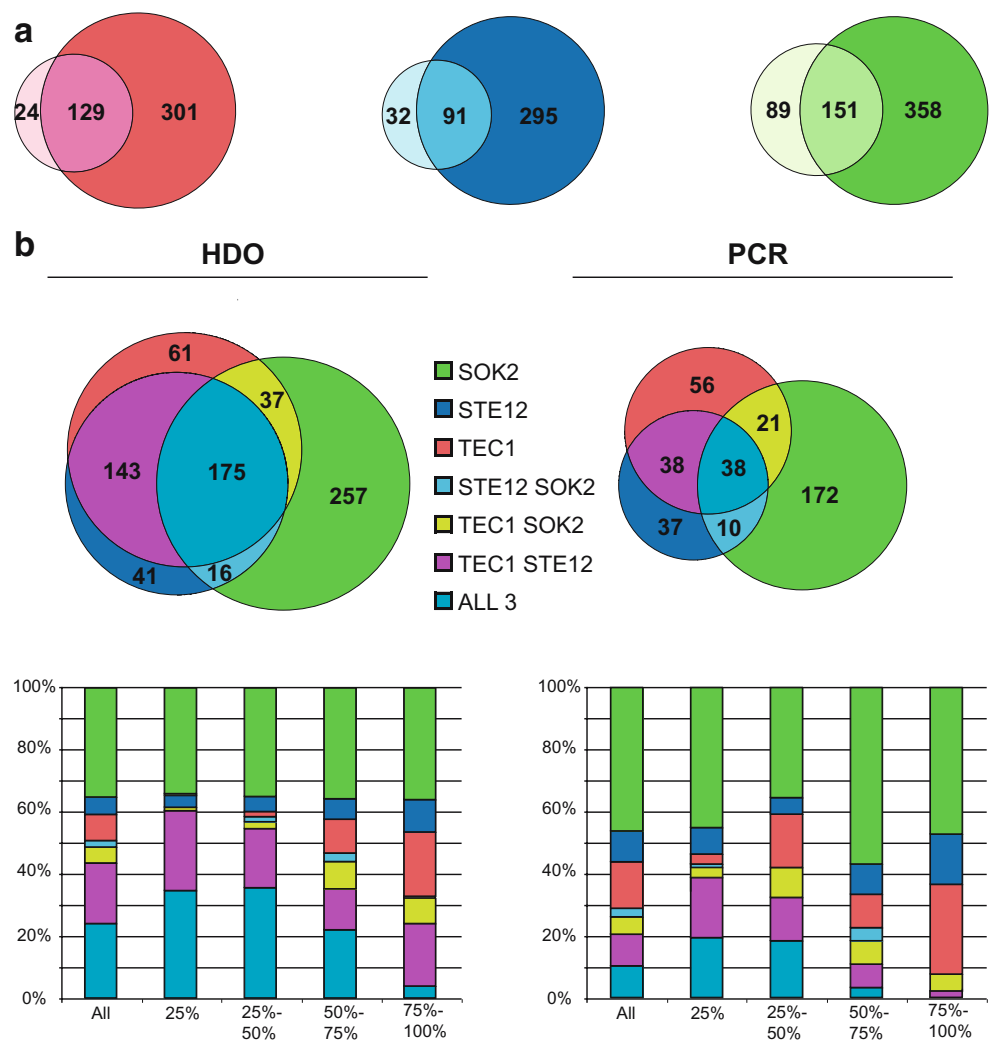
HDO arrays display higher sensitivity for detecting transcription factor binding

Given that the HDO lists contained far more targets than the PCR lists, we next sought to determine whether the overlap between the two lists was enriched for higher confidence

targets and also if the apparent overlap between the two types of arrays could be increased by simply lowering the cut off originally used for the PCR experiments to imitate the apparent increased sensitivity of the HDO arrays. The PCR results for each factor were therefore ranked by binding significance and compared to the results from the HDO arrays in a stepwise fashion proceeding from those with the highest confidence, with the total number of overlapping hits recorded with each step (Fig. 5a). For each factor, it was noted that, while the significance of the PCR hits was directly proportional to the amount of overlap, this was not a smooth transition, with major points of inflection apparent in the data. These inflection points are indicative of a major decrease in the level of congruence between the two platforms occurring after a certain signal strength, and interestingly, these points were shown to occur at or near the cut off used to score hits in the PCR data (Borneman et al. 2006).

The results of the target list comparisons showed that, while the two platforms possessed a majority of common

Fig. 4 Comparison of the targets determined by HDO and PCR arrays. **a** The overlap of target lists from chIP chip analysis of Tec1 (red), Ste12 (blue), and Sok2 (green) from HDO (darker color) and PCR arrays (lighter color). **b** Comparison of target lists for Tec1 (red), Ste12 (blue), and Sok2 (green) within platforms. The proportion of targets within each class for each platform is also shown as a function of their rank. Each bin group was calculated from the descending rank order of the targets, such that the 25% group represents the top 25% of hits and the 75–100% represents the bottom 25% of hits



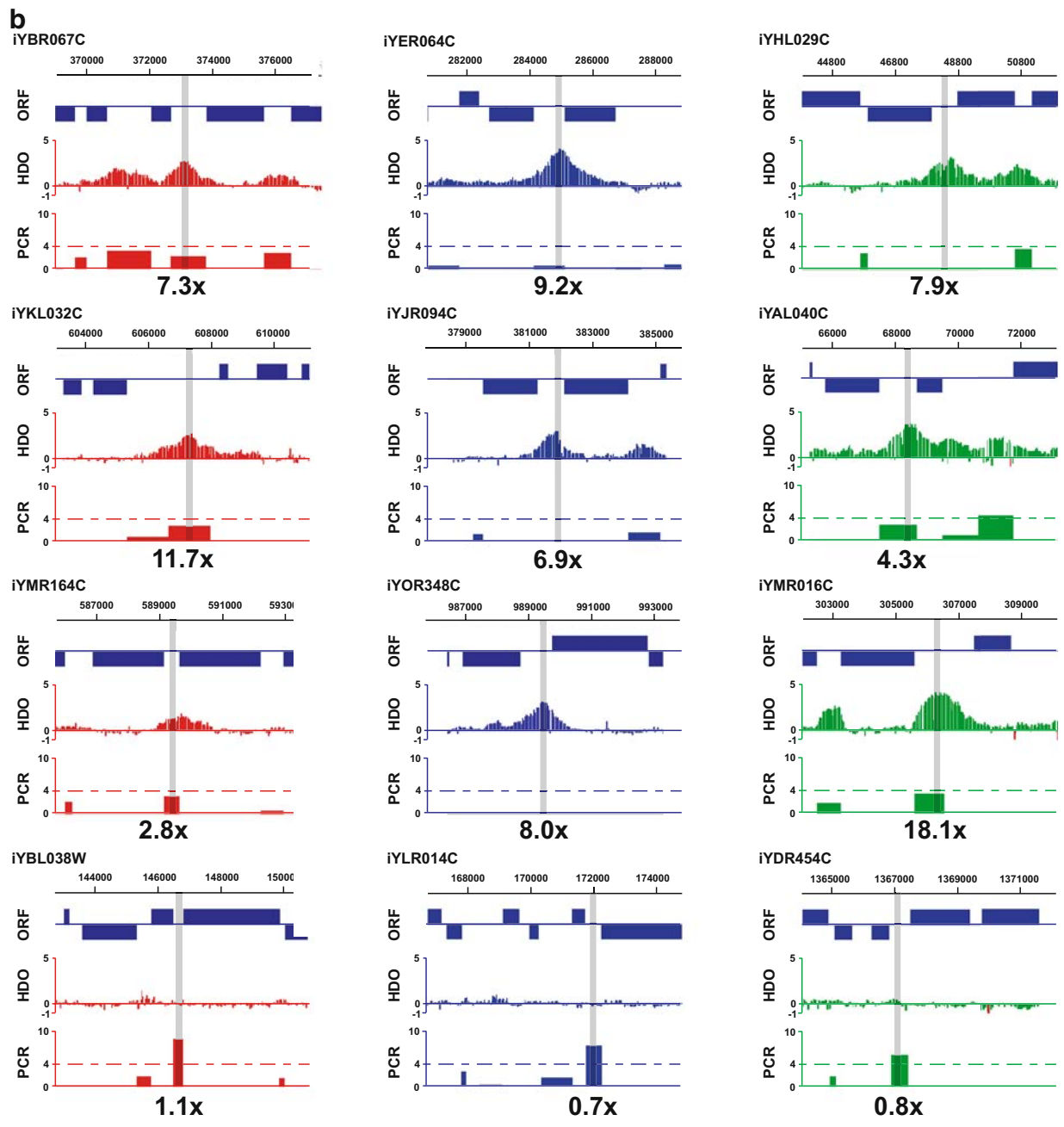
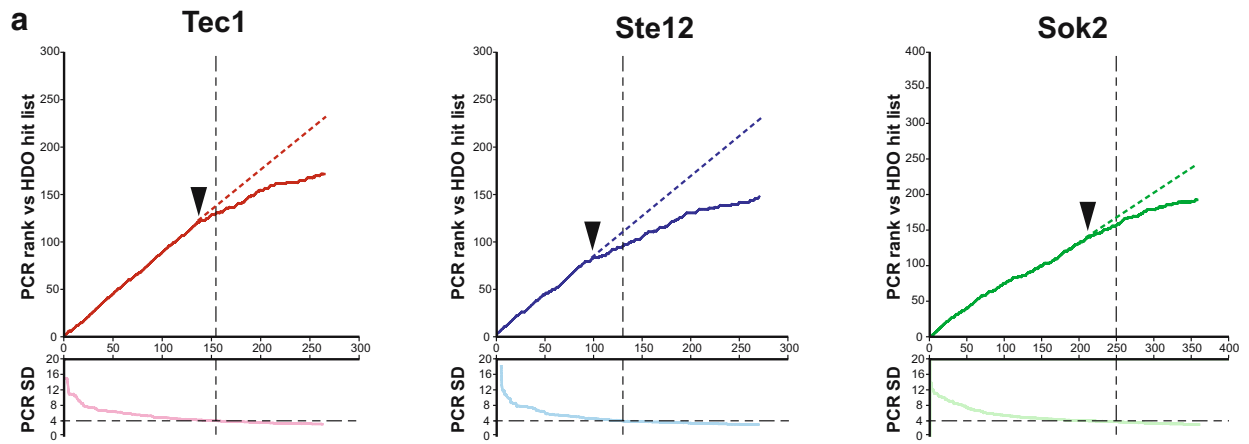


Fig. 5 HDO arrays allow for the detection of more binding targets than PCR arrays. **a** Potential gene targets were ranked by their significance of binding determined from chIP chip analysis of Tec1 (red), Ste12 (blue), or Sok2 (green) using PCR arrays. These ranked lists were compared to the HDO target lists of the same factor in a step-wise manner, with the number of common genes presented as line graphs. The significance of binding for each intergenic region is also presented in rank order, with the cut off used for the PCR experiments ($SD \geq 4$) indicated (black dotted line). The point at which the average level of overlap as a function of rank begins to significantly decrease is indicated by the black arrows, with the colored dotted line representing the expected path if no decrease in congruence was encountered. **b** Intergenic regions displaying either HDO-specific or PCR-specific enrichment by Tec1 (red), Ste12 (blue), and Sok2 (green) were analyzed by qPCR for confirmation of the observed effect. qPCR primers were designed to amplify the sequence shaded in gray from each intergenic region with the qPCR results presented below each graph (fold change tagged/untagged)

binding sites, there were still a considerable number of targets that were unique to either platform, with their numbers being inversely proportional to the strength of their binding. To test which of these two platforms were the most accurate with respect to these conflicted targets, qPCR was used as a third, independent measure of chIP enrichment. Unlike the HDO or PCR arrays, which rely on competitive hybridization to quantify differences in the relative amount of DNA present in two samples, qPCR uses differences in the total amount of DNA amplified by PCR to calculate relative amounts of starting DNA. Thirty-six intergenic regions (12 for each factor), which were shown to be targets by the HDO but not PCR arrays, and eight intergenic regions (two for Tec1, three for Ste12, and Sok2), which were called as targets by the PCR but not HDO arrays, were chosen for confirmation of enrichment by qPCR (Fig. 5b). All of the regions tested that were HDO but not PCR targets were shown to also display enrichment by qPCR (see Supplemental Table 4 online). This indicates that these regions represent actual binding targets of the factors and can therefore be classified as false negatives in the PCR data. For these regions, the level of enrichment calculated from the qPCR was also shown to generally be proportional to the level of signal detected in the HDO arrays, indicating that the enrichment values obtained by the HDO can be considered quantitative with respect to enrichment strength (see Supplemental Table 4 online).

Unlike the results observed for the HDO-specific regions, those targets that were specific to the PCR platform failed to show enrichment by qPCR suggesting that these represent false positives in the PCR datasets (see Supplemental Table 4 online). These results therefore indicate that the HDO arrays offer increased levels of accuracy over their PCR-based counterparts. In addition, while the PCR experiments do allow for the detection of the majority of binding sites for these factors, they suffer from the presence of both false negatives and of false positives, which

increase in number proportional to their proximity to the statistical cut off.

Discussion

Our studies indicate that HDO arrays offer many advantages over PCR-based arrays for the detection of protein binding by chIP chip. The comparison of chromatin immunoprecipitation data prepared for Tec1, Ste12, and Sok2 all showed the HDO arrays to exceed the sensitivity, resolution, and accuracy of PCR arrays, even when using fewer replicate experiments (three versus five). While these comparisons were made using biological rather than technical replicates between microarray platforms and the presence of variation due to experimental differences unique to each experiment cannot be absolutely discounted, accurate comparisons between the data sets can still be made due to a number of reasons. The use of multiple biological replicates for each experiment, which were themselves performed across several different weeks, combined with the high reproducibility of the biological replicates (apparent from the HDO data) suggests that there is far more variation between microarray platforms than is present in the individual immunoprecipitated DNA samples. This is further supported by work that has shown that chIP chip data obtained using these same factors and HDO arrays is highly reproducible (over 95% overlap of target genes) between experiments (each consisting of another three biological replicates) performed several months apart (Borneman et al. 2007). Array experiments that require replication after an extended period should still produce very similar results to the original experiments, the time at which the experiments were performed should not contribute significantly to the overall results. This is exactly what we observed with the HDO arrays.

The high-resolution and unbiased nature of the HDO arrays also allowed for the detection of binding events that would be extremely difficult using PCR arrays such as detecting multiple independent binding events in one intergenic region and also of binding within ORFs. The detection of binding within ORFs is particularly surprising as this has not been observed previously in *S. cerevisiae* and raises the possibility that these intra-ORF binding sites represent “internal” promoter/enhancer sequences. The ORF region in which the binding occurs may therefore act either as a promoter for its own expression or, more likely, for regulating the expression of flanking genes during conditions when the bound ORF is not expressed. Further experiments are required, however, to determine if these sequences have any effects on gene expression.

The increased sensitivity of the HDO arrays resulted in the detection of approximately three times more target

regions per factor than observed using the PCR arrays, indicating that each factor binds to and therefore has the possibility of regulating almost 5% of the total yeast genome. While this number is high, it is not biologically unfeasible. These results, in addition to those published previously (Borneman et al. 2006), have shown that there are a large number of common gene targets between the three factors, with only 730 distinct regions bound by the three proteins. This, combined with the drastic changes that occur in both metabolism and morphology during the pseudohyphal response that Tec1, Ste12, and Sok2 regulate, makes this a practical estimate of the number of possible targets (Gimeno et al. 1992; Liu et al. 1993; Ward et al. 1995; Gavrias et al. 1996). At this stage, however, it cannot be determined whether the weaker binding sites detected by the HDO arrays reflect a level of binding that would not mediate significant changes in gene expression. Nonetheless, the qPCR results demonstrated that these are real binding events regardless of the downstream effects and may contribute to additive or subtle effects on gene activity.

HDO arrays likely represent the future of microarray platforms. With current technology, they enable the use of at least an order of magnitude more features than traditional spotted PCR/oligonucleotide arrays and can be manufactured in a more controlled and reproducible fashion. As this technology matures, it will be possible to produce even more features on a single array. This will allow for tiling at even greater resolutions, down to individual base pairs for small genomes or regions, or for tiling extremely large sequences, such as the human genome at resolutions presently available for small genomes, greatly improving the quality of microarray data that can be produced for these systems.

Acknowledgments We would like to thank Elsa Eysteinsdottir and Chloe Lepar for expert technical assistance and Dan Gelperin for critical reading of this manuscript. This work was supported by Burroughs Wellcome and NIH grants to M.S. and M.G.

References

- Bertone P, Gerstein M, Snyder M (2005) Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. *Chromosome Res* 13:259–274
- Bieda M, Xu X, Singer MA, Green R, Farnham PJ (2006) Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome. *Genome Res* 16:595–605
- Borneman AR, Leigh-Bell JA, Yu H, Bertone P, Gerstein M, Snyder M (2006) Target hub proteins serve as master regulators of development in yeast. *Genes Dev* 20:435–448
- Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Sringhaus MR, Wang JY, Gerstein M, Snyder M (2007) Divergence of transcription factor binding sites across related yeast species. *Science*, in press
- Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122:947–956
- Buck MJ, Lieb JD (2004) ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* 83:349–360
- Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammanna H, Helt G, Struhl K, Gingeras TR (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116:499–509
- Euskirchen G, Royce TE, Bertone P, Martone R, Rinn JL, Nelson FK, Sayward F, Luscombe NM, Miller P, Gerstein M, Weissman S, Snyder M (2004) CREB binds to multiple loci on human chromosome 22. *Mol Cell Biol* 24:3804–3814
- Gavrias V, Andrianopoulos A, Gimeno CJ, Timberlake WE (1996) *Saccharomyces cerevisiae* *TEC1* is required for pseudohyphal growth. *Mol Microbiol* 19:1255–1263
- Gimeno CJ, Ljungdahl PO, Styles CA, Fink GR (1992) Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* 68:1077–1090
- Hahn JS, Hu Z, Thiele DJ, Iyer VR (2004) Genome-wide analysis of the biology of stress responses through heat shock transcription factor. *Mol Cell Biol* 24:5249–5256
- Hayashi M, Katou Y, Itoh T, Tazumi M, Yamada Y, Takahashi T, Nakagawa T, Shirahige K, Masukata H (2007) Genome-wide localization of pre-RC sites and identification of replication origins in fission yeast. *EMBO J* 26:1327–1339
- Horak CE, Luscombe NM, Qian J, Bertone P, Piccirillo S, Gerstein M, Snyder M (2002) Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev* 16:3017–3033
- Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR, Kobayashi S, Davis C, Dai H, He YD, Stephanians SB, Cavet G, Walker WL, West A, Coffey E, Shoemaker DD, Stoughton R, Blanchard AP, Friend SH, Linsley PS (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19:342–347
- Isogai Y, Takada S, Tjian R, Keles S (2007) Novel TRF1/BRF target genes revealed by genome-wide analysis of *Drosophila* Pol III transcription. *EMBO J* 26:79–89
- Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, Brown PO (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409:533–538
- Lieb JD, Liu X, Botstein D, Brown PO (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat Genet* 28:327–334
- Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ (1999) High density synthetic oligonucleotide arrays. *Nat Genet* 21:20–24
- Liu H, Styles CA, Fink GR (1993) Elements of the yeast pheromone response pathway required for filamentous growth of diploids. *Science* 262:1741–1744
- Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M (2003) Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc Natl Acad Sci USA* 100:12247–12252
- Nuwaysir EF, Huang W, Albert TJ, Singh J, Nuwaysir K, Pitas A, Richmond T, Gorski T, Berg JP, Ballin J, McCormick M, Norton J, Pollock T, Sumwalt T, Butcher L, Porter D, Molla M, Hall C, Blattner F, Sussman MR, Wallace RL, Cerrina F, Green RD (2002) Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. *Genome Res* 12:1749–1755

- Pokholok DK, Harbison CT, Levine S, Cole M, Hannett NM, Lee TI, Bell GW, Walker K, Rolfe PA, Herbolzheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA (2005) Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122:517–527
- Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP, Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* 290:2306–2309
- Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
- Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang SW, Margueron R, Reinberg D, Green R, Farnham PJ (2006) Suz12 binds to silenced regions of the genome in a cell-type-specific manner. *Genome Res* 16:890–900
- Ward MP, Gimeno CJ, Fink GR, Garrett S (1995) SOK2 may regulate cyclic AMP-dependent protein kinase-stimulated growth and pseudohyphal development by repressing transcription. *Mol Cell Biol* 15:6854–6863
- Xu W, Aparicio JG, Aparicio OM, Tavaré S (2006) Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in *S. cerevisiae*. *BMC Genomics* 7:276
- Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M (2007) Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. *Genes Dev* 21:385–390
- Zhang ZD, Rozowsky JS, Lam HYK, Snyder M, Gerstein MB (2007) Telescope: online analysis pipeline for high-density tiling microarray data. *Genome Biol*, in press
- Zheng Y, Josefowicz SZ, Kas A, Chu TT, Gavin MA, Rudensky AY (2007) Genome-wide analysis of Foxp3 target genes in developing and mature regulatory T cells. *Nature* 445:936–940