Classification of *Saccharomyces cerevisiae* promoter regions into distinct chromatin classes reveals the existence of nucleosome-depleted hotspots of transcription factor occupancy

Junbai Wang,[1,3] Lucas D. Ward[1], and Harmen J. Bussemaker[1,2]

[1]Department of Biological Sciences, Columbia University, New York, New York 10027, U.S.A.
[2]Center for Computational Biology and Bioinformatics, Columbia University, New York, New York 10032, U.S.A.
[3]Department of Pathology, The Norwegian Radium Hospital, Montebello 0310, Oslo, Norway.

Draft, March 19, 2007
Revised, Jan 18, 2008
Formatted as Letter for *Nature Genetics*

Abstract
Transcription factors (TF) play an essential role in the cell as locus- and condition-specific recruiters of transcriptional machinery or chromatin-modifying complexes. However, predicting the *in vivo* profile of TF occupancy along the genome, which depends on complex interactions with other chromatin-associated proteins, from the DNA sequence remains a major challenge. Through careful reanalysis of ChIP-chip data for 138 TFs obtained in rich media, we were able to classify the upstream promoter regions of *S. cerevisiae* into 15 distinct chromatin types. One of these encompasses 5% of all promoters and is unique in that it is highly occupied by (essentially) all TFs expressed in rich media. These "hotspots" of TF occupancy are strongly nucleosome-depleted and preferentially targeted by chromatin-remodeling complexes and the origin-of-replication complex (ORC). They are also the only chromatin type enriched for predicted Rap1p and Pdr1p binding sites, which we found to work cooperatively with AAA/TTT motifs, known to affect local DNA structure, to reduce nucleosome occupancy. Taken together, our results reveal and characterize a new type of local chromatin structure in yeast.

Introudction
Attempts to understand the interplay between chromatin structure and gene expression regulation have a long tradition. A classic distinction is that between (active) euchromatin and (silent) heterochromatin. More recently, however, it has become clear that this classification is in need of refinement, and that local chromatin state is being carefully modulated. The mechanisms controlling nucleosome positioning and histone modification are the focus of intense current experimental and computational analysis. While a few studies have addressed the relationship between chromatin state and TF binding [1,2] [Kurdistani, Pokholok, Shamir-natgenet, Lieb-natgenet], our mechanistic understanding of these regulatory processes is still very incomplete.

To investigate the possible variation in local chromatin structure among the yeast promoter regions, we carefully reanalyzed raw yeast genome-wide occupancy data in rich media condition [3]. An importance difference between the current study and the origin publication [3]is that we avoided mean center of ChiP enrichment across all TFs for each promoter and did not average over replicated experiments for the same TF. The resulting dataset contains a "chromatin signature" (fold-enrichment over 354 ChiP-chip experiments) for each of 4050 promoter regions.

Results
By using unsupervised machine learning algorithms[4], we performed unbiased classification of the 4050 promoter regions into 15 chromatin types based on their chromatin signature. Independently, the 354 Chip-chip experiments were classified into 7 distinct TF types according to their fold-enrichment signature over all promoters please refer to Figures S1 and S2. Figure 1 showing how random blocks of raw Chip-chip ratios for each of the 7 TF groups varies with 15 chromatin types, where most of chromatin types are only enriched by a couple of TF blocks. However, one of the chromatin types (CT15) was highly enriched for occupancy by TFs from all but one (FT1) of the TF types (~70% of all TFs enriched in CT15). For that reason, we first asked whether differences in protein expression level in the rich median conditions used by [3] could explain this. Figure S3 suggests that the cellular protein abundance for TFs of type FT1 is indeed significantly lower than for other factors. Therefore, we hypothesized that a unique and shared property of the CT15 promoter regions caused the occupancy by all TFs present in the nucleus to be high, when averaged over time and the cell population. Then we looked at

nucleosome density of each chromatin type because the nucleosomes are known to exert a major influence on TF occupancy such as the large size and the competition between nucleosomes and TFs for binding at the same DNA [5]. Using genome-wide nucleosome binding profiles recently measured by {Lee, 2007 #85, we tested whether promoters of type CT15 showed a significant difference in occupancy. In Figure 2, the CT15 promoters are indeed significantly depleted for nucleosomes as well as promoters of CT6, CT9, CT11 and CT12. Though CT11 shows a similar protein co-localization pattern as the CT15, other three nucleosome depleted chromatin types (CT6, CT9 and CT12) are only occupied by a small subset of TFs (see Figure 1). Consistent with the observed nucleosome depletion of the CT6, CT11, CT12 and CT15 promoters, we found that the genes they control have higher mRNA expression levels than the rest of the genome (Figure 2). In addition, we found that both CT11 and CT15 are the only two chromatin types are preferentially targeted by both components of the chromatin-remodeling complex and replication origin related proteins (Figure 2).

To further study the local nucleosome variations among 15 chromatin types, we collected predicted yeast nucleosome distribution on the DNA sequence from [6] and computed mean percent of no-nucleosome, fuzzy-nucleosome and nucleoseome occupancy within each of 15 chromatin types (see Table 1). From Table 1, we discovered that promoters of three (CT6, CT9 and CT12) nucleosome depleted chromatin types indeed have fewer nucleosomes (~23% to ~33% with nucleosomes but ~52% to ~62% no-nucleosomes) than the rest of promoters. Interestingly, nucleosome depleted CT11 (Figure 2) shows marginal nucleosome occupancy on the promoters (~35% with nucleosomes but ~45% no-nucleosome) but CT15 shows an equal distribution for both nucleosome (~40%) and no-nucleosome (~40%) occupancy on the promoters. Particularly, the percent of nucleosomes in CT15 is almost the same as CT2, CT5, CT7, CT8 and CT 10 (~41% to ~47% with nucleosomes but ~39% to ~45% no nucleosomes) but the later five chromatin types are not nucleosome depleted in Figure 2. Taken together, these results indicate that while the decreased occupancy by nucleosomes is the striking feature of CT15 – protein co-localization, alone does not sufficiently characterize chromatin type CT15. Especially, the nucleosome depletion of CT15 may inbreed an unknown mechanism that generates open chromatin within a relatively nucleosome dense region such as chromatin type CT15.

Differences in local chromatin structure must have their origin in differences in DNA sequence. We therefore performed an unbiased search for DNA sequence features that can explain how chromatin type dependents on promoter sequence. Firstly, using a library of position weight matrices [7] derived from the data of [3], we screened for differences in the distribution of predicted binding affinity over the promoter regions between the 15 different chromatin types. The results shown in SFigure 4 which indicate that various chromatin types are significantly enriched or depleted for binding sites of specific TFs. For example, Rap1p and Pdr1p are the most significant in a small group of TFs (Fhl1p, Sfp1p, Gcr1, Met31p and Rlr1p) whose binding sites are significantly over-represented in CT15, but not in any of the other chromatin types. We hypothesized that these proteins are plausible candidates for a mechanistic role in the observed nucleosome delpletion in the “hotspots” –CT15. Here we also noticed that there is a correlation between predicted TF binding affinity (SFigure 4) and predicted local nucleosome density (Table 1): for example, nucleosome dense chromatins (i.e. CT2, CT5, CT7, CT8 and CT10 in SFigure 4) contain many significantly enriched TF binding sites but nucleosome poor chromatins contain

over-represented binding sites for almost all of TFs such as CT6 in SFigure 4. Nevertheless, the majority of chromatin types only contain a small number of over-represented TF binding sties in the promoters such as CT1, CT3, CT4, CT11, CT12, CT13, CT14 and CT15 in SFigure 4. Then, by computing correlation coefficient between predicted sequence affinity scores and the mean raw Chip-chip binding affinity of each TF type, we tried to quantify the contribution of various sequence affinity profiles to the enrichment of TF binding affinity in 7 distinct TF types. The results are displayed in Figure 3 where the color scale represents T-values to the correlation coefficients. From this analysis, we found that several TFs' (i.e. Rap1p, Pdr1p, Rlr1 and Met31p see Figure 3) sequence affinity profiles showing significant positive correlation with the mean Chip-chip binding affinity of all 7 types chromatins. It is unsurprising that these TFs also showing strong positive correlation with the total mean of all Chip-chip occupancy datasets. In a similar analysis, by using a set of weight matrices from TRANSFAC, we identified the same pattern of strong positive correlation between the predicted sequence affinity (i.e. Mcm1p, Rap1p and Abf1p see SFigure 5) and the total mean Chip-chip ratios of all Chip-chip ratios. In other words, the sequence-specific binding of these proteins such as Rap1p and Pdr1p may play a key role in forming yeast "hotspots" and the corresponding nucleosome depletion. Finally, we tested for the overrepresentation of compact oligomer motifs and found that the motifs AAA amd TCA, and their reverse complements, are significantly overrepresented in CT15 promoter regions (see Table 2 and STable 1 for details). Unlike the binding sites of Rap1p and Pdr1p, however, these motifs are also enriched in other chromatin types such as CT2, CT4, CT5, CT8, CT10 and CT15. Interestingly, these chromatin types are all belong to nucleosome dense regions (i.e. percent of measured nucleosome calls in the promoters are between 41% and 47% see Table 1 for details).

In addition to the early analysis, we also tried to look for histone modification patterns that may relevant to our defined 15 chromatin types. By using the most recent measurements of genome-wide map of nucleosome acetylation and methylation in yeast [2], we performed T-test for each chromatin type against the rest of genome. The results T-values are represented by a color map in SFigure 6, where we sorting probe orders according to their position in a gene (i.e. upstream, start, middle and end of a gene). From this figure, we found that the global nucleosome occupancy such as H3 and H4 share a clear anti-correlation with the nucleosome acetylation and methylation statues. However, transcription factors such as Gcn4p, Gcn5p and Esa1p that control the histone modifications are positively correlated with nucleosome acetylation or methylation state of several chromatin types such as CT6, CT11, CT12 and CT15 in Table 3. Interestingly, HeK36Me3 showing a strong negative association with global nucleosome occupancy on chromatins such as CT6, CT11, CT12 and CT15 in SFigure 6 but positive correlation with the TFs occupancy on the corresponding chromatins see SFigures 7 and 8, respectively. To further verify this finding, we computed correlation coefficients between gene upstream H3K36Me3 profiles and the total mean of nucleosome density (measurements of both [6] and [8]) or the total mean of raw Chip-chip occupancy data (354 raw Chip-chip experiments by [3]), respectively. These results are showing in SFigures 9 and 10, which also suggests a strong positive correlation between He36Me3 and the total mean of Chip-chip occupancy data (p=2,8e-07) but a clear negative association with the nucleosome density (p=6.3e-144). Thus, yeast protein co-localization may contribute positively to the local histone modification but effect negatively to the local nucleosome depletion such as we detected a strong anti-correlation between the sum of

354 raw Chip-chip binding ratios [3] and the total mean of all available nucleosome densities [6 8] (p=1.6e-5 see SFigure 11 for details).

In the last test, we asked whether the genome-wide protein co-localization and the correlation between Chip-chip data and predicted sequence affinities for TFs exist in mean centered Chip-chip measurements [3]. We first classified the mean centered Chip-chip ratios (204 TFs with 6138 yeast probed promoter regions), the data as provided by [3], into 7 clusters. The results are shown in a 2-D color map SFigure 12, which does not reveal genome-wide protein co-localization in any small subset of promoter regions as we found in the early classification of raw Chip-chip ratios SFigure 1. Then we computed correlation coefficient between predicted sequence affinities for Rap1p (Pdr1p), the two most significant TFs that correlated with raw Chip-chip ratios in Figure 3, and the raw Chip-chip ratios or the mean centered Chip-chip ratios. For Rap1p, a scatter plot of above two types of correlation coefficients is shown in Figure 4 where we can see that a significant genome-wide correlation between Chip-chip data and the predicted sequence affinity ($p<1e-4$) is absent when a normalization across TF is performed for each promoter region as was done by [3]. For Pdr1p, there is a similar effect of normalization across Chip-chip experiments see SFigure 13 details.

Discussion and Conclusion

In this work, we have classified yeast promoter regions into 15 distinct chromatin types Figure 1 according to the raw genome-wide Chip-chip occupancy data from [3]. Additionally, for each chromatin type, we carefully studied the genomic properties of their promoter regions such as mRNA abundance, nucleosome density, binding occurrence of nucleosome-remodeling complex and replication original proteins and enrichment of predicted sequence affinities see details in Figures 2 and 3 and Tables 1 and 2. Based on these results, we found that these 15 yeast chromatin types can be further grouped into 5 classes which with various sequence structure properties and genomic regulation mechanisms please see details in Table 3: for example, 1) Chromatins CT6, CT9 and CT12 are all nucleosomes depleted with both low Chip-chip nucleosome density and small percent of predicted nucleosome occupancy on the DNA sequences, as well as no overrepresented TCA/AAA motifs in the promoters. In addition, CT6 and CT9 contain a large number of predicted TFs binding sites in the promoters and highly occupied by nucleosome remolding complex, histone modification related controllers such as H3k36Me3 and Gcn5p/Gcn4p/Esa1p. Therefore, the first class chromatin (CT6, CT9 and CT12) may represent **the constantly active regions – active euchromatin** in the yeast genome which with large open chromatin structure and abundant protein binding sites in the promoters that may easily be accessed by various TFs for activating or repressing other transcriptional regulations. 2) The second class chromatin includes CT1, CT3, CT13 and CT14 which may belong to the typical silent regions in the yeast genome because these chromatins have condensed nucleosome packing, small number of predicted protein binding sites on the DNA sequences and low transcriptome levels. However, the promoters of these chromatins are highly occupied by nucleosome remolding complex. So we cannot exclude a possibility that these chromatins may be controlled by an unknown process through the nucleosome remolding complex to modify the local chromatin structure in order to participating gene regulations. So the second class chromatin may be called **pseudo-heterochromatin controlled by nucleosome remolding complex.** 3) Chromatins CT4, CT5 CT7 and CT8 are the third class chromatin which have similar high nucleosome density in the promoters as the second one. Nevertheless, there is a

significant difference between the second and the third class chromatin in the DNA sequence level -- the later one contains overrepresented TCA/AAA motifs and abundant protein binding sites in the promoters. The TCA motifs is a partial protein binding site for Abf1p which is known for its active role in opening local chromatin structure [9,10]. Particularly, the poly (dA.dT) sequences exist as rigid DNA structures and contribute also to unwinding the local DNA double structure because it makes DNA sequence hardly to fold on the histone [11]. Therefore, we hypothesize that the third class chromatin may modify the local chromatin structure without recruiting the nucleosome remolding complex but by following collaborative competition between histone and multiple TFs to dissociate histone from the DNA sequence [5] which is also evidenced by low occupancy of nucleosome remolding complex in the third class chromatin. Thus, this class may be called **pseudo-heterochromatin modified by collaborative competition.** 4) Chromatins CT11 and CT15 are the fourth class chromatin, which are the only two chromatin types that show Chip-chip enrichment for almost all TFs in the yeast rich media condition. Though the forth class chromatin has the same high percent of predicted nucleosome call in the promoters as the early two classes, its measured Chip-chip nucleosome density is extremely low when compared with the rest of chromatins. Additionally, there are high levels of origin replication protein, nucleosome remolding complex, transcriptome and histone modification related regulators (i.e. Gcn4p/Gcn5p/Esa1p and Hek36Me3) in the promoters of the forth class chromatin. This leads us to feel the forth class chromatin may be the most interesting chromatin class because it actively participates many kinds of regulations with a highly compact chromatin structure. Particularly, chromatin CT15 – nucleosome-depleted yeast hotspots of transcription factor occupancy -- contains very few overrepresented protein binding sites in the promoters such as Rap1p and Pdr1p but these binding motifs are not enriched in any other chromatin types. And Rap1p is well known for its function in opening chromatin [12], so we suspect that the forth class chromatin, which with a compact chromatin structure and rigid DNA sequence structures, may frequently open or close the chromatin by a combinatorial regulation among few essential TFs (i.e. Rap1p), replication origin protein and nucleosome remolding complex for achieving its active role in the genome such as high transcriptome levels. We may call such class chromatin as **a dynamical or active heterochromatin.** 5) The fifth class chromatin includes CT2 and CT10 which has the same DNA sequence structure (i.e. enriched AAA/TCA motifs) and sequence features (i.e. large number of predicted protein binding sites in the promoters) as the third class chromatin. However, for the last chromatin class, all tested genome regulations are very low except for high nucleosome density and high origin replication complex occupancy in the promoters. For that reason, we may say this class chromatin is **the most silent heterochromatin** in the yeast genome because it lacks of the majority of genomic activities.

Overall, our present results implicate the classic distinction of chromatin structure – euchromatin and heterochromatin -- may be refined into five distinct classes (active euchromatin, pseudo-heterochromatin controlled by nucleosome remolding complex, pseudo-heterochromatin modified by collaborative competition, active (dynamic) heterochromatin and silent heterochromatin) according to the local DNA sequence property and different modification roles for the regions. Particularly, the classic heterochromatin has more heterogeneous modification mechanisms than the euchromatin, which may reflect the nucleosomes are highly dynamic in response to the process of transcription. One of the most interesting new classes of heterochromatin is the active (dynamic) heterochromatin – nucleosome-depleted hotspots of

transcription factor occupancy – which may share a similar functional role as the active euchromatin in genomic regulations but with a highly compact chromatin structure. Below is a possible module for such dynamic modification of condensed chromatin structure – CT15-- in yeast: following a recent report that nucleosomal DNA under going spontaneous unwrapping and rewrapping which allows any protein rapid access even to buried stretches of the DNA, then other factors such as remolding factors can be recruited to particular nucleosomes on a biological relevant timescale [13], we may understand the current finding in chromatin type CT15 -- a couple of master regulators (i.e. Rap1 and Abf1p) can get access to their DNA binding sits which buried inside nucleosomse. And with nearby structurally rigid DNA sequences (dA:dT) that increase the occupancy of a neighboring transcription factor binding site [11,14,15], other sequence-specific or non-sequence specific binding (yeast hotpots) creates a region of open chromatin near its binding sites [10,12,16,17]. Additionally, these general regulators such as Abf1p are associated with replication [10,16]and involving the organizing the chromatin structure of the yeast ARS1 region by reducing the number of nucleosomal positions [9] and activating transcription [18]. Thus, a disassociation of these essential proteins (i.e. Rap1p and Abf1p) from their binding sites may quickly result an opposite process – for example, binging of origin replication complex then recruiting other factors such as Rpd3p and Sin3p to generate silent heterochromatin again [19-23]. These dynamic changes in chromatin structure also provide the underlying mechanism to link transcription and replication [24,25]. Interestingly, we also found a similar mechanism in the fly hotspots [26], where binding motif of general regulator GAGA factor is highly enriched in protein co-localized fly hotspots and binding of the GAGA transcription factor on existing nucleosomes leads nucleosome disruption and chromatin remolding [27]. Additionally, origin recognition complex is enriched in the fly hotspots too which is known for forming silent heterochromatin in higher eukaryotes [28]. Thus, the new class chromatin – dynamic (active) heterochromatin may play a pivotal role in the genome regulation as well as the active euchromatin.

Materials and Methods

CHIP-chip dataset: We downloaded raw CHIP-chip datasets of Harbison et al [3] from the ArrayExpress[29] database (accession number E-WMIT-10), where we only found 354 raw GPR files with the right format that can be used for further data analysis. Pre-processing of raw GPR file and missing value imputation for all experimental datasets were performed by the MArray[30] and the LSimput[31] software packages, respectively. The final dataset contains 4050 yeast intergenic probes with less than 10 percent of missing values across 354 CHIP-chip experiments (138 yeast transcription factors). High level of microarray data analysis was performed in the MATLAB by using the SOM toolbox[4].

Weight matrix of yeast transcription factors: We collected 42 weight matrixes for yeast transcription factors from the TRANSFAC[32] database. Another 125 yeast transcription factor weight matrixes were obtained from the publication of Maclsaac et al[7]. Prediction of transcription factor affinity profiles and DNA sequences analysis (i.e. oligo–analysis) were carried out by either the MATRIXReduce[33] software or the RSA[34] tools.

Other datasets: For a whole yeast genome nucleosome density measurement, we used publicly available datasets from Lee et al[35] (H3 and H4) and Bernstein et al[8] (H2b and H3). Absolute transcriptome expression level in yeast was taken from the publication of Holstege et al [36].

Experimental measurements for genome-wide distribution of ORC and MCM proteins in the yeast were taken from the paper by Wyrick et al [37]. Protein expression data in the yeast was collected from the publication of Ghaemmaghami et al[38]. Experimental measurements of genome-wide location and regulated recruitment of the nucleosome-remodeling complex in the yeast were obtained from paper by Ng et al[39]. To evaluate the enrichment of various measurements in the yeast genome, we computed T-values for each type of measurement by using the T-test. Such statistical test was performed in the MATLAB and the T-values for all tests were clustered and visualized by the Clustering [40] and the Java TreeView[41] software, respectively.

References

1. **Kurdistani, S.K., Tavazoie, S. & Grunstein, M. Mapping global histone acetylation patterns to gene expression. *Cell* 117, 721-33 (2004).**
2. **Pokholok, D.K. et al. Genome-wide map of nucleosome acetylation and methylation in yeast. *Cell* 122, 517-27 (2005).**
3. **Harbison, C.T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* 431, 99-104 (2004).**
4. **Wang, J., Delabie, J., Aasheim, H., Smeland, E. & Myklebost, O. Clustering of the SOM easily reveals distinct gene expression patterns: results of a reanalysis of lymphoma study. *BMC Bioinformatics* 3, 36 (2002).**
5. **Miller, J.A. & Widom, J. Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol* 23, 1623-32 (2003).**
6. **Lee, W. et al. A high-resolution atlas of nucleosome occupancy in yeast. *Nat Genet* 39, 1235-44 (2007).**
7. **MacIsaac, K.D. et al. An improved map of conserved regulatory sites for Saccharomyces cerevisiae. *BMC Bioinformatics* 7, 113 (2006).**
8. **Bernstein, B.E., Liu, C.L., Humphrey, E.L., Perlstein, E.O. & Schreiber, S.L. Global nucleosome occupancy in yeast. *Genome Biol* 5, R62 (2004).**
9. **Venditti, P., Costanzo, G., Negri, R. & Camilloni, G. ABFI contributes to the chromatin organization of Saccharomyces cerevisiae ARS1 B-domain. *Biochim Biophys Acta* 1219, 677-89 (1994).**
10. **Yarragudi, A., Miyake, T., Li, R. & Morse, R.H. Comparison of ABF1 and RAP1 in chromatin opening and transactivator potentiation in the budding yeast Saccharomyces cerevisiae. *Mol Cell Biol* 24, 9152-64 (2004).**
11. **Suter, B., Schnappauf, G. & Thoma, F. Poly(dA.dT) sequences exist as rigid DNA structures in nucleosome-free yeast promoters in vivo. *Nucleic Acids Res* 28, 4083-9 (2000).**
12. **Morse, R.H. RAP, RAP, open up! New wrinkles for RAP1 in yeast. *Trends Genet* 16, 51-3 (2000).**
13. **Li, G., Levitus, M., Bustamante, C. & Widom, J. Rapid spontaneous accessibility of nucleosomal DNA. *Nat Struct Mol Biol* 12, 46-53 (2005).**
14. **Kornberg, R.D. & Lorch, Y. Chromatin rules. *Nat Struct Mol Biol* 14, 986-8 (2007).**
15. **Peckham, H.E. et al. Nucleosome positioning signals in genomic DNA. *Genome Res* 17, 1170-7 (2007).**

16. **Yarragudi, A., Parfrey, L.W. & Morse, R.H. Genome-wide analysis of transcriptional dependence and probable target sites for Abf1 and Rap1 in Saccharomyces cerevisiae. *Nucleic Acids Res* (2006).**
17. **Lascaris, R.F., Groot, E., Hoen, P.B., Mager, W.H. & Planta, R.J. Different roles for abf1p and a T-rich promoter element in nucleosome organization of the yeast RPS28A gene. *Nucleic Acids Res* 28, 1390-6 (2000).**
18. **Buchman, A.R. & Kornberg, R.D. A yeast ARS-binding protein activates transcription synergistically in combination with other weak activating factors. *Mol Cell Biol* 10, 887-97 (1990).**
19. **Aparicio, J.G., Viggiani, C.J., Gibson, D.G. & Aparicio, O.M. The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in Saccharomyces cerevisiae. *Mol Cell Biol* 24, 4769-80 (2004).**
20. **Bell, S.P. The origin recognition complex: from simple origins to complex functions. *Genes Dev* 16, 659-72 (2002).**
21. **Laurenson, P. & Rine, J. Silencers, silencing, and heritable transcriptional states. *Microbiol Rev* 56, 543-60 (1992).**
22. **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, 583-92 (1995).**
23. **Henikoff, S. Nucleosome destabilization in the epigenetic regulation of gene expression. *Nat Rev Genet* 9, 15-26 (2008).**
24. **Schwaiger, M. & Schubeler, D. A question of timing: emerging links between transcription and replication. *Curr Opin Genet Dev* 16, 177-83 (2006).**
25. **Thoma, F. Nucleosome positioning. *Biochim Biophys Acta* 1130, 1-19 (1992).**
26. **Moorman, C. et al. Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc Natl Acad Sci U S A* 103, 12027-32 (2006).**
27. **Tsukiyama, T., Becker, P.B. & Wu, C. ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. *Nature* 367, 525-32 (1994).**
28. **Pak, D.T. et al. Association of the origin recognition complex with heterochromatin and HP1 in higher eukaryotes. *Cell* 91, 311-23 (1997).**
29. **Parkinson, H. et al. ArrayExpress--a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res* 35, D747-50 (2007).**
30. **Wang, J., Nygaard, V., Smith-Sorensen, B., Hovig, E. & Myklebost, O. MArray: analysing single, replicated or reversed microarray experiments. *Bioinformatics* 18, 1139-40 (2002).**
31. **Bo, T.H., Dysvik, B. & Jonassen, I. LSimpute: accurate estimation of missing values in microarray data with least squares methods. *Nucleic Acids Res* 32, e34 (2004).**
32. **Wingender, E., Dietze, P., Karas, H. & Knuppel, R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 24, 238-41 (1996).**
33. **Foat, B.C., Morozov, A.V. & Bussemaker, H.J. Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* 22, e141-9 (2006).**
34. **van Helden, J. Regulatory sequence analysis tools. *Nucleic Acids Res* 31, 3593-6 (2003).**

35. **Lee, C.K., Shibata, Y., Rao, B., Strahl, B.D. & Lieb, J.D. Evidence for nucleosome depletion at active regulatory regions genome-wide.** ***Nat Genet*** **36, 900-5 (2004).**

36. **Holstege, F.C. et al. Dissecting the regulatory circuitry of a eukaryotic genome.** ***Cell*** **95, 717-28 (1998).**

37. **Wyrick, J.J. et al. Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins.** ***Science*** **294, 2357-60 (2001).**

38. **Ghaemmaghami, S. et al. Global analysis of protein expression in yeast.** ***Nature*** **425, 737-41 (2003).**

39. **Ng, H.H., Robert, F., Young, R.A. & Struhl, K. Genome-wide location and regulated recruitment of the RSC nucleosome-remodeling complex.** ***Genes Dev*** **16, 806-19 (2002).**

40. **de Hoon, M.J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software.** ***Bioinformatics*** **20, 1453-4 (2004).**

41. **Saldanha, A.J. Java Treeview--extensible visualization of microarray data.** ***Bioinformatics*** **20, 3246-8 (2004).**

Figure Legends

Figure 1: (**A**) Overview of our computational approach. (**B**) Results of the unbiased classification of ChIP-chip data for 354 yeast transcription factors into 15 chromatin types (CT1-CT15) and 7 factor groups (FG1-FG7) based on their promoter occupancy in rich media conditions. An asterisk marks the TF localization "hotspots" (CT15), which are occupied by the majority of TFs present in the cell. Within each TF block, the columns are randomly ordered.

Figure 2: Characterization of each chromatin type using functional genomics data. For each chromatin type, color-coded t-values quantify the difference in mean between the genes in the chromatin type and all other genes, for: (**A**) absolute mRNA abundance; (**B**) occupancy by various components of the origin of replication complex (ORC); (**C**) occupancy by nucleosome-remolding complex; (**D**) nucleosome occupancy as represented by various histone proteins;

Figure 3: Clustering of T-values to correlation coefficients between predicted sequence affinity [7] and the mean of raw Chip-chip ratios within each TF type: Factors are sorted by the last column – the total mean of raw Chip-chip ratios. The results suggest that protein binding sites of Rap1p and Pdr1p have significant positive correlation with almost all seven TF types.

Figure 4: The effect of normalization across ChIP-chip experiments. Shown is the Pearson correlation coefficient between the predicted affinity for RAP1p and the raw CHIP-chip ratios versus that based on mean-centered CHIP-chip ratios. The dashed lines represent correlations with P-values smaller than 10^{-4}.

Tables

Table 1

Percent of predicted nucleosome calls of 15 chromatin types based on experiment data from Lee et al Nat Gen 2007.

Chromatin type	nucleosome %	Fuzzy nucleosome %	No nucleosome %	Percent Nucleosome occupancy
CT 6	**0.234**	**0.144**	**0.620**	**Low**
CT 9	**0.336**	**0.097**	**0.565**	**Low**
CT 12	**0.309**	**0.168**	**0.521**	**Low**
CT 14	0.384	0.095	0.520	Marginal
CT 3	0.365	0.150	0.48	Marginal
CT 1	0.379	0.152	0.468	Marginal
CT 13	0.393	0.127	0.479	Marginal
CT 4	0.430	0.115	0.453	Marginal
CT 8	0.419	0.130	0.450	Marginal
CT 7	0.44	0.123	0.433	Marginal
CT 5	0.410	0.167	0.421	Marginal
CT 11	**0.347**	**0.199**	**0.452**	**Marginal**
CT 15	**0.404**	**0.186**	**0.409**	**High**
CT 2	**0.470**	**0.133**	**0.395**	**High**
CT 10	**0.437**	**0.169**	**0.393**	**High**

Table 2

Over-represented k-mer (k=1 to 8) oligomeric motifs in chromatin type 15 based on RSA tools.

Oligomer length	Over-represented sequence	Occurrence probability	E-value for occurrences
1	a/t	6e-13	1.2e-12
2	aa/tt	1e-09	1.0e-08
3	**tca/tga** **aaa/ttt**	**1.7e-08** **3.2e-05**	**5.4e-07** **1.0e-03**
4	aaat/attt tgaa/ttca tcaa/ttga	3.4e-07 8.2e-07 1.9e-06	4.6e-05 1.1e-04 2.6e-04
5	aaaat/atttt ttcaa/ttgaa aaatt/aattt	3e-07 8.4e-07 1.4e-05	1.5e-04 4.3e-04 7.0e-03
6	aaaatt/aatttt attgaa/ttcaat	1.1e-05 1.7e-05	2.2e-02 3.6e-02
7	caatttc/gaaattg tcaatga/tcattga	8.1e-05 8.4e-05	6.6e-01 6.9e-01
8	NA	NA	NA

Table 3 A table represents results of all data analysis that have been done in this study.

Chromatin type	Percent Nucleosome occupancy	Chip-chip Nucleosome density	Origin replication complex	Nucleosome remolding complex	Enriched TCA/AAA motifs	Predicted binding sites	Transcriptome level	H3/H4	Gcn5p/Gcn4p/Esa1p	H3K36Me3	Yeast Hotspots	Rap1p/Pdr1p	Chromatin Class
CT 6	**L**	**L**	L	**H**	--	Almost All	**H**	**L**	**H**	**H**	No	M	1
CT 12	**L**	**L**	L	**H**	--	**Few**	**H**	**L**	**H**	**H**	No	M	1
CT 9	**L**	**L**	L	**H**	--	Many	L	M	M	L	No	L	1
CT 14	M	H	L	**H**	--	**Few**	L	H	L	L	No	L	2
CT 3	M	H	L	**H**	--	**Few**	L	H	L	L	No	L	2
CT 1	M	H	L	**H**	--	**Few**	L	H	L	L	No	L	2
CT 13	M	H	L	M	--	**Few**	L	H	L	L	No	L	2
CT 4	M	H	L	L	**TCA,AAA**	**Few**	L	H	L	L	No	L	3
CT 8	M	H	L	L	**TCA,AAA**	Many	L	H	L	L	No	L	3
CT 7	M	H	L	L	TCA	Many	L	H	L	L	No	L	3
CT 5	M	H	**H**	L	**TCA,AAA**	Many	M	**L**	M	M	No	L	3
CT 11	**M**	**L**	**H**	**H**	**AAA**	**Few**	**H**	**L**	**H**	**M**	**Partially Yes**	L	4
CT 15	**H**	**L**	**H**	**H**	**TCA,AAA**	**Few**	**H**	**L**	**H**	**H**	**Yes**	**H**	4
CT 2	H	H	H	L	TCA,AAA	Many	L	L	L	L	No	L	5
CT 10	H	H	H	L	TCA,AAA	Many	L	L	L	L	No	L	5

L means low level, M represents marginal level and H represents high level; 1 is the active euchromatin, 2 is the pseudo-heterochromatin controlled by nucleosome remolding complex, 3 is the pseudo-heterochromatin modified by collaborative competition, 4 is the dynamic or active heterochromatin and 5 is the silent heterochromatin

Figures

Figure 1 Results of the unbiased classification of ChIP-chip data for 354 yeast transcription factors into 15 chromatin types (CT1-CT15) and 7 factor groups (FG1-FG7) based on their promoter occupancy in rich media conditions. An asterisk marks the TF localization "hotspots" (CT15), which are occupied by the majority of TFs present in the cell.

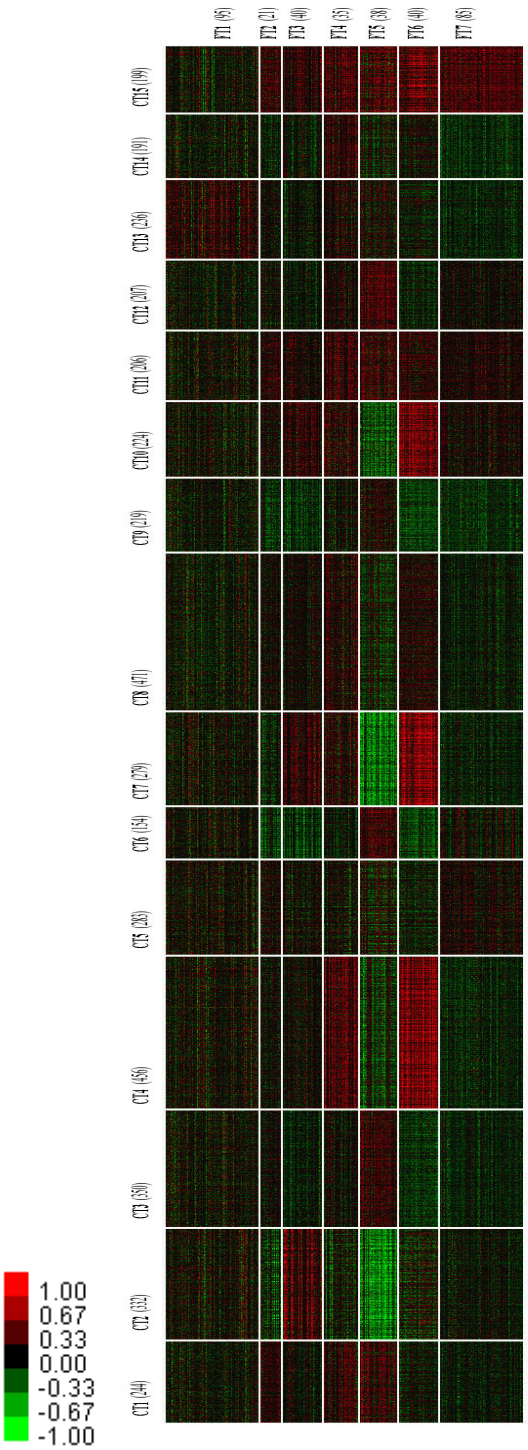


Figure 2 For each chromatin type, color-coded T-values quantify the difference in mean between the genes in the chromatin type and all other genes, for: (A) absolute mRNA abundance; (B) occupancy by various components of the origin replication complex (ORC); (C) occupancy by nucleosome-remodeling complex; (D) nucleosome occupancy as represented by various histone proteins.

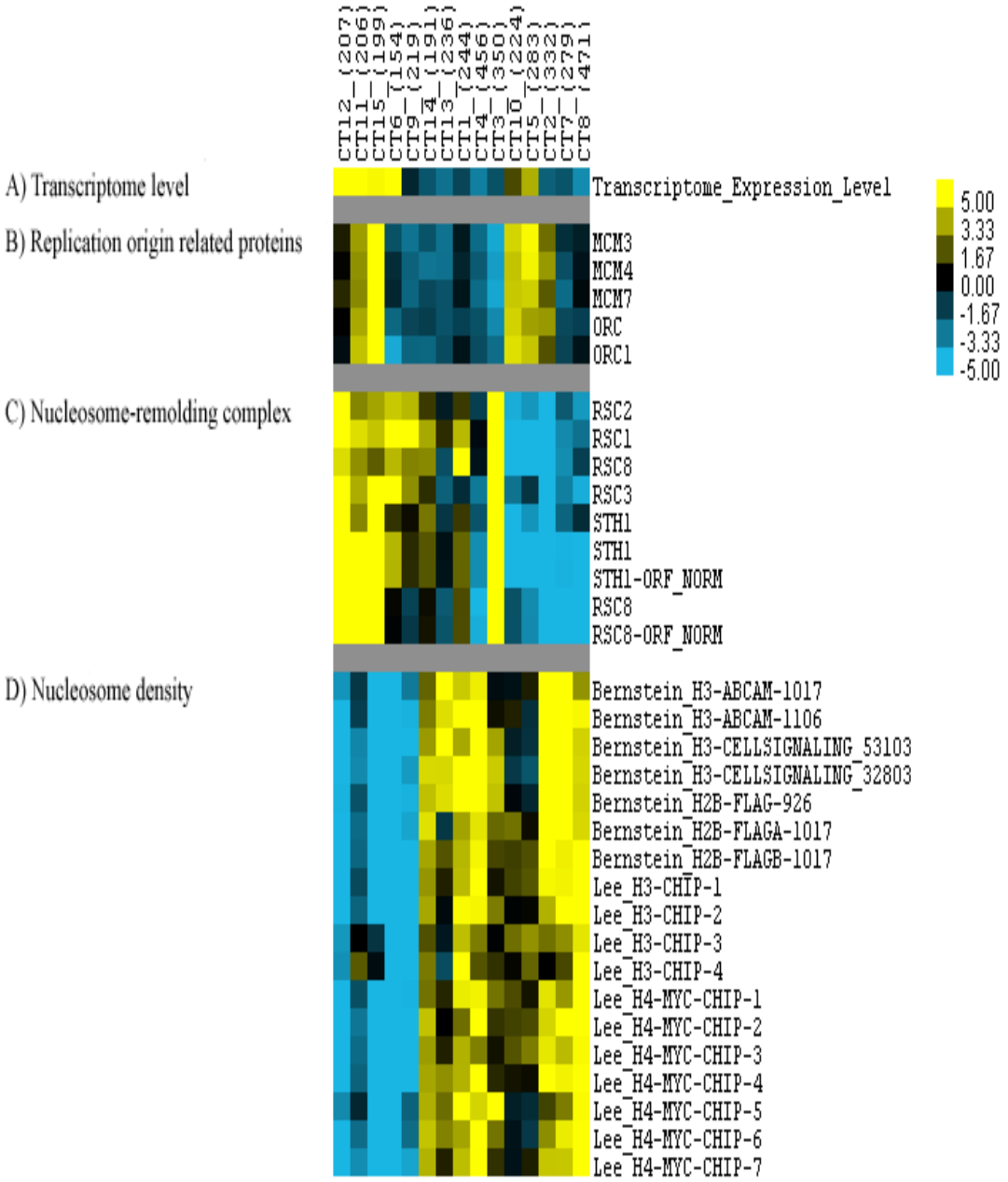


Figure 3 Correlation between sequence affinity scores of yeast TFs with mean CHiP-chip profiles of 7 factor groups, which suggests that protein binding motifs of RAP1, PDR1 and MET31 have significant positive correlation with almost all TFs.

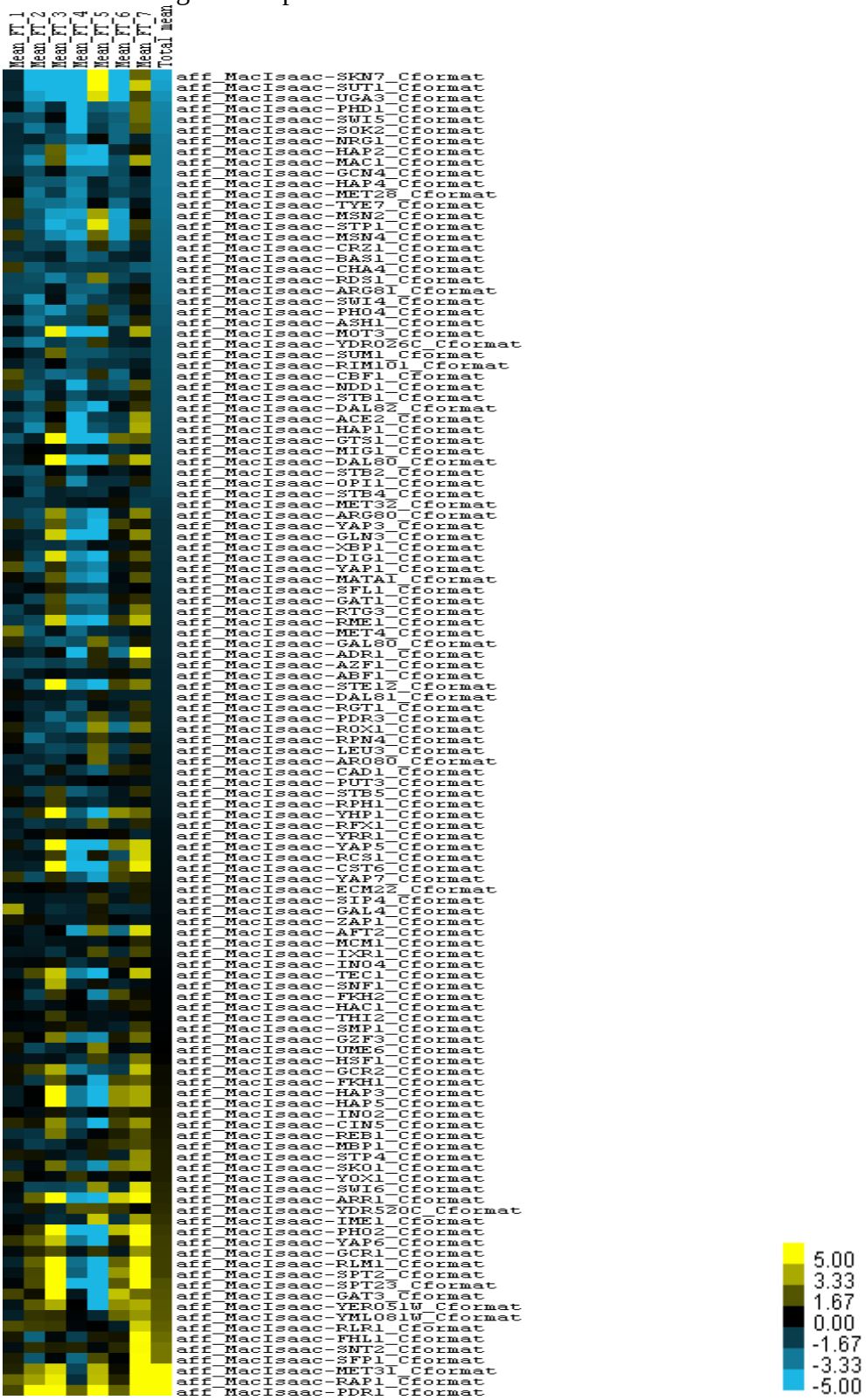
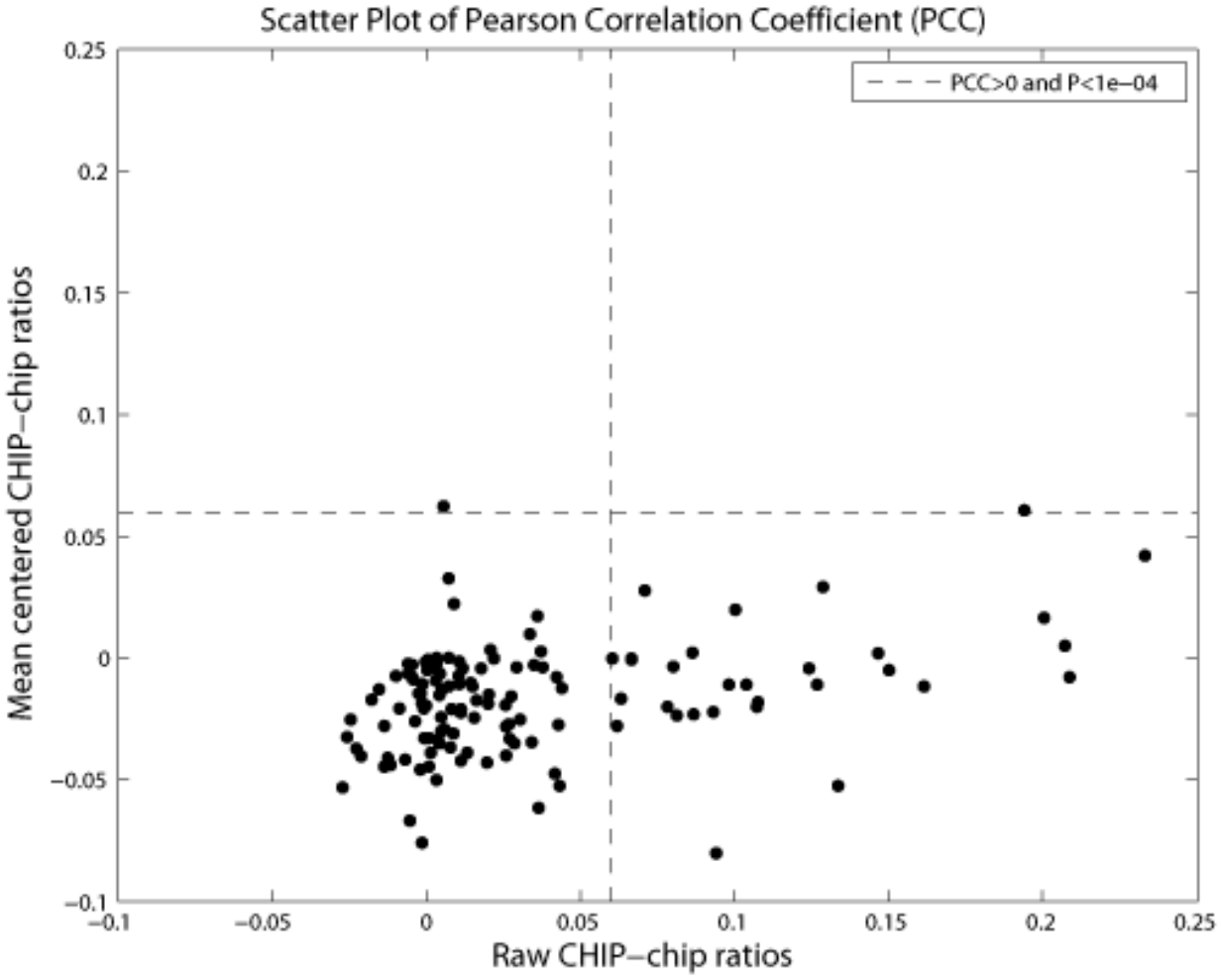
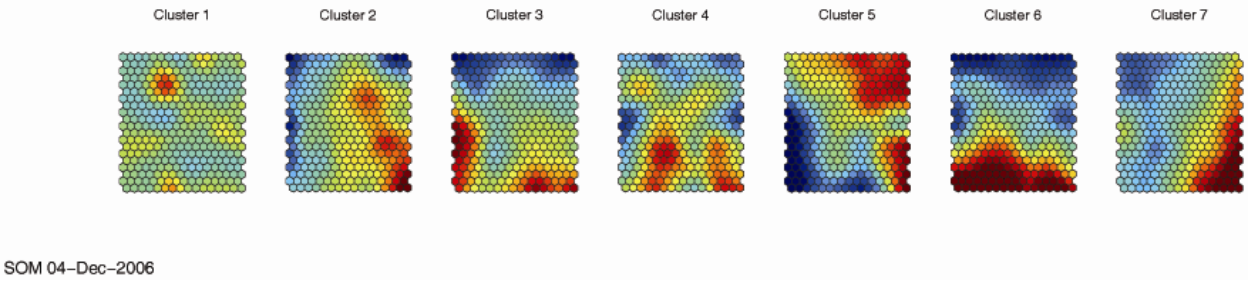


Figure 4 Pearson correlation coefficient between the predicted affinity for Rap1p and the raw CHIP-chip ratios (X-axis) vs. that based on mean-centered CHIP-chip ratios (Y-axis). The dashed lines are thresholds for correlation coefficients with P-values smaller than 10E-4. For Pdr1p, there is a similar effect of normalization across CHIP-chip experiments.

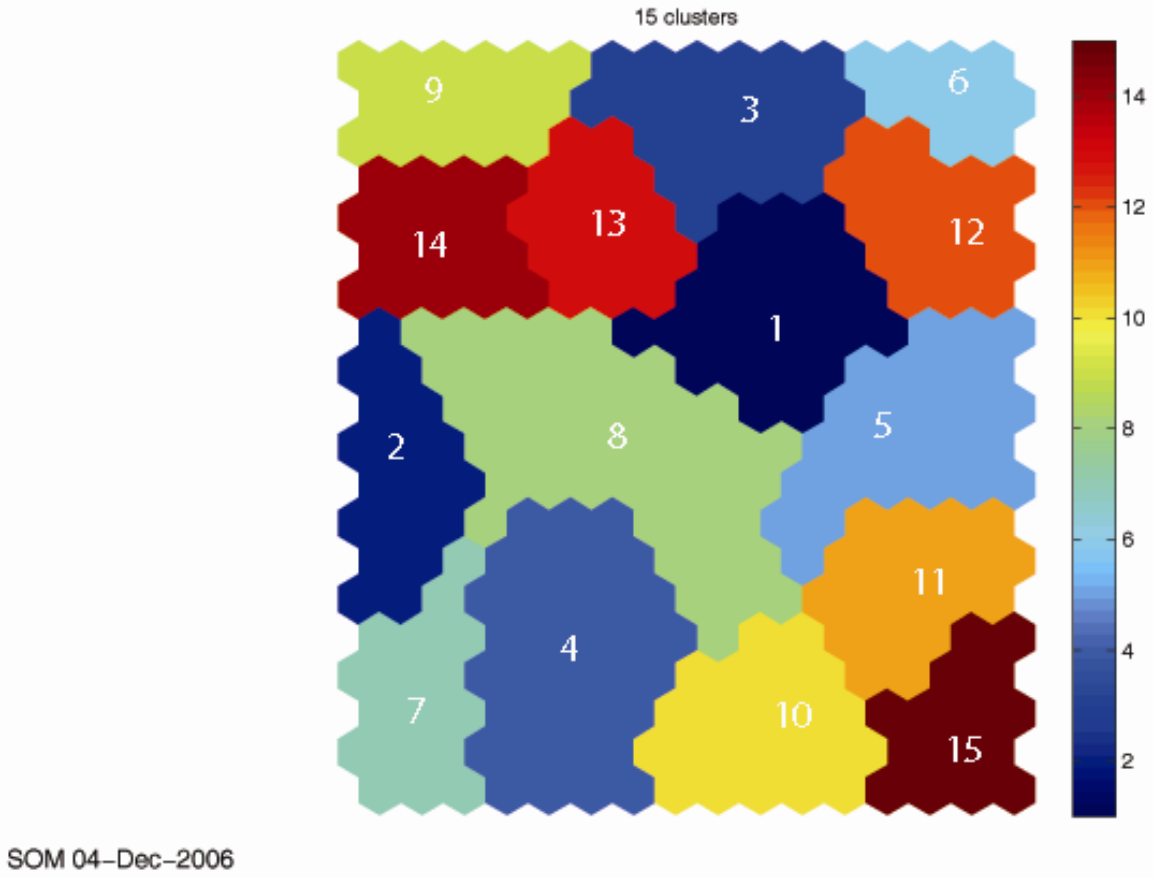


Supplementary

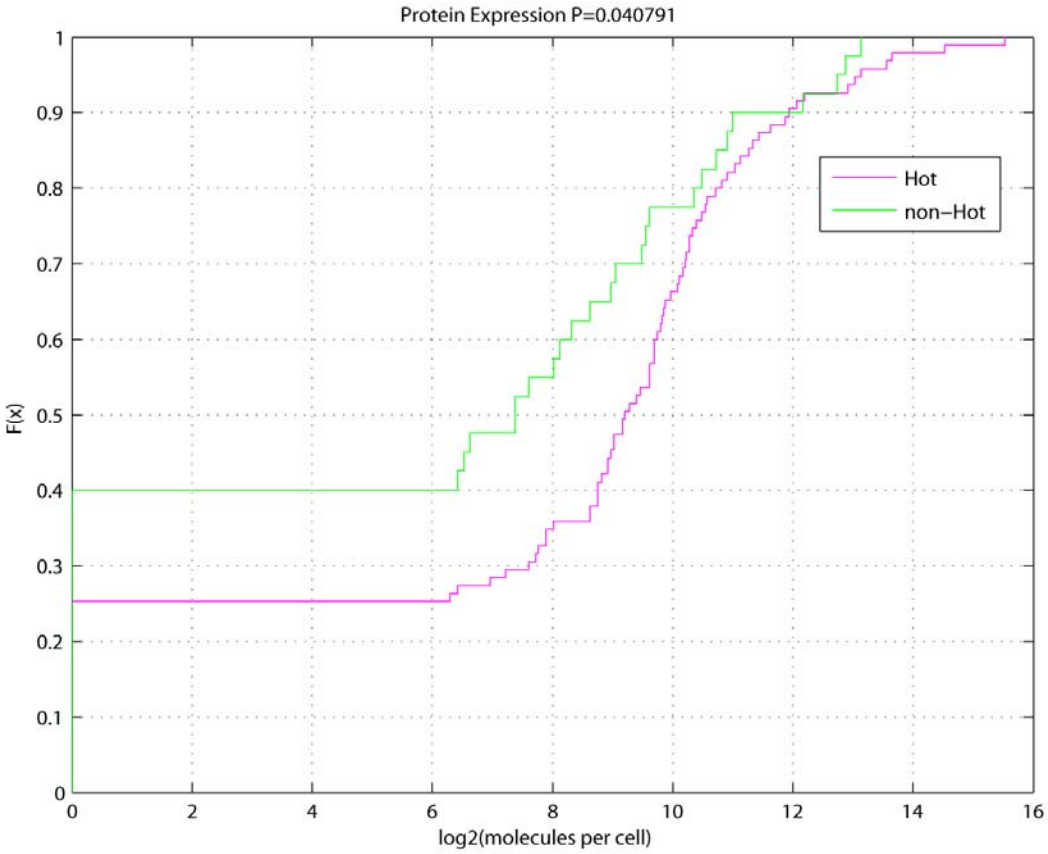
SFigure 1 SOM complement plan (map size 20x16) for 7 sample clusters based on raw CHIP-chip ratios (354 genome-wide TF occupancy profiles with 4050 yeast probed promoter regions).



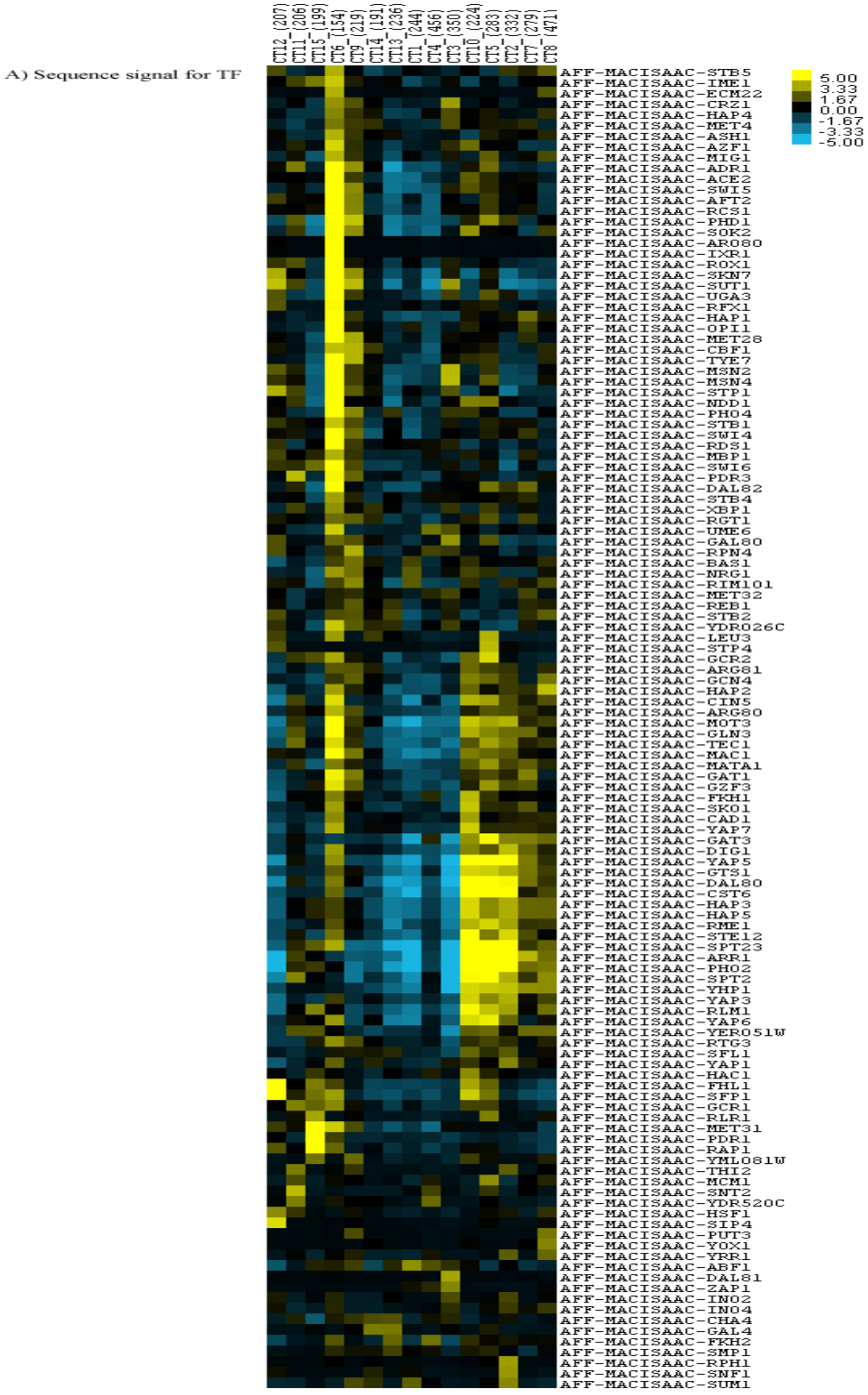
SFigure 2 K-means clusters for SOM component plan (15 clusters for 320 SOM codebooks).



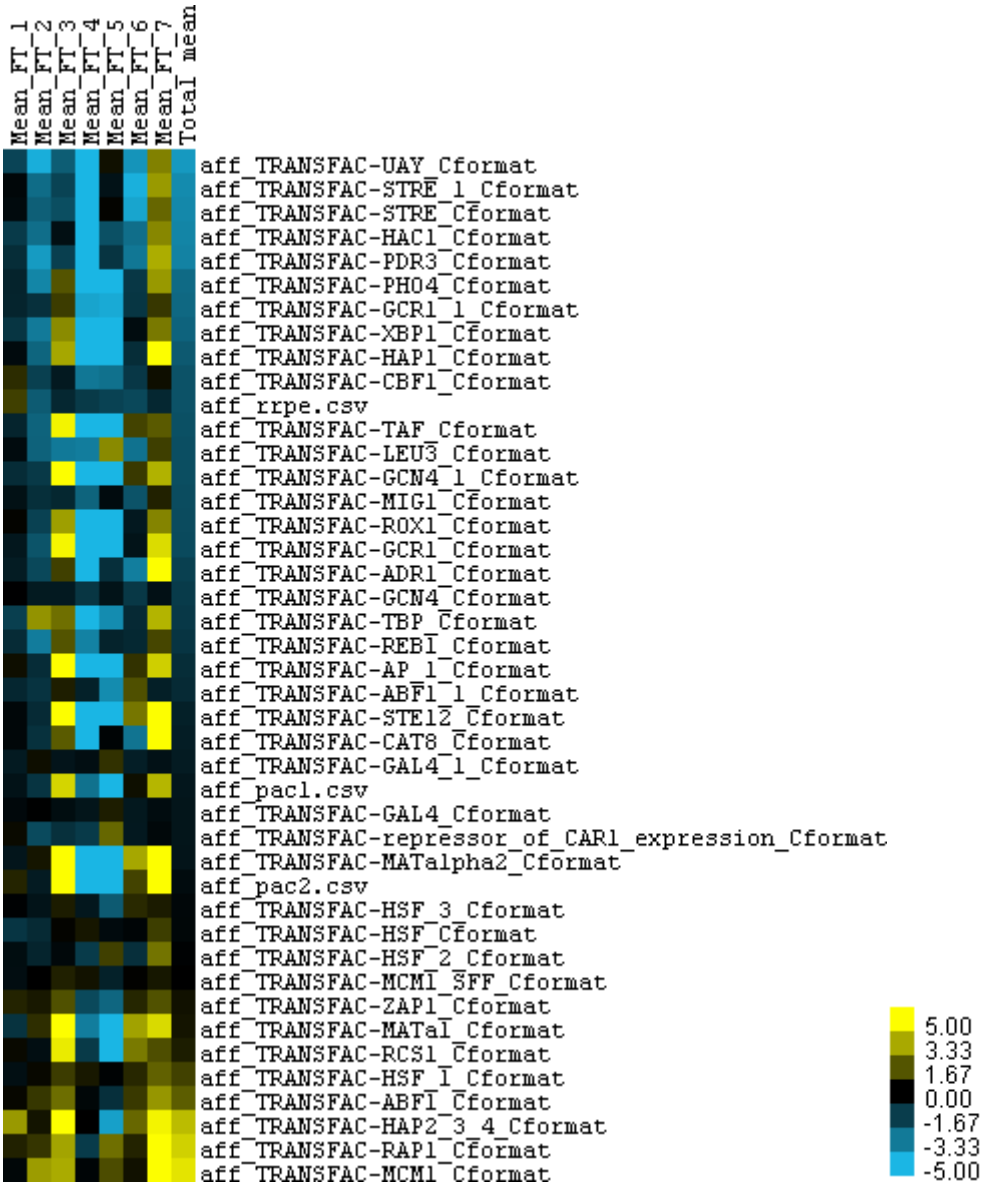
SFigure 3 Empirical cumulative distribution of protein expression level (rich media conditions only based on measurements by [38] for co-localized transcription factors (95 hot proteins) and not co-localized transcription factors (40 non-hot proteins), P-value to the t-test is 0.04.



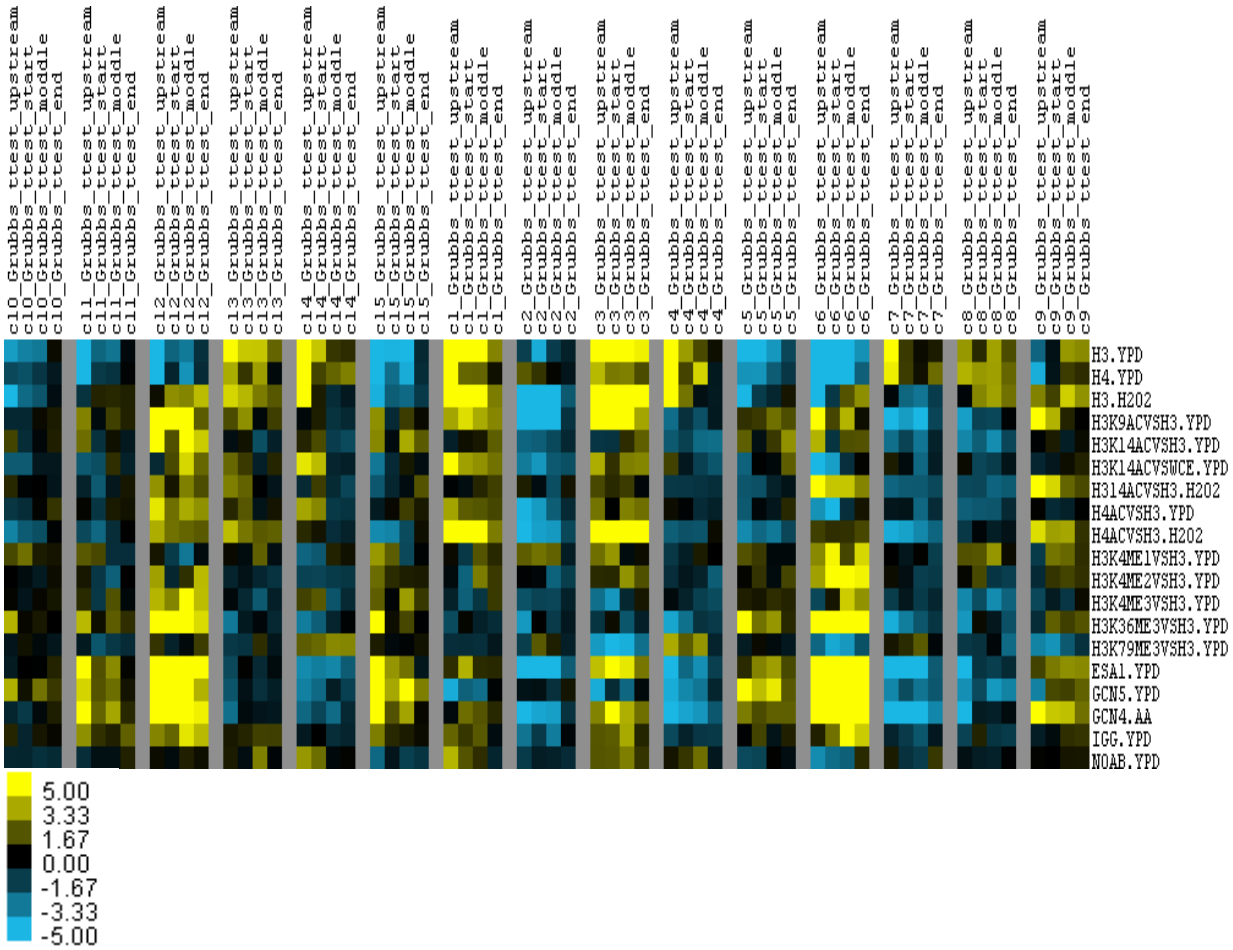
SFigure 4 Enrichment (yellow) and depletion (blue) of predicted binding sites for various TFs in the 15 different chromatin types. Factors whose binding sites are significantly enriched in the "hotspots" but not in any other chromatin type are Rap1p and Pdr1p.



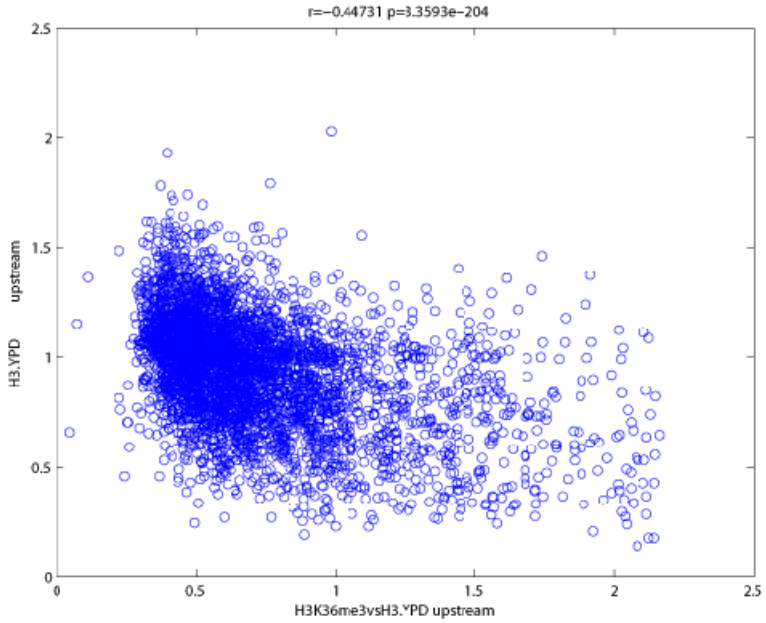
SFigure 5. Correlation between sequence affinity scores of yeast TFs (TRANSFAC) with mean CHiP-chip profiles of 7 factor groups, which suggests that protein binding motifs of RAP1, MCM1 and ABF1 have significant positive correlation with almost all TFs' CHiP-chip profiles.



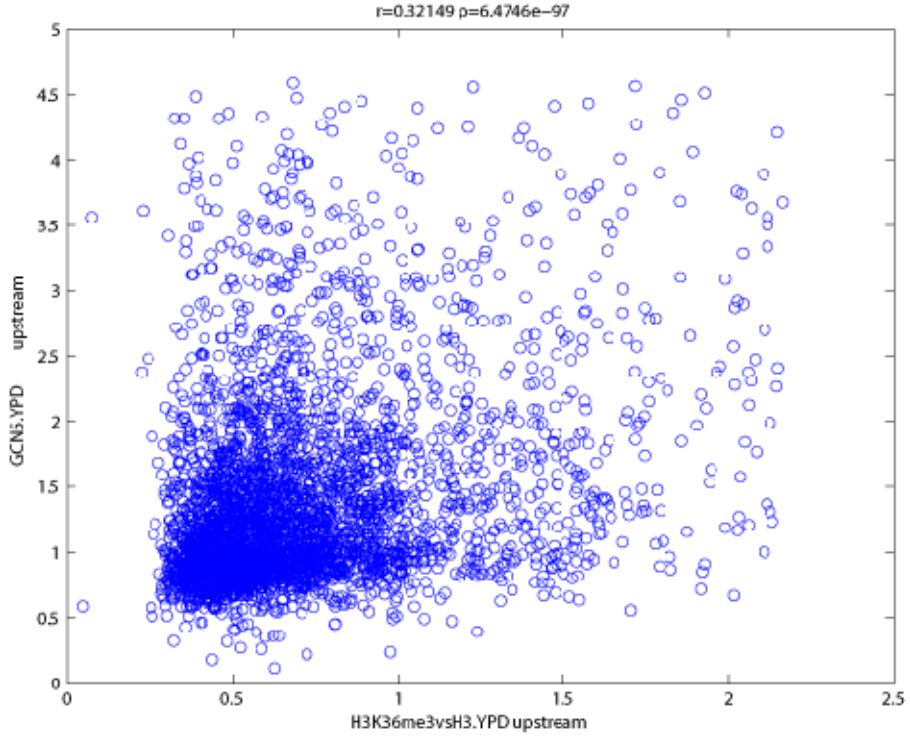
SFigure 6 Clusters of T-values for Pokholok's histone modification datasets based on our current classification of 15 chromatin types: probe orders are sorted by their position in a gene such as upstream, start, middle and end of a gene.



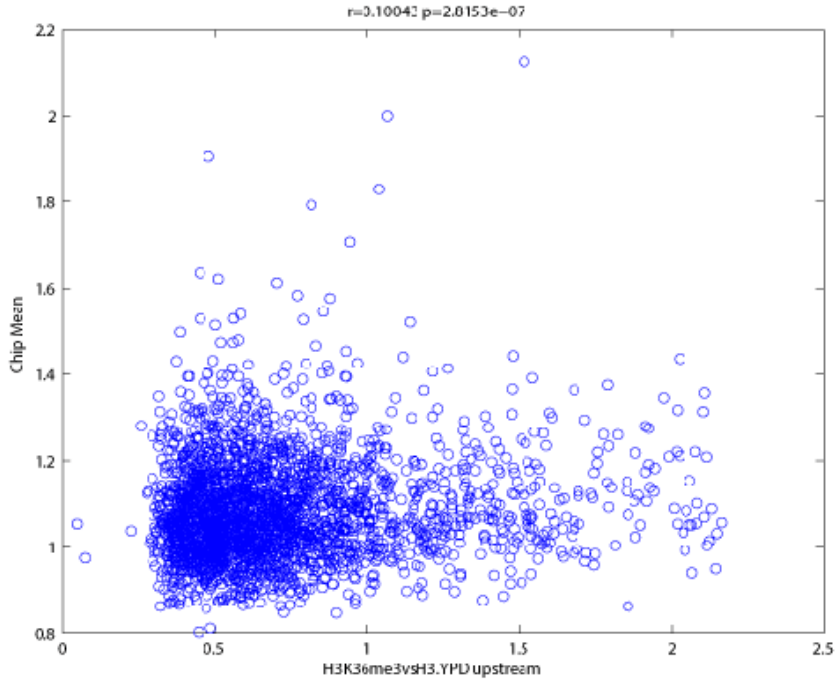
SFigure 7 Scat plot of H3 profiles vs. H3K36Me3, which shows a strong negative correlation between H3K36me3 and histon such as H3 profiles.



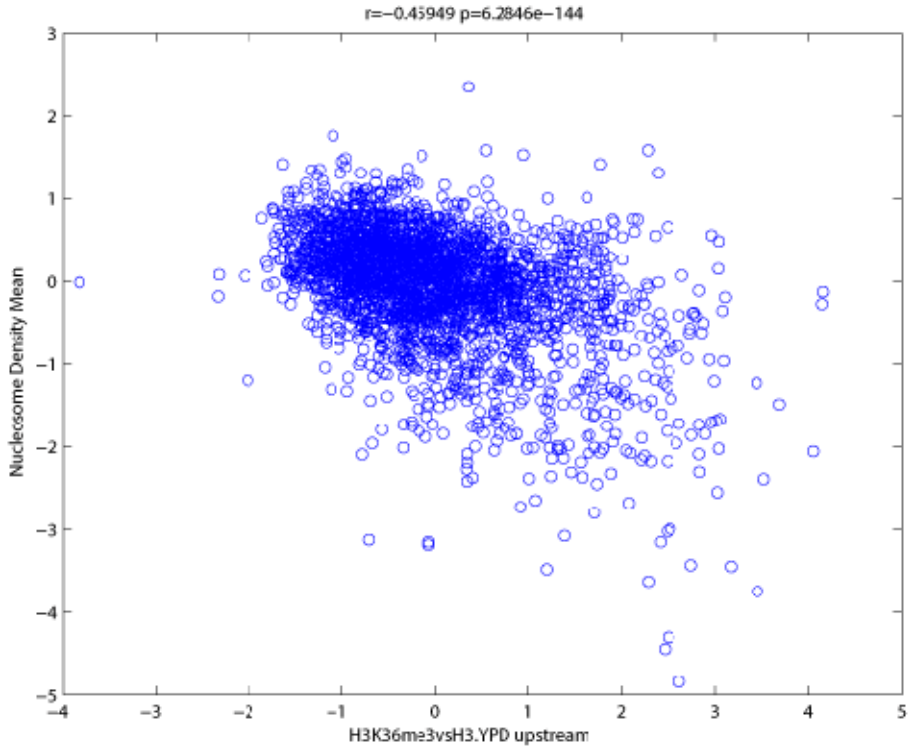
SFigure 8 Scat plot of GCN5 binding vs. H3K36Me3 profiles, which shows a strong positive correlation between H3K36me3 and TF binding such as GCN5 here.



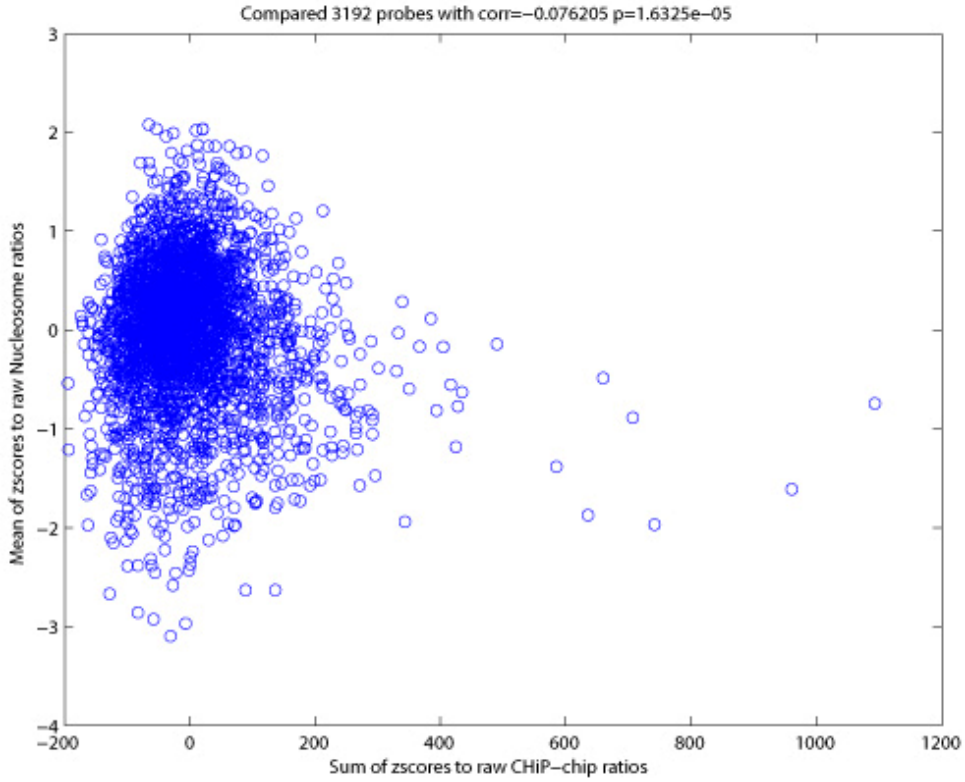
SFigure 9 Scat plot of meanChip vs Pokholok H3K36Me3, which suggested a significant positive correlation between mean Chip raw ratios and upstream H3K36Me3 ratios.



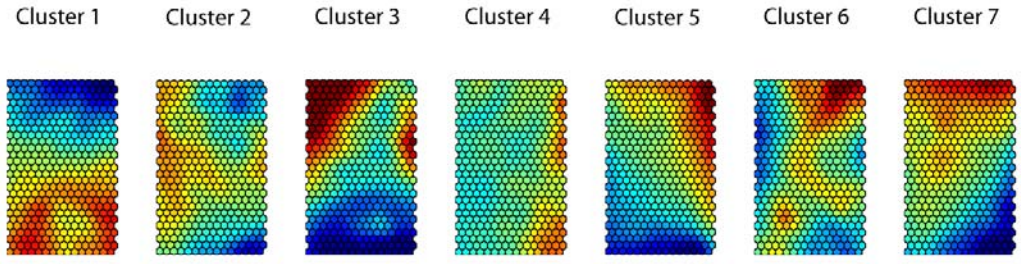
SFigure 10 Scat plot of nucleosome Density vs H3K36Me3, where nucleosome density mean is the mean nucleosome density of both Lee and Bernstein nucleosome datasets, which shows a strong negative correlation between H3K36me3 and nucleosome density.



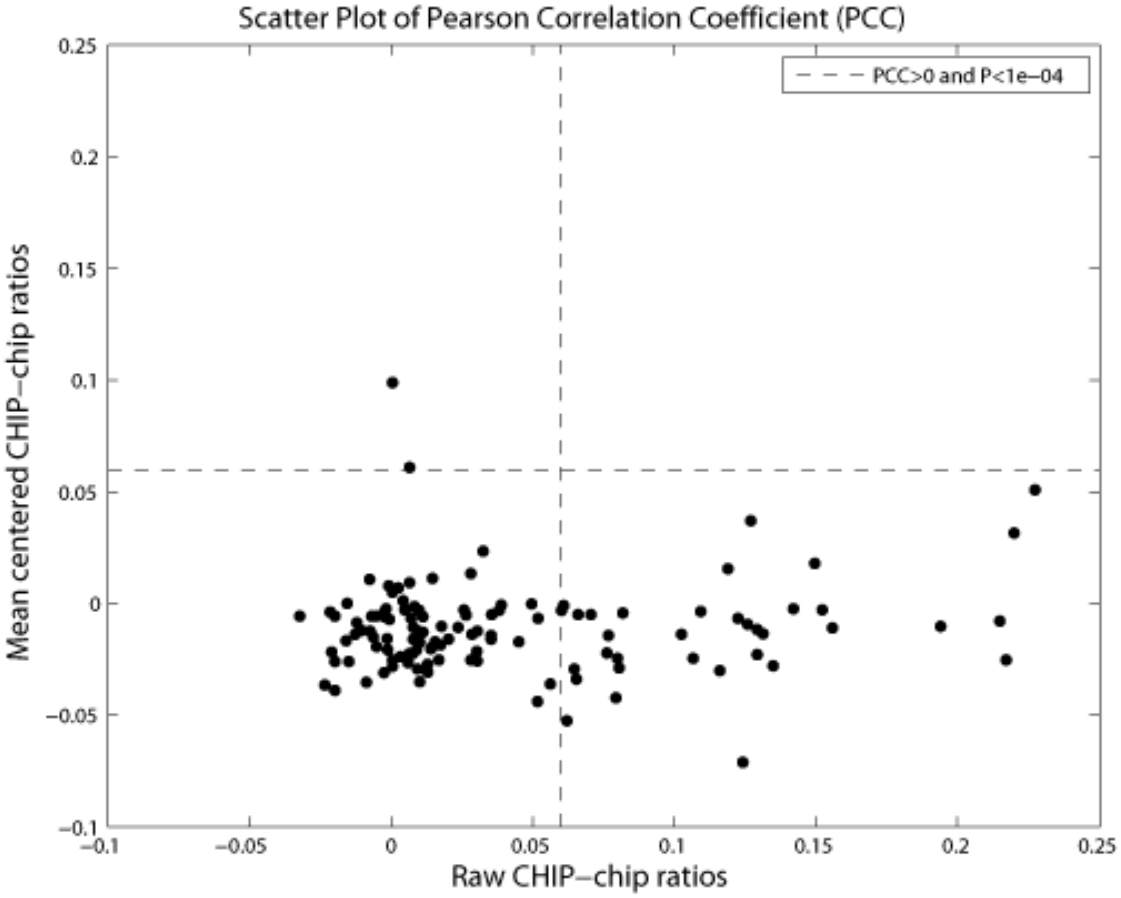
SFigure 11 Scat plot of the sum of all ChiP-chip binding ratios vs. mean of nucleosome density, where nucleosome density mean is the mean nucleosome density of both Lee and Bernstein nucleosome datasets, which suggests a strong negative correlation between syngenetic CHiP-chip occupancies and the local nucleosome density.



SFigure 12 SOM complement plan (map size 27x15) for 7 sample clusters based on mean centered CHIP-chip ratios (the data as provided by Harbison et al (2004), 204 transcription factors with 6138 yeast probed promoter regions).



SFigure 13 Pearson correlation coefficient between the predicted affinity for PDR1p and the raw CHIP-chip ratios (X-axis) vs. that based on mean-centered CHIP-chip ratios (Y-axis). The dashed lines are thresholds for correlation coefficients with P-values smaller than 10E-4.



STable 1 Over-represented 3-mer DNA sequences in 15 chromatin types based on RSA tools.

Chromatin type	Over-represented sequence	Occurrence probability	E-value for occurences
CT 1	gca|tgc	2.00E-07	6.20E-06
	gcc|ggc	6.70E-07	2.10E-05
	agc|gct	1.20E-05	3.80E-04
	caa|ttg	0.00017	5.50E-03
	aag|ctt	0.00069	2.20E-02
	cca|tgg	0.00121	3.90E-02
	cga|tcg	0.02591	8.30E-01
CT 2	**tca|tga**	**9.80E-16**	**3.20E-14**
	aaa|ttt	**2.30E-15**	**7.40E-14**
	caa|ttg	1.10E-13	3.40E-12
	aat|att	4.70E-10	1.50E-08
	atg|cat	2.40E-05	7.60E-04
	act|agt	0.00014	4.50E-03
	gaa|ttc	0.0004	1.30E-02
	aag|ctt	0.00108	3.50E-02
	aac|gtt	0.00304	9.70E-02
	taa|tta	0.02264	7.20E-01
	atc|gat	0.02958	9.50E-01
CT 3	gcc|ggc	5.60E-19	1.80E-17
	cgc|gcg	3.00E-17	9.60E-16
	gca|tgc	1.60E-13	5.10E-12
	ccg|cgg	6.70E-12	2.10E-10
	agc|gct	1.10E-10	3.70E-09
	acg|cgt	2.50E-07	7.80E-06
	cga|tcg	1.90E-06	6.10E-05
	cac|gtg	4.00E-05	1.30E-03
	cag|ctg	4.10E-05	1.30E-03
	aag|ctt	4.70E-05	1.50E-03
	ccc|ggg	7.60E-05	2.40E-03
	cca|tgg	0.00024	7.70E-03
	gga|tcc	0.00166	5.30E-02
	gac|gtc	0.00217	7.00E-02
	agg|cct	0.00397	1.30E-01
CT 4	caa|ttg	7.70E-09	2.50E-07
	aag|ctt	5.10E-08	1.60E-06
	tca|tga	**4.00E-06**	**1.30E-04**
	gaa|ttc	0.00073	2.30E-02
	agc|gct	0.00173	5.50E-02
	aaa|ttt	**0.00356**	**1.10E-01**
	atg|cat	0.00657	2.10E-01
	aac|gtt	0.01049	3.40E-01
	atc|gat	0.02814	9.00E-01
CT 5	**aaa|ttt**	**3.40E-14**	**1.10E-12**
	aag|ctt	2.30E-09	7.30E-08
	gaa|ttc	5.20E-07	1.70E-05
	aga|tct	0.00053	1.70E-02
	atg|cat	0.00066	2.10E-02
	aat|att	0.0099	3.20E-01
	tca|tga	**0.01023**	**3.30E-01**
	taa|tta	0.01848	5.90E-01
CT 6	ccg|cgg	2.50E-61	8.10E-60
	gcc|ggc	1.40E-44	4.50E-43
	cgc|gcg	2.60E-31	8.30E-30
	ccc|ggg	6.90E-28	2.20E-26
	gca|tgc	2.30E-25	7.20E-24
	acg|cgt	6.80E-15	2.20E-13
	cac|gtg	5.80E-13	1.90E-11
	agc|gct	2.40E-11	7.70E-10

	gga|tcc	8.40E-11	2.70E-09
	cag|ctg	1.10E-09	3.50E-08
	cga|tcg	6.30E-09	2.00E-07
	ctc|gag	6.60E-08	2.10E-06
	agg|cct	3.30E-07	1.10E-05
	cca|tgg	5.60E-07	1.80E-05
	acc|ggt	1.80E-06	5.90E-05
	aga|tct	0.00345	1.10E-01
	gaa|ttc	0.01249	4.00E-01
	aag|ctt	0.01349	4.30E-01
CT 7	**tca|tga**	**8.80E-14**	**2.80E-12**
	caa|ttg	2.10E-09	6.70E-08
	gaa|ttc	5.60E-09	1.80E-07
	aag|ctt	1.10E-05	3.50E-04
	aac|gtt	0.00012	3.90E-03
	act|agt	0.00039	1.20E-02
	cag|ctg	0.00392	1.30E-01
CT 8	**aaa|ttt**	**9.40E-09**	**3.00E-07**
	gaa|ttc	1.70E-08	5.50E-07
	aag|ctt	4.30E-07	1.40E-05
	caa|ttg	7.80E-07	2.50E-05
	tca|tga	**7.80E-05**	**2.50E-03**
	aga|tct	0.00012	3.80E-03
	aac|gtt	0.01339	4.30E-01
CT 9	gcc|ggc	1.80E-22	5.90E-21
	ccg|cgg	1.20E-14	3.80E-13
	gca|tgc	6.50E-13	2.10E-11
	cgc|gcg	3.80E-10	1.20E-08
	cag|ctg	1.80E-09	5.80E-08
	agc|gct	1.10E-07	3.60E-06
	cca|tgg	2.90E-07	9.40E-06
	acg|cgt	4.20E-06	1.30E-04
	cac|gtg	8.30E-06	2.70E-04
	cga|tcg	1.70E-05	5.50E-04
	gac|gtc	1.80E-05	5.70E-04
	ccc|ggg	3.80E-05	1.20E-03
	agg|cct	0.0002	6.40E-03
	gga|tcc	0.00053	1.70E-02
	ctc|gag	0.00187	6.00E-02
	acc|ggt	0.00296	9.50E-02
	aag|ctt	0.01978	6.30E-01
CT 10	**aaa|ttt**	**1.20E-18**	**3.90E-17**
	caa|ttg	4.30E-09	1.40E-07
	tca|tga	**5.10E-09**	**1.60E-07**
	gaa|ttc	1.20E-06	3.90E-05
	aat|att	4.40E-06	1.40E-04
	aac|gtt	1.40E-05	4.50E-04
	atg|cat	2.00E-05	6.30E-04
	aag|ctt	8.60E-05	2.80E-03
	taa|tta	0.0004	1.30E-02
CT 11	**aaa|ttt**	**4.30E-05**	**1.40E-03**
	gaa|ttc	9.00E-05	2.90E-03
	agg|cct	0.005	1.60E-01
	aag|ctt	0.01202	3.80E-01
	aac|gtt	0.01262	4.00E-01
CT 12	ccg|cgg	1.10E-14	3.40E-13
	gcc|ggc	2.00E-13	6.50E-12
	cgc|gcg	3.00E-13	9.60E-12
	gca|tgc	1.20E-09	3.80E-08
	agc|gct	6.80E-08	2.20E-06
	cca|tgg	8.80E-08	2.80E-06
	ccc|ggg	3.80E-07	1.20E-05

	ctc|gag	7.20E-06	2.30E-04
	gga|tcc	0.0003	9.80E-03
	cac|gtg	0.00036	1.10E-02
	acc|ggt	0.00098	3.10E-02
	cag|ctg	0.00538	1.70E-01
	aga|tct	0.00803	2.60E-01
	acg|cgt	0.01586	5.10E-01
CT 13	agc|gct	7.50E-08	2.40E-06
	cag|ctg	3.10E-06	9.90E-05
	aag|ctt	2.50E-05	8.00E-04
	agg|cct	0.00041	1.30E-02
	gca|tgc	0.00068	2.20E-02
	gaa|ttc	0.0009	2.90E-02
	gcc|ggc	0.00334	1.10E-01
	ctc|gag	0.02061	6.60E-01
	caa|ttg	0.02922	9.40E-01
CT 14	aag|ctt	2.30E-05	7.40E-04
	aga|tct	0.00047	1.50E-02
	agc|gct	0.00308	9.90E-02
	gca|tgc	0.00361	1.20E-01
	gaa|ttc	0.00528	1.70E-01
CT 15	**tca|tga**	**1.70E-08**	**5.50E-07**
	aaa|ttt	**3.10E-05**	**1.00E-03**
	aat|att	0.00057	1.80E-02
	atg|cat	0.00182	5.80E-02
	caa|ttg	0.00416	1.30E-01
	gaa|ttc	0.00691	2.20E-01
	cca|tgg	0.01911	6.10E-01
	act|agt	0.02434	7.80E-01

STable 2

Top 10 of TFs with sequence affinity scores significantly correlated with H3K36Me3 profiles.

TF Name	H3K36me3vsH3.YPD (T -value)	r - correCoef	p value
SFP1	13.2682	0.21265	1.43E-38
FHL1	12.8507	0.20625	2.45E-36
RAP1	**8.6405**	**0.14032**	**1.67E-17**
SKN7	4.9085	0.080251	1.21E-06
SUT1	4.6924	0.076739	3.48E-06
UGA3	4.3978	0.071947	1.36E-05
PUT3	3.3464	0.054807	0.000928
HSF1	3.219	0.052726	0.001444
GCR1	3.1614	0.051785	0.001755
RGT1	3.1004	0.050788	0.002152

STable 3 Result of a reanalysis TE elements in the DAMID fly dataset, which suggests that gene, mRNA, rescure fragement, and TE insertion site are highly enriched in fly hotspots.

Feature	HYPEGe-P (1kb flank)	Hypergeometric P (2kb flank)	t-test P (1 kb flank))	t-test T (1 kb flank))	t-test P (2KB flank)	t-test T (2 kb flank)
Gene	1.75E-09	2.20E-08	1.93E-08	5.63E+00	2.75E-07	5.15E+00
Mrna	7.17E-09	1.93E-07	6.20E-08	5.42E+00	1.62E-06	4.80E+00
rescue_fragment	1.88E-06	6.03E-07	1.63E-09	6.05E+00	3.37E-10	6.30E+00
transposable_element_insertion_site	6.42E-04	5.13E-06	1.61E-03	3.16E+00	9.49E-06	4.43E+00
aberration_junction	9.94E-06	6.50E-06	1.32E-09	6.08E+00	1.10E-09	6.11E+00
Exon	2.57E-04	1.48E-05	8.73E-04	3.33E+00	5.49E-05	4.04E+00
syntenic_region	4.04E-04	2.96E-04	1.76E-03	3.13E+00	7.54E-01	-3.13E-01
five_prime_utr	1.65E-02	4.18E-04	7.12E-02	1.80E+00	9.49E-04	3.31E+00
cds	3.29E-04	1.40E-03	1.07E-03	3.27E+00	5.84E-03	2.76E+00
regulatory_region	3.40E-02	5.81E-03	2.17E-03	3.07E+00	3.06E-05	4.17E+00
protein	1.61E-03	9.25E-03	6.88E-03	2.70E+00	4.81E-02	1.98E+00
orthologous_region	7.57E-03	1.09E-02	3.73E-02	2.08E+00	5.95E-02	1.88E+00
trna	2.12E-01	2.42E-02	4.01E-01	8.40E-01	1.46E-02	2.44E+00
Point_mutation	8.63E-02	3.23E-02	2.06E-01	1.26E+00	6.42E-02	1.85E+00
pcr_product	5.16E-02	5.23E-02	5.17E-01	6.48E-01	4.83E-01	7.01E-01
Three_prime_utr	7.29E-02	5.29E-02	6.38E-01	4.71E-01	4.39E-01	7.74E-01
oligonucleotide	3.89E-02	5.78E-02	3.08E-01	-1.02E+00	6.58E-01	-4.43E-01
transposable_element	1.08E-01	8.11E-02	8.30E-01	2.15E-01	5.90E-01	5.38E-01
Polya_site	3.38E-01	1.49E-01	5.82E-01	5.51E-01	1.77E-01	1.35E+00
protein_binding_site	3.56E-01	1.96E-01	6.89E-01	4.01E-01	3.34E-01	9.66E-01
sequence_variant	2.55E-01	2.19E-01	6.92E-01	-3.97E-01	1.30E-03	3.22E+00