[Cell Cycle 4:12, 1788-1797, December 2005]; ©2005 Landes Bioscience

Report

Deciphering Transcriptional Regulatory Elements that Encode Specific Cell Cycle Phasing by Comparative Genomics Analysis

Chaim Linhart[1]
Ran Elkon[2]
Yosef Shiloh[2]
Ron Shamir[1]

[1]School of Computer Science; [2]The David and Inez Myers Laboratory for Genetic Research; Department of Molecular Genetics and Biochemistry; Sackler School of Medicine; Tel Aviv University; Tel Aviv, Israel

*Correspondence to: Ron Shamir; School of Computer Science; Tel Aviv University; Tel Aviv, 69978 Isreal; Tel: +972.3.640.5383; Fax: +972.3.640.5384; Email: rshamir@tau.ac.il

Received 08/18/05; Accepted 09/13/05

Previously published online as a Cell Cycle E-publication:
http://www.landesbioscience.com/journals/cc/abstract.php?id=2173

KEY WORDS

E2F, NF-Y, CHR, B-Myb, cell cycle regulation, transcriptional network

ACKNOWLEDGEMENTS

R. E. is a Joseph Sassoon Fellow. This study was supported in part by a research grant from the Ministry of Science and Technology, Israel. R.S was also supported in part by the Israel Science Foundation (grant 309/02).

NOTE

Supplementary Material can be found at: http://www.landesbioscience.com/cc/supplement/linhartCC4-12-sup.pdf

ABSTRACT

Transcriptional regulation is a major tier in the periodic engine that mobilizes cell cycle progression. The availability of complete genome sequences of multiple organisms holds promise for significantly improving the specificity of computational identification of functional elements. Here, we applied a comparative genomics analysis to decipher transcriptional regulatory elements that control cell cycle phasing. We analyzed genome-wide promoter sequences from 12 organisms, including worm, fly, fish, rodents and human, and identified conserved transcriptional modules that determine the expression of genes in specific cell cycle phases. We demonstrate that a canonical E2F signal encodes for expression highly specific to the G_1/S phase, and that a cis-regulatory module comprising CHR-NF-Y elements dictates expression that is restricted to the G_2 and G_2/M phases. B-Myb binding site signatures occur in many of the CHR-NF-Y target genes, suggesting a specific role for this triplet in the regulation of the cell cycle transcriptional program. Remarkably, E2F signals are conserved in promoters of G_1/S genes in all organisms from worm to human. The CHR-NF-Y module is conserved in promoters of G_2/M regulated genes in all analyzed vertebrates. Our results reveal novel modules that determine specific cell cycle phasing, and identify their respective putative target genes with remarkably high specificity.

INTRODUCTION

The eukaryotic cell cycle is driven by a periodic, tightly controlled network of accumulation and destruction of key regulators and effectors. Precise coordination of cell cycle processes of DNA replication and chromosome segregation is required to ensure that daughter cells receive the requisite complement of genetic material. The fidelity of the cell cycle engine operation is tightly supervised by an intricate checkpoints mechanism acting during different phases of the division cycle. The mobilization of this engine is regulated at three major layers: transcriptional regulation of gene expression, post-translational modulation of protein activity, and modulation of protein stability.[1-3]

In this study, we focus on the transcriptional program associated with cell cycle progression. Prominent among the regulators of this program are the members of the E2F family of transcription factors (TFs). E2F1-3 are positive regulators of cell cycle progression while E2F4-6 play an inhibitory role.[4] Traditionally, the regulatory function of E2F was linked to the G_1 and S phases, but recent studies pointed to the involvement of this family in other cell cycle phases as well.[5-7] Other TFs and regulatory elements were shown to play an important role in driving the cell cycle transcriptional program. The CCAAT binding TF NF-Y was linked to the regulation of G_2/M progression by several studies: NF-Y controls the expression of several key regulators of this phase, including *CDC2*,[8] *CCNB1*[9] and *CCNB2*.[9,10] Furthermore, p53-mediated activation of the G_2/M checkpoint is executed through its inhibition of NF-Y induction of these target genes.[11] CDE (cell cycle dependent element) and CHR (cell cycle homology region) cis-regulatory elements were found in promoters of several cell cycle genes, including *CDC25C*,[12] *CDC2*,[13] *CCNB1*,[14] *CCNB2*,[15] *AURKB* (encoding aurora kinase B)[16] and *PLK*,[17] suggesting that these elements too play a role in controlling G_2/M progression. The B-Myb TF is an E2F-regulated gene induced at G_1/S, whose activity is enhanced during S phase through phosphorylation by cyclin A/Cdk2.[18,19] The transcriptional activity of B-Myb is required for cell cycle progression and it was recently suggested to play a role, together with E2F, in linking the G_1/S and G_2/M transcriptional programs.[14] Another TF, FOXM1, was recently shown to be required for execution of mitosis.[20,21]

While conventional biological studies focus on specific isolated components within a network of interest, the availability of essentially complete genome sequences in many organisms, and the maturation of novel functional genomics technologies, enable systems-level analysis of cellular networks. In a previous study, we applied computational promoter analysis to publicly available cell cycle related functional genomics datasets, delineating on a genomic scale regulatory mechanisms that control the human cell cycle transcriptional program.[22] We identified a significant statistical over-representation of several TF binding site (BS) signatures on promoters of cell cycle regulated genes. Among the most significant observation was the enrichment of E2F and NF-Y signatures in G_1/S and G_2/M promoters, respectively. Here, we employ comparative genomics to further elucidate the cell cycle network regulated by these regulators, and to pinpoint, with high accuracy, the target genes that they control.

A major challenge in computational promoter analysis is the typically very short (8–14 bp) and highly flexible nature of cis-regulatory elements recognized and bound by TFs: Most positions within the binding site motif are not strictly limited to a particular nucleotide, so genome-wide computational scans for putative TF binding sites (TFBSs) inevitably yield many false positive hits[23,24] (we use the term *hit* to refer to computationally-identified putative binding sites). The availability of sequences of many genomes in addition to the human genome greatly boosts the specificity of in silico identification of regulatory elements embedded in the genome.[25,26] Because higher selective pressure imposed on functional elements makes them more conserved than their surrounding non-functional DNA, scanning for evolutionarily conserved elements, an approach called phylogenetic footprinting, markedly reduces false-positive hit rates.[27,28]

We searched for conserved transcriptional regulators of cell cycle progression by integrating several sources of information: promoter sequences from twelve organisms, ranging from worm to human; orthology relationships among genes of these organisms; models of BSs of known TFs; and genome-wide gene expression profiles. We find that E2F signals are conserved in G_1/S regulated genes in all organisms from worm to human, and show that a canonical E2F signal is associated with gene expression that is highly specific to the G_1/S phase. In addition, we define a novel cis-regulatory module comprising CHR-NF-Y cis-elements, demonstrate that it dictates an expression pattern that is tightly restricted to the G_2 and G_2/M phases, and identify with very high specificity the genes that it putatively regulates. We show that the CHR-NF-Y module is conserved in cell cycle regulated promoters in vertebrates. We also observe that B-Myb signature appears in many of the CHR-NF-Y target promoters, suggesting that B-Myb cooperates with CHR-NF-Y, together constituting a triplet with specific functional roles in the regulation of G_2 and M phases.

METHODS

Extraction of promoter sequences. Putative promoter sequences were extracted based on gene transcription start site (TSS) annotation from the genome sequences of twelve organisms: two worms (*C. elegans* and *C. briggsae*), two insects (the fruit fly, *Drosophila melanogaster* and a mosquito, *Anopheles gambiae*), three fish (the zebrafish, *Danio rerio*; Fugu, *Fugu rubripes*; and Tetraodon, *Tetraodon nigroviridis*), chicken (Gallus, *gallus*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), dog (*Canis familiaris*) and human. Promoters were extracted using a Perl script based on the application programming interface provided by the Ensembl project.[29] All sequences were extracted from version 27 of Ensembl genome release (Dec 2004), except for the *C. briggsae* promoters that were extracted from the v22 release (this worm was not included in later Ensembl releases). Ortholog genes between these organisms were determined using the EnsMart utility.[30] Only one-to-one mapped genes were taken into account when constructing the orthology maps. Specifically, the human-mouse orthology set contained a total of 16,299 genes.

Models for transcription factor binding sites. TFBSs are commonly modeled by position weight matrices (PWMs). PWMs for known human TFBSs were obtained from the TRANSFAC database (release 8.2, June 2004).[31] Typically, promoter sequences of a set of coregulated genes are scanned using a given PWM, and each subsequence is assigned a score that indicates how similar it is to the PWM; subsequences whose score is above some threshold are counted as hits, i.e., putative BSs. A judicious choice of the threshold value is essential in order to find a good balance between the rates of false positives and false negatives. Hits for the TATA-box cis-element were detected using the PRIMA software that we developed in a previous study, which sets the threshold by scanning randomly generated sequences with similar statistical characteristics to those of the genomic promoters.[22] PRIMA is available for download as part of the EXPANDER gene-expression analysis and visualization software (http://www.cs.tau.ac.il/~rshamir/expander/).[32] The TRANSFAC matrix M00252 was used by PRIMA as the TATA-box model. For ease of implementation and in order to ensure efficient performance, the other TFs in this study were modeled using regular expressions, which were composed and fine-tuned manually, based on TRANSFAC PWMs and a small selected list of known BSs. The following TFBS models were used (the models are written using the IUPAC nucleotide base code, e.g., Y stands for [CT]; "|" denotes "or"):

- E2F:TTT{E2F-core}NNN|(TTN|TNT|NTT){E2F-core}(ANN|NAN|NNA), where E2F-core is (SS|AG)CGSS|SSCG(SS|CT); Canonical E2F hits are those matching TTT{E2F-core}AAN (based on TRANSFAC matrix M00516 and BSs in *E2F1*, *CDC2*, *ORC1*)[33]
- NF-Y: ((RN|NR)CCAATSR)|(RRCCAAT(SN|NR)) (based on M00185 and BSs in *CCNB1*, *CDC2*,[14] *CCNB2*)[10]
- CHR: BNNRTTTRAAH (based on seven CHR BSs summarized in Kimura et al[16] and in *CCNB1*, *CDC2*)[14]
- B-Myb: (MNR|NNY)AACB(NYY|GHB) (based on M00004 and BSs in *CCNB1*, *CDC2*)[14]

Scanning promoters for TF hits. Promoter sequences were scanned by searching for matches of each regular expression in both strands of a predefined interval around the TSS. The intervals used were (positions are relative to the TSS, negative positions are upstream to the TSS): E2F: from -300 to +100; NF-Y: from -400 to 0; CHR: from -400 to +50; B-Myb: from -300 to +200. Each match of a regular expression was considered a hit of the corresponding TF. A hit of a module consisting of a pair of TFs was declared if the scanned promoter interval contained a match for both regular expressions, and the distance between the two matches was at most 200 bp. Hits of the triplet CHR-NF-Y-B-Myb were found by intersecting the promoters containing the pair CHR-NF-Y with those containing CHR-B-Myb. When scanning for evolutionarily conserved hits, additional constraints were applied, as explained below. Hits of the TATA-box element for Supplementary Fig. 2 were located using PRIMA, as described in Elkon et al.[22]

Phylogenetic footprinting constraints. Given the human promoters and their orthologs in one or more other species (typically mouse), each set of orthologous promoters was scanned for conserved hits. In the case of a single TF, a match was considered a hit if the following conservation criteria were fulfilled: (1) Each of the orthologous promoters contained a match for the regular expression in the corresponding interval; (2) All matches were on the same strand; (3) The locations of the matches in each of the non-human promoters differed by at most 100 bp from the location of the match in the human promoter; (4) The Hamming distance (i.e., number of different nucleotides) between each of the non-human matches and the human match was at most H, where H was set on a per-TF basis, as follows: $H = 4$ for E2F, $H = 3$ for NF-Y and CHR and $H = 2$ for B-Myb. For a module of two TFs, two additional constraints were applied: (5) The order of the TFs was

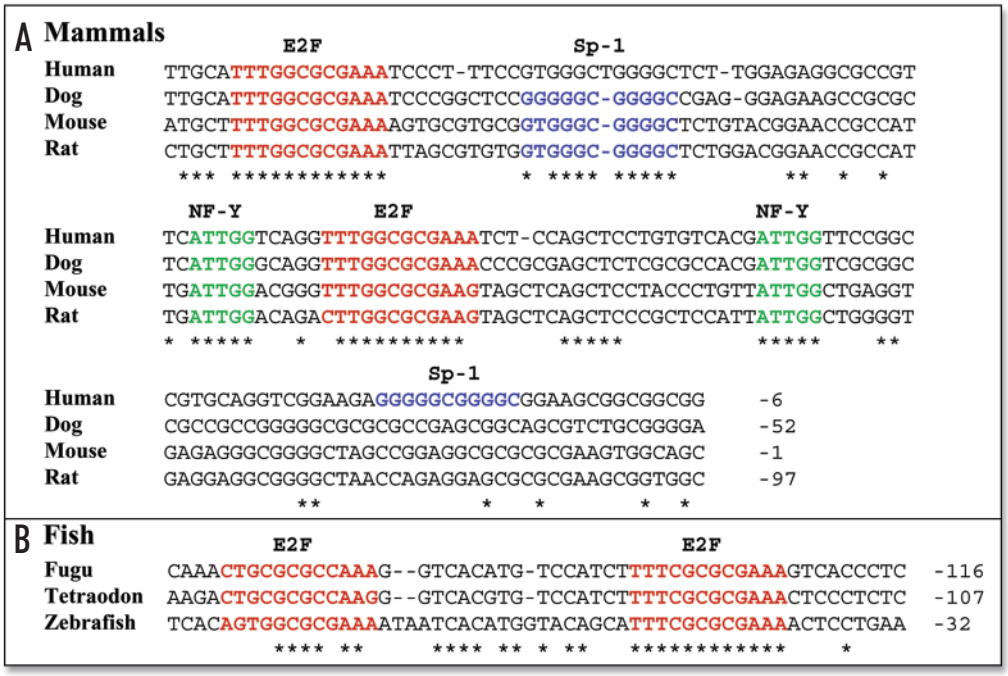

Figure 1. The power of phylogenetic footprinting. Alignment of promoter sequences from multiple species of the G_1/S-induced gene *MCM6* demonstrates the strength of comparative genomics in boosting computational identification of cis-regulatory elements. E2F elements (red), which have been validated experimentally,[36] are perfectly conserved across mammals (A) and fish (B). The alignment also points to other putative functional sites corresponding to NF-Y (green) and Sp1 (blue). The sequences flanking the TFBSs show lower conservation. Interestingly, the Sp1 site in human may have shifted downstream. Numbers next to the sequences indicate their location relative to the TSS (negative means upstream).

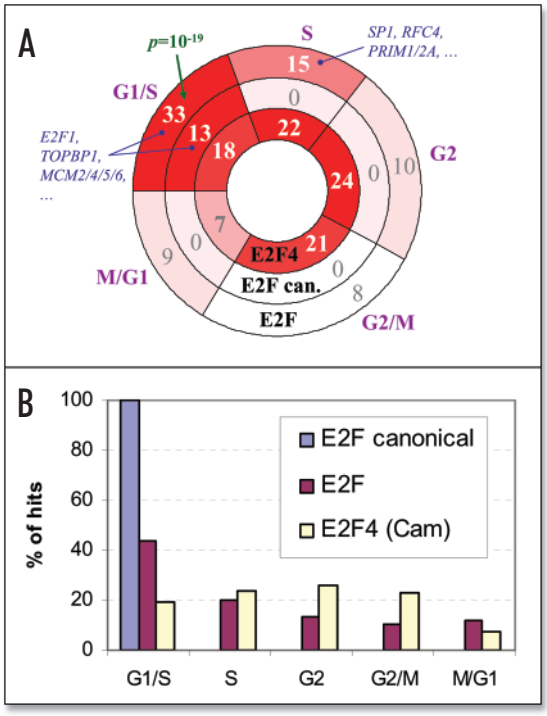

Figure 2. Distribution of E2F signatures and binding among cell cycle phases. (A) The number of promoters that contain a hit for the general E2F signature (outer circle) and canonical E2F signature (middle), and those that were shown to bind E2F4 in quiescent cells (inner) are indicated within each sector of the cycles. Each cycle is partitioned into five sectors corresponding to G_1/S, S, G_2, G_2/M and M/G_1 as defined by Whitfield et al.[41] Color intensities correspond to enrichment *p*-values of the corresponding set relative to all non-cell cycle genes: The general E2F signature is significantly enriched in G_1/S ($p = 10^{-19}$) and, to a lesser extent, in S; the canonical E2F signature is enriched in G_1/S. In contrast, E2F4 binding is highly enriched in each of the first four phases ($p = 10^{-12}$). Interesting representative targets are listed next to selected sectors (blue). (B) Distribution of the cell cycle targets across the five phases, for E2F general and canonical signatures, and for E2F4-bound promoters: The canonical E2F signature appears exclusively in G_1/S promoters, whereas binding of E2F4 is distributed uniformly across the first four phases.

identical in all organisms; (6) The distance between the matches of the two TFs in each non-human promoter differed by at most 30 bp from their distance in the human promoter. In promoters that contained several matches for the same TF, all matches were checked.

Enrichment score and factor. The standard hypergeometric score was used to determine whether a certain TF, or module of TFs, is over-represented in a given set of genes. Specifically, let *TS* be a given gene set of interest of size *T* (in this study, a cell cycle phase or the entire cell cycle set), and let *BS* denote a large background set of size *B* (in our case, all the genes that are not included in the cell cycle set). Let *t* and *b* denote the number of promoters in *TS* and *BS*, respectively, in which a hit was identified (either in the single- or multi-species case). Assuming that *TS* is randomly chosen out of *BS*, the probability, or *p*-value, of observing at least *t* hits in *TS* is:

$$p = \sum_{i=t}^{\min\{b+t,T\}} \frac{\binom{T}{i}\binom{B}{b+t-i}}{\binom{B+T}{b+t}}$$

The enrichment factor, denoted by *f*, of a given TF or module is the ratio between its frequency in a specified set of promoters and its frequency in the rest of the genome, i.e., $f = (t/T)/(b/B)$.

Cooccurrence of a pair of TFs. Given a pair of TFs, a cooccurrence score was computed in order to ascertain whether their hits tend to appear together in the same promoters significantly more often than expected by chance. Denote by *m* the number of analyzed genes, let f_a and f_b be the number of promoters that contain a hit for the each TF, and let f_{ab} be the number of promoters with a hit for both TFs. Using the hypergeometric score, the *p*-value for observing f_{ab} or more promoters containing hits for both TFs is:

$$p = \sum_{i=f_{ab}}^{\min\{f_a,f_b\}} \frac{\binom{f_a}{i}\binom{m-f_a}{f_b-i}}{\binom{m}{f_b}}$$

The set of E2F4-bound promoters. Cam et al.[7] used ChIP-on-chip to identify promoters bound by E2F4 in quiescent cells, which were arrested using three methods: mitogen depletion, contact inhibition, and $p16^{INK4A}$ induction. They reported a very high overlap (roughly 80%) between the results obtained by all three methods. Their microarray contained approximately

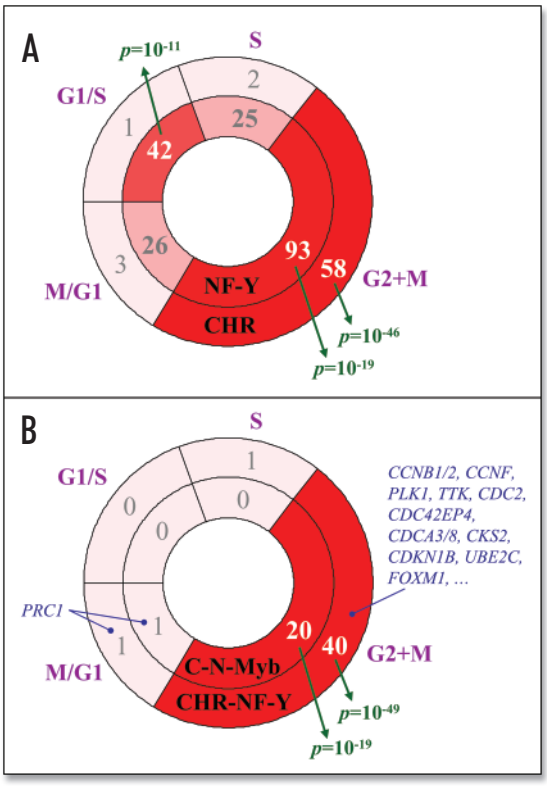

Figure 3. Distribution of NF-Y and CHR signatures among cell cycle phases. As in Figure 2, but here unifying the G_2 and G_2/M sectors, the number of promoters that contain a hit for NF-Y (A, inner circle) and CHR (A, outer) are indicated in each sector; and for the CHR-NF-Y pair (B, outer circle) and the CHR-NF-Y-B-Myb triplet (B, inner). NF-Y is enriched especially in G_1/S and G_2+M, whereas CHR is highly specific to G_2+M. The targets of the CHR-NF-Y and CHR-NF-Y-B-Myb modules are almost exclusive to G_2+M.

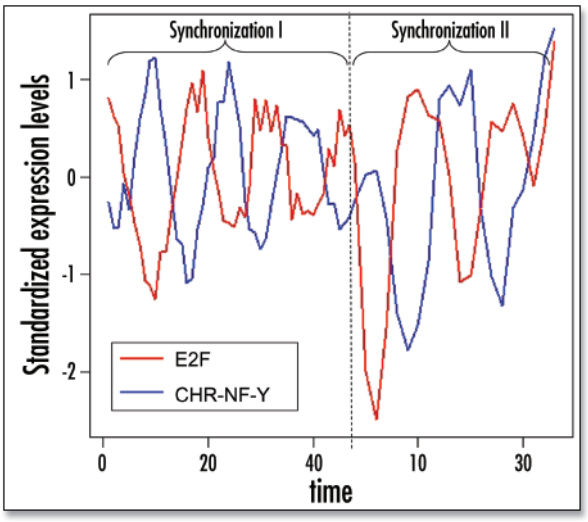

Figure 4. The canonical E2F signature and the CHR-NF-Y module dictate distinct and specific cell cycle phasing. Mean expression patterns over cell cycle progression (Whitfield et al. dataset) of genes containing the canonical E2F hits (13 cell cycle genes) and the CHR-NF-Y module (42 cell cycle genes) sharply peak at G_1/S and G_2/M phases, respectively. Expression levels of each gene were standardized to mean 0 and SD 1 before averaging over gene sets (in order to focus on the pattern rather than on the magnitude of expression). Y-axis represents standardized expression levels. Two synchronization methods were used by Whitfield et al.: Cells were arrested either in S phase using double thymidine block (synchronization I), or in M phase with a thymidine-nocodazole block (synchronization II).

13,000 sequences corresponding to promoter regions from -700 to +200 relative to the TSS. The set of E2F4-bound promoters used in this study consists of 271 promoters that were bound by E2F4 in at least one of the three methods (using a binding threshold of $p < 0.001$), and that are included in our human promoters set.

RESULTS

In this study, applying wide-scale computational promoter analysis, we sought to further elucidate transcriptional mechanisms that drive cell cycle progression. Our main objectives were to identify major cis-regulatory signals that dictate phase-specific expression, and to pinpoint, with high specificity, target genes that are under the control of these promoter elements. Several approaches have been proposed in an effort to increase the specificity of computational search for TFBSs. Phylogenetic footprinting[34,35] builds on the fact that TFBSs play an important biological role and have therefore evolved at a slower rate than non-functional intergenic sequences. Consequently, hits that are conserved across orthologous promoters in related species are more likely to be active BSs. Figure 1 illustrates the power of phylogenetic footprinting. The figure shows aligned promoter sequences of the gene *MCM6*, which encodes a subunit of the replication licensing complex, whose expression peaks at G_1/S,[36] in several mammals and fish. Evidently the promoters of *MCM6* are quite variable—most positions are not perfectly conserved across all species of each group. Remarkably, however, most of the conserved positions reside in contiguous blocks of 5-12bp in length, most of which match signatures of known TFs, namely E2F, NF-Y and Sp1. The role of all three TFs in cell cycle regulation is well established.[4,9,37,38] In the promoters presented in Figure 1, biologically active BSs emerge as islands of conservation, easily distinguishable from the surrounding sequences. Unfortunately, in most cases the identification of TFBSs is more difficult, either because of differences between the orthologous BSs, or because the BSs lie within long stretches of highly conserved promoter regions.

Transcriptional regulation in eukaryotes is to a large extent combinatorial, that is, the spatio-temporal conditions under which a gene is expressed are encoded by the specific combination of cis-regulatory elements embedded in its promoter region (and in the more distant regulatory regions, the enhancers and silencers). Therefore, a second common approach for reducing the rate of false positives in a TFBS scan is to search for a *module* of TFs, that is, a group of TFs whose joint binding activity has a specific transcriptional effect.[22,39,40] Identifying BSs of several TFs that tend to co-occur in the same promoters, possibly in a fixed order or at conserved distances, can eliminate many of the false hits that turn up when searching for each individual TF separately. Again, Figure 1 illustrates this idea: The order of the various BSs and the distances between them are highly conserved within each group of organisms; the only exception is the Sp1 BS, which seems to have drifted downstream in the human promoter.

Based on the aforementioned ideas, we sought to identify evolutionarily conserved transcriptional modules that control cell cycle progression. We first focused on E2F and performed a genome-wide scan for E2F signatures that are conserved between orthologous human-mouse promoters (see Methods). We found a conserved hit for E2F in 595 promoters out of the 16,299 orthologous promoter pairs included in our analysis. Next, we examined whether these hits are biased for cell cycle regulated promoters, using the cell cycle gene expression dataset published by Whitfield et al.[41] That study employed microarrays to profile gene expression throughout progression of the cell cycle in human Hela cells, and reported 872 cell cycle oscillating genes with periodic expression profiles. Our set of orthologous human-mouse promoters contains promoter sequences for 697 out of these 872 genes. We found that the overlap between the sets of promoters with conserved hits for E2F and the set of cell cycle oscillating genes (hereafter referred to as the 'cell cycle set') contains 75 genes, a statistically highly significant enrichment ($p=2 \times 10^{-17}$). The list of cell cycle regulated promoters on

Table 1 **CHR-NF-Y putative target genes whose expression peaks in the G_2 or G_2/M phases of the cell cycle#**

Symbol	Ensembl ID	Description
ARHGAP19	ENSG00000187122	ARHGAP19 (Rho GTPase activating protein 19) is involved in the regulation of members of the Rho GTPase family, that, among other roles, regulate chromosome alignment and cytokinesis.
ATF7IP	ENSG00000171681	Activating transcription factor 7 interacting protein
CCNB1*+	ENSG00000134057	Cyclin B1 complexes with CDC2 (Cdk1) to form the maturation-promoting factor (MPF), a master regulator of G_2/M phase.
CCNB2*+	ENSG00000157456	Cyclin B2 complexes with CDC2 (Cdk1) to form the M-phase-promoting factor (MPF), a master regulator of G_2/M phase.
CCNF	ENSG00000162063	Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin-cyclin interaction.[42]
CDC2*+	ENSG00000170312	CDC2 is a catalytic subunit of the M-phase promoting factor (MPF), which is essential for G_1/S and G_2/M phase transitions of eukaryotic cell cycle.
CDC42EP4 (BORG4)	ENSG00000179604	This protein is a member of the CDC42-binding protein family. Members of this family interact with Rho family GTPases and regulate the organization of the actin cytoskeleton.
CDCA3 (Tome-1)*+	ENSG00000111665	Tome-1 (trigger of mitotic entry 1) mediates the destruction of the mitosis-inhibitory kinase, Wee1, via the E3 ligase, SCF.
CDCA8+	ENSG00000134690	A component of the mitotic spindle.
CDKN1B (p27)	ENSG00000111276	CDKN1b binds to and prevents the activation of cyclin E-CDK2 or cyclin D-CDK4 complexes, and thus controls the cell cycle progression at G_1.
CENPF	ENSG00000117724	CENPF associates with the centromere-kinetochore complex and may play a role in chromosome segregation during mitosis.
CKS2+	ENSG00000123975	CKS2 is required for the first metaphase/anaphase transition of mammalian meiosis.[43]
DEPDC1	ENSG00000024526	DEP domain containing 1.
DEPDC1B	ENSG00000035499	DEP domain containing 1B.
ECT2+	ENSG00000114346	ECT2 is related to Rho-specific exchange factors and regulates the activation of CDC42 in mitosis.
FOXM1+	ENSG00000111206	FoxM1 is a transcription factor that is required for execution of the mitotic programme and chromosome stability.[21]
GTSE1 (G-2 and S-phase expressed 1)	ENSG00000075218	GTSE1 is only expressed in the S and G_2 phases of the cell cycle, where it colocalizes with cytoplasmic tubulin and micro tubules. In response to DNA damage, the encoded protein accumulates in the nucleus and binds the tumor suppressor protein p53, shuttling it out of the nucleus and repressing its ability to induce apoptosis
H2AFX+	ENSG00000188486	H2A histone family, member X
HMGB2+	ENSG00000164104	This gene encodes a member of the non-histone chromosomal high mobility group protein family, which are chromatin-associated and ubiquitously distributed in the nucleus of higher eukaryotic cells. HMGB2 was demonstrated to associate with mitotic chromosomes.[52]
HMGB3+	ENSG00000029993	This gene encodes a member of the non-histone chromosomal high mobility group protein family, which are chromatin-associated and ubiquitously distributed in the nucleus of higher eukaryotic cells.
HMMR (RHAMM)	ENSG00000072571	The receptor for hyaluronan mediated motility has been reported to mediate migration, transformation, and metastatic spread of murine fibroblasts. Its over-expression results in structural centrosomal abnormalities and and mitotic defects.[53]
KPNA2	ENSG00000182481	Karyopherin-α2 protein interacts with Chk2 and contributes to its nuclear import.[54]
LRRC17	ENSG00000128606	Leucine rich repeat containing 17.
MKI67	ENSG00000148773	The cell proliferation-associated antigen of antibody Ki-67 is widely used in routine pathology as a "proliferation marker" to measure the growth fraction of cells in human tumors.[55]
NUSAP1	ENSG00000137804	NuSAP (nucleolar and spindle associated protein 1) is primarily nucleolar in interphase, and localizes prominently to central spindle microtubules during mitosis. Depletion of NuSAP by RNA interference resulted in aberrant mitotic spindles, defective chromosome segregation, and cytokinesis.[56]
PLK1*+	ENSG00000166851	Polo-like kinase 1 (Plk1) is a key regulator of centrosome maturation, mitotic entry, sister chromatid cohesion, the anaphase-promoting complex/cyclosome (APC/C), and cytokinesis.
SFPQ	ENSG00000116560	splicing factor proline/glutamine rich.
SGOL2 (shugoshin-like 2)	ENSG00000163535	Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells.[57]
STK17B	ENSG00000081320	Serine/threonine kinase 17b (apoptosis-inducing).
TACC3	ENSG00000013810	TACC3 is a centrosomal/mitotic spindle-associated protein that is highly expressed in a cell cycle dependent manner in hematopoietic lineage cells.[44]
TMPO (LAP2)	ENSG00000120802	Lamina-associated polypeptide (LAP) 2 is suggested to play a role in targeting mitotic vesicles to chromosomes and reorganizing the nuclear structure at the end of mitosis.[58]

Table 1 **CHR-NF-Y putative target genes whose expression peaks in the G_2 or G_2/M phases of the cell cycle# (continued)**

Symbol	Ensembl ID	Description
TOP2A	ENSG00000131747	This nuclear enzyme is involved in processes such as chromosome condensation, chromatid separation, and the relief of torsional stress that occurs during DNA transcription and replication.
TTK+	ENSG00000112742	TTK was demonstrated to be dynamically distributed from the kinetochore to the centrosome, as cell enters into anaphase, and to phosphorylate the centrosomal protein TACC2 in mitosis.[59]
UACA+	ENSG00000137831	Uveal autoantigen with coiled-coil domains and ankyrin repeats.
UBE2C+	ENSG00000175063	This gene encodes a member of the E2 ubiquitin-conjugating enzyme family that is required for the destruction of mitotic cyclins and for cell cycle progression.[48]

*Genes whose promoter is already reported to be regulated by a CHR element. +Genes whose promoter was found to contain a strong conserved hit for B-Myb (in addition to CHR-NFY hit). #Four additional genes that do not have an official HUGO symbol are not included here (but included in Supplementary Table C).

which a conserved E2F hit was identified is provided in Supplementary Table A. Whitfield et al. partitioned the cell cycle oscillating genes into five clusters —G_1/S, S, G_2, G_2/M, and M/G_1—according to the phase in which their expression peaked. Hereafter we refer to the set of genes assigned either to the G_1/S or to the S clusters as the G_1+S set, and to the set of the genes assigned either to the G_2 or to the G_2/M clusters as the G_2+M set. In agreement with current biological knowledge and with results we previously reported,[22] we also observed here, but this time for human-mouse evolutionarily conserved hits, a strong bias of E2F signature for promoters of cell cycle regulated genes that peak at G_1/S phase ($p = 5 \times 10^{-19}$ relative to all the genes that are not in the cell cycle set) and, to a lesser extent, at S phase ($p = 7 \times 10^{-6}$) (Fig. 2A).

Recent functional genomics studies showed that the regulatory role of the E2F family on cell cycle progression extends also to G_2 and M phases.[5,6] To examine this point more closely, we analyzed the dataset published by Cam et al.[7] that used the combination of chromatin immunoprecipitation and promoter microarrays, also known as 'ChIP-on-chip', to identify promoters that are bound by the inhibitory E2F4 in quiescent cells. We checked the overlap between the set of 271 genes, whose promoters are bound by E2F4 (see Methods), and the cell cycle set. We found a highly significant overlap of 92 common genes ($p = 10^{-67}$). Surprisingly, these genes were not biased to G_1/S phase but were distributed almost uniformly across the first four phases- G_1/S, S, G_2, and G_2/M (Fig. 2A and B). This apparent discrepancy between the near-uniform distribution of E2F4 targets and the strong G_1/S bias of the E2F signature can be explained in several ways. It is possible that the inhibitory E2F4 is recruited to many G_2+M promoters by physical association with other DNA binding TFs rather than by its direct binding to the DNA. Another option is that E2F binding elements on G_2+M phase promoters are variants, perhaps with lower binding affinity, of the canonical E2F signature (which was originally defined using mainly G_1 and S phase E2F target promoters). To check this hypothesis, we performed a genome-wide scan for promoters that contain human-mouse conserved E2F signatures, with a strict requirement of adherence to the canonical E2F BS consensus (see Methods). Only 22 promoters met this stringent genome-wide scan (Supplementary Table B); 13 of them are contained in the cell cycle set. Remarkably, all 13 genes peak at G_1/S phase ($p = 4 \times 10^{-22}$) (Fig. 2).

Our previous computational cell cycle analysis, as well as other experimental studies, indicated that NF-Y plays a major role in regulating cell cycle progression in general, and is especially linked to G_2 and G_2/M phases.[9,11,22] Using our current approach, we identified 1,754 promoters with human-mouse conserved NF-Y hits, of which 186 are in the cell cycle set ($p = 2 \times 10^{-33}$), reflecting that NF-Y regulates a variety of biological processes, and its key function in cell cycle. In agreement with our previous results, NF-Y hits are enriched in all five phases ($p \leq 10^{-4}$ in each phase compared to the non-cell cycle genes), and most prominently in G_1/S (42 genes, $p = 6 \times 10^{-11}$) and in G_2+M (93 genes, $p = 10^{-19}$) (Fig. 3A).

Zhu et al.[14] recently validated functional NF-Y and CHR elements in the promoters of both *CDC2* and *CCNB1*, the master regulators of G_2 and G_2/M phases. Based on this observation, we tested whether this pair constitutes a recurrent cis-regulatory module. First, scanning for conserved CHR hits, we detected a striking bias for the G_2 + M set (Fig. 3A). We next searched for targets of the pair CHR-NF-Y with some distance constraints (see Methods). In the entire genome (16K genes), only 71 promoters met our criteria for human-mouse conserved hits of this module (Supplementary Table C), and 42 of them are contained in the cell cycle set ($p=2 \times 10^{-39}$). Remarkably, 40 of these genes are assigned to G_2 + M ($p = 9 \times 10^{-50}$, Fig. 3B). Moreover, this bias is not explained merely by the hit distributions of each individual TF within the G_2+M genes—the co-occurrence of the CHR and NF-Y elements is way above the expected rate given the prevalence of each TF separately ($p = 4 \times 10^{-13}$). Thus, CHR-NF-Y emerges as a major regulatory module of the G_2+M transcriptional program. This module dictates a highly phase-specific expression pattern, which is strongly anti-correlated with the expression imposed by the canonical E2F signature (Fig. 4).

Table 2 **Evolutionary conservation of cell cycle TFs**

		Organism											
		Tetrapods					Fish			Insects		Worms	
TFs	Cell cycle Phases	Human	Mouse	Rat	Dog	Chicken	Zebrafish	Fugu	Tetraodon	Fly	Mosquito	*C. elegans*	*C. briggsae*
E2F	G_1+S	6×10^{-20} (268)	3×10^{-23} (248)	4×10^{-12} (237)	3×10^{-4} (247)	2×10^{-9} (206)	4×10^{-13} (181)	7×10^{-7} (203)	2×10^{-4} (208)	3×10^{-4} (110)	1×10^{-7} (112)	8×10^{-4} (102)	2×10^{-4} (88)
CHR-NF-Y	G_2+M	4×10^{-38} (350)	2×10^{-31} (334)	9×10^{-11} (320)	2×10^{-9} (322)	2×10^{-15} (269)	3×10^{-8} (252)	3×10^{-3} (270)	5×10^{-4} (271)	N.E. (132)	N.E. (120)	N.E. (106)	N.E. (99)

The table shows enrichment p-values across 12 organisms of the E2F signature and the CHR-NF-Y module in promoters of genes whose human orthologs have an expression profile that peaks at G_1 + S and G_2 + M phases, respectively. "N.E." denotes "not enriched" ($p > 0.1$). E2F is enriched in G_1 + S across all tested species, whereas CHR-NF-Y is enriched in G_2+M only in vertebrates. The total number of genes in each set is written in parentheses (e.g., our data contains 203 Fugu genes, whose human orthologs are expressed in G_1 + S).

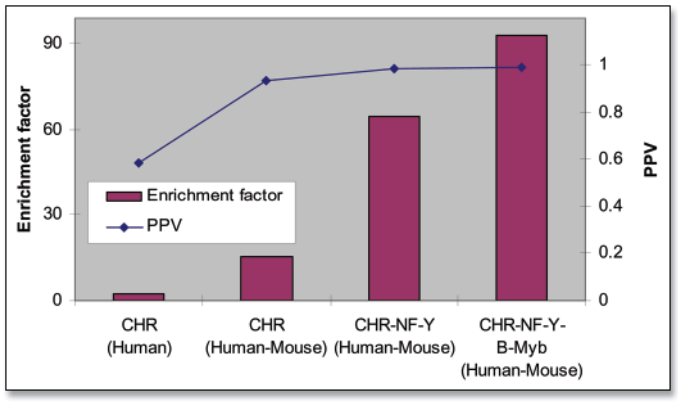

Figure 5. Improving TFBS detection by utilizing comparative genomics and searching for TF modules. The graph shows the dramatic improvement in enrichment factor (bars) and PPV values (curve) when searching for CHR-related hits in G_2 + M genes. Searching for targets of CHR in human yields an enrichment factor f = 2.4 and PPV = 0.58, indicating a low rate of true hits. Utilizing human-mouse conservation criteria improves the performance to f = 15.5, PPV = 0.93. Scanning for conserved modules results in an additional increase in the specificity, reaching a remarkable enrichment factor of 93 and PPV = 0.99 for CHR-NF-Y-B-Myb.

The forty G_2+M putative CHR-NF-Y targets include several genes that have already been shown to be controlled by CHR and NF-Y elements, but the majority of the targets are reported here for the first time (Table 1). The utilization of phylogenetic footprinting and the fact that these genes were experimentally demonstrated to peak at G_2 + M phases greatly boost the confidence that the hits reported here are biologically significant. Known CHR targets among the G_2+M hits include *CDC2*, *CCNB1* and *CCNB2*,[13-15] which constitute the Cyclin-CDK complex of the G_2/M phase; and *PLK1*,[17] which plays a major role in controlling centrosome maturation, mitotic entry, sister chromatid cohesion, the anaphase-promoting complex/cyclosome (APC/C), and cytokinesis. The proteins encoded by the novel targets putatively regulated by CHR-NF-Y participate in all major activities that are carried out during G_2 and M phases. Prominent among them are CCNF, which regulates the nuclear localization of cyclin B1 through a cyclin-cyclin interaction;[42] CKS2, which regulates CDKs activity during mitosis and meiosis;[43] CENPF, which associates with the centromere-kinetochore; the mitotic spindle-associated protein TACC3 that functions in chromosome segregation;[44] and the CDCA8, NUSAP1 (nucleolar and spindle associated protein 1) and TTK regulators of the mitotic spindle. Correct alignment of sister chromatide during metaphase and their balanced segregation during anaphase are critical processes executed by intricate complexes of cohesins, the centrosome-kinetochore at centromeres, and the bipolar structure of the mitotic spindle composed of microtubules and associated motor proteins. Recently, the Cdc42 member of the Rho GTPases family and its effector mDia3 were shown to regulate chromosome alignment by stabilizing kinetochore-microtobule attachment.[45,46] The Rho member of this GTPase family is known to regulate cytokinesis by controlling the assembly and the contraction of the myosin-actin network that comprises the contractile ring that is attached to the plasma membrane.[47] Importantly, ECT2, a major regulator of these GTPases-mediated pathways, and two additional proteins (CDC42EP4/Borg4 and the Rho GTPase activating protein ARHGAP19) that are tightly involved in them, are among our putative CHR-NFY targets. We also observed that protein products of many of the known and putative CHR-NF-Y targets are targeted for degradation by the anaphase-promoting complex/cyclosome (APC/C). In this regard, it is noteworthy that *UBE2C*, which encodes for an E2 ubiquitin-conjugating enzyme required for the destruction of mitotic cyclins,[48] is among the CHR-NF-Y putative targets.

A few salient examples of evolutionarily conserved CHR-NF-Y hits are shown in Supplementary Figure 1. As in the case of Figure 1, the BSs appear as islands of conservation along the promoter sequences. We observed that many of the G_2 + M putative CHR-NF-Y targets also contain a conserved cis-element that resembles the signature of B-Myb, as demonstrated in the promoters of *PLK1* (Supplementary Fig. 1A) and *UBE2C* (Supplementary Fig. 1B). Functional B-Myb sites were also identified in promoters of *CDC2* and *CCNB1*.[14] This suggests that B-Myb cooperates with CHR-NF-Y, together constituting a triplet with a specific functional role in regulating G_2 and M phases. We located a total of 31 conserved hits of this triplet; 21 of them are in the cell cycle set ($p = 4 \times 10^{-22}$), and of these 20 are in G_2 + M (see Table 1). Of note, the single cell cycle target gene of the triplet that is not in G_2 + M, PRC1 (protein regulator of cytokinesis 1), is also closely involved in regulation of the mitotic spindle and cytokinesis.[49] Interestingly, in 12 of the 21 promoters, the order of the hits within the triplet is NF-Y, CHR and B-Myb (from 5' to 3' on the coding strand), suggesting a possible structural preference of this module.

Our analysis highlights two major conserved transcriptional regulators of cell cycle progression—E2F and CHR-NF-Y—with key roles in G_1 + S and G_2 + M, respectively. We sought to trace the conservation of these signals along metazoan evolution. To this aim, we first extracted genome-wide promoter sequences of 12 organisms, including worms, insects, fish, chicken, rodents, dog and human (see Methods). In order to ensure that the quality of the annotated TSS in all organisms suffices for the detection of cis-regulatory elements, we verified that the TATA-box signal peaks at the correct location, in the very proximity of the TSS, in each of the tested species (e.g., the peak value is 7.9 standard deviations in human, and 11.0 in mouse) (Supplementary Fig. 2). Strikingly, we found that the E2F signature is conserved in G_1 + S genes across all organisms, from worm to human: scanning each organism separately for hits of E2F, we found a strong enrichment in the orthologs of human G_1 + S genes across all species ($p \leq 8 \times 10^{-4}$) (Table 2). In contrast, the CHR-NF-Y module, as defined using our BSs models, apparently evolved in vertebrates, as it is enriched in G_2+M in all vertebrates but in none of the other species.

DISCUSSION

Advances in functional genomics provide broad systems-level views of biological networks for the first time. In this study we conducted computational promoter analysis using genome sequences from multiple organisms and cell cycle gene expression profiles in order to comprehensively delineate the cell cycle transcriptional program. We demonstrate that E2F signals are conserved in G_1 + S regulated genes in all organisms from worm to human, and that a canonical E2F signal encodes for an expression at a very precise timing during cell cycle progression. In addition, we define a novel cis-regulatory module made up of CHR-NF-Y cis-elements, and demonstrate that it determines an expression pattern that is tightly restricted to the G_2 + M phase. Our analysis identifies with high specificity forty G_2+M genes that are putatively regulated by this module, thereby substantially extending current knowledge on the role of the CHR element in cell cycle regulation. We show that the CHR-NF-Y module is conserved in cell cycle regulated promoters in vertebrates.

TFBS detection has been the subject of numerous studies, but remains a difficult challenge. Existing BSs models do not contain enough information to locate functional BSs accurately. Typically, when promoter sequences are scanned using thresholds that allow recovering a large percentage of the true sites, many false positive hits are also reported.[24] Another difficulty lies in the evaluation of the results: Since there are no large validated gene sets in which the entire list of active BSs of the studied TF have been completely mapped, the specificity and sensitivity values are hard to assess. In this study, we searched for TFs and regulatory modules that are not only over-represented in the set of cell cycle promoters, but are also

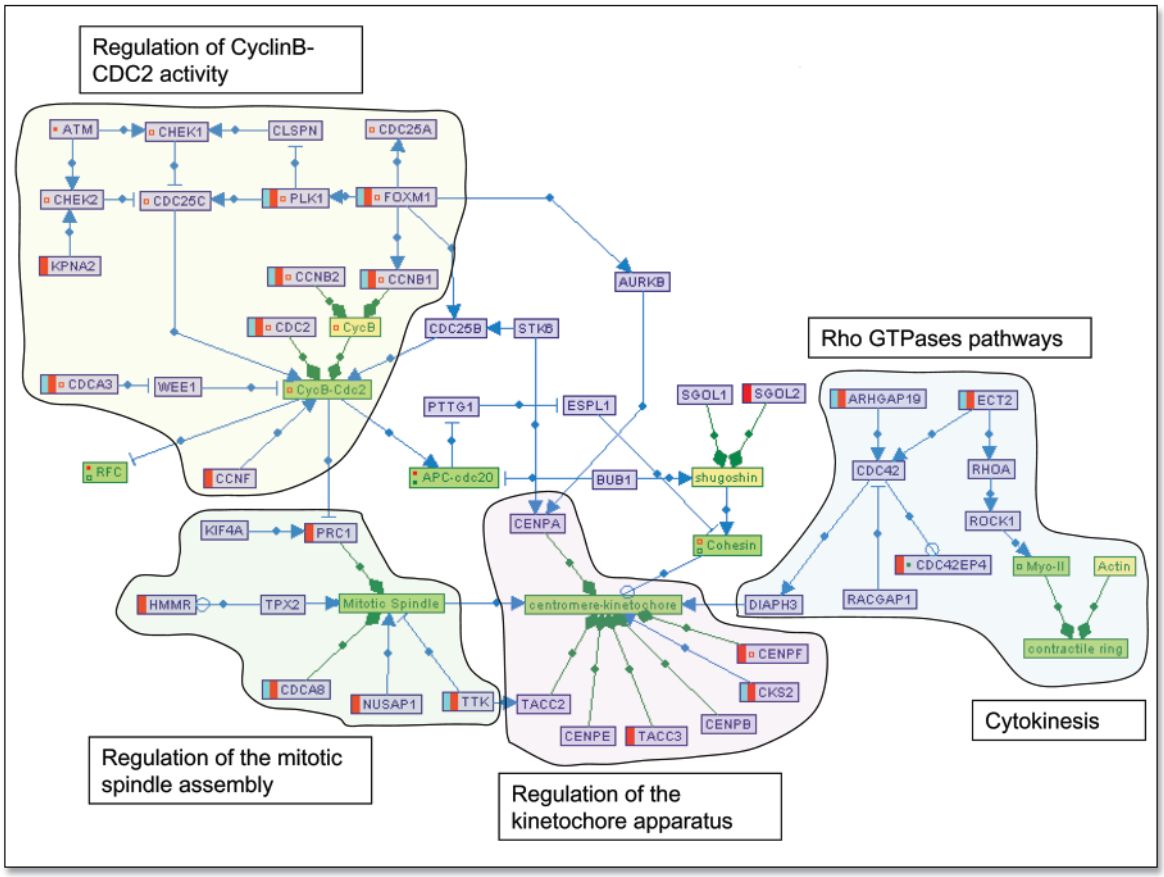

Figure 6. Putative roles for the CHR-NF-Y module in regulation of the G_2 and M phases. The interaction map contains nodes of three types: gray nodes represent single genes (denoted by the official HUGO symbol), yellow nodes represent gene families (e.g., Cyclin B), and green nodes represent protein complexes (e.g., CyclinB-CDC2). Blue edges denote regulation relations (→ for 'activation', ——| for 'inhibition') and green edges denote containment relations among nodes (e.g., CDC2 is contained in the CyclinB-CDC2 complex). Genes whose promoter contains a conserved CHR-NF-Y hit are marked by a red bar to the left of their node; an additional blue bar marks putative targets of the CHR-NF-Y-B-Myb triplet. CHR-NF-Y putative targets participate in all major activities that are carried out during G_2 and M phases, including modulation of CyclinB-CDC2 activity, control of sister chromatide alignment by the centrosome-kinetochore, control of chromosome segregation by the mitotic spindle apparatus, and regulation of the contractile ring assembly for the execution of cytokinesis. The figure was created using our SHARP software and knowledgebase for signaling pathways (http://www.cs.tau.ac.il/~sharp/). A red dot within a node indicates that the node has additional regulations in the SHARP database that are not displayed in the current map. Similarly, a green dot indicates that not all containment relations in which the node is involved are displayed.

highly biased to specific phases. Measuring the phase specificity of the TFs that we identified enables us to approximate their overall specificity, as false hits are expected to be distributed randomly among the genes in all phases. The TFs and modules we report are exceedingly phase-specific: All 13 promoters with a conserved canonical E2F site are G_1/S genes, which constitute a mere 19% of the entire cell cycle set; 95% (40 out of 42) of the promoters with conserved CHR-NF-Y hits are in the G_2 + M phases, which contain 48% of the cell cycle regulated genes.

Another approach to evaluating the accuracy of our results is to compute each TF's or module's *enrichment factor*, denoted by *f*—the ratio between its frequency in a given set of promoters and its frequency in the rest of the genome (see Methods). Using the latter frequency as an upper-bound estimate of the false-positives rate, the value *1 - 1/f* approximates the positive predictive value, or PPV, which is the fraction of true-positive hits out of all the identified hits in the promoter set (see Tompa et al.).[24] For example, the CHR-NF-Y-B-Myb module has 10 putative targets out of 15,602 genes that are not in the cell cycle set. Thus, we estimate that our scan reports up to one false-positive hit per 1,560 promoters. Searching for the same module within the set of 334 G_2 + M genes yields 20 targets. Using the 1:1560 false-positives rate, we expect the number of false targets in this set to be no more than 0.21 (=334/1560). In other words, at least 19 (or, more accurately, 19.79, which is 99% of 20) of the 20 identified targets should be true hits (i.e., PPV = 0.99).

Remarkably, the enrichment factor *f* increases as more sources of information are added into the scan algorithm (Fig. 5). For instance, searching for hits of CHR in the human genome yields f = 2.4 (PPV = 0.58) for G_2 + M phases; utilizing phylogenetic footprinting—requiring each hit to match human-mouse conservation constraints—increases the enrichment factor to f = 15.5; and searching for human-mouse conserved modules improves this even further: f = 64.4 for CHR-NF-Y, and f = 93 for CHR-NF-Y-B-Myb. The latter enrichment factor implies that PPV = 0.99, that is, 99% of the reported G_2 + M hits are expected to be true BSs, as explained above. These enrichment factors and PPV's exemplify the dramatic improvement in TFBS detection accuracy gained by applying comparative genomics techniques and by searching for modules of cooperative TFs.

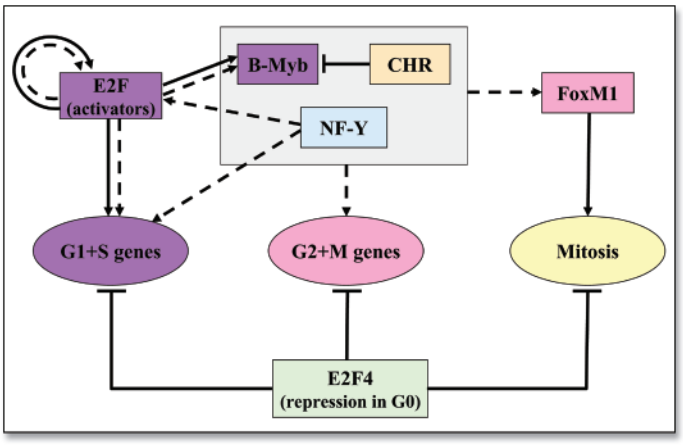

Figure 7. A model of the cell cycle transcriptional program. Integrating current knowledge and our computational results, a model for the propagation of the cell cycle transcriptional program emerges. Interactions supported by experimental data are presented by solid arrows, while those based on our computational analysis are marked by dashed ones. Genes expressed in G_1 or S phases are colored purple (dark gray), and G_2 + M genes are pink (light gray). Missing pieces in this puzzle can be expected to be discovered by combinations of biochemical experiments, functional genomics studies, and computational analyses.

In agreement with current biological knowledge, but without being biased by it, our computational analysis identified E2F as the major transcriptional regulator of the G_1 + S transcriptional network. However, recent studies extended the role of the E2F family in cell cycle regulation beyond the G_1 and S phases. Indeed, we show that the promoters reported to bind the inhibitory E2F4 in quiescent cells are not biased to any specific cell cycle phase. This apparent discrepancy may suggest that E2F regulation on G_2 and M promoters is mediated by a variant of the canonical E2F signature, possibly with lower binding affinity (or even, for some G_2 + M promoters, by non-direct DNA binding). The absolute assignment of the 13 canonical E2F targets to the G_1/S phase supports this hypothesis.

Our analysis points to the CHR-NF-Y cis-module as the major regulator of gene expression in G_2 + M phases. Several cell cycle regulated promoters were reported to be regulated by the CHR element, including *Cyclin A*,[8] CDC25C,[8] CDC2,[8] Cyclin B2,[10] Aurora B,[16] B-Myb[50] and PLK1.[17] However, the importance of the CHR-NF-Y module as a key regulator of G_2 and M phases is not widely appreciated, and is put in the spotlight by our results. We report, with high specificity, 42 mitotic genes that are putatively regulated by this module. Examination of these putative targets suggests that the CHR-NF-Y module regulates all known major activities that are carried out in G_2 and M phases, including modulation of CCNB-CDC2, and the assembly of the kinetochore-centrosome complexes, of the mitotic spindle and its associated motor proteins, and of cytokinesis effectors. A portion of the intricate network putatively modulated by the CHR-NF-Y module is depicted in Figure 6. Given its apparent pivotal role, it is intriguing that the protein that binds the CHR element is yet to be identified.[13] The list of putative CHR hits we provide could guide the empirical identification of this protein. Experimental analysis of CHR elements on several cell cycle-regulated promoters showed that these elements exert a repressive effect on the expression of their target genes. This suggests a model in which CHR and NF-Y play antagonistic roles, with the former acting as a repressor and the latter as an activator of G_2 + M promoters. This model requires experimental examination in which it will also be interesting to study whether the CHR and NF-Y elements are occupied simultaneously by their respective binding TFs or at different times during cell cycle progression.

We observed that many of the promoters that contain a hit for the CHR-NF-Y module also contain a conserved signature of B-Myb, suggesting a combinatorial role for the triplet CHR-NF-Y-B-Myb module. The promoter of B-Myb itself is regulated by E2F and is activated in late G_1/early S phase.[50] B-Myb is known to cooperate with E2F in the activation of *CDC2* and *CCNB1*.[14] In addition, a repressive CHR element was defined in the B-Myb promoter.[50] Furthermore, FOXM1, a TF recently demonstrated to be required for the execution of the mitotic program,[21] is among our twenty putative targets of the CHR-NF-Y-B-Myb triplet. Taken together, a picture of an intricate regulatory network maintained among the transcriptional regulators of cell cycle progression emerges (Fig. 7).

While preparing this manuscript, a computational paper analyzing cell cycle regulation was published by Zhu et al.[51] These authors too pointed out CHR-NF-Y (together with the CDE element) as a major transcriptional regulatory module of G_2+M genes.

Our methodology and results demonstrate the power of computational analysis applied to functional genomics data in delineating novel aspects of the architecture of the transcriptional network that controls cell cycle progression. High false-positive rates are often a major limiting factor of computational binding site predictions, gravely hampering their experimental examination. Therefore, the significant improvement that we achieved in the specificity of the putative targets can potentially make their empirical validation much more focused and efficient.

References

1. Murray AW. Recycling the cell cycle: Cyclins revisited. Cell 2004; 116:221-34.
2. Castro A, Bernis C, Vigneron S, Labbe JC, Lorca T. The anaphase-promoting complex: A key factor in the regulation of cell cycle. Oncogene 2005; 24:314-25.
3. Fung TK, Poon RY. A roller coaster ride with the mitotic cyclins. Semin Cell Dev Biol 2005; 16:335-42.
4. Dimova DK, Dyson NJ. The E2F transcriptional network: Old acquaintances with new faces. Oncogene 2005; 24:2810-26.
5. Ren B, Cam H, Takahashi Y, Volkert T, Terragni J, Young RA, Dynlacht BD. E2F integrates cell cycle progression with DNA repair, replication, and G(2)/M checkpoints. Genes Dev 2002; 16:245-56.
6. Ishida S, Huang E, Zuzan H, Spang R, Leone G, West M, Nevins JR. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. Mol Cell Biol 2001; 21:4684-99.
7. Cam H, Balciunaite E, Blais A, Spektor A, Scarpulla RC, Young R, Kluger Y, Dynlacht BD. A common set of gene regulatory networks links metabolism and growth inhibition. Mol Cell 2004; 16:399-411.
8. Zwicker J, Lucibello FC, Wolfraim LA, Gross C, Truss M, Engeland K, Muller R. Cell cycle regulation of the cyclin A, *cdc25C* and *cdc2* genes is based on a common mechanism of transcriptional repression. EMBO J 1995; 14:4514-22.
9. Manni I, Mazzaro G, Gurtner A, Mantovani R, Haugwitz U, Krause K, Engeland K, Sacchi A, Soddu S, Piaggio G. NF-Y mediates the transcriptional inhibition of the cyclin B1, cyclin B2, and cdc25C promoters upon induced G2 arrest. J Biol Chem 2001; 276:5570-6.
10. Bolognese F, Wasner M, Dohna CL, Gurtner A, Ronchi A, Muller H, Manni I, Mossner J, Piaggio G, Mantovani R, Engeland K. The cyclin B2 promoter depends on NF-Y, a trimer whose CCAAT-binding activity is cell-cycle regulated. Oncogene 1999; 18:1845-53.
11. Imbriano C, Gurtner A, Cocchiarella F, Di Agostino S, Basile V, Gostissa M, Dobbelstein M, Del Sal G, Piaggio G, Mantovani R. Direct p53 transcriptional repression: In vivo analysis of CCAAT-containing G_2/M promoters. Mol Cell Biol 2005; 25:3737-51.
12. Lucibello FC, Liu N, Zwicker J, Gross C, Muller R. The differential binding of E2F and CDF repressor complexes contributes to the timing of cell cycle-regulated transcription. Nucleic Acids Res 1997; 25:4921-5.
13. Liu N, Lucibello FC, Korner K, Wolfraim LA, Zwicker J, Muller R. CDF-1, a novel E2F-unrelated factor, interacts with cell cycle-regulated repressor elements in multiple promoters. Nucleic Acids Res 1997; 25:4915-20.
14. Zhu W, Giangrande PH, Nevins JR. E2Fs link the control of G_1/S and G_2/M transcription. EMBO J 2004; 23:4615-26.

15. Wasner M, Haugwitz U, Reinhard W, Tschop K, Spiesbach K, Lorenz J, Mossner J, Engeland K. Three CCAAT-boxes and a single cell cycle genes homology region (CHR) are the major regulating sites for transcription from the human cyclin B2 promoter. Gene 2003; 312:225-237.
16. Kimura M, Uchida C, Takano Y, Kitagawa M, Okano Y. Cell cycle-dependent regulation of the human aurora B promoter. Biochem Biophys Res Commun 2004; 316:930-6.
17. Uchiumi T, Longo DL, Ferris DK. Cell cycle regulation of the human polo-like kinase (PLK) promoter. J Biol Chem 1997; 272:9166-74.
18. Joaquin M, Watson RJ. Cell cycle regulation by the B-Myb transcription factor. Cell Mol Life Sci 2003; 60:2389-401.
19. Schubert S, Horstmann S, Bartusel T, Klempnauer KH. The cooperation of B-Myb with the coactivator p300 is orchestrated by cyclins A and D1. Oncogene 2004; 23:1392-404.
20. Costa RH. FoxM1 dances with mitosis. Nat Cell Biol 2005; 7:108-10.
21. Laoukili J, Kooistra MR, Bras A, Kauw J, Kerkhoven RM, Morrison A, Clevers H, Medema RH. FoxM1 is required for execution of the mitotic programme and chromosome stability. Nat Cell Biol 2005; 7:126-36.
22. Elkon R, Linhart C, Sharan R, Shamir R, Shiloh Y. Genome-wide in silico identification of transcriptional regulators controlling the cell cycle in human cells. Genome Res 2003; 13:773-80.
23. Stormo GD. DNA binding sites: Representation and discovery. Bioinformatics 2000; 16:16-23.
24. Tompa M, Li N, Bailey TL, Church GM, De Moor B, Eskin E, Favorov AV, Frith MC, Fu Y, Kent WJ, Makeev VJ, Mironov AA, Noble WS, Pavesi G, Pesole G, Regnier M, Simonis N, Sinha S, Thijs G, van Helden J, Vandenbogaert M, Weng Z, Workman C, Ye C, Zhu Z. Assessing computational tools for the discovery of transcription factor binding sites. Nat Biotechnol 2005; 23:137-44.
25. Duret L, Bucher P. Searching for regulatory elements in human noncoding sequences. Curr Opin Struct Biol 1997; 7:399-406.
26. The ENCODE (ENCyclopedia Of DNA Elements) Project. Science 2004; 306:636-40.
27. Lenhard B, Sandelin A, Mendoza L, Engstrom P, Jareborg N, Wasserman WW. Identification of conserved regulatory elements by comparative genome analysis. J Biol 2003; 2:13.
28. Sandelin A, Wasserman WW, Lenhard B. ConSite: Web-based prediction of regulatory elements using cross-species comparison. Nucleic Acids Res 2004; 32(Web Server issue):W249-252.
29. Birney E, Andrews TD, Bevan P, Caccamo M, Chen Y, Clarke L, Coates G, Cuff J, Curwen V, Cutts T, Down T, Eyras E, Fernandez-Suarez XM, Gane P, Gibbins B, Gilbert J, Hammond M, Hotz HR, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Lehvaslaiho H, McVicker G, Melsopp C, Meidl P, Mongin E, Pettett R, Potter S, Proctor G, Rae M, Searle S, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Ureta-Vidal A, Woodwark KC, Cameron G, Durbin R, Cox A, Hubbard T, Clamp M. An overview of Ensembl. Genome Res 2004; 14:925-8.
30. Kasprzyk A, Keefe D, Smedley D, London D, Spooner W, Melsopp C, Hammond M, Rocca-Serra P, Cox T, Birney E. EnsMart: A generic system for fast and flexible access to biological data. Genome Res 2004; 14:160-9.
31. Wingender E. TRANSFAC, TRANSPATH and CYTOMER as starting points for an ontology of regulatory networks. In Silico Biol 2004; 4:55-61.
32. Sharan R, Maron-Katz A, Shamir R. CLICK and EXPANDER: A system for clustering and visualizing gene expression data. Bioinformatics 2003; 19:1787-99.
33. Kel AE, Kel-Margoulis OV, Farnham PJ, Bartley SM, Wingender E, Zhang MQ. Computer-assisted identification of cell cycle-related genes: New targets for E2F transcription factors. J Mol Biol 2001; 309:99-120.
34. Tagle DA, Koop BF, Goodman M, Slightom JL, Hess DL, Jones RT. Embryonic epsilon and gamma globin genes of a prosimian primate (Galago crassicaudatus). Nucleotide and amino acid sequences, developmental regulation and phylogenetic footprints. J Mol Biol 1988; 203:439-55.
35. Blanchette M, Schwikowski B, Tompa M. Algorithms for phylogenetic footprinting. J Comput Biol 2002; 9:211-23.
36. Ohtani K, Iwanaga R, Nakamura M, Ikeda M, Yabuta N, Tsuruga H, Nojima H. Cell growth-regulated expression of mammalian *MCM5* and *MCM6* genes mediated by the transcription factor E2F. Oncogene 1999; 18:2299-309.
37. Rotheneder H, Geymayer S, Haidweger E. Transcription factors of the Sp1 family: Interaction with E2F and regulation of the murine thymidine kinase promoter. J Mol Biol 1999; 293:1005-15.
38. Chang YC, Illenye S, Heintz NH. Cooperation of E2F-p130 and Sp1-pRb complexes in repression of the Chinese hamster *dhfr* gene. Mol Cell Biol 2001; 21:1121-31.
39. Pilpel Y, Sudarsanam P, Church GM. Identifying regulatory networks by combinatorial analysis of promoter elements. Nat Genet 2001; 29:153-9.
40. Sudarsanam P, Pilpel Y, Church GM. Genome-wide cooccurrence of promoter elements reveals a cis-regulatory cassette of rRNA transcription motifs in *Saccharomyces cerevisiae*. Genome Res 2002; 12:1723-31.
41. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. Mol Biol Cell 2002; 13:1977-2000.
42. Kong M, Barnes EA, Ollendorff V, Donoghue DJ. Cyclin F regulates the nuclear localization of cyclin B1 through a cyclin-cyclin interaction. EMBO J 2000; 19:1378-88.
43. Spruck CH, de Miguel MP, Smith AP, Ryan A, Stein P, Schultz RM, Lincoln AJ, Donovan PJ, Reed SI. Requirement of Cks2 for the first metaphase/anaphase transition of mammalian meiosis. Science 2003; 300:647-50.
44. Piekorz RP, Hoffmeyer A, Duntsch CD, McKay C, Nakajima H, Sexl V, Snyder L, Rehg J, Ihle JN. The centrosomal protein TACC3 is essential for hematopoietic stem cell function and genetically interfaces with p53-regulated apoptosis. EMBO J 2002; 21:653-64.
45. Yasuda S, Oceguera-Yanez F, Kato T, Okamoto M, Yonemura S, Terada Y, Ishizaki T, Narumiya S. Cdc42 and mDia3 regulate microtubule attachment to kinetochores. Nature 2004; 428:767-71.
46. Narumiya S, Oceguera-Yanez F, Yasuda S. A new look at Rho GTPases in cell cycle: Role in kinetochoremicrotubule attachment. Cell Cycle 2004; 3:855-7.
47. Glotzer M. The molecular requirements for cytokinesis. Science 2005; 307:1735-9.
48. Townsley FM, Aristarkhov A, Beck S, Hershko A, Ruderman JV. Dominant-negative cyclin-selective ubiquitin carrier protein E2-C/UbcH10 blocks cells in metaphase. Proc Natl Acad Sci USA 1997; 94:2362-7.
49. Jiang W, Jimenez G, Wells NJ, Hope TJ, Wahl GM, Hunter T, Fukunaga R. PRC1: A human mitotic spindle-associated CDK substrate protein required for cytokinesis. Mol Cell 1998; 2:877-85.
50. Liu N, Lucibello FC, Zwicker J, Engeland K, Muller R. Cell cycle-regulated repression of B-myb transcription: Cooperation of an E2F site with a contiguous corepressor element. Nucleic Acids Res 1996; 24:2905-10.
51. Zhu Z, Shendure J, Church GM. Discovering functional transcription-factor combinations in the human cell cycle. Genome Res 2005; 15:848-55.
52. Pallier C, Scaffidi P, Chopineau-Proust S, Agresti A, Nordmann P, Bianchi ME, Marechal V. Association of chromatin proteins high mobility group box (HMGB) 1 and HMGB2 with mitotic chromosomes. Mol Biol Cell 2003; 14:3414-26.
53. Maxwell CA, Keats JJ, Belch AR, Pilarski LM, Reiman T. Receptor for hyaluronan-mediated motility correlates with centrosome abnormalities in multiple myeloma and maintains mitotic integrity. Cancer Res 2005; 65:850-60.
54. Zannini L, Lecis D, Lisanti S, Benetti R, Buscemi G, Schneider C, Delia D. Karyopherin-alpha2 protein interacts with Chk2 and contributes to its nuclear import. J Biol Chem 2003; 278:42346-51.
55. Schluter C, Duchrow M, Wohlenberg C, Becker MH, Key G, Flad HD, Gerdes J. The cell proliferation-associated antigen of antibody Ki-67: A very large, ubiquitous nuclear protein with numerous repeated elements, representing a new kind of cell cycle-maintaining proteins. J Cell Biol 1993; 123:513-22.
56. Raemaekers T, Ribbeck K, Beaudouin J, Annaert W, Van Camp M, Stockmans I, Smets N, Bouillon R, Ellenberg J, Carmeliet G. NuSAP, a novel microtubule-associated protein involved in mitotic spindle organization. J Cell Biol 2003; 162:1017-29.
57. McGuinness BE, Hirota T, Kudo NR, Peters JM, Nasmyth K. Shugoshin prevents dissociation of cohesin from centromeres during mitosis in vertebrate cells. PLoS Biol 2005; 3:e86.
58. Furukawa K. LAP2 binding protein 1 (L2BP1/BAF) is a candidate mediator of LAP2-chromatin interaction. J Cell Sci 1999; 112:2485-92.
59. Dou Z, Ding X, Zereshki A, Zhang Y, Zhang J, Wang F, Sun J, Huang H, Yao X. TTK kinase is essential for the centrosomal localization of TACC2. FEBS Lett 2004; 572:51-6.