CANCER GENOMICS & PROTEOMICS 5: 55-62 (2008)

Modelling Gene Regulation Networks *via* Multivariate Adaptive Splines

XIANG CHEN and HEPING ZHANG

Department of Epidemiology and Public Health and Collaborative Center for Statistics in Science Yale University School of Medicine, New Haven, CT 06520-8034, U.S.A.

Abstract. *After the completion of sequencing for dozens of genomes, as well as the draft of human genome, a major challenge is to characterize genome-wide transcriptional regulation networks. Identification of regulatory functions for transcription factor binding sites in eukaryotes greatly enhances our understanding of the networks, as it has been done extensively under various physiological conditions in yeast. We propose a novel approach based on multivariate adaptive splines to modelling regulatory roles of motifs in gene expression time series data. By applying the proposed approach on two meiotic datasets, we identified well-documented motifs as well as some novel putative motifs that are involved in the transcriptome reprogramming. In addition to identifying single regulatory motifs, we also modelled and unravelled motifs that manifest their effects through coupling with others in regulatory networks. Our findings reveal the potential of multivariate adaptive splines in deciphering complex and important transcriptional regulatory networks in eukaryotes.*

Gene transcription is primarily regulated by selective binding of transcription factors to sequences in cis-regulatory regions of genes (transcription factor binding sites, or motifs) (1). The selective binding between the transcription factors and their corresponding binding sites in gene regulatory regions is fundamental for complicated gene expression regulation in living organisms. Thus, we expect that genes directly regulated by a certain transcription factor should contain its binding sequence in their regulatory regions. With the development of high throughput technology, increasing amount of sequence data is available and dozens of genomes have been completely sequenced. This advancement enables systematic search for putative regulatory motifs by analyzing the regulatory regions of genes (2, 3). Genes with the same regulatory motif are likely to exhibit similar expression profiles. With revolutionary microarray technology, global transcription analyses allow simultaneous measures of the expression levels of thousands of genes (possibly over time) in a single experiment (4-6), which could potenitally capture the response of the entire transcriptome for a signaling process.

Mitosis, meiosis, dauxic shift and many other cellular processes involve dramatic reprogramming of the whole transcriptome in hours. Global measuring of the transcription levels at different times in the process provides abundant information for dissecting the control network. Many researches have focused on correlating the gene expression level with the existence of motifs (7-9). Despite the successes in identifying important motifs and motif combinations, most approaches do not account for the time-dependent impact of different motifs, although it is well established that the binding of transcription factors and their corresponding motifs in cellular processes are transient and precisely regulated.

In this report we propose an approach to comprehensively analyzing the time-dependent motif impact in these datasets using adaptive splines technique to model the relationship between temporal gene expression and the presence of regulatory motifs. This approach employs nonparametric regression modelling framework for identifying systematic variation from random variation in the gene expression profile. It contrasts the expression levels between the set of genes with a certain combination of covariates (regulation motifs here) and the rest of genes in the genome. The resulting regression models provide an accurate description for the underlying regulation role of motifs.

The multivariate adaptive splines for the analysis of longitudinal data (MASAL), originally designed for analysis of longitudinal data, uses the technique of the multivariate regression splines (MARS) (10) for the estimation of mean curves (11, 12). The key difference between MASAL and MARS is that MASAL accommodates correlated data (such as time course data) whereas MARS assumes independent

Correspondence to: Heping Zhang, Department of Epidemiology and Public Health and Collaborative Center for Statistics in Science Yale University School of Medicine, New Haven, CT 06520-8034, U.S.A.

Key Words: Transcription factor binding sites, temporal gene expression, MASAL.

observations. MASAL allows unrestricted interactions among covariates and between time and covariates, making it convenient for characterizing gene expression time series data. The method fits easily interpretable models to the expression data.

Materials and Methods

Data acquisition. Occurrences of known and putative regulatory motifs in upstream regulatory region of yeast genes were downloaded from Pilpel *et al.* (2), which includes 43 binding sites for known regulatory proteins and 313 putative motifs that are overrepresented in the regulatory regions. To eliminate sporadic findings, the temporal expression profiles for yeast sporulation were obtained from two independent sources (4, 5).

Data processing. We created three binary variables to indicate whether the motif is in a gene promoter or not, whether in the coding strand or not, and whether in the complement strand or not. For motif k and gene i, we use x_{ki0}, x_{ki1}, and x_{ki2} to denote the three variables, respectively. When the meaning was clear, we used x_{ki} to replace x_{ki0} in the text. Genes without motifs in their promoter regions are irrelevant to the goal of the present work, and hence their expressions are not analyzed.

A transformed expression level was obtained by taking the logarithm of the ratio of the expression level at the current time to that at the baseline point for every gene. At each time point, the log-transformed gene expression level was centralized by subtracting the mean over all genes from it. This centralization is to remove potential overall trends of expression during the large transcriptional reprogramming, allowing us to focus on the contribution of individual regulatory motif to transcription profiles. For convenience, we refer to the transformed and centralized expression levels simply as expression levels.

Search for co-occurring motifs. The co-occurrence of motifs in the yeast genome was assessed by a cumulative hypergeometric distribution (13). The co-occurrence of the two motifs were considered significant beyond chance if the probability of observing at least c co-occurrences is less than $1/N_p$ (N_p is the number of motif pairs tested, which is 63190 in the present case).

MASAL application. The effects of motifs, interaction between motifs, and interaction between motif and time were estimated using MASAL program (11, 12), which is available online at: http://peace.med.yale.edu/pub/masal/. We refer to Zhang (11, 12) for detailed discussions of MASAL.

Motif ranking. To measure the importance of active motifs in a process, we introduce a relative importance measure

$$R_k = \frac{RVR_k}{Freq_k} = \frac{1 - RSS_k / RSS_0}{G_k / G},$$

where RVR_k is the rate of reduction in the sum of residual squares by motif k (including k_0, k_1 and k_2), $Freq_k$ is the frequency of motif in the genome (counted at most once for each gene). RSS_k is the sum of residual squares when motif k (including k_0, k_1 and k_2) is a single predictor in a MASAL model, RSS_0 is the sum of squares of the responses, G_k is the number of genes in the dataset containing motif k, and G is the total number of genes in the dataset.

Table I. *Results for simulated data.*

	Category 1			Category 2			Category 3		
	NAM	TP	FP	NAM	TP	FP	NAM	TP	FP
MASAL									
FWER									
0.01		190	1		430	1		662	0
0.05		201	4		492	8		703	3
	294			778			1265		
FDR									
0.01		196	2		497	11		725	4
0.05		224	28		547	58		829	34
Reg-Tree		129	247		206	269		273	295

NAM: Total number of active motifs in 100 simulations; TP: total number of true positive; FP: total number of false positive; FWER: family-wise error rate; FDR: false discovery rate.

Control of false-discovery in MASAL. In ordinary regression, variables are usually selected based on a significance test that depends on RVR_k. Due to the adaptive nature of MASAL, the validity of a significance test is difficult to prove. Instead, we carried out a simulation study to determine the threshold of RVR_k so that the false discovery is controlled at 1 (with a less than 5% false discovery rate) or the false discovery rate is controlled at 5%.

Data simulation. To evaluate the effectiveness of MASAL in revealing the underline effects of regulatory effects, a dataset with 500 genes was simulated. The simulated dataset has 49 candidate regulatory motifs and an expression profile of 7 time points (t = 0, 1, 2, ..., 6). We temporarily ignore non-linear interaction among motifs and assume the effects of different motifs are additive. The motifs were randomly assigned to each gene with different occurrence (1%, 2%, 4%, 8%, 16%, 32% and 64%, 7 motifs for each occurrence). A binary variable x_{ij} was created to indicate whether motif j occurs in the promoter of gene i. Next, k (Category 1: 1 – 5; Category 2: 6 – 10; Category 3: 11 – 15) active motifs (M_1, M_2, ..., M_k) were randomly selected and the log-transformed expression level of the ith gene is modeled as:

$$Y_{it} = \sum_{j=1}^{k} x_{iM_j} A_j \left(\frac{t - B_j}{\max(t - B_j)} \right)^2 + \varepsilon_{it}$$

$$\varepsilon_{it} \sim N(0,1)$$

where A_j is randomly selected from $\{\pm 1, \pm 3, \pm 5\}$ and B_j is randomly selected from $\{0 : 9\}$. Finally the simulated dataset was centralized as before.

The simulation was repeated 100 times for each category. The threshold for each simulated data was determined as before.

Table II. *Top ranked motifs for the yeast meiotic datasets.*

Dataset	Motif	R_k	Motif	R_k
Meiosis_Chu	Ndt80(MSE)	1.721	m_RPE6	0.146
	Ume6(URS1)	0.557	m_RPE72	0.138
	m_meiosis_orfnum2SD_n3	0.287	m_RPE34	0.131
	m_amino-acid_metabolism_orfnum2SD_n14	0.196	MCB	0.100
	RAP1	0.162	m_RPE49	0.079
Meiosis_Primig	Ndt80(MSE)	1.478	PAC	0.027
	Ume6(URS1)	0.368	mRRPE(M3A)	0.016
	m_meiosis_orfnum2SD_n3	0.183		

Results

Simulation. In order to evaluate the effectiveness of MASAL in revealing the underlying regulatory role of motifs, a simulated dataset of 500 genes with 49 regulatory motif (combined with basic promoter elements) and expression level at 7 time points was generated. We also ran the RegTree, a multivariate regression tree method from (9) on the same simulated data and compared the result with ours. Table I summarizes the result. It clearly demonstrated that under all conditions tested, our approach outperformed the RegTree algorithm by identifying more true active motifs with fewer false positive errors.

Single motif based analysis. Having shown the effectiveness of the proposed method, we examined the contribution of single motifs to the transcriptional reprogramming during yeast meiosis process.

We identified 57 motifs active in the meiosis process from the Meiosis_Chu dataset. The top 10 motifs for the Meiosis_Chu dataset were listed in Table II. A full list of active motifs is displayed in Table III. Among these active motifs, MSE, URS1, RAP1, m_RPE72, MCB, RPN4, SFF, m_regulation_of_amino-acid_metabolism_orfnum2SD_n11, PAC, mRRPE and m_other_protein-destination_activities_orfnum2SD_n7 were among the 20 top ranked motifs in (9), where the same dataset was analyzed by the RegTree program.

Only 5 active motifs were identified in the Meiosis_Primig dataset, all of them active in the Meiosis_Chu dataset. The small number of active motifs could be largely explained by the fact that the W303 strain used in this experiment exhibited lower degree of synchrony (5) compared to the SK1 strain used in the Meiosis_Chu dataset. All 5 active motifs in the Meiosis_primig dataset were included in Table II.

The top three ranked motifs (Table II) were identified in both datasets: URS1, binding site for Ume6p (14), MSE, binding site for Ndt80p (15) and a putative motif, m_meiosis_orfnum2SD_n3. This result suggested their vital roles in meiosis, as it was unlikely that they were coincidentally selected as the top ranked motifs by chance from two independent datasets. Figure 1a compared the modeled expression profile of a gene in the Meiosis_Chu dataset that is regulated by MSE with those do not have MSE in their promoter. Figure 1b illustrated the density plots estimated from the dataset for genes with MSE in their promoter and those without MSE. It is clear from both the modeled profiles and the actual density plots that most genes with MSE in their promoters had an elevated expression after 7 hours.

Table IV showed the modeled expression profiles for genes controlled by motifs URS1, MSE, and m_meiosis_orfnum2SD_n3 in meiosis. As can be seen, genes controlled by these motifs were up-regulated. Comparing the models derived from the two datasets, we can see that the onsets of various meiosis events were delayed in the W303 strain compared to the SK1 strain, which is consistent with the literature (5). The selection of m_meiosis_orfnum2SD_n3 motif provided the first direct evidence that this motif is also involved in the control of gene transcription during meiosis, which is suggested by the MIPS annotated name (16).

It is noteworthy that the RegTree program (9) failed to identify motif m_meiosis_orfnum2SD_n3. Furthermore, even though RegTree identified motifs MSE and URS1, they reported a low relative importance for these two motifs (ranked 10th and 8th, respectively). In contrast, we found these two motifs to be the most vital motifs in meiosis. Current biological knowledge (17) supports our finding. Compared to the method used by RegTree, our method provided explicit information as to when a motif is active and how it regulates the gene expression (Table IV).

Effects of motif pairs. Eukaryotes exhibit highly complex regulation of gene expression, which usually occurs through multiple transcription factor binding events (18, 19). In

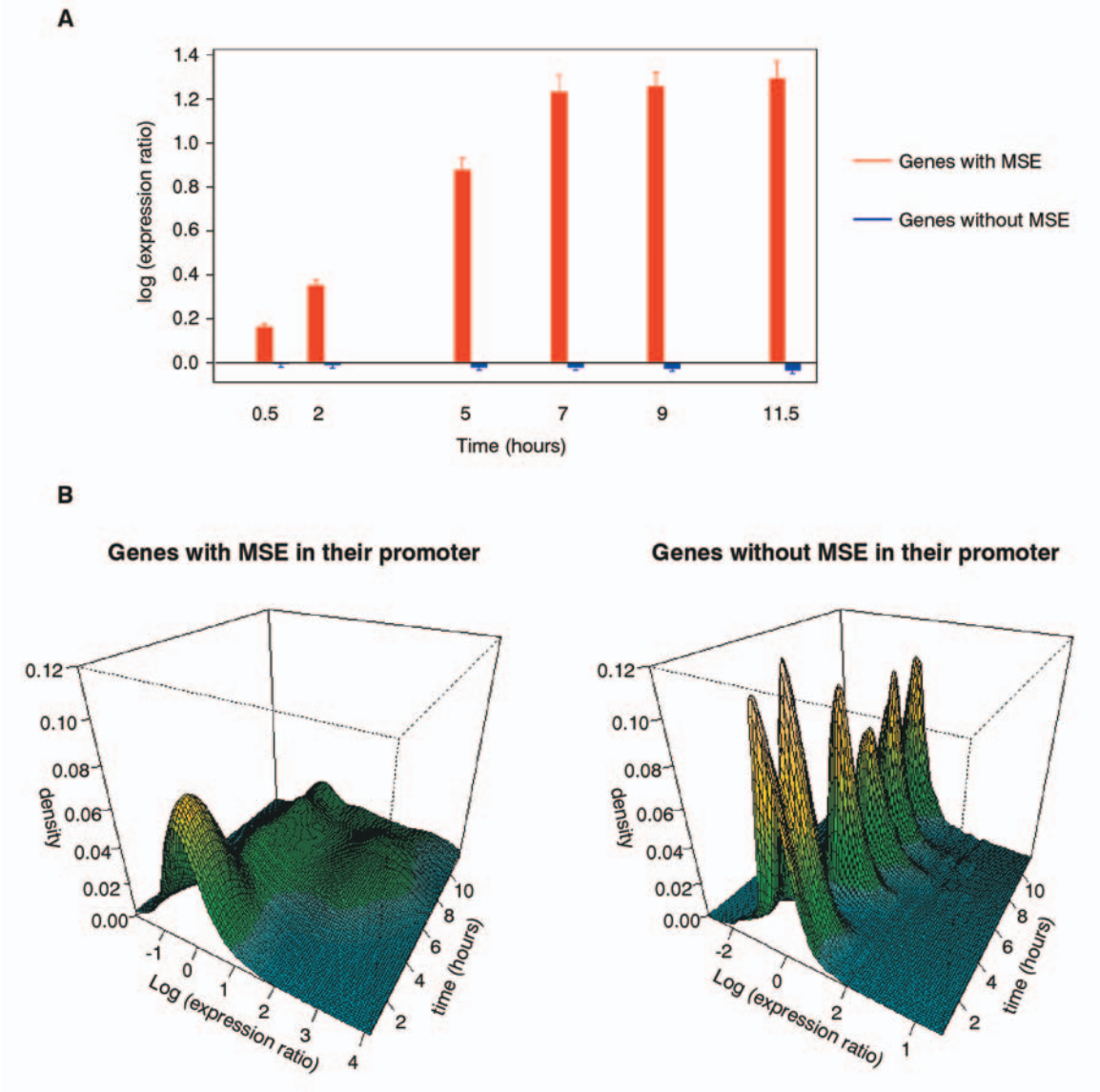

Figure 1. *Expression profile of genes regulated by MSE in the Meiosis_Chu dataset. A) Modelled expression profile (with 95% confidence band) of genes with MSE (red bars) and without MSE (blue bars) in their promoter. B) Density plot estimated from actual expression profile of genes with MSE (left) and without MSE (right) in their promoter.*

addition, it has been suggested in the literature that genes with common pairs of motifs are much more likely to have similar expression profiles than those with a single common motif (2, 20). As we have shown that time-dependent active motifs could be identified by the single motif based approach, we further examined the possible co-regulations of motif pairs to fully exploit MASAL's capability of capturing potential time-varying interactions among motifs.

The potential combinatorial effects of significantly co-occurring motif pairs could be divided into several categories: 1) both motifs are active individually and show additive effects when combined; 2) both motifs are active individually and show synergistic effects when combined; 3) one motif does not have effects by itself but has an add-on effect when combined with an active motif; and 4) both motifs are inactive individually but have combinatorial effects.

Using the Meiosis_Chu dataset, we successfully identified pairs of motifs for each of three categories mentioned above (except for the last category) and Table V summarizes the models for examples of each category.

The mRRPE and PAC pair was also identified previously (2, 7, 9). The RAP1 and m_RPE6 pair was reported in the supplementary table for (7) as well. The co-regulation of a pair of a mitotic motif (SCB) and a meiototic motif (URS1)

Table III. *Active motifs for the Meiosis_Chu dataset.*

Motif	R_k	Motif	R_k
Ndt80(MSE)	1.720623	m_c-compound_and_carbohydrate_metabolism_orfnum2SD_n8	0.041197
Ume6(URS1)	0.556861	m_regulation_of_amino-acid_metabolism_orfnum2SD_n11	0.040407
m_meiosis_orfnum2SD_n3	0.287138	m_metabolism_of_energy_reserves_orfnum2SD_n8	0.038027
m_amino-acid_metabolism_orfnum2SD_n14	0.196123	m_MERE4	0.03773
RAP1	0.162146	m_lipid_and_fatty-acid_binding_orfnum2SD_n13	0.036808
m_RPE6	0.145759	m_fermentation_orfnum2SD_n18	0.034295
m_RPE72	0.137831	PAC	0.034138
m_RPE34	0.1311	m_stress_response_orfnum2SD_n4	0.033674
MCB	0.0997	m_homeostasis_of_metal_ions_orfnum2SD_n20	0.029755
m_RPE49	0.079377	m_cell_death_orfnum2SD_n16	0.028616
m_RPE69	0.075252	m_anion_transporters_orfnum2SD_n4	0.028091
m_organization_of_centrosome_orfnum2SD_n5	0.074267	mRRPE	0.026354
m_c-compound_and_carbohydrate_utilization_orfnum2SD_n31	0.073677	m_amino-acid_degradation_orfnum2SD_n24	0.025683
m_fermentation_orfnum2SD_n5	0.070476	m_amino-acid_transport_orfnum2SD_n3	0.024604
m_other_cation_transporters_orfnum2SD_n5	0.070011	m_OCSE15	0.024459
m_stress_response_orfnum2SD_n24	0.068762	Gcr1	0.024096
RPN4	0.068317	m_amino-acid_transport_orfnum2SD_n20	0.021139
m_other_energy_generation_activities_orfnum2SD_n20	0.066691	m_ion_transporters_orfnum2SD_n3	0.019908
m_sugar_and_carbohydrate_transporters_orfnum2SD_n6	0.066119	m_amino-acid_transporters_orfnum2SD_n11	0.019676
m_MERE11	0.06387	m_anion_transporters_orfnum2SD_n22	0.018758
m_RPE11	0.05846	m_nucleotide_transport_orfnum2SD_n9	0.018729
HAP234	0.057319	m_glyoxylate_cycle_orfnum2SD_n7	0.018273
m_RPE68	0.054025	m_regulation_of_lipid_fatty-acid_and_isoprenoid_biosynthesis_orfnum	0.016749
m_tricarboxylic-acid_pathway_orfnum2SD_n3	0.052528	m_anion_transporters_orfnum2SD_n15	0.016356
m_cell_rescue_defense_cell_death_and_ageing_orfnum2SD_n20	0.049419	m_other_cell_growth_cell_division_and_dna_synthesis_activities_orfn	0.015998
m_nitrogen_and_sulphur_metabolism_orfnum2SD_n17	0.046735	m_other_protein-destination_activities_orfnum2SD_n7	0.01281
m_assembly_of_protein_complexes_orfnum2SD_n23	0.046647	m_glyoxylate_cycle_orfnum2SD_n8	0.012414
m_MERE17	0.045806	SFF	0.002611
STRE	0.044788		

in meiosis process was not reported previously although Honigberg and Purnapatre noted that many early meiotic genes do contains an SCB box in their protmoters (21). The exact mechnisms of co-regulation between URS1 and SCB warrant further investigation.

In summary, these results indicated that MASAL was capable of identifying critical motifs, and they also demonstrated that the analyses based on single and multiple motifs complement each other by providing different insights.

Discussion

After completion of sequencing for dozens of genomes as well as the draft of human genome, a major challenge is to illustrate the genome-wide transcriptional regulation network. Multiple transcription factors together with short consensus DNA sequences in the promoter (motifs) that act as their binding targets constitute fundamental components of the network. Identification of regulatory functions for these motifs would greatly enhance our understanding of the whole network and has been examined extensively under different physiological conditions in yeast (4, 5, 22). Current approaches (7-9) have successfully identified critical motifs presented but have fallen short in revealing the direct evidence of relationship between the presence of a certain motif and the gene expression change in the process.

We have presented a new and easily interpretable method to directly link transcription binding sites to a certain biological process from the expression data. By using simulated dataset as well as meiotic datasets as studying

Table IV. *Estimates of MASAL models for different motifs in meiosis.*

Motif (X)	Data sets					
	Meiosis_Chu			Meiosis_Primig		
	Motif or time-motif interaction	Regression Coefficient	SE	Motif or time-motif interaction	Regression coefficient	SE
X=MSE	Intercept	–0.0027	0.0056	Intercept	0.0002	0.0106
	X_0*t	0.1661	0.0060	X_0*(t-3.48)+	0.3096	0.0130
	X_0*(t-7)+	–0.1440	0.0208	X_0*(t-9.18)+	–0.3672	0.0375
	T	–0.0032	0.0009	(t-2.01)+	–0.0055	0.0010
				X_1*(t-3.48)+	0.0485	0.0129
X=URS1	Intercept	–0.0130	0.0076	Intercept	–0.0144	0.0098
	X_0	0.1954	0.0596	X_0*t	0.2240	0.0153
	X_0*t	0.3120	0.0210	X_1*(t-5.01)+	–0.2166	0.0234
	X_0*(t-3.34)+	–0.2283	0.0276	X_2*t	–0.0535	0.0107
	X_2*t	–0.1247	0.0213			
	X_1*(t-3.34)+	–0.1295	0.0299			
X=m_meiosis_orfnum2SD_n3	Intercept	–0.0162	0.0077	Intercept	–0.0162	0.0099
	X_0	0.1670	0.0467	X_0*t	0.1528	0.0115
	X_0*t	0.2139	0.0158	X_0*(t-5.24)+	–0.1470	0.0182
	X_0*(t-3.40)+	–0.1725	0.0211	X_2*t	–0.0418	0.0084
	X_2*t	–0.0721	0.0156			
	X_1*(t-3.40)+	–0.0807	0.0221			

(t-a)+ = max(t-a, 0). X_0 =1 if motif X is present in the gene promoter and 0 otherwise. X_1 =1 if Motif X is present in + chain of the gene promoter and 0 otherwise. X_2 =1 if Motif X is present in - chain of the gene promoter and 0 otherwise.

cases, we have demonstrated that 1) compared to other approaches, our approach has higher sensitivity as well as higher specificity; 2) important regulatory motifs that are critical to these processes could be identified from the constructed models; and 3) their effects in these processes and timing of the effects are revealed by the easily interpretable models.

Unlike many genome-wide expression analyses (4, 5, 9), we did not exclude those genes with no significant changes in mRNA level during the time course of studies. While it creates analytic challenges to consider genes with relatively stable expressions, inclusion of all genes more realistically reflects the underlying physiological condition.

A prominent advantage of the MASAL is to allow unrestricted interactions among covariates and time, which is important in motif analysis because it is reported that motif combinations have significantly stronger co-regulations of downstream genes (2, 20, 23). Furthermore, our approach can be used to explore higher order combinatorial regulations of motifs.

In the following, we discuss further how we can gain biological insights from MASAL models based on what we reported in the Results section.

URS1 is one of related binding target sequences for Ume6p (14) (which further recruit Ime1p) and located in the promoter region of many of early meiosis-specific genes (24), consistent with the MASAL model (Table IV). Besides the URS1 motif, many early meiosis genes also contain MCB and SCB in their promoter regions. While transcription level of some genes in meiosis is known to be regulated by MCB box (25), which is in consistent with our results, the effects of SCB box in meiosis is less obvious. Our result confirmed that the effects of SCB itself in meiosis transcription regulation are not significant. However, genes with URS1 combined with SCB showed additional induction in the early stage of meiosis compared to genes with URS1 alone (Table V). A potential transcription factor involved in this case is Sok2p, an apparently negative transcription regulator, which binds SCB box and represses gene expression in the presence of glucose. When yeast cells are growing in sporulation medium where acetate is the sole carbon source, Ime1p associates with the N-terminal domain of Sok2p, result in relieving its repression activity and further converting Sok2p into a weak activator (26). There is currently no conclusive evidence of whether Ume6p/Ime1p interacts with Sok2p/Ime1p and our result invites for investigations.

In conclusion, we have presented in this paper a novel and important approach for analyzing the regulatory role of DNA motifs at genome level. Given the high-throughput

Table V. *Models examples of combinatorial regulation of motif pairs in the Meiosis_Chu dataset.*

Pairs of motifs	Motif(s) or time-motif(s) interaction	Regression coefficient	SE
RAP1	Intercept	0.0141	0.0076
m_RPE6	$RAP1_2$	–0.2534	0.0536
(Category 1)	$RAP1_2$**t*	–0.1737	0.0161
	$RAP1_2$**(t-3.28)*$_+$	0.2374	0.0218
	m_$RPE6_0$**t*	–0.0570	0.0091
	m_$RPE6_0$**(t*-6.43)$_+$	0.1126	0.0215
MRRPE	Intercept	0.0369	0.0082
PAC	$mRRPE_0$	–0.1255	0.0242
(Category 2)	mRRPE0 *PAC_0	–0.2083	0.0471
	$mRRPE_0$ *(*t*-7)$_+$	0.0610	0.0118
	$mRRPE_0$ *(*t*-5)$_+$	–0.0419	0.0080
	PAC_2	–0.1739	0.0436
	PAC_2**t*	0.0204	0.0027
SCB	Intercept	–0.0142	0.0076
URS1	$URS1_0$	0.1892	0.0589
(Category 3)	$URS1_0$**t*	0.2787	0.0213
	$URS1_0$**(t-3.37)*$_+$	–0.1873	0.0284
	$URS1_2$**t*	–0.1278	0.0212
	$URS1_0$*SCB_2**t*	0.1307	0.0230
	$URS1_1$**(t-3.37)*$_+$	–0.1347	0.0299
	SCB_1* SCB_2**(t-3.37)*$_+$	0.0212	0.0052
	$URS1_0$*SCB0**(t-3.37)*$_+$	–0.1462	0.0344
	$URS1_0$*SCB1**(t-3.37)*$_+$	0.1375	0.0332

(*t-a*)+ = max(*t* – *a*, 0); X_0 =1 if motif X is present in the gene promoter and 0 otherwise; X_1 =1 if motif X is present in + chain of the gene promoter and 0 otherwise; X_2 =1 if motif X is present in - chain.

technologies for studying transcriptome, such approaches would be extremely useful in decipher the complex transcriptional regulation network.

Acknowledgements

This research is supported in part by grants K02DA017713 and R01DA016750 from the National Institutes on drug abuse.

References

1 Mitchell PJ and Tjian R: Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science *245*: 371-378, 1989.
2 Pilpel Y, Sudarsanam P and Church GM: Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet *29*: 153-159, 2001.
3 Zavolan M *et al*: SMASHing regulatory sites in DNA by human-mouse sequence comparisons. Proc IEEE Comput Soc Bioinform Conf *2*: 277-286, 2003.
4 Chu S *et al*: The transcriptional program of sporulation in budding yeast. Science *282*: 699-705, 1998.
5 Primig M *et al*: The core meiotic transcriptome in budding yeasts. Nat Genet *26*: 415-423, 2000.
6 Cho RJ *et al*: Transcriptional regulation and function during the human cell cycle. Nat Genet *27*: 48-54, 2001.
7 Hvidsten TR *et al*: Discovering regulatory binding-site modules using rule-based learning. Genome Res 15: 856-866, 2005.
8 Das, D, Banerjee N and Zhang MQ: Interacting models of cooperative gene regulation. Proc Natl Acad Sci USA *101*: 16234-16239, 2004.
9 Phuong TM, Lee D and Lee KH: Regression trees for regulatory element identification. Bioinformatics *20*: 750-757, 2004.
10 Friedman JH: Multivariate Adaptative Regression Splines. Annuals of Statistics *19*: 1-141, 1991.
11 Zhang H: Multivariate adaptive splines for longitudinal data. J Comput Graphic Stat *6*: 74-91, 1997.
12 Zhang H: Analysis of infant growth curves using multivariate adaptive splines. Biometrics *55*: 452-459, 1999.
13 Lee TI *et al*: Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*: 799-804, 2002.
14 Buckingham LE *et al*: Nucleotide sequence and promoter analysis of SPO13, a meiosis-specific gene of Saccharomyces cerevisiae. Proc Natl Acad Sci USA *87*: 9406-9410, 1990.
15 Chu S and Herskowitz I: Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. Mol Cell *1*: 685-696, 1998.
16 Mewes HW *et al*: MIPS: analysis and annotation of proteins from whole genomes. Nucleic Acids Res *32*: D41-44, 2004.
17 Vershon AK and Pierce M: Transcriptional regulation of meiosis in yeast. Curr Opin Cell Biol *12*: 334-339, 2000.
18 Kel OV *et al*: A compilation of composite regulatory elements affecting gene transcription in vertebrates. Nucleic Acids Res *23*: 4097-4103, 1995.
19 Halfon MS *et al*: Ras pathway specificity is determined by the integration of multiple signal-activated and tissue-restricted transcription factors. Cell *103*: 63-74, 2000.
20 Beer MA and Tavazoie S: Predicting gene expression from sequence. Cell *117*: 185-198, 2004.
21 Honigberg SM and Purnapatre K: Signal pathway integration in the switch from the mitotic cell cycle to meiosis in yeast. J Cell Sci *116*: 2137-2147, 2003.
22 Wang Y *et al*: Ras and Gpa2 mediate one branch of a redundant glucose signaling pathway in yeast. PLoS Biol *2*: E128, 2004.
23 Segal E, Yelensky R and Koller D: Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. Bioinformatics *19(Suppl 1)*: i273-282, 2003.
24 Strich R *et al*: UME6 is a key regulator of nitrogen repression and meiotic development. Genes Dev *8*: 796-810, 1994.
25 Raithatha SA and Stuart DT: Meiosis-specific regulation of the Saccharomyces cerevisiae S-phase cyclin CLB5 is dependent on MluI cell cycle box (MCB) elements in its promoter but is independent of MCB-binding factor activity. Genetics *169*: 1329-1342, 2005.
26 Leem SH *et al*: Meiotic role of SWI6 in Saccharomyces cerevisiae. Nucleic Acids Res *26*: 3154-3158, 1998.

Received October 30, 2007
Accepted November 14, 2007