Published online September 23, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 17 ***4955–4961***
doi:10.1093/nar/gkh816

In silico identification of transcriptional regulators associated with c-Myc

Ran Elkon, Karen I. Zeller[2], Chaim Linhart[1], Chi V. Dang[2,3], Ron Shamir[1] and Yosef Shiloh*

The David and Inez Myers Laboratory for Genetic Research, Department of Human Genetics, Sackler School of Medicine, [1]School of Computer Science, Tel Aviv University, Israel, [2]Division of Hematology, Department of Medicine and [3]The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD 21212, USA

Received July 22, 2004; Revised and Accepted August 21, 2004

ABSTRACT

The development of powerful experimental strategies for functional genomics and accompanying computational tools has brought major advances in the delineation of transcriptional networks in organisms ranging from yeast to human. Regulation of transcription of eukaryotic genes is to a large extent combinatorial. Here, we used an *in silico* approach to identify transcription factors (TFs) that form recurring regulatory modules with c-Myc, a protein encoded by an oncogene that is frequently disregulated in human malignancies. A recent study identified, on a genomic scale, human genes whose promoters are bound by c-Myc and its heterodimer partner Max in Burkitt's lymphoma cells. Using computational methods, we identified nine TFs whose binding-site signatures are highly overrepresented in this promoter set of c-Myc targets, pointing to possible functional links between these TFs and c-Myc. Binding sites of most of these TFs are also enriched on the set of mouse homolog promoters, suggesting functional conservation. Among the enriched TFs, there are several regulators known to control cell cycle progression. Another TF in this set, EGR-1, is rapidly activated by numerous stress challenges and plays a central role in angiogenesis. Experimental investigation confirmed that c-Myc and EGR-1 bind together on several target promoters. The approach applied here is general and demonstrates how computational analysis of functional genomics experiments can identify novel modules in complex networks of transcriptional regulation.

INTRODUCTION

With the completion of the sequencing of the human genome and genomes of many other organisms, research efforts are shifting to functional genomics, i.e. deciphering how components of physiological networks interact and regulate each other in cellular systems. The most notable achievements to date have been the delineation of transcriptional regulatory networks, which became feasible with the advent of genome-wide experimental techniques that specifically shed light on this layer of cellular systems (1–4). These techniques include gene expression microarrays and the combined strategy of chromatin immunoprecipitation (ChIP) and promoter microarrays (also termed 'ChIP-on-chip'). The former enables the simultaneous recording of expression levels of thousands of genes (5), and the latter enables identification on a genomic scale of promoters that are bound by specific transcription factors (TFs) under certain conditions, in a single experimental assay (6,7). These novel techniques generate huge volumes of biological data, the exploitation of which is largely dependent on the development of appropriate computational tools.

The ChIP-on-chip approach was recently used to map global TF-promoter binding relationships in yeast under standard growth conditions (6). It was also used in mammalian cells to identify genome-wide direct targets of multiple TFs (8–10). Seeking direct targets of c-Myc, Li *et al*. (8) identified 776 human genes in Burkitt's lymphoma cells whose promoters are bound by the oncoprotein c-Myc and its heterodimer partner Max. c-Myc regulates cell-cycle proliferation, apoptosis and differentiation. Overexpression of c-Myc is one of the most common alterations in human cancer, yet it is not clear how it promotes malignant transformation (11–14). It is widely accepted that the transcriptional regulation activity of c-Myc is critical for the development of malignancy associated with it, but the target genes that mediate this process remain elusive.

Regulation of transcription of eukaryotic genes is to a large extent combinatorial, i.e. the conditions under which a gene is expressed are determined using an intricate interplay of multiple positive and negative transcriptional regulators that recognize and bind to *cis*-regulatory elements within and beyond the gene's promoter region. Therefore, a major goal in deciphering transcriptional regulation networks is to identify combinations of TFs that functionally cooperate to form regulatory modules. Such modules can be recognized by the co-occurrence of the corresponding TF-binding sites in the same promoters (1,15,16). Our aim in this work is to identify TFs that form recurrent *cis*-regulatory modules with c-Myc. To this end, we computationally analyzed the promoters

*To whom correspondence should be addressed at Department of Human Genetics and Molecular Medicine, Sackler School of Medicine, Room 1002, Tel Aviv University, Ramat Aviv 69978, Israel. Tel: +972 3 6409760; Fax: +972 3 6407471; Email: yossih@post.tau.ac.il

reported by Li *et al*. (8) to be bound by c-Myc. We report on nine TFs whose binding signatures are significantly over-represented in this set of promoters, including several TFs known to play important roles in the regulation of cell cycle progression. The results point to possible functional links between these TFs and c-Myc, and offer an additional explanation for the positive effect of c-Myc on cellular proliferation. Another TF whose binding signature was highly enriched in the c-Myc target promoters is EGR-1, which is rapidly activated by many types of stress, including hypoxia, DNA damage and vascular injury, and has a central role in angiogenesis (17,18). We demonstrate experimentally that c-Myc and EGR-1 bind together several common target promoters. The approach applied here is general and demonstrates the potential of computational promoter analysis of ChIP-on-chip data for the identification of novel transcriptional modules.

MATERIALS AND METHODS

Extraction of putative promoter sequences from human and mouse genomes

Putative promoter sequences corresponding to all known human and mouse genes were extracted from the human and mouse genome sequences (Ensembl, version 13, December 2002) using a Perl script based on the application programming interface provided by the Ensembl project (19). For each gene flagged as a 'known gene', a genomic sequence was extracted that spanned 1000 bp upstream to 200 bp downstream of the gene's putative transcription start site (TSS). The extracted sequences were masked for repetitive elements. A total of 19 351 and 18 748 putative promoters were extracted from the human and mouse genome, respectively. The sequences of these promoter sets can be downloaded from http://www.cs.tau.ac.il/~rshamir/prima/PRIMA.htm.

To avoid biases due to highly similar promoters, we constructed for each organism a non-redundant set of promoters by running all-against-all BLAST comparisons. For every promoters pair with BLAST *E*-score $<10^{-50}$, we excluded one of the members from the non-redundant set. The non-redundant human and mouse promoter sets contain 17 390 and 15 521 promoters, respectively.

Computational identification of enriched TFs

The computational analysis used PRIMA software that is described in detail in (16) and is available at http://www.cs.tau.ac.il/~rshamir/prima/PRIMA.htm. In short, given target and background sets of promoters, PRIMA performs statistical tests aimed at identifying TFs whose binding sites are significantly more abundant in the target set than in the background set. PRIMA uses position weight matrices (PWMs) as models for regulatory sites that are bound by TFs. A total of 300 PWMs that represent human or mouse TF-binding sites were obtained from the TRANSFAC database (20). The entire collection of non-redundant human promoters was used as the background set in PRIMA tests. For the analysis of the mouse orthologs, the entire collection of non-redundant mouse promoters was used as the background set. Human–mouse homolog genes were determined using the EnsMart utility provided by Ensembl (21). Default parameters were used in all PRIMA runs. The program scanned both strands of each promoter for putative binding sites, i.e. for sites with high similarity score to the PWM. Throughout the text, we term such high scoring sites as 'hits' of the corresponding TF. Thresholds of similarity scores for declaring hits correspond to an average rate of one hit per 10 000 bp of random promoter sequences (16).

ChIP assays

The cell line used for all ChIP analyses was the human myelogenous leukemia K562 cells. The cells were grown in RPMI 1640 medium supplemented with 10% fetal bovine serum and antibiotics. The cells growing in log phase were formaldehyde cross-linked and chromatin was immunoprecipitated as described previously (22). An aliquot of 2 μg of rabbit polyclonal c-Myc and HGF antibodies (anti-Myc sc-764, anti-HGF sc-7949; Santa Cruz Biotechnology, Inc., Santa Cruz, CA) and 5 μg of rabbit polyclonal Egr-1 antibody (sc-189; Santa Cruz Biotechnology, Inc.) were used to precipitate chromatin from 2×10^7 cells.

Real-time PCR

For real-time PCR, a SYBR green core reagents kit (Applied Biosystems) was utilized. Known quantities of total input DNA were used to generate a standard curve for determining the percentage of total input for each ChIP sample. All amplifications were carried out in the linear range of standard curves.

RESULTS

Enriched TF signatures in promoters bound by c-Myc/Max

The data set recently published by Li *et al*. (8) contained 876 human promoters that were bound by c-Myc, 931 promoters that were bound by Max and 776 promoters that were bound by both TFs. To reduce false positive discoveries, we focused on the set of promoters that were bound by both c-Myc and Max, hereafter referred to as the c-Myc/Max target set. As a first step, we extracted from the human genome sequences that correspond to promoters of all known genes (for details see Materials and Methods), limiting the analysis to sequences spanning 1000 bp upstream to 200 bp downstream of the putative TSSs of the corresponding genes. We determined the region of sequence around the putative TSS in which to search for transcriptional regulatory elements by examining the location distribution of 1075 empirically validated TF-binding sites in human promoters [data from TRANSFAC database (20)]. Eighty percentage of these elements were located within the region of our analysis (data not shown). Clearly, present experimental data on TF-binding sites is biased toward binding sites located at short distances from the TSS. Although certain regulatory elements were demonstrated to act over great distances, up to several kilobases from the TSS, it is clear that ample information on mechanisms of transcriptional regulation resides in sequences in close proximity to the TSS.

Our collection of human promoter sequences contains a total of 19 351 promoters, of which a subset of 17 390 promoters is non-redundant (see Materials and Methods). The full and non-redundant promoter sets include sequence data for 615 and 519 genes, respectively, out of the 776 genes in the Myc/Max target set. We applied PRIMA, a promoter analysis software that we developed recently (16), to scan the 519 non-redundant promoters of the Myc/Max target set for overrepresented TF-binding-site signatures. This means that we searched for TFs whose binding-site signatures are significantly more abundant on this promoter set than expected by chance, given their abundance on the entire collection of non-redundant human promoters (for details on PRIMA see Materials and Methods). Such overrepresentation suggests the existence of functional links between the over-abundant TFs and c-Myc/Max. PRIMA uses PWMs as models of regulatory sites that are bound by TFs. A total of 300 PWMs that represent mammalian TF-binding sites (obtained from the TRANSFAC database) were tested, and nine of them were identified by PRIMA to be significantly enriched in the c-Myc/Max target set [$P < 10^{-5}$, and after conservative adjustment for multiple testing $P < 0.05$; Table 1, (A)].

Previously, we used PRIMA to analyze human genes whose expression is cell-cycle-dependent (16). Interestingly, most of the TFs whose signatures were enriched in the cell-cycle-dependent promoter set were also enriched in the c-Myc/Max target set (E2F, NF-Y, Sp1, Nrf1, ETF, CREB and AhR/Arnt). We therefore checked the overlap between the cell cycle and the c-Myc/Max target sets. The cell cycle set contains 568 genes, only 30 of which are common to the c-Myc/Max set. When we analyzed the c-Myc/Max set after deleting these 30 genes, the overrepresentation of all the TFs reported in Table 1 remained significant (data not shown). Thus, the overlap between the results obtained on the two datasets is not explained by common genes, and is probably due to a general role of these regulators in cell cycle progression, where they control different sets of genes in different cell types. The mitogen- and stress-induced ELK-1 and EGR-1 TFs were also enriched on the Myc/Max target set.

Li *et al*. (8) reported, somewhat surprisingly, that only some 25% of the promoters bound by c-Myc/Max contain a core E-box motif (CACGTG), which is directly recognized and bound by the c-Myc/Max heterodimer. In accordance with this observation, PRIMA found that PWMs with canonical E-box core motifs are only slightly enriched on the Myc/Max target set, and they did not pass our significance threshold. However, another PWM that models c-Myc/Max binding sites, whose core motif is a variant of the E-box (CAYGYG, Y = {C,T}), was significantly enriched on this target set (PWM M00322 in TRANSFAC DB) [Table 1, (A)]. It cannot be ruled out that we failed to detect the E-box overrepresentation as a result of the fact that a fraction of these regulatory elements are located outside the analyzed promoter region.

Conservation of TF-binding-site enrichment on mouse ortholog promoters

The enrichment of binding-site signatures of specific TFs in the c-Myc/Max target set raises the possibility that c-Myc/Max and the TFs maintain functional relationships and together form recurrent transcriptional regulation modules that control the expression of numerous genes. In order to strengthen this *in silico*-derived hypothesis, we repeated the tests for TF-binding site enrichment, this time on the set of promoters comprising the mouse orthologs of the human c-Myc/Max target set. Extracting promoter sequences for all known mouse genes, we collected 18 478 mouse promoter sequences, of which 15 521 were non-redundant. The non-redundant set includes sequence data for 407 mouse orthologs of the human Myc/Max target set. Applying PRIMA to this set, we found that seven out of the nine PWMs that were enriched in the human set were also enriched in the mouse set [Table 1, (B)]. The enrichment of the same TF-binding sites on the ortholog sets suggests that the functional relationships between these TFs are conserved between mice and humans.

TFs associated with c-Myc/Max through direct DNA binding

As noted above, a canonical E-box is found in only ~25% of the promoters identified by Li *et al*. (8) as direct targets of c-Myc/Max. Therefore, it was suggested that the c-Myc/Max heterodimer controls its target by two different modes: in the first one, it directly binds its target promoters through its classical E-box element or a variant thereof; and in the second mode, it participates in the regulation of target promoters

Table 1. TF-binding site signature enrichments in c-Myc/Max target sets

TF [PWM accession ID in TRANSFAC DB]	(A) *P*-value for PWM hits' abundance on c-Myc/Max target set	(B) *P*-value for PWM hits' abundance on mouse ortholog set	(C) *P*-value for PWM hits' abundance on mode 1 subset
ETF [M00695]	1.2×10^{-15}	6.8×10^{-13}	3.2×10^{-7}
Sp1 [M00196]	1.7×10^{-14}	8.9×10^{-12}	6.3×10^{-10}
Nrf-1 [M00652]	6.5×10^{-14}	7.9×10^{-11}	3.2×10^{-7}
NF-Y [M00185]	3.2×10^{-12}	Not enriched	1.5×10^{-5}
CREB [M00177]	4.7×10^{-8}	1.4×10^{-7}	Not enriched
c-Myc/Max [M00322]	1.5×10^{-7}	2.8×10^{-6}	5.7×10^{-62}[a]
Egr-1 [M00243]	2.4×10^{-7}	6.7×10^{-5}	3.4×10^{-8}
Elk-1 [M00025]	3.9×10^{-7}	5.5×10^{-11}	Not enriched
E2F [M00516]	5.8×10^{-7}	6.6×10^{-5}	Not enriched
AhR/Arnt [M00237]	6.4×10^{-7}	Not enriched	Not enriched
E-box	8.4×10^{-4}	4.1×10^{-5}	6.5×10^{-7}

[a]c-Myc/Max signature is markedly enriched on 'mode 1' subset, since by definition this set comprises the genes whose promoters were found to have a hit for c-Myc/Max. Not enriched indicates $P > 0.0001$ (*P*-values are before correction for multiple testing).

without binding directly to the DNA, but by physically interacting with other sequence-specific TFs or with components of the general transcription machinery.

We attempted to identify TFs that form *cis*-regulatory modules with c-Myc/Max via the first mode. To this end, we identified a subset of the c-Myc/Max target set, comprising genes whose promoters contain high-scoring putative binding sites (*hits*) for c-Myc/Max. As the c-Myc/Max variant binding element represented by PWM M00322 was more enriched than the canonical E-box, we scanned the c-Myc/Max target set for promoters with hits for this PWM. We identified hits for M00322 in 134 out of 615 promoters, which were good candidates for being regulated by c-Myc/Max through its direct DNA binding (Supplementary Table A). We refer to this subset as 'c-Myc/Max mode1 subset'. Since it is smaller than the complete c-Myc/Max target set, we expected statistical phenomena associated with it to be less significant. However, we observed that one TF, EGR-1, was even more significantly enriched in the mode1 subset than in the complete set [Table 1, (C)]. This makes EGR-1 a strong candidate for being c-Myc/Max's partner in the regulation of mode1 target promoters. Of note, E2F, a pivotal regulator of the transcriptional program associated with cell cycle progression was not enriched in the mode1 subset, suggesting that the c-Myc/Max cooperation with E2F is maintained mainly through mode 2.

EGR-1 binding site enrichment in the GC-rich c-Myc/Max mode1 subset

As noted above, the EGR-1 binding site signature was more enriched in the c-Myc/Max mode1 subset than in the complete c-Myc/Max set. Inspection of EGR-1 PWM (M00243) showed that it has a high-GC-content with a GCGTGGG core. This led us to compare the GC-content of the mode1 subset, the c-Myc/Max target set and the full collection of non-redundant human promoters (Table 2). We observed that the c-Myc/Max target set is more GC-rich than the full set of human promoters (57 versus 53%, respectively, *Z*-score = 9.4), and that the c-Myc/Max mode1 subset is even more GC-rich (60.1%, *Z*-score = 5.8 when compared to GC-content of the c-Myc/Max target set). This raised the question whether the abundance of hits for EGR-1 on the mode 1 subset could be explained merely by its high-GC-content. To address this concern, we generated random PWMs based on the EGR-1 PWM in a way that preserved its GC-content. We generated the random PWMs by permuting the columns of the original PWM, and randomly interchanging A with T and G with C. We then compared the enrichment of five permuted and the original EGR-1 PWM on the c-Myc/Max mode1 subset. Significantly, the original EGR-1 PWM yielded an enrichment score that was far more significant than the scores obtained by the random PWMs generated based on it (Table 3). For further examination, we sorted all TRANSFAC mammalian PWMs according to their GC-content and recorded their enrichment in the mode1 subset. Supplementary Table B lists the results for the top 25 GC-rich PWMs. This list shows that the over-representation of EGR-1 in the mode1 subset is not a mere reflection of its high-GC-content, as there are many PWMs at least as GC-rich as EGR-1 that are not at all enriched on this subset.

Table 2. High-GC-content of the c-Myc/Max mode1 subset

	Number of promoters	%A	%C	%G	%T
Non-redundant human promoters	17 390	23.1	26.4	27.0	23.5
c-Myc/Max target set	615	21.5	28.3	28.7	21.6
c-Myc/Max mode 1 subset	134	19.5	30.4	30.5	19.7

Table 3. Enrichment of the original and permuted EGR-1 PWM on the c-Myc/Max mode 1 subset

PWM	EGR-1	EGR-1 Rand1	EGR-1 Rand2	EGR-1 Rand3	EGR-1 Rand4	EGR-1 Rand5
Enrichment score	4.5×10^{-7}	0.17	0.28	0.25	0.088	0.0034

High-scoring hits for c-Myc/Max (M00322) and EGR-1 (M00243) were found in 134 and 167 promoters, respectively, out of a total of 615 promoters in the c-Myc/Max target set. A total of 54 promoters contained strong hits for both c-Myc/Max and EGR-1 (Supplementary Table C). Finally, we examined the location distribution of c-Myc/Max and EGR-1 hits on the promoters of the c-Myc/Max target set. For both TFs, the computationally identified binding sites are significantly concentrated in the proximity of the TSS and their density drops downstream of it (Figure 1). In contrast, the distribution of hits identified for a random PWM generated by permuting the EGR-1 PWM was quite uniform between −650 and +200 bp with respect to the TSS, as expected for random hits. The fact that the hit distributions for c-Myc/Max and EGR-1 show a prominent peak in the anticipated position is an additional indication of the quality of the human genome TSS annotations (such peaks would have not been obtained if significant deviations had existed between the locations of the annotated and real TSS in a large proportion of the genes), and for the specificity of the hits identified by PRIMA (high false positive rates for hits identified by PRIMA would have obscured the peak of true hits).

Experimental confirmation for co-binding of Myc and EGR-1 in selected promoters

A total of 54 promoters of the c-Myc/Max mode1 subset contained strong hits for EGR-1. To experimentally test our *in silico* derived hypothesis that these two TFs together form a recurrent transcriptional *cis*-regulatory module, we selected six of these genes and examined the binding of both c-Myc and EGR-1 to their promoters using ChIP assays. For all six genes examined (*RAP2B*, *KHSRP*, *PolH*, *PTPN1*, *PP* and *KPNA3*), the signals obtained for both c-Myc and EGR-1 were above the background level ($P < 0.025$, one-tailed *t*-test; Figure 2). For a negative control, we chose from the c-Myc/Max target set the *MCCC2* gene, whose promoter did not contain any hit for EGR-1. In agreement, the signal obtained by the ChIP assay for EGR-1 binding to this promoter was very close to background level. These results demonstrate that c-Myc and EGR-1 co-binding occurs on multiple promoters. Further experiments will be required to establish direct functional link between these two transcriptional regulators.

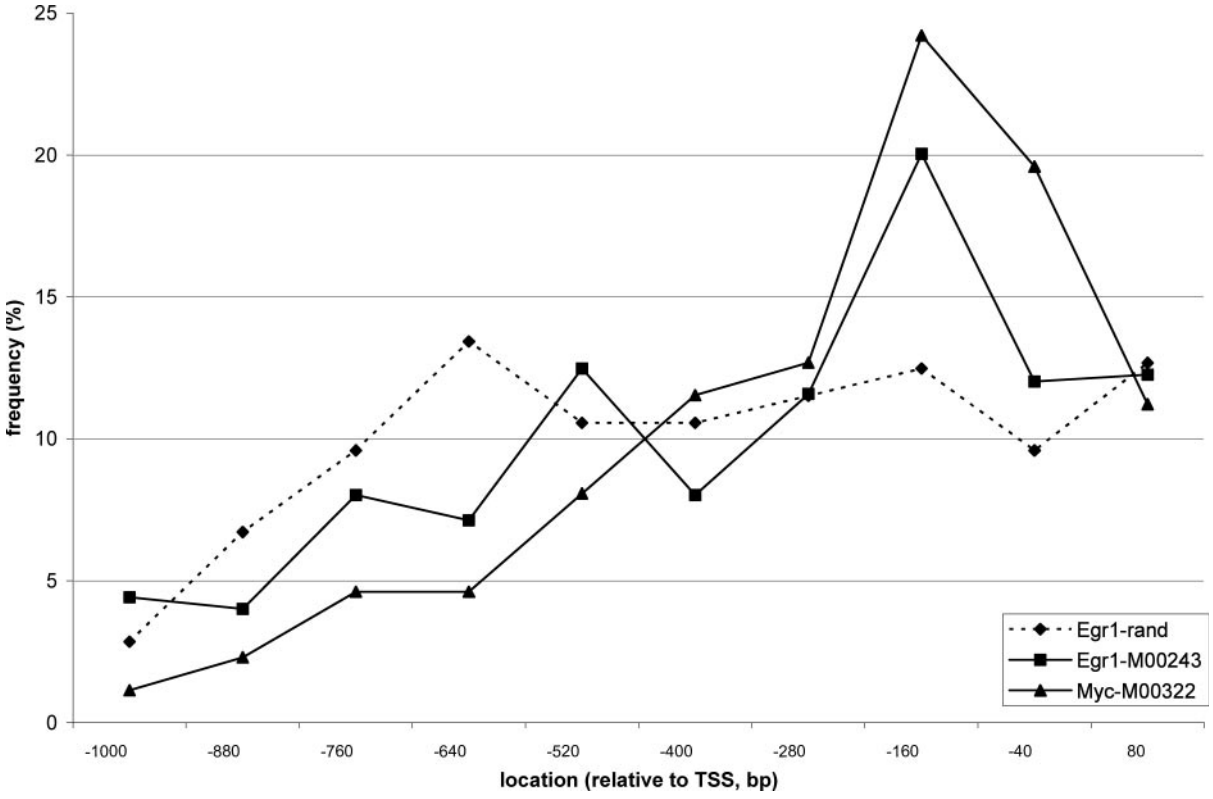

Figure 1. Location distribution of hits for c-Myc/Max target set and EGR-1 on promoters of the c-Myc/Max target set. The promoter region spanning 1000 bp upstream to 200 bp downstream of the TSS was divided into 10 bins of 120 bp. The graph represents the relative frequency of hits over the bins for c-Myc/Max (M00322), EGR-1 (M00243) and for a random PWM derived from the EGR-1 PWM as explained in the text. The number of hits in each bin was normalized by the effective sequence length scanned in the bin (effective lengths can be different in different bins due to masking of repetitive elements in promoters). Multiple hits in promoters were counted as such.

DISCUSSION

The mechanisms by which c-Myc promotes cell growth and transformation are poorly understood, but several lines of evidence suggest that its function as transcriptional regulator is required for its ability to promote malignancy. Since transcriptional regulation is often the result of combinatorial interplay among various TFs, we attempted to identify TFs that form recurring *cis*-regulatory modules with c-Myc. To this end, we analyzed a publicly available dataset of human promoters that are bound by c-Myc/Max in Burkitt's lymphoma cells. Our computational analysis identified nine TFs whose binding-site signatures are significantly overrepresented in this set of promoters. Most of these TFs were also enriched in the set of mouse orthologs of c-Myc target promoters, suggesting a conserved biological function mediated by the cooperation of c-Myc and these TFs.

Among the TFs enriched in the c-Myc/Max target promoter set were the pivotal regulators of the transcriptional program associated with cell cycle progression, E2F and NF-Y. The enrichment of the binding-site signatures of these factors in promoters that are bound by c-Myc/Max suggests functional links between them and c-Myc. Functional links between c-Myc and E2F are well documented (23–26). Myc promotes the cell cycle progression by coordinate activation of cell cycle driving genes (e.g. *Cdc25A*, *Cdk4*, and *Cyclins D2*, *E* and *A*), and by the suppression of cell cycle arrest genes (such as *p15*, *p21*, *p27* and *GADD*s) (24). Some of the cell cycle promoting genes are common targets of c-Myc and members of the E2F family TFs (27). Recently, a physical link between c-Myc and E2F-6 was reported, with a role in maintaining cellular quiescence (28). As for the link between c-Myc and NF-Y, physical interaction between c-Myc and the NF-YB and NF-YC subunits of the NF-Y trimer has been demonstrated previously (29), and in particular, these interactions are important for the down-regulation of a well-established c-Myc target, PDGFRb (30). Computational identification of *cis*-modules cannot decipher the regulatory effect of the module. The finding on the negative effect of the c-Myc/NF-Y module on PDGFRb transcription underscores the importance of further studies, which are beyond the scope of our current study, to establish whether the functional interplay between predicted TFs with c-Myc results in negative or positive regulation of target genes.

In addition to the examples outlined above, we sought to determine whether EGR-1 binds promoters, which are predicted *in silico* to contain EGR-1 consensus sites along with c-Myc binding sites. Given that promoter-reporter transient transfection assays, with specific reference to c-Myc, are inherently noisy and dependent on the cell types as well as cell density and other factors, we chose ChIP assays to determine whether our *in silico* prediction of EGR1 binding could be validated. The experimental data substantiate the prediction that these two TFs together form a transcriptional *cis*-regulatory module that recurs on multiple promoters.

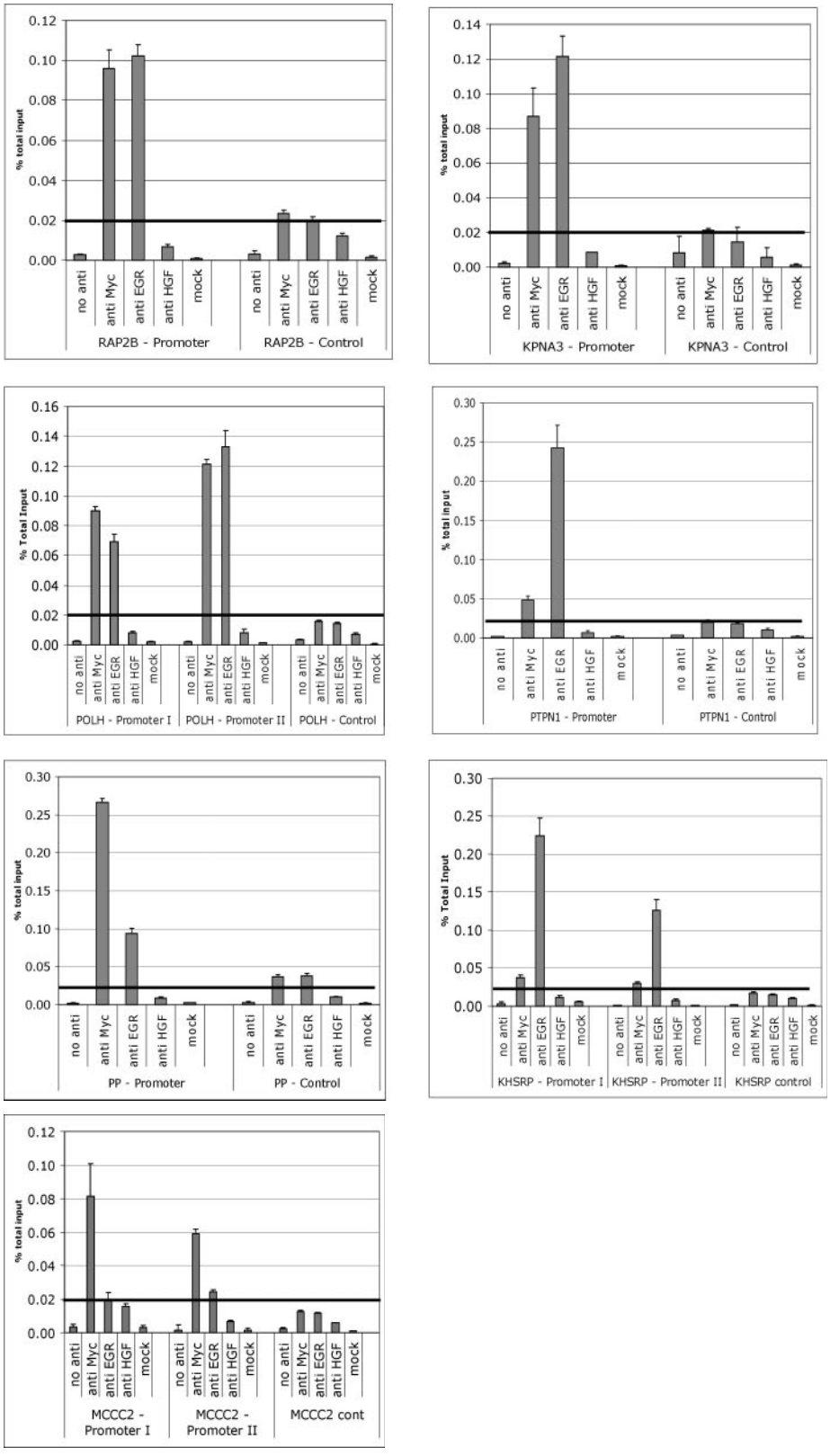

Figure 2. ChIP of c-Myc and EGR-1 targets in K562 cells. Each graph represents real-time PCR amplification of the promoter and control regions of each gene using anti-Myc, anti-EGR-1, anti-HGF and no antibody precipitated chromatin as template. Bars represent the percentage of total input DNA for each ChIP sample averaged over three PCRs. Error bars represent 1 SD. The horizontal solid line represents 0.02% total input DNA, the background signal for this assay. The signals obtained for the binding of c-Myc and EGR-1 to the promoter regions of all six genes examined (*RAP2B*, *KHSRP*, *PolH*, *PTPN1*, *PP* and *KPNA3*), but not for the negative control, MCCC2, were above the background level ($P < 0.025$, one-tailed *t*-test). Primer sequences and positions, and locations of the putative hits for c-Myc and EGR-1 on the examined promoters are given in Supplementary Table D.

The possibility for functional links between c-Myc and EGR-1 is intriguing because of the pivotal role of EGR-1 in angiogenesis, the formation of new blood vessels from pre-existing vasculature. Uncontrolled angiogenesis plays an important role in tumor growth, and the sprouting of new blood vessels into tumors suggests that angiogenesis is necessary for the progression of malignancy. Recent reports underscored the critical roles of both EGR-1 and c-Myc in angiogenesis. Fahmy *et al*. (17) reported that the inhibition of EGR-1 expression repressed neovascularization and blocked angiogenesis and tumor growth in mouse and rat models. Baudino *et al*. (31) reported that c-Myc is also required for the proper expression of several major angiogenic factors, and that c-Myc(−/−) ES cells are dramatically impaired in their ability to form tumors in immune-compromised mice, and the small tumors that do develop are poorly vascularized. Here, we proposed a possible synergistic action between c-Myc and EGR-1 in transcriptional regulation of the target genes, and experimentally demonstrated the binding of both c-Myc and EGR-1 to several target promoters.

The approach applied here is general, and demonstrates the power of computational analysis to elucidate novel functional links in transcriptional networks, based on functional genomics data.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

R.E. is a Joseph Sassoon Fellow. This work was partially supported by a research grant from the Ministry of Science and Technology, Israel (Y.S. and R.S.) and NIH grants R01CA057341 (C.V.D.) and T32HL07525 (K.I.Z.). This work was carried out in partial fulfillment of the requirements for the PhD degree of R.E.

REFERENCES

1. Pilpel,Y., Sudarsanam,P. and Church,G.M. (2001) Identifying regulatory networks by combinatorial analysis of promoter elements. *Nature Genet*., **29**, 153–159.
2. Segal,E., Shapira,M., Regev,A., Pe'er,D., Botstein,D., Koller,D. and Friedman,N. (2003) Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nature Genet*., **34**, 166–176.
3. Segal,E., Yelensky,R. and Koller,D. (2003) Genome-wide discovery of transcriptional modules from DNA sequence and gene expression. *Bioinformatics*, **19** (Suppl. 1), i273–i282.
4. Tavazoie,S., Hughes,J.D., Campbell,M.J., Cho,R.J. and Church,G.M. (1999) Systematic determination of genetic network architecture. *Nature Genet*., **22**, 281–285.
5. Lockhart,D.J. and Winzeler,E.A. (2000) Genomics, gene expression and DNA arrays. *Nature*, **405**, 827–836.
6. Lee,T.I., Rinaldi,N.J., Robert,F., Odom,D.T., Bar-Joseph,Z., Gerber,G.K., Hannett,N.M., Harbison,C.T., Thompson,C.M., Simon,I. *et al*. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science*, **298**, 799–804.
7. Ren,B., Robert,F., Wyrick,J.J., Aparicio,O., Jennings,E.G., Simon,I., Zeitlinger,J., Schreiber,J., Hannett,N., Kanin,E. *et al*. (2000) Genome-wide location and function of DNA binding proteins. *Science*, **290**, 2306–2309.
8. Li,Z., Van Calcar,S., Qu,C., Cavenee,W.K., Zhang,M.Q. and Ren,B. (2003) A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. *Proc. Natl Acad. Sci. USA*, **100**, 8164–8169.
9. Ren,B., Cam,H., Takahashi,Y., Volkert,T., Terragni,J., Young,R.A. and Dynlacht,B.D. (2002) E2F integrates cell cycle progression with DNA repair, replication, and G_2/M checkpoints. *Genes Dev*., **16**, 245–256.
10. Martone,R., Euskirchen,G., Bertone,P., Hartman,S., Royce,T.E., Luscombe,N.M., Rinn,J.L., Nelson,F.K., Miller,P., Gerstein,M. *et al*. (2003) Distribution of NF-kappaB-binding sites across human chromosome 22. *Proc. Natl Acad. Sci. USA*, **100**, 12247–12252.
11. Nilsson,J.A. and Cleveland,J.L. (2003) Myc pathways provoking cell suicide and cancer. *Oncogene*, **22**, 9007–9021.
12. Pelengaris,S., Khan,M. and Evan,G. (2002) c-MYC: more than just a matter of life and death. *Nature Rev. Cancer*, **2**, 764–776.
13. Cole,M.D. and McMahon,S.B. (1999) The MYC oncoprotein: a critical evaluation of transactivation and target gene regulation. *Oncogene*, **18**, 2916–2924.
14. Oster,S.K., Ho,C.S., Soucie,E.L. and Penn,L.Z. (2002) The myc oncogene: MarvelouslY Complex. *Adv. Cancer Res*., **84**, 81–154.
15. Sudarsanam,P., Pilpel,Y. and Church,G.M. (2002) Genome-wide co-occurrence of promoter elements reveals a *cis*-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. *Genome Res*., **12**, 1723–1731.
16. Elkon,R., Linhart,C., Sharan,R., Shamir,R. and Shiloh,Y. (2003) Genome-wide *in silico* identification of transcriptional regulators controlling the cell cycle in human cells. *Genome Res*., **13**, 773–780.
17. Fahmy,R.G., Dass,C.R., Sun,L.Q., Chesterman,C.N. and Khachigian,L.M. (2003) Transcription factor Egr-1 supports FGF-dependent angiogenesis during neovascularization and tumor growth. *Nature Med*., **9**, 1026–1032.
18. Thiel,G. and Cibelli,G. (2002) Regulation of life and death by the zinc finger transcription factor Egr-1. *J. Cell. Physiol*., **193**, 287–292.
19. Birney,E., Andrews,T.D., Bevan,P., Caccamo,M., Chen,Y., Clarke,L., Coates,G., Cuff,J., Curwen,V., Cutts,T. *et al*. (2004) An overview of Ensembl. *Genome Res*., **14**, 925–928.
20. Matys,V., Fricke,E., Geffers,R., Gossling,E., Haubrock,M., Hehl,R., Hornischer,K., Karas,D., Kel,A.E., Kel-Margoulis,O.V. *et al*. (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res*., **31**, 374–378.
21. Kasprzyk,A., Keefe,D., Smedley,D., London,D., Spooner,W., Melsopp,C., Hammond,M., Rocca-Serra,P., Cox,T. and Birney,E. (2004) EnsMart: a generic system for fast and flexible access to biological data. *Genome Res*., **14**, 160–169.
22. Zeller,K.I., Jegga,A.G., Aronow,B.J., O'Donnell,K.A. and Dang,C.V. (2003) An integrated database of genes responsive to the Myc oncogenic transcription factor: identification of direct genomic targets. *Genome Biol*., **4**, R69.
23. Leone,G., Sears,R., Huang,E., Rempel,R., Nuckolls,F., Park,C.H., Giangrande,P., Wu,L., Saavedra,H.I., Field,S.J. *et al*. (2001) Myc requires distinct E2F activities to induce S phase and apoptosis. *Mol. Cell*, **8**, 105–113.
24. Gartel,A.L. and Shchors,K. (2003) Mechanisms of c-myc-mediated transcriptional repression of growth arrest genes. *Exp. Cell Res*., **283**, 17–21.
25. Santoni-Rugiu,E., Duro,D., Farkas,T., Mathiasen,I.S., Jaattela,M., Bartek,J. and Lukas,J. (2002) E2F activity is essential for survival of Myc-overexpressing human cancer cells. *Oncogene*, **21**, 6498–6509.
26. Baudino,T.A., Maclean,K.H., Brennan,J., Parganas,E., Yang,C., Aslanian,A., Lees,J.A., Sherr,C.J., Roussel,M.F. and Cleveland,J.L. (2003) Myc-mediated proliferation and lymphomagenesis, but not apoptosis, are compromised by E2f1 loss. *Mol. Cell*, **11**, 905–914.
27. Stevens,C. and La Thangue,N.B. (2003) E2F and cell cycle control: a double-edged sword. *Arch. Biochem. Biophys*., **412**, 157–169.
28. Ogawa,H., Ishiguro,K., Gaubatz,S., Livingston,D.M. and Nakatani,Y. (2002) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G_0 cells. *Science*, **296**, 1132–1136.
29. Izumi,H., Molander,C., Penn,L.Z., Ishisaki,A., Kohno,K. and Funa,K. (2001) Mechanism for the transcriptional repression by c-Myc on PDGF beta-receptor. *J. Cell. Sci*., **114**, 1533–1544.
30. Mao,D.Y., Barsyte-Lovejoy,D., Ho,C.S., Watson,J.D., Stojanova,A. and Penn,L.Z. (2004) Promoter-binding and repression of PDGFRB by c-Myc are separable activities. *Nucleic Acids Res*., **32**, 3462–3468.
31. Baudino,T.A., McKay,C., Pendeville-Samain,H., Nilsson,J.A., Maclean,K.H., White,E.L., Davis,A.C., Ihle,J.N. and Cleveland,J.L. (2002) c-Myc is essential for vasculogenesis and angiogenesis during development and tumor progression. *Genes Dev*., **16**, 2530–2543.