Classification of *Saccharomyces cerevisiae* Promoter Regions into Distinct Chromatin Types Reveals the Existence of Nucleosome-Depleted Hotspots of Transcription Factor Occupancy

Junbai Wang[1] & 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.

Abstract

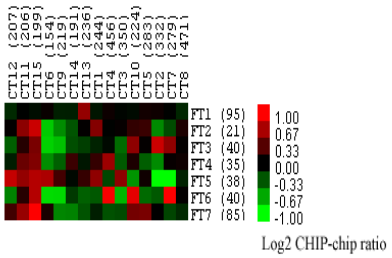
Transcriptional regulation is a complex process involving sequence-specific interactions between transcription factor (TF) proteins and DNA as well as protein-protein interactions with other TF's and nucleosomes. In a recent study (Moorman *et al.*, 2006) we used DamID to map the occupancy by 14 TF's and other chromatin-associated factors of diverse physiological function at ~1kb resolution in a 2.9 Mb region of the genome of *Drosophila melanogaster*. We found that ~5% of the genome consists of "hotspots" of size 1-5kb that are highly occupied by most of the profiled proteins.

Here we report that, surprisingly, this phenomenon also occurs in the yeast *Saccharomyces cerevisiae*. We re-analyzed the raw ChIP-chip data of Harbison *et al.* (2004), resulting in a set of 354 genomewide TF occupancy profiles (up to 3 replicates for each of 138 TF's, rich media conditions only). We then used unsupervised classification algorithms to partition the set of 4050 probed promoter regions into 15 clusters, corresponding to distinct chromatin types. One of these clusters, consisting of 199 "hotspot" promoter regions, is unique in that (i) it is characterized by a high ChIP enrichment ratio for ~70% of all TFs, and (ii) the DNA sequence of the hotspot regions is strongly enriched for predicted binding by Rap1p.

Consistent with a known function of Rap1p, the hotspots are found to be strongly nucleosome-depleted, as are four other chromatin types. Genes in the vicinity of hotspots were enriched for components of the ribosome, and more highly expressed than average. Significantly, when we repeat our analysis using the data as provided by Harbison *et al.* (2004), who normalized by the average across all TF's for each probe, the hotspot chromatin subtype can no longer be detected.

Classification of Promoter Regions into Distinct Chromatin Types

Understanding the association between chromatin structure and gene expression regulation is a topic of great current interest. We investigated the variation in local chromatin structure among the promoter regions in the yeast *Saccharomyces cerevisiae* through a careful reanalysis of the genome-wide occupancy data Harbison *et al.* (2004) in rich media conditions.

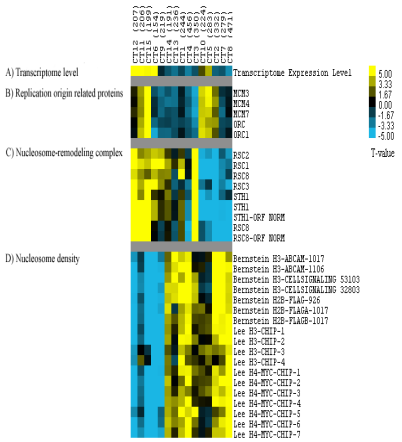


Unbiased classification (using SOMs and K-means clustering) of 354 yeast CHIP-chip experiments into 15 chromatin types (CT1-CT15) and 7 factor groups (FT1-FT7) based on their promoter occupancy in rich media condition. Chromatin type 15 (CT15) corresponds to the TF localization "hotspots".

Characterization of Each Chromatin Type Using Independent Functional Genomics Data

Based on the chromatin signature, we could classify 4050 promoter regions into 15 chromatin types. Then, we attempted to look for potential associations between the chromatin variation and the other functional genomics data. This may help us understand the variation in local chromatin structure among the promoter regions in yeast.

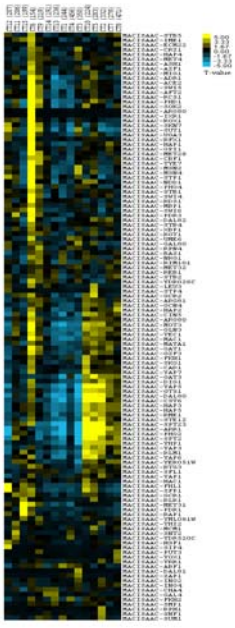
Characterization of Each Chromatin Type Using Independent Functional Genomics Data (cont.)



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.

Enrichment and Depletion of Predicted Binding Sites for Various Transcription Factors in the 15 Chromatin Types

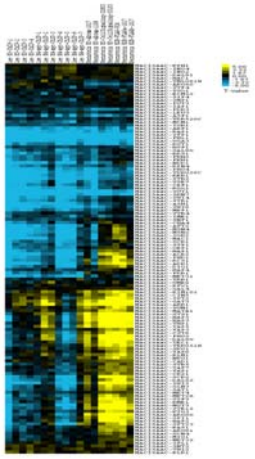
Differences in local chromatin structure must have their origin in differences in DNA sequence. Therefore, we performed an unbiased search for DNA sequence features that may explain the dependence of chromatin type on promoter sequence. A library of position weight matrices generated by MacIsaac *et al.* (2006) was used to screen for differences in predicted TF binding affinity across 15 chromatin types.



Transcription factors whose binding sites are significantly enriched in the "hotspots" (CT15) but not in any other chromatin types include Rap1p and Pdr1p. MatrixREDUCE software (Foat *et al.*, 2006) was used to predict promoter affinity for 125 yeast transcription factors whose sequence specificity had been modeled as weight matrixes by MacIsaac *et al.* (2006). For each chromatin type, color-coded T-values quantify the difference in mean between the promoter regions of genes of a specific chromatin type and those all other genes.

Correlation between Nucleosome Occupancy and the Predicted Affinity for Specific TFs

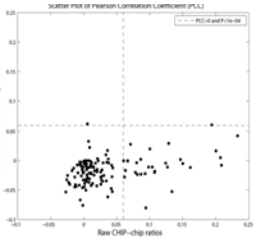
In the above analysis, we found these "hotspots" (CT15) of TF occupancy are strongly nucleosome-depleted and predicted affinities for Rap1p and Pdr1p are only enriched in CT15. Therefore, we were interested to see whether there is a direct link between nucleosome occupancy and the predicted affinity for specific TFs.



Transcription factors Rap1p and Abf1p show a strong negative correlation with nucleosome occupancy. Shown is the Pearson correlation coefficient between predicted transcription factor affinity and nucleosome density, transformed to T-value.

The Effect of Normalization across CHIP-chip Experiments

In the previous figures, we observed a large number of TFs showing significant genome-wide correlation between CHIP-chip data and predicted affinity for TFs. Here we asked whether similar correlation exists for the mean-centered CHIP-chip data provided by Harbison *et al.* (2004).



Shown is the 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.

Conclusions

- In yeast, 5% of all promoter regions are "hotspots" of TF occupancy. These regions are strongly nucleosome-depleted, and preferentially targeted by chromatin-remodeling complex and origin-replication complex (ORC).
- The "hotspots" are also the only chromatin type that is significantly enriched for predicted Rap1p and Pdr1p binding sites.
- A significant genome-wide correlation exists between CHIP-chip enrichment and the predicted affinity for Rap1p and Pdr1p. This correlation is absent when a normalization across CHIP-chip experiments is performed for each promoter region.

References

- Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J, Jennings EG, Zeitlinger J, Pokholok DK, Kellis M, Rolfe PA, Takusagawa KT, Lander ES, Gifford DK, Fraenkel E, Young RA (2004). Transcriptional regulatory code of a eukaryotic genome. *Nature* 431(7004):99-104.
- Foat BC, Morozov AV, Bussemaker HJ (2006). Statistical mechanical modeling of genome-wide transcription factor occupancy data by MatrixREDUCE. *Bioinformatics* 22(14):e141-9.
- Moorman C, Sun LV, Wang J, de Wit E, Talhout W, Ward LD, Greil F, Lu XJ, White KP, Bussemaker HJ, van Steensel B (2006). Hotspots of transcription factor colocalization in the genome of Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* 103(32):12027-32.
- MacIsaac KD, Wang T, Gordon DB, Gifford DK, Stormo GD, Fraenkel E (2006). An improved map of conserved regulatory sites for *Saccharomyces cerevisiae*. *BMC Bioinformatics* 7:113.