

# MINIREVIEW

## Identification of Target Genes of Oncogenic Transcription Factors (44425)

KATHRYN E. BOYD AND PEGGY J. FARNHAM<sup>1</sup>

*McArdle Laboratory for Cancer Research, University of Wisconsin Medical School, Madison, Wisconsin 53706*

---

**Abstract.** Disregulation of many transcription factors is associated with the development of human neoplasia. Transcription factors regulate cell growth, differentiation, and apoptosis by binding to specific DNA sequences within the promoter regions of growth-regulatory genes and modulating expression of these genes. This simple model is complicated by the fact that mammalian transcription factors are often members of large protein families that bind to similar DNA sequences. This raises the question as to whether members of a particular family regulate expression of overlapping or unique sets of genes. This review is focused on addressing this question using the Ets, Myc, and E2F transcription factor families as examples. Deregulated activity of some, but not all, members of these families is observed in cancer. Here, we summarize the data illustrating the concept that binding of individual members of these families of factors can result in promoter-specific responses and review the studies that have provided some insight into how target gene specificity is achieved. Since, for all of these oncogenic transcription factors, it remains unclear exactly which target genes are important in neoplasia, we have also reviewed the many approaches researchers are using to identify target genes of the various Ets, Myc, and E2F family members.

[P.S.E.B.M. 1999, Vol 222]

---

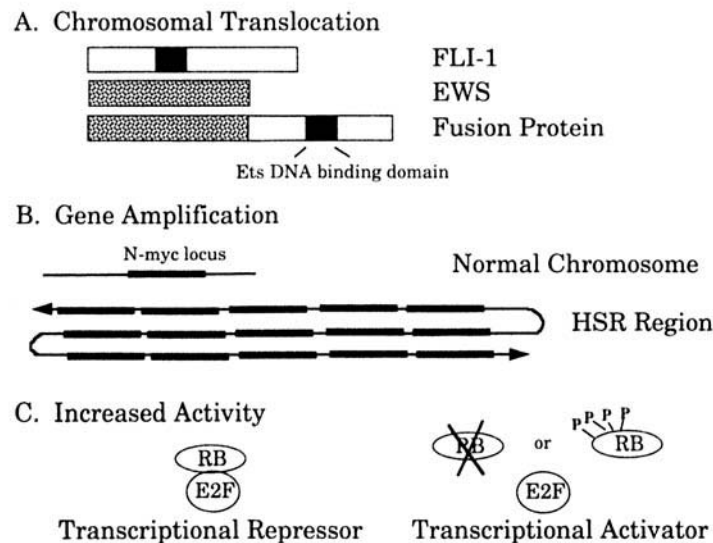
Most nuclear oncogenes have been identified as promoter-specific transcription factors that can cause neoplastic transformation when their activity is increased to inappropriately high levels. These promoter-specific transcription factors regulate the activity of the basal transcription complex (which includes RNA polymerase II and initiation factors TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, and TFIIH) formed near the transcription start site of the gene. Promoter-specific transcription factors can control the switch from inactive to active chromatin, stabilize the transcription complex on the promoter DNA, recruit other factors to the complex, or mediate the transition from initiation to elongation.

Promoter-specific transcription factors usually have two modular domains, a DNA binding domain and a regulatory domain. Several classes of site-specific transcription factors have been identified based on their DNA binding and/or dimerization domains. Commonly found dimerization structures include helix-loop-helix motifs and leucine zippers. DNA binding domains include basic regions, which are often found in combination with helix-loop-helix and/or leucine zippers, homeobox domains, which consist of a helix-turn-helix structure, and the zinc finger motif that consists of pairs of cysteine-cysteine or histidine-cysteine residues that bind zinc and form finger-like structures that interact with specific DNA sequences (1). There are multiple examples of transcription factors that use each type of DNA binding motif.

Increased activity of promoter-specific factors is achieved in many ways in human cancers (Fig. 1). For example, leukemia-associated chromosomal rearrangements can result in expression of a novel transcription factor (2); many of these rearrangements involve members of the Ets

---

<sup>1</sup>To whom requests for reprints should be addressed at McArdle Laboratory for Cancer Research, 1400 University Avenue, Madison, WI 53706. E-mail: farnham@oncology.wisc.edu



**Figure 1.** The activity of oncogenic transcription factors can be modulated in a variety of ways in human tumors. (A) Many leukemias and solid tumors are characterized by the presence of a chromosomal translocation that creates a novel fusion protein. In the example shown, a member of the Ets family, the FLI-1 protein, is fused to the EWS protein to create a novel transcription factor that retains the ability to bind to Ets sites. (B) The increased levels of transcription factors in tumor cells has been shown to be due to gene amplification. The extra copies of the genes encoded in the amplicons can be found as extrachromosomal elements called double minutes or as integrated chromosomal structures called homogeneously staining regions (HSRs). In the example shown, the HSRs contain the N-myc gene and surrounding chromosomal DNA. In neuroblastomas, more than 100 copies of the N-myc gene have been found. (C) Functional loss of a negative regulator can increase the activity of pre-existing transcription factors. In the example shown, the activity of pre-existing E2F family members is increased due to loss of the negative regulator RB or due to the increased phosphorylation state of RB caused by alterations in the abundance or activity of cyclin-dependent kinases.

family of transcription factors. Other cancers, such as Burkitt's lymphoma and neuroblastoma, are thought to be due to increased levels of structurally normal members of the Myc family of transcription factors (3, 4). Finally, in certain cases such as retinoblastomas, mutations in upstream regulators result in increased activity of members of the E2F family of transcription factors (5). Each of these three families of transcriptional regulators, Ets, Myc, and E2F, is composed of multiple members, and all members of a given family contain highly related DNA binding domains. In general, *in vitro* DNA binding assays indicate that factors having similar DNA binding motifs bind to similar DNA sequences. In fact, it is this commonality of DNA binding motifs that creates the problem that will be discussed in this review. If a group of different transcription factors can all bind to the same sequence, how does one determine which member of the family regulates a particular target gene? Of particular importance to cancer is the corollary question, how can increased levels or activity of one member of a family of transcription factors cause neoplastic transformation whereas high levels of other, very similar, family members do not lead to the loss of cell growth control?

In this review, we will briefly describe the members of three different families of oncogenic transcription factors; the Ets family, the Myc family, and the E2F family. We will use these examples to illustrate the concept that different members of a family of factors can cause different responses when bound to different target genes, and we will review the studies that have provided some insight into how this target gene specificity is achieved. However, it is not

yet clear that the known target genes of these transcription factors are the relevant targets in neoplastic transformation. Therefore, we will also describe a variety of methods by which investigators are currently attempting to discover novel target genes of oncogenic transcription factors. We will conclude by presenting a general approach that can serve as a rudimentary guide to understanding the neoplastic properties of many different oncogenic transcription factors.

## Oncogenic Transcription Factors are Members of Multigene Families

**The Ets Family of Transcription Factors.** Members of the Ets family of transcription factors play important roles in organogenesis, hematopoiesis, B-cell development, and signal transduction (for a review see Ref. 6). These proteins all bind to a purine-rich consensus sequence (C/A GGA A/T) and transcriptionally activate a number of genes that contain these Ets binding sites. The Ets DNA binding domain is composed of 85 amino acids; structural analysis indicates the presence of three  $\alpha$  helices and a four-stranded  $\beta$  sheet similar to structures of helix-turn-helix motifs found in several other mammalian and bacterial transcription factors. Disregulation of Ets family members is often associated with neoplasia. For example, *v-ets*, the first member of the Ets gene family to be identified, is co-transduced with the *v-myb* gene in the E26 retrovirus that expresses a GAG-MYB-ETS fusion protein. Similarly, a newly identified Ets protein, ESX, becomes overexpressed at an early stage of human breast cancer development (7). Members of the Ets family of transcription factors are often found as fusion

proteins in leukemias and other solid tumors. In accordance with a postulated role for Ets proteins in regulating cell growth control, overexpression of various Ets proteins (e.g. Ets1 and Ets2) can lead to neoplastic transformation of cells in culture and can cause tumors in nude mice.

Ets family members provide many interesting examples of how chromosomal rearrangements can lead to neoplastic transformation. In particular, almost all leukemias are characterized by a particular DNA rearrangement that results in fusion of a transcription factor to another protein. In some cases, the second protein is also a transcription factor; in other cases, the function of the fusion partner is as yet unknown. An Ets family member that is a common target of gene rearrangements in human leukemias is a site-specific DNA binding transcription factor called *tel* (translocation, *ets*, leukemia). *Tel* is rearranged in acute myelogenous leukemia (AML), chronic myelomonocytic leukemia (CMML), myelodysplastic syndrome (MDS), and acute lymphoblastic leukemia (ALL). There are now more than 20 different chromosomal translocations in human leukemias involving the *tel* gene. One aspect of the *tel* gene fusions is that different regions of the *tel* gene are retained in the different leukemias. For example, in the TEL/MN1 fusion, the TEL DNA binding domain is fused to the MN1 protein. This is thought to result in deregulation of Ets target genes. In contrast, in the TEL/AML1 fusion, the N-terminal helix-loop-helix domain of TEL (but not the DNA binding domain), is fused to a nearly complete AML1 protein. The TEL/AML1 fusion has an extremely high occurrence frequency in ALL, the most common malignancy of childhood (8); thus, the TEL/AML1 fusion is the most common gene rearrangement in any pediatric malignancy. AML1 normally binds to the DNA sequence TGTGGT, which is found in the transcriptional regulatory region of the T-cell antigen receptor and in a variety of cytokines and their receptors. Thus, for the TEL/AML1 fusion protein, neoplastic transformation is thought to be due to deregulated transcriptional activity of AML1 target genes. It appears as if the fusion protein does not activate transcription, but rather inhibits basal promoter activity (9). Thus, it has been proposed that the TEL/AML1 fusion protein is a dominant interfering protein that can efficiently downregulate all AML-regulated target genes and that it is the loss of AML-activated gene expression that causes leukemia.

EWS/FLI-1 and EWS/ERG are fusion genes characteristically found in Ewing's sarcomas and primitive neuroectodermal tumors of childhood. EWS/FLI-1 and EWS/ERG are chimeric proteins consisting of the amino terminus of EWS, a putative RNA-binding protein, fused to the carboxyl terminal DNA binding domain of one of two different Ets family members, FLI-1 or ERG. The mechanism of transformation by fusion proteins that incorporate Ets DNA binding domains is not known, but likely includes inappropriate activation of genes containing Ets binding sites in their promoter regions. Evidence in support of this hypothesis includes the finding that all tumors containing an EWS/

FLI-1 fusion have an intact FLI DNA binding domain and deletion of this Ets DNA binding domain abolishes the ability of EWS/FLI-1 to transform NIH 3T3 cells. There are many Ets family members in a cell. Although the EWS/FLI-1 rearrangement provides one additional factor that can bind to an Ets site, the other normal Ets family members are still present. Therefore, one might assume that the fusion protein will be in competition with other normal Ets proteins. It is unclear how simply increasing the amount of one particular DNA binding factor, in the continued presence of the multitude of other factors present in the cell that bind to the same sequence, can dramatically influence gene regulation. However, studies suggest that EWS/FLI-1 and FLI-1 may regulate distinct subsets of target genes although they have the same DNA binding domain (10). These experiments, as well as others that suggest that target gene specificity among the Ets family members (both normal and oncogenic varieties) does exist, are discussed below.

**The Myc Family of Transcription Factors.** There are three very similar mammalian *myc* genes, *c-myc* (which can encode three separate proteins called Myc1, Myc2, and MycS), *N-myc*, and *L-myc* (11). All Myc proteins share a carboxyl terminal basic region helix-loop-helix-leucine zipper (bHLH-LZ) domain that mediates sequence-specific DNA binding and dimerization. Myc dimerizes with Max, another bHLH-LZ protein, and these heterodimers bind to the core DNA sequence CACGTG. Two predominant Max proteins exist that differ by a nine amino acid insertion amino-terminal to the basic domain. Other minor splice variants of Max have also been identified. Although Max protein levels are constitutive under all growth conditions, Myc protein levels are high in proliferating cells and low in quiescent or differentiated cells. The Myc proteins (except for MycS) have a transactivation domain at the amino terminus. Max does not contain a transactivation domain and thus Max/Max homodimers bind to, but do not activate transcription from, the same sites as Myc/Max heterodimers. Another protein, called B-Myc, is thought to be encoded by a duplicated second exon of the *c-myc* gene. This protein contains the transactivation domain but is missing the third exon of *c-myc* and thus does not have a nuclear localization signal or a DNA binding/heterodimerization motif (12).

The *c-myc* gene is probably the most extensively studied nuclear oncogene. The oncogenic potential of *myc* can be activated by chromosomal translocation, retroviral insertion, or gene amplification. For example, rearrangements of the *c-myc* gene, to a locus downstream of the immunoglobulin enhancer, is found in every case of Burkitt's lymphoma. One of the earliest discoveries of the role of *c-Myc* in carcinogenesis was the finding that avian leukosis virus (ALV) inserts near the *c-myc* gene in chickens and controls its aberrant expression, leading to the development of lymphomas. Amplified *c-myc* is found in carcinomas of the colon, lung, breast, and ovary, as well as other tumor types. The effects of over- and underexpression of *c-Myc* on cell

growth have been examined in several experimental systems. Overexpressed c-Myc protein can immortalize primary cells and transform established cell lines. Increased expression of c-Myc can also induce apoptosis in cells growth-arrested by a variety of means and at various points in the cell cycle. Several studies have employed c-Myc knockout cell lines and c-Myc antisense RNA to examine the effects of inhibiting *myc* expression. In general, artificially raising the amount of c-Myc can transform a cell and lowering the amount of c-Myc slows cell growth and delays entrance into S phase. The N-*myc* gene is expressed in cells of primarily neuronal lineages during embryogenesis (13). N-Myc heterodimerizes with Max and binds to the same DNA sequence element as c-Myc/Max heterodimers. However, unlike c-*myc* which is expressed in many cell types and displays deregulation in a variety of tumors, the oncogenic activation of N-Myc occurs only in neuroblastomas and in other types of neuroectodermal tumors. In neuroblastoma, the most common alteration of N-Myc levels is mediated by gene amplification that can be correlated with a poor prognosis. In some cases, the N-*myc* gene has been found to be amplified more than 100-fold. Fewer studies have been performed on the L-*myc* gene. Amplification of L-*myc* has been seen in cases of small cell lung cancer. Due to the high degree of homology between c-*myc* and L-*myc*, it is generally assumed that overexpressed L-Myc will function in a manner similar to overexpressed c-Myc in a transformation assay; however, L-Myc has been shown to co-transform rat embryo cells with a lower efficiency than c-Myc (14).

Although it is clear that overexpression of Myc family members can lead to neoplastic transformation, the mechanism by which Myc mediates this response is controversial. One hypothesis put forth is that transformation by Myc is regulated by interaction between Myc and other cellular proteins. The basic-HLH region of Myc has been shown to interact with a variety of proteins (other than the heterodimeric partner, Max) such as BRCA1, YY1, AP2, NMI-1, and Miz1. Two other protein interaction domains that are found in the N-terminus of all Myc family members, called Myc Box 1 (MB1) and Myc Box 2 (MB2) have been implicated in Myc transforming activity. MB2 is a contact site for a cellular protein called TRRAP that participates positively in Myc-mediated neoplastic transformation (15). A cellular protein called BIN1 has recently been shown to bind to MB1 (16). This protein can inhibit transformation mediated by Myc in colony formation assays. It is likely that the oncogenic potential of Myc family members is related to their ability to regulate gene expression. The identification of target genes for the Myc family of transcriptional regulators has been the focus of much study for almost two decades. To date, two dozen or so genes have been suggested to be regulated by c-Myc (11). For most, it is unclear how increased or decreased levels of the putative target protein are critically important for Myc's oncogenic

capacity. It has been proposed that a Myc target gene should a) have a binding site for Myc that is crucial for transcriptional regulation; b) show a pattern of growth regulation similar to that of the Myc protein (or opposite that of Myc for a Myc-repressed target gene); and c) encode a protein that, when deregulated, could contribute to neoplasia. Some genes that have been suggested to be activated by Myc include *cad* (required for pyrimidine biosynthesis), *cdc25a* (a cell cycle phosphatase), *odc* (required for polyamine biosynthesis), *elF-4E* (a translation factor), and *ISGF3 $\gamma$*  (a transcription factor). Evidence also exists to suggest that Myc represses target genes that negatively influence cell proliferation; such genes may include *C/EBP $\alpha$*  (a transcription factor), and *gadd153/CHOP* (a transcription factor). The molecular mechanisms by which c-Myc mediates transcriptional repression are not well defined; however, it is believed that c-Myc interacts with proteins other than Max to repress transcription and that DNA binding sites distinct from Myc/Max binding sites may be required. It was recently demonstrated that expression of many of these putative c-Myc target genes remains unchanged in c-Myc null cells, with the exception of *cad* and *gadd45* (17). Therefore, it is likely that many true transcriptional targets of c-Myc have yet to be identified.

The relative importance of Myc as a transcriptional regulator in mediating neoplastic transformation is highly debated. Much of this controversy comes from the difficulty in clearly defining a true Myc target gene. Since other proteins that can bind to the same DNA element (which is called an E box) as the Myc/Max heterodimers also exist in cells, it has been difficult to classify cellular promoters as being activated specifically by Myc. A group of proteins that includes Mad, Mxi1, Mad3, and Mad4 can also dimerize with Max and bind to E boxes. Mad/Max heterodimers are repressors of transcription due to interactions with the mSin3 family of transcriptional repressors. It is postulated that Mad/Max complexes function in differentiation pathways by inhibiting transcription from Myc-activated promoters. Yet another partner of Max called Mnt (or Rox) has recently been cloned. Mnt/Max also functions as a transcriptional repressor in a complex with mSin3 proteins, and it has been postulated that Mnt/Max complexes may be involved in repression of transcription in quiescent cells. Other proteins such as USF1 and USF2, as well as TFE3 and TFEB, can bind to and transactivate *via* an E box. Overexpression assays often show that a putative Myc target is also regulated by USF and Mad family members. In addition to being in competition with other E box-binding proteins, Myc can also bind to another sequence element called an initiator. This element is located at the transcription start site of some cellular genes and is thought to be a site at which both positive and negative regulatory proteins can interact with the basal transcriptional machinery. Myc has been shown to negatively regulate some promoters that have particular initiator elements (18) and to interact directly with another protein called TFII-I that also binds to initiator elements

(19). It is proposed that Myc heterodimerizes with TFII-I instead of Max to mediate transcriptional repression. Several experiments have highlighted the significance of transcriptional repression by Myc. For example, Myc proteins that have lost the transcriptional repression function due to mutation of a region of the N-terminus also lose the ability to transform cells (18). Although this suggests that the repression function of Myc is critical for transformation, others have shown that these same Myc constructs have also lost transcriptional activation at certain promoters (20). Recent studies of MycS, a naturally occurring N-terminal-truncation of c-Myc which can only repress transcription, may provide some insight into this problem (21). MycS can mimic Myc in certain assays, such as stimulation of proliferation and allowing for anchorage-independent growth. However, unlike full length Myc, the N-terminally truncated MycS protein cannot transform primary cells in cooperation with Ras. These results suggest that the repression function of Myc may be important in imparting some, but not all, of the biological phenotypes seen in tumors having high levels of Myc. It is likely that critical Myc target genes will include both those that are activated by and those that are repressed by Myc family members.

As with the Ets family, it is unclear why deregulation of Myc family members is commonly associated with human cancer when in the same cells there exists highly abundant factors such as USF that can bind to and activate transcription from the same DNA elements. Studies of mice that have been engineered to lack different USF factors show that whereas USF1 and USF2 may encode redundant functions, mice lacking both USF1 and USF2 are embryonic lethal. Thus, Myc family members cannot substitute for USF family members. Similarly, mice lacking either N-Myc or c-Myc die *in utero* (mice lacking L-Myc are normal), suggesting that N-Myc and c-Myc may have nonredundant transcriptional activities and that the abundant and ubiquitous USF proteins cannot substitute for either c-Myc or N-Myc (22–27). Another unsolved question is why deregulation of different Myc family members is associated with different types of tumors. c-Myc, N-Myc, and L-Myc can all cooperate with a mutated Ras to transform cells neoplastically in culture and transgenic mice overexpressing c-Myc, N-Myc, or L-Myc *via* an immunoglobulin enhancer develop malignancies (28–30). However, N-Myc is associated only with neuroblastomas and L-Myc with small cell lung cancers, whereas c-Myc is associated with a variety of human tumors. Does N-Myc activate a distinct set of target genes whose deregulation is only critical in certain cells or is N-Myc overexpression detrimental to all cells in the body except for neuronal cells? It is likely that the answers to all of these questions will require an understanding of target gene specificity of these DNA binding transcription factors. A review of our general understanding of Myc family member target gene specificity is provided below in Section III.

**The E2F Family of Transcription Factors.** The E2F family is a collection of transcription factors that func-

tion as heterodimers that bind to and regulate transcription from a consensus sequence (TTTSSCGC, where S is C or G) known as an E2F site (31–33). To date six different mammalian E2Fs (E2F1, E2F2, E2F3, E2F4, E2F5, and E2F6) have been cloned; each of them can heterodimerize with either DP1 or DP2. The E2F components contain a central DNA binding and dimerization domain and (except for E2F6) a C terminal transactivation domain. A subset of the E2Fs (E2F1, E2F2, and E2F3) also contains a conserved N-terminal protein interaction domain. One protein that can bind to the N terminal domain is cyclin A. Studies have shown that cyclin A-dependent kinase activity, but not cyclin B-, E-, or D-dependent kinase activity, can negatively regulate the *in vitro* DNA binding activity of heterodimers containing E2F1, E2F2, or E2F3. The DP proteins contribute to the DNA binding activity of the heterodimer, but do not contain a transactivation domain. Nested within the transactivation domain of the E2F proteins is a protein interaction domain that mediates contact with the Retinoblastoma (Rb) family of proteins. E2F1, E2F2, and E2F3 all bind preferentially to Rb; E2F4 and E2F5 bind preferentially to the Rb-related proteins p107 and p130. E2F1, E2F2, and E2F3 have nuclear localization signals and are predominantly nuclear at all times. However, E2F4 and E2F5 lack nuclear localization signals; they are brought into the nucleus *via* interaction with either DP2 or a splice variant of DP1, both of which have nuclear localization signals. E2F4 and E2F5 can also be brought into the nucleus by interaction with p107 or p130.

Rb, p107, and p130 can all repress transcription of a promoter that contains an E2F site. However, it is clear that the binding of the Rb family of proteins to E2F family members does not inhibit their DNA binding activity. Therefore, the Rb family of proteins must influence E2F transcriptional activity. The E2F transactivation domain has been shown to interact with several components of the basal RNA polymerase II transcriptional machinery, such as TBP, TFIIF, and CBP. Since E2F cannot bind to both Rb and the other transcription factors at the same time, it is believed that Rb may repress E2F-mediated transcriptional activation by interfering with functional transcription complex formation. Because Gal4/Rb fusion proteins can block transcription when artificially brought to the DNA of a variety of promoters (34), Rb must also have a functional repression domain that is distinct from the E2F binding domain. Insight into the mechanism by which Rb can repress transcription comes from the recent findings that Rb, p107, and p130 can all associate with a histone deacetylase (35). It is postulated that the recruitment of the histone deacetylase to E2F-regulated promoters represses transcription through changes in chromatin structure. Rb (or p107 or p130) must be released from the E2F/promoter complex for the gene to be transcribed. The disruption of the E2F/Rb protein complex is thought to be primarily due to the action of cell cycle-regulated kinases. Each of the Rb family members contains multiple sites for phosphorylation by these kinases.

The current model is that cyclin D-dependent kinases initiate the phosphorylation events and that cyclin E- and cyclin A-dependent kinases complete the hyperphosphorylation. Hyperphosphorylated Rb (as well as hyperphosphorylated p107 and p130) cannot bind to E2F proteins. Thus, the increased cyclin-dependent kinase activity that results as cells progress through G<sub>1</sub> into S phase of the cell cycle results in release of Rb from E2F and activation of E2F target genes. One of these E2F target genes is E2F1 that leads to a positive autoregulation loop. The action of the cyclin-dependent kinases is kept under control by cdk inhibitors of the p16 and p21 families. In turn, p21 is controlled by the activity of the p53 tumor suppressor protein. Deregulation of the different components of this pathway (e.g., by decreasing levels of Rb, p53, or the cdk inhibitors or by increasing the levels of the cyclins) can result in upregulation of E2F activity.

Unlike members of the Ets and Myc families of transcription factors, E2F family members have not been found to be mutated in human cancers. However, the factors that control the activity of E2F proteins are mutated in a large number of different types of human tumors. Many studies have shown that Rb and p53 are lost in a variety of human tumors and that cyclin D1 and cyclin E are upregulated in human tumors. The loss of Rb and the increased cyclin-dependent kinase activity (with the resultant increased phosphorylation of Rb) both result in the conversion of the E2F/Rb transcriptional repressor complex to the E2F transcriptional activator. The frequent deregulation of the Rb/cyclin signal transduction pathway in human cancers suggests that E2F activity should be increased in tumor cells. Several recent studies have indicated increased E2F activity in tumors having mutated Rb protein (5, 36). These studies, taken in combination with the fact that known E2F target genes encode proteins whose activities are required for DNA synthesis (e.g., dihydrofolate reductase, DNA polymerase- $\alpha$ , and thymidine kinase) or cell cycle progression (e.g., cyclin E and B-*myb*), suggest that overexpression of E2F proteins should provide an environment conducive to DNA replication and neoplastic transformation. Accordingly, several studies have demonstrated the ability of E2F family members to transform cells grown in culture (37). Also, recent studies have shown that E2F1 can cooperate with Ras to induce tumors in keratinocytes in transgenic mice (38, 39).

Several strains of mice have been created that are nullizygous for a particular E2F family member. In general, these mice are viable for at least several weeks after birth (40–42) suggesting that E2F family members can at least partially substitute for each other in the regulation of many E2F target genes. However, E2F1 null mice develop tumors in a limited number of organs after about a year, and E2F5 null mice develop hydrocephalus caused by excessive secretion of cerebrospinal fluid, suggesting that these E2Fs are not completely redundant with the other E2F family members. Additional evidence in support of the hypothesis that

different E2F complexes regulate different target genes will be described below. The apparent contradiction between the overexpression studies of E2F1 (which suggests it is an oncogene) and the nullizygous phenotype (which suggests that E2F1 may be a tumor suppressor) might be explained by the complex regulation of E2F activity. The increased amounts of E2F1 upon overexpression in tissue culture and in transgenic mice will change the ratio of E2F1/Rb complexes versus free E2F in the cell. If enough E2F1 is provided, it can sequester the Rb protein, then E2F target promoters will be occupied by free E2F and derepression can occur. The loss of E2F1 can also result in removal of Rb from E2F1-specific promoters, again leading to derepression of certain E2F target genes. It is also possible that it is the loss of activation of a subset of E2F target genes in the E2F1 null mouse, rather than a loss of repression, that leads to tumor formation in certain tissues. Similar to the situation with the Myc oncogene, it is likely that the set of important E2F target genes will include genes that are activated by E2F family members as well as genes that are repressed by E2F family members.

### Oncogenic Transcription Factors Show Target Gene Specificity

As described above, families of oncogenic transcription factors are grouped according to their ability to bind to the same DNA consensus element; accordingly, the DNA binding domains in the various members of a family of transcription factors are highly conserved. *In vitro* analyses have shown that, in general, similar DNA binding domains will recognize similar DNA sequences. However, there is biological evidence suggesting that different Ets, Myc, and E2F family members have different functions. This indicates that there must be some target gene specificity among different, yet highly similar, family members. This apparent contradiction between initial *in vitro* DNA binding studies and biological phenotypes can be reconciled by a variety of mechanisms. For example, it is possible that some DNA binding specificity can be determined by subtle differences in the structure of two highly related factors. This may allow two factors to bind to certain common elements but also allow a subset of sites to be bound only by one of the factors. In this case, the specificity is likely to be retained in assays that use isolated DNA binding sites. Another mechanism to engender target gene specificity would be to allow for the context of the binding site within a promoter region to influence the affinity or activity of some, but not all, members of a family of transcription factors at a particular promoter. In this case, specificity may only be observed when longer DNA fragments are used and when additional proteins are included in the assay. As described below, both of these modes of specificity are used to provide target gene specificity for members of the Ets, Myc, and E2F families of oncogenic transcription factors.

**Differences Can Be Determined by the Specific Target Site.** As noted above, levels of c-Myc are quite

low in a normal cell but are increased in human tumors. This has led to the hypothesis that the increased levels of c-Myc lead to neoplastic transformation of cells by deregulating transcription of Myc target genes. However, this hypothesis does not take into account the fact that other proteins that bind to the same sequence, in particular USF1 and USF2, are ubiquitous and highly abundant in both normal and tumor cells. One possible hypothesis to explain why Myc, but not USF, can transform cells is that USF might bind to only a subset of the E boxes that are bound by Myc. Therefore, several studies have focused on determining if Myc and USF binding sites are identical or if the related proteins bind overlapping subsets of sites. In general, two different types of experiments have been performed that are directed toward a more detailed understanding of Myc versus USF binding sites. In the first type of experiment, binding sites are selected from a pool of random sequences using the DNA binding properties of purified protein. The development of a consensus sequence allows for identification of a high affinity site, whereas the obvious exclusion of a particular nucleotide at a particular position from all captured sequences suggests the identification of a disfavored nucleotide. Another type of experiment involves direct mutational analysis of the sequences flanking the consensus hexamer CACGTG of a particular promoter. Using a selection procedure, Solomon *et al.* (43) found that a c-Myc/Max heterodimer fails to bind to a CACGTG hexamer when the core is flanked by a 5' T or a 3' A. In contrast, Bendall and Molloy (44) selected for USF binding sites and found little preference for particular sequences flanking the core binding site. A comparison of the ability of USF and c-Myc to bind to the selected USF binding sites showed that a T at the -4 position was inhibitory to Myc binding. Similarly, direct mutation of the sequences flanking the CACGTG hexamer in the hamster *cad* promoter indicated that both USF and Myc could bind to the wildtype E box in the promoter but that mutation of the flanking sequences could abolish Myc, but not USF, binding (45). Thus, several studies have shown that although USF is relatively insensitive to the specific sequences flanking the CACGTG of the E box, alterations of one or two base pairs adjacent to the core hexamer can greatly influence the affinity of Myc for the binding site. Although most of the analysis of the different E boxes has been performed using *in vitro* DNA binding assays, several studies have assayed for E box function within the cell. For example, when the sequences flanking the E box in the hamster *cad* promoter were altered to contain a 5' T and a 3' A, transcriptional regulation of the promoter was abolished in mouse cells (45). Similarly, E boxes containing a 5' T and a 3' A were also unable to support activation by c-Myc in yeast cell assays (46).

In summary, USF appears to bind to a wider variety of E boxes than does c-Myc. However, the collection of sites to which a factor binds may be more complex than initial studies indicate. For example, although *in vitro* selection of Myc binding sites identified the CACGTG consensus as a

high affinity site, a series of non-canonical binding sites was also obtained (47). Furthermore, Hann *et al.* (48) have shown that Myc1 (a slightly longer form of c-Myc that is abundant in certain growth-inhibited cells) can bind to and transactivate promoters that contain a C/EBP consensus element (TTATGCAAT). Such studies, which indicate that a factor can bind to a variety of different sites, taken in combination with the multitude of experiments indicating that promoter context can influence binding (see below), suggest that a better approach for identification of target genes would be to identify sites to which a transcription factor is bound in genomic DNA. These methods are discussed below.

There is also much evidence that different Ets family members have different biological properties. Similar to the studies of Myc and USF binding, a comparison of DNA binding properties of different Ets proteins has revealed that not all Ets proteins bind to exactly the same sequences. For example, binding site selection using the Ets1 DNA binding domain identified a consensus sequence that differs from the binding site consensus for the Ets domain protein E74A (49). Similarly, a comparison of protein binding to two Ets binding sites in the interleukin-2 promoter indicated that Elf-1, but not Ets1 or Ets2, could bind to these particular Ets sites (50). Some of the documented differences in target gene specificity between the various Ets family members are due to the slight differences in the two DNA binding domains. For example, DNA binding site selection indicates that the Ets domain in Elk-1 exhibits a more stringent DNA-binding site specificity than does the corresponding domain of Sap-1 (51). Elk-1 selects sites conforming to the consensus sequence ACCGGAAGTR, whereas Sap-1 selects the more degenerate sequence A(C/t)CGGA(A/t)(G/a)(T/c)N. Thus, the Ets domain of Sap-1 can bind to a series of sites to which the Ets domain of Elk-1 cannot bind. Further studies have identified the regions of the Elk-1 and Sap-1 DNA binding domains that mediate the differences in target site selection. Mutational analyses have shown that two amino acids found at particular positions in Sap-1 allow a greater degree of flexibility in binding to DNA than do the two amino acids found in similar positions in Elk-1 (51).

In addition to amino acid changes in DNA binding domains conferring target gene specificity, DNA binding specificity at isolated Ets consensus sites can also be modulated by other regions of the Ets protein. One such example comes from a comparison of FLI-1 and EWS-FLI-1 (52). Both of these proteins contain the exact same Ets DNA binding domain, however, they demonstrate different DNA binding specificity. FLI-1 can bind to the Ets site adjacent to the serum response element of the *fos* promoter, but only in the presence of another protein called serum response factor (SRF). However, the EWS-FLI-1 fusion protein binds to the same site in the absence of other protein-protein interactions. Deletional analysis of FLI-1 revealed the presence of an inhibitory domain in the N terminus of FLI-1 (which is missing in the fusion protein) that prevents autonomous

binding of FLI-1 to the Ets site in the SRE. Thus, differences in protein structure outside of the highly conserved DNA binding domains of family members can contribute to target gene specificity. In summary, at least some of the target gene specificity observed with the Ets family of transcription factors can be explained by specific differences in various domains of particular Ets proteins. However, as will be described below, target gene specificity in the Ets family can also be conferred by methods other than selective DNA binding.

The DNA binding domains of all the E2Fs are very similar, and *in vitro* DNA binding assays suggest that most, if not all, of the E2F/DP heterodimers bind with similar affinity to the same collection of target sites. These *in vitro* studies are complicated by the fact that E2F4 is very abundant in comparison to the other E2Fs and comprises the majority of the E2F activity in nuclear extracts of most types of cells. Therefore, most *in vitro* DNA binding assays are mainly a monitor of the ability of E2F4 to bind to an E2F site. However, there is one report that demonstrates differential binding of different E2Fs to a particular site. Liu *et al.* (53) have reported that a site in the cyclin A promoter is weakly recognized by E2F4, but clear binding is seen by E2F1 and E2F3. In contrast to most of the *in vitro* assays that show little specificity of different E2Fs, a comparison of the ability of the different E2Fs to activate a panel of genes upon overexpression in quiescent cells using an adenoviral construct suggests that target gene specificity does exist *in vivo* (54). Further analysis suggests that some of these differences may be due to the fact that the E2F/DP heterodimers also interact with members of the Rb family. *In vitro* casting experiments suggest that the optimal consensus site for the trimolecular complex E2F1/DP1/Rb is different from the optimal consensus for E2F1/DP1 (55). For example, the trimolecular complex containing Rb preferentially bound to two inverted, overlapping E2F sites (TTTc/gGCGCg/cAAA) whereas the E2F1/DP1 dimer preferred a single site (TTTCCCGC). A requirement for binding to an inverted, overlapping site (as opposed to a single E2F site) would greatly restrict the number of target genes that could be regulated by E2F1 when bound to Rb. Therefore, the Rb/E2F1/DP1 complex may regulate only a subset of genes that can be bound by E2F1/DP1. All combinations of Rb family members with different E2F/DP heterodimers have not yet been tested for DNA site specificity. However, it is possible that the addition of Rb family members to the E2F/DP heterodimers may cause protein conformational changes that allow a divergence of binding specificities of the different E2Fs.

In summary, members of a family of transcription factors recognize a very similar set of DNA binding sites. However, subtle differences inherent in the structure of the different family members and/or conformational changes caused by interaction with another protein can result in different family members binding to overlapping, but not identical, sets of DNA sequences. As shown in Figure 2, USF is

Binding Site	Proteins Bound
<u>CCACGTGG</u>	Myc/Max; USF
(c/t) <u>CACGTG</u> (g/a)	USF
<u>ACCGGAAGTR</u>	Elk-1; Sap1
A(c/t) <u>CGGA</u> (a/t) (g/a) (t/c)N	Sap1
<u>TTT</u> (c/g) <u>GCGC</u> (g/c) AAA	E2F1/DP1/RB
<u>TTTCCCGC</u>	E2F1/DP1

**Figure 2.** Sequences flanking core consensus elements can determine family member specificity. Shown are examples of how alterations in the nucleotides flanking Myc, Ets, and E2F binding sites can affect the binding affinity of specific members of a family of transcription factors (see text for details). For the top and middle panels, the lower oligonucleotide represents a less stringent consensus that can be bound by some, but not all, of the DNA binding protein complexes that represent a particular family of transcription factors. For the bottom panel, the inclusion of a third protein in the E2F1/DP1 complex alters the DNA binding preference of the E2F family member. In each case, the sequence identical in the pairs of binding sites is underlined.

allowed a greater flexibility in the sequences that flank an E box than is Myc. Similarly, Sap-1 binds to a less stringent consensus sequence than does Elk-1. Finally, the interaction of E2F1/DP1 heterodimers with Rb can restrict the subset of possible E2F sites available to E2F1.

**Differences Can Be Determined by Promoter Context.** Target gene specificity can also be determined by the context of the DNA binding site in relation to other promoter elements. For example, one of the most intriguing aspects of the Ets family of proteins is that they do not seem to associate as homo- or heterodimers, unlike many other families of transcription factors. Instead they form complexes with transcription factors of unrelated families. On their own, Ets family members display only weak transactivation properties but when combined with other factors they can robustly stimulate transcription. Thus, different protein-protein interaction domains can result in different Ets factors regulating different subsets of target genes. An example of how critical protein-protein interactions can be in the activation of Ets target genes comes from the analysis of the effect of Erg, an Ets family member, on two different Ets target genes (56). Both collagenase and stromelysin have multiple Ets consensus sequences in their promoters, and both can be activated by certain Ets family members such as Ets2. However, cotransfection experiments indicated that Erg can activate the collagenase, but not the stromelysin-1, promoter. This indicated that Erg can only regulate a subset of Ets target genes. Mutational analysis identified the Ets binding sites in the collagenase promoter that were critical for Erg-mediated transcriptional activity. However, DNA binding assays indicated that Erg could not bind to an isolated oligonucleotide bearing this site. It was

noted that an AP1 site (a binding site for Fos/Jun heterodimers) was located very near the Ets site. Further mutational analysis indicated that the AP1 site was also critical for Erg-mediated transcriptional regulation. Because mutational analysis suggests that the AP1 site and Ets site must cooperate, the authors postulated that a Fos/Jun heterodimer bound nearby stabilizes Erg binding to the Ets element. In support of this hypothesis, DNA binding assays indicated that a Fos/Jun heterodimer formed a direct physical interaction with Erg. Additionally, inclusion of Fos and Jun in a DNA binding assay using an oligonucleotide expanded to contain the AP1 site as well as the Ets site showed that Erg could now bind to the DNA probe. Inspection of the stromelysin promoter indicated that the nearest AP1 and Ets binding sites were 130 base pairs apart, explaining why Erg could not activate stromelysin. Although interaction with AP1 is critical in determining Erg target gene specificity, not all Ets family members require this specific protein-protein interaction. For example, Elk cannot be recruited to the Ets binding site in the *fos* promoter unless the serum response factor (SRF) is bound to the adjacent serum response element (57). Thus, classification of Ets site-containing genes as targets of particular Ets family members may first require the identification of other promoter elements that are commonly found near an Ets site.

Although studies of the effects of Myc and USF on the transcriptional activity of putative target genes have been difficult due to the modest *in vivo* transcriptional activity of these proteins, the amino terminal transactivation domains of USF and Myc have been shown to display some target gene specificity. Evidence that the differences observed in transcriptional assays can be mediated *via* mechanisms other than differential binding affinity for specific E boxes comes from several different experiments using Gal4 fusion proteins. Boyd *et al.* (58) have shown that USF fused to the Gal4 DNA binding domain cannot activate the *cad* promoter containing a Gal4 site. In contrast, the Myc transactivation domain can increase *cad* promoter activity when fused to Gal4. Since the Gal4 DNA binding domain is mediating the protein/DNA interactions in both cases, the transactivation domain of Myc may make a critical protein-protein contact with some component of the transcriptional machinery with which USF cannot interact. Other evidence suggests that the USF activation domains are very sensitive to core promoter structure. Luo and Sawadogo (59) have shown that a domain called USR that is well conserved in USF1 and USF2 can only activate transcription in the presence of both a TATA box and an initiator element. Although the *cad* promoter lacks a TATA box, it does contain a consensus initiator. Therefore, one might predict that the addition of a consensus TATA box would alter the *cad* promoter such that it could now be activated by USF. However, the addition of a TATA box does not convert the *cad* promoter to a promoter that can be activated by USF (Boyd, unpublished results). Evidence supports the hypothesis that not all initiator elements are bound by the same proteins.

Replacement of the *cad* initiator (which has 93% homology to the consensus initiator element) with two different sequences, each having about a 90% homology to the consensus, indicated that one, but not the other, replacement initiator could direct transcription from the *cad* promoter *in vivo* (60). Therefore, it remains possible that Myc and USF are active only in the context of specific initiator elements. If so, then interaction between Myc or USF with distinct initiator binding proteins may provide some aspects of target gene specificity. Other experiments also suggest that Myc may make specific interactions with some component of the basal transcription complex that cannot be reproduced by USF. Studies of Desbarats *et al.* (20) suggest that Myc can activate gene expression over a longer distance than can USF. It is possible that strong protein-protein interactions between Myc and a component of the basal machinery allow this long-distance activation *via* a DNA-looping mechanism. It has been proposed that the position-independent activation properties of Myc can allow it to activate a larger set of target genes than can USF. Other studies have shown that different cellular promoters will display different extents of position independence in their activation by Myc. For example, the *dhfr* promoter is very sensitive to the position of the bound Myc but the *cad* promoter is less sensitive (58). Therefore, it is likely that a combination of DNA binding specificity and protein-protein interactions combine to create different subsets of Myc and USF target genes.

Most E2F transactivation studies have indicated that overexpression of any of the E2Fs can lead to activation of a particular target gene. Some of this lack of target gene specificity may be due to the fact that several of the E2F genes are themselves regulated by E2F sites. Therefore, introduction of a single E2F into a cell may lead to increased levels of several different E2Fs. It is also likely that overexpression of a single E2F can obscure promoter specificities that do occur under normal physiological conditions. One of the few cases in which target gene specificity has been observed when different E2Fs are used in transactivation assays is a study involving the cyclin D1 promoter. Watanabe *et al.* (61) found that E2F1 negatively regulates the cyclin D1 promoter but that E2F4 is an activator of the same promoter construct. Interestingly, they also showed that a nearby Sp1 site contributed to the repression mediated by E2F1. Others have shown a direct interaction between Sp1 and E2F1, E2F2, or E2F3, but not E2F4 or E2F5 (62). This raises the possibility that the observed target gene specificity between E2F1 (which represses the cyclin D1 promoter) and E2F4 (which activates the cyclin D1 promoter) may be due to E2F1 recruitment of Sp1. Although specific E2Fs were not tested, two other experiments have found functional relationships between Sp1 and E2F sites. For example, the murine thymidine kinase promoter contains an Sp1 binding site spaced seven base pairs upstream from an E2F binding site. It was shown that mutation of either of the two sites in the context of a stably integrated

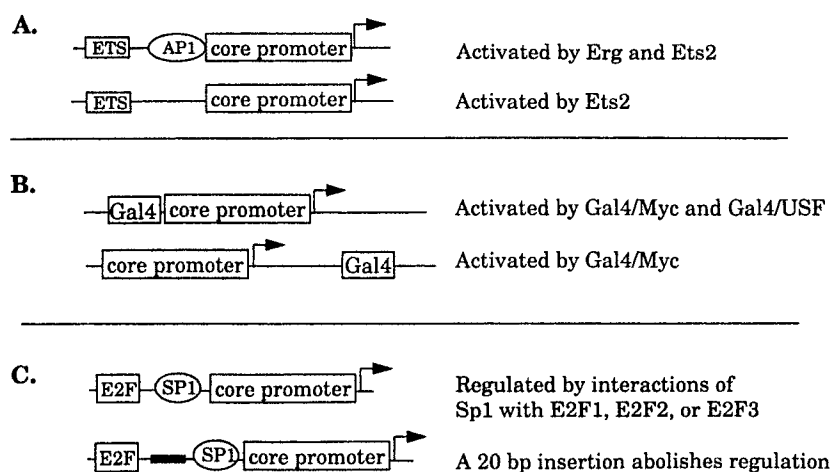
promoter construct abolished the *in vivo* footprint at both sites. Functional analyses indicated that increasing the separation of the Sp1 and E2F sites by an additional 20 base pairs also abolished the cell cycle stage-specific promoter activity, suggesting that an interaction between the proteins binding to these two sites was critical (62). Thus, a prediction of this work is that the murine thymidine kinase promoter would be a target gene of only a subset of E2Fs that bind Sp1 (i.e., E2F1, E2F2, and E2F3). The mouse *dhfr* promoter is also extremely sensitive to the position of the E2F site. The mouse *dhfr* promoter contains four binding sites for the transcription factor Sp1 located from -50 to -210 and an E2F site located at the transcription start site. Mutational analysis indicates that the E2F site is the critical determinant in specifying cell cycle stage-specific transcriptional regulation of this promoter (63). The positional requirements for the E2F site were investigated by inserting DNA fragments just upstream of the E2F site. Shifting the E2F site downstream of the start site by about 66 base pairs abolished the increase in transcriptional activity that normally occurs from this promoter as cells enter the S phase (64). These studies did not identify the reason why movement of the E2F site was detrimental to transcriptional regulation of the mouse *dhfr* promoter. Other studies have shown that E2F1 can cooperate with Sp1 to activate the hamster *dhfr* promoter in cotransfection assays performed using *Drosophila* cells. This cooperation required the region of the N terminus of the E2F1 protein that is conserved in E2F2 and E2F3 (65), again supporting the hypothesis that promoters containing Sp1 binding sites may be regulated by a subset of E2Fs. However, the Sp1 sites in the mouse *dhfr*

promoter are fairly far apart (50 base pairs), leaving open the possibility that it is the proximity to another promoter element that is critical for regulation of the mouse *dhfr* gene.

In summary, promoter context can be a critical determinant of the degree to which a particular transcription factor can regulate a given promoter. As shown in Figure 3, members of the Ets, Myc, and E2F families can be influenced both by the presence of other site-specific DNA binding proteins and by core promoter elements. For example, certain Ets factors require cooperation with proteins bound to a nearby AP1 site to achieve transcriptional activation. Likewise, in certain promoters the E2F site must be located within a very close proximity to an Sp1 site in order for E2F proteins to regulate transcription. In contrast, specific activation by Myc versus USF appears to be determined by the precise arrangement of core promoter elements and the bound transcription factors.

### Methods That Can Aid in the Identification of Target Genes

Based on the examples provided above, it seems likely that each member of a multi-gene family will have at least some target gene specificity when compared to other members of the same family. Therefore, it is important to have in our arsenal a collection of methods by which we can identify a set of putative target genes. First, this can be done by the use of assays, such as cDNA filter arrays, which compare the expression of a fairly random panel of cellular genes in the presence and absence of high levels of the factor of interest (described below). Unfortunately, arrays are limited to the study of previously identified cellular



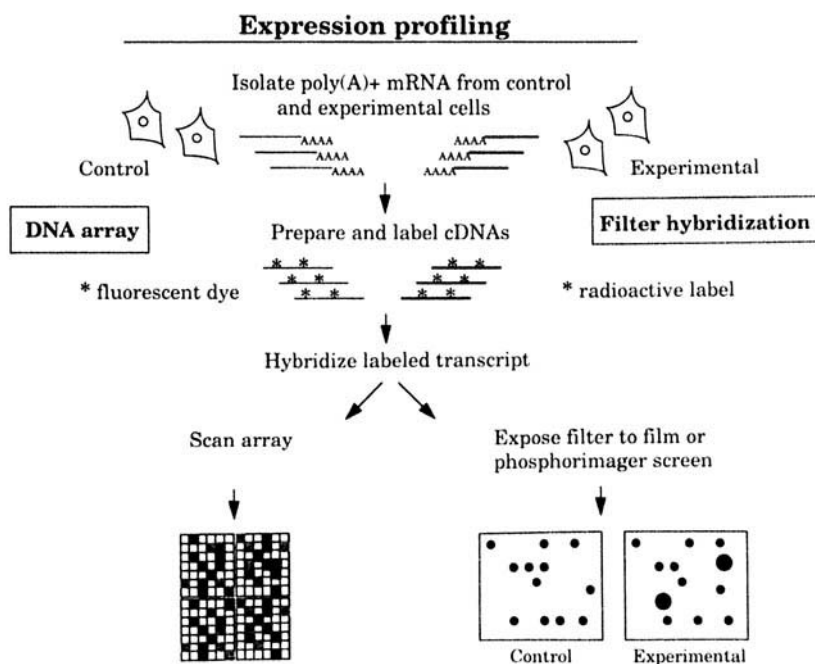
**Figure 3.** Promoter context can influence the activity of oncogenic transcription factors. (A) Shown are schematics that represent a promoter (e.g., collagenase) that contains both an Ets and an AP1 binding site versus a promoter (e.g., stromelysin) that does not contain an AP1 site adjacent to the Ets site. Although both of the Ets family member proteins Erg and Ets2 can activate the collagenase promoter, only Ets2 can activate the stromelysin promoter. This is due to the requirement for Fos and Jun binding to the AP1 site to recruit Erg to the Ets site (see text for details). (B) Shown are schematics of a minimal promoter containing a Gal4 binding site that has been placed just upstream of the core promoter and a minimal promoter containing a Gal4 site that has been placed several kilobases downstream of the core promoter. Although both Gal4USF and Gal4Myc fusion proteins can activate the top construct, only Gal4Myc can activate the bottom construct from the downstream Gal4 site (see text for details). (C) Shown is an example in which movement of an E2F site relative to other promoter elements can alter transcriptional regulation. Insertion of 20 bp between the Sp1 site and the E2F site in the *tk* promoter can abolish the increase in transcriptional activity that is normally observed in the S phase (see text for details). One hypothesis that may account for these results is the requirement for a physical interaction between Sp1 and an E2F family member. Since only a subset of E2F family members interacts with Sp1, promoters containing adjacent E2F and Sp1 sites may be regulated by only a subset of E2F family members.

genes that are currently in the database. Other approaches, like Representational Difference Analysis, which have the advantage that the resultant target genes can be derived from all cellular mRNAs, not just previously cloned mRNAs, are described below. However, identified mRNAs may not be directly regulated by the transcription factor but instead may be downstream of the factor in a signal transduction cascade. Therefore, a third method for identifying target genes is described below in which genomic fragments that are bound directly by an oncogenic transcription factor are isolated and characterized. We will discuss each experimental method, citing examples where these techniques have been used to identify Myc, Ets, and E2F target genes. A general outline of each type of approach is also presented in Figures 4–6.

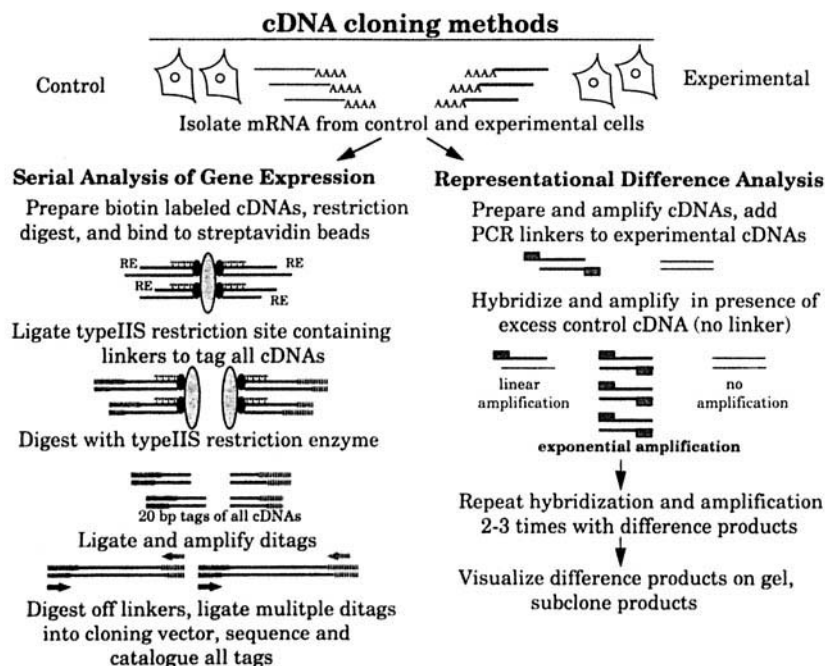
**Expression Profiling Using DNA Arrays.** A time-honored technique for the analysis of gene expression is to prepare a probe specific for one gene, then monitor the expression of that gene using Northern hybridization, RNase protections, or Reverse Transcription-Polymerase Chain Reaction assays. Although under certain circumstances, the expression level of several known genes of interest can be successfully monitored simultaneously, these assays are fairly tedious and time-consuming if one wishes to compare the expression level of a multitude of genes. In general, the number of genes easily monitored using these

assays is quite low. An improvement to these assays has been the fairly recent development of cDNA filter arrays and DNA microchip. As outlined in Figure 4, a comparison of the expression level of all genes on a filter or microchip can be achieved by preparing a probe from mRNA obtained from the control (e.g., normal cells) and test (e.g., cells having high amounts of a transcription factor) samples. This allows the investigator to prepare only two probes to monitor hundreds to thousands of mRNAs. Although the advantages over single gene analysis are obvious, the information obtained in these experiments is limited by the choice of genes that are currently available on commercially obtained filters (66, 67). Even so, investigators are currently using such methods to identify genes whose expression is altered upon overexpression of oncogenic transcription factors.

Arrayed filters, containing DNAs complementary to hundreds of mRNAs, have been applied to the search for both c-Myc and N-Myc target genes. Although investigators are limited to examining the expression levels of those mRNAs represented on the commercially available filters, the variety of such filters is increasing. For example, it is now possible to purchase filters containing cDNAs from human, mouse, and rat cells. Specialty filters and custom filters are also becoming available; there are now filters containing genes relevant to cancer and to apoptosis. It is likely that the utility of this method will increase in the next



**Figure 4.** Identification of target genes with Expression Profiling. Shown is a schematic of the DNA Array and Filter Hybridization approaches that can be used to identify previously cloned genes as direct and indirect targets of oncogenic transcription factors. For both techniques, Poly(A)+ messenger RNA is isolated from control and experimental cells and converted into cDNA. Probes are then generated by labeling the cDNA populations with fluorescent dye or  $^{32}\text{P}$ . (Left) DNA arrays that contain thousands of cDNA sequences are hybridized with both probes (labeled with two different dyes) simultaneously. Hybridized signals, detected by scanning microscopy, fluoresce in a range between the two dyes indicating how much of either probe is hybridized (e.g., white boxes indicate cDNAs that are represented in the control sample, black boxes indicate cDNAs that are represented in the experimental sample, and gray are common to both). (Right) Hybridization to the cDNA filters, containing hundreds of cDNAs, can be carried out in parallel using the control and experimental probes on two identical filters. Relative signal intensities can be compared between control and experimental filters to define cDNAs that increase or decrease in response to treatment.



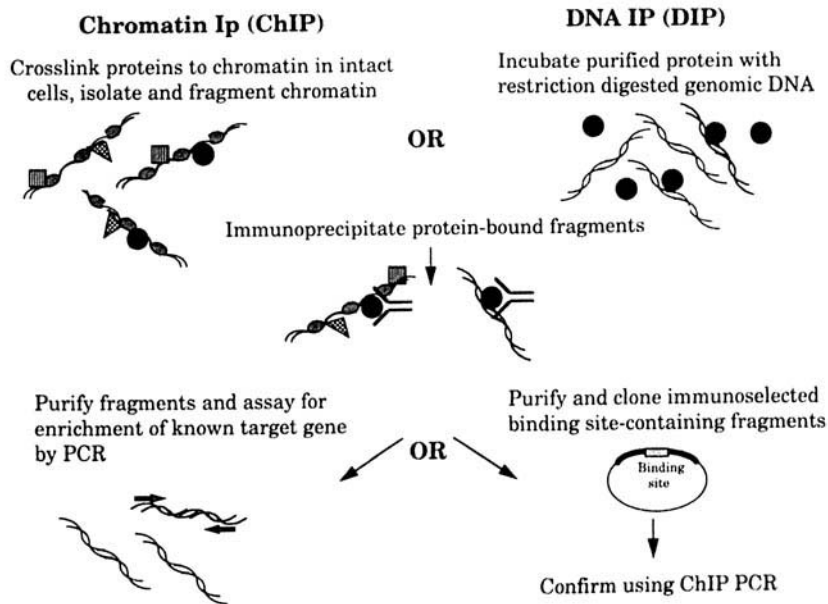
**Figure 5.** cDNA-based methods used to identify target genes. Shown is a schematic of the Serial Analysis of Gene Expression (SAGE) and Representational Difference Analysis (RDA) approaches that can be used to clone both direct and indirect target genes of oncogenic transcription factors. For both techniques Poly(A)<sup>+</sup> messenger RNA is isolated from both control and experimental cells. (Left) In SAGE, mRNAs (only shown for experimental sample) are converted into double-stranded cDNAs using biotin-labeled (small black circles) oligo(dT) primers. Labeled cDNAs are digested with a restriction endonuclease and then bound to streptavidin beads (large gray ovals). The sample is divided in half and ligated to two different linkers (gray and striped lines) containing a recognition site for a type IIS restriction endonuclease (which cut at a defined number of bases downstream of the recognition site); this serves to tag a short fragment of each cDNA present in the population. Following digestion with the type IIS enzyme, the cDNA fragments are blunt-ended, ligated together, and then amplified by PCR. The PCR products are digested with the original restriction enzyme to remove the linkers. Digested products are pooled and ligated into a sequencing vector. Clones (usually containing multiple tagged inserts) are sequenced to identify each tag. Computer analysis is performed to match sequence tags to known cDNAs or ESTs and to catalogue the expression frequency of each sequence tag within a population. (Right) In the RDA approach, control and experimental mRNA populations are converted into cDNAs and amplified by PCR. To find those mRNA species that are increased in the experimental sample (shown in the schematic), the experimental cDNAs are ligated to PCR linkers, denatured, and then hybridized in the presence of excess control cDNA. Cross-hybrids between control and experimental cDNAs (i.e., mRNA common to both populations) will not become amplified. Resulting hybrids that contain linkers on both strands (i.e., mRNAs that are more abundant in the experimental sample) will become amplified by PCR. Amplified cDNAs, called the first round difference product, are then used in two to three more rounds of hybridization and amplification. To find mRNAs that are decreased in the experimental sample, control cDNAs are ligated to PCR linkers and then hybridized in the presence of excess experimental cDNA (not shown). The final difference products are visualized on an agarose gel and subcloned. Not shown in this figure is the differential display technique that is clearly outlined in a previous report (73).

several years when investigators have a wider variety of different filters from which to choose. Using currently available filters, several groups have investigated the effects of Myc family members on gene expression. For example, liver-specific overexpression of c-Myc in transgenic mice results in the rapid development of hepatocellular adenomas. Hybridization of radiolabeled probes from wild type and c-Myc transgenic liver mRNA to a pair of Clontech Atlas Mouse cDNA filters identified nine genes that are differentially expressed in the liver of Myc transgenic animals (S. Kim, personal communication). Northern analysis revealed that these nine genes were expressed at moderate to high levels in the cell, suggesting that the original profiles could not detect low abundance mRNAs. This group hypothesized that a probe obtained after a single round of representational difference analysis (see below and Figure 5 for details of this technique) would enhance detection of low-level messages. Accordingly, an additional nine differentially regulated genes were identified using the new

probes. Clontech's Atlas Human Array and Cancer Array filters have been used by another group to screen for N-Myc targets using a neuroblastoma cell line having a tetracycline-regulated N-Myc construct (S. Mac, personal communication). Preliminary results from this study identified 10 mRNAs that were increased in the presence of high levels of N-Myc and five that were decreased (e.g.,  $\alpha$ -prothymosin was increased up to 6-fold in cells that overexpress N-Myc).

Microarray technology has been developed by a number of laboratories for a more comprehensive analysis of the expression of known genes. These arrays typically consist of high-density grids of unique cDNA clones (or ESTs) or oligonucleotides that correspond to thousands of different transcripts (68, 69). Similar to the cDNA filter arrays, the relative expression levels of a large number of mRNA species can then be assessed in parallel by generating cDNA probes from a population of mRNA obtained from control and test samples, followed by hybridization of these probes

## ChIP and DIP cloning of binding sites



**Figure 6.** Cloning genomic fragments bound by transcription factors through ChIP and DIP techniques. Shown in this figure is a schematic of the chromatin immunoprecipitation (ChIP) and DNA immunoprecipitation (DIP) approaches that can be used to identify and clone direct binding targets of oncogenic transcription factors. (Left) For the ChIP procedure, intact cells are treated with formaldehyde or UV light to covalently cross-link DNA-bound factors to their cognate binding site. The cells are then lysed, and the nuclei are sonicated to fragment the cross-linked chromatin. These DNA fragments contain nucleosomes (gray ovals), other transcription factors (squares and triangles), and the transcription factor of interest (black circles). (Right) For the DIP procedure, genomic DNA is purified from cells and digested with a restriction enzyme. Prepared DNA is then incubated with a preparation of the purified protein of interest (black circles). For both ChIP and DIP, protein-DNA complexes are collected by immunoprecipitation with an antibody specific for the protein of interest. Immunoselected chromatin or DNA fragments are then purified and cloned into a vector (to identify novel binding targets) or analyzed by PCR using gene-specific primers (to verify binding to a putative target gene).

against duplicate arrays. From the hybridization signals, the relative expression levels of all genes represented on the array can be ascertained. Although this technique has several advantages over the filter arrays (e.g., the number of genes examined can be in the thousands), problems remain. For example, each of the two different microchip methods requires expensive machinery, software, and reagents. This limits the access of these techniques to only a handful of investigators. Second, each of the two microarray techniques requires that the sequence of a gene be in the database, as either a cDNA or an expressed sequence tag, before it can be monitored in this assay. Thus, if the key target genes of a particular transcription factor have not yet been cloned, this type of analysis may be both expensive and nonproductive.

Although very few experiments using microarrays have been reported, one experiment has provided data concerning the *c-myc* oncogene. Primary human fibroblasts containing inducible Myc protein were used to examine gene expression in quiescent cells and in cells progressing toward S-phase after c-Myc activation. Of approximately 6,000 genes present on the chips, 27 showed consistent upregulation and 9 downregulation in three separate experiments (in addition, a number of genes showed induction in two out of three experiments). Among the known Myc target genes, *odc* showed the most robust and reproducible changes. The others all represent new potential Myc target genes, and several

were confirmed by Northern blot analysis (H. Collier, C. Grandori, R. Eisenman, and T. Golub, personal communication). The relatively small number of affected genes observed in these experiments suggests that many c-Myc target genes are not yet represented on the chips or that many Myc targets are low abundance messages that undergo only slight changes in response to c-Myc and are difficult to detect in this assay.

In summary, several investigators are attempting to identify Myc target genes using DNA arrays. There are currently ongoing attempts to use both filters and microchips to identify N-Myc (B. Carroll, personal communication), Ets (C. Denny, personal communication) and E2F (J. Nevins, personal communication) target genes. Many of the preliminary results obtained to date have not been extremely encouraging. First, there is a high rate of false positives using these methods. It is extremely important to use a different method (e.g., Northern analysis, RNase protection, or RT-PCR) to confirm the differential expression seen using the arrays. Second, in the c-Myc and N-Myc studies, several of the genes identified using the filters were already previously known to be regulated by Myc family members. For example, one of the mRNAs that was dramatically unregulated by c-Myc using DNA microarrays was *odc*, and the gene that was the most dramatically regulated by N-Myc using filter arrays was found to be  $\alpha$ -*prothymosin*. Both of these genes had previously been reported to be regulated by

both c-Myc (70, 71) and N-Myc (72). However, despite the many difficulties of these new approaches, a handful of novel c-Myc target genes have been identified in a fairly rapid fashion using DNA microarrays.

**cDNA Cloning Methods.** As noted above, the methods using DNA filter arrays and DNA microchips have been slow in providing new insight into the identification of target genes of Myc, E2F, or Ets family members. One different approach that researchers are taking to identify such target genes is to clone those mRNA species that are increased or decreased in response to deregulated activity of a particular member of these families. Methods based on this approach include subtractive hybridization, Differential Display (DD), Representational Difference Analysis (RDA), and Serial Analysis of Gene Expression (SAGE). For these techniques (outlined in Fig. 5), mRNAs isolated from control and experimental cells are used to generate cDNAs that are then compared to each other in a variety of ways. The obvious benefit of any of these cDNA cloning techniques over the current cDNA arrays is that they can identify both known and novel genes as targets of oncogenic transcription factors.

The general principle behind subtractive hybridization is that cDNAs common to both control and experimental samples are removed or suppressed to enrich for differentially expressed mRNA species. cDNAs generated from the experimental sample are hybridized to control cDNA, duplexes are removed, and the remaining nonsubtracted cDNA clones are screened to identify those clones that display differential expression. This approach has proven to be difficult, time consuming, and not highly sensitive since low abundance mRNAs are not easily cloned. To attempt to circumvent the sensitivity problems inherent in subtractive hybridization methods, a PCR-based approach was developed. Differential Display is a PCR-based approach that allows for a rapid, broad search of large expression differences. With this technique, subsets of cDNAs from both the experimental and control samples are amplified using a set of random primer pairs, and the PCR products are compared directly on a sequencing gel (73). Although this technique is less time-consuming, it still lacks sensitivity and can have a high false positive rate. Therefore, RDA, which combines both subtractive hybridization and PCR amplification, is now a commonly used approach. RDA allows for the rapid identification of both slight and significant changes in a broad range of message levels. By performing successive rounds of hybridization between experimental and control cDNA populations followed by selective PCR amplification of unique DNA sequences, one can generate PCR products corresponding to differentially expressed sequences (74, 75). These difference products can be visualized easily in an ethidium-stained agarose gel, isolated, and cloned. In contrast to RDA, which selects out differentially expressed mRNAs for analysis, the SAGE approach directly catalogues all messages present in a population of cells by tagging, cloning, and sequencing short segments of the 3' end

of all cDNAs generated from an mRNA sample (76). SAGE provides a highly sensitive measure of the relative expression levels of mRNA species under specific cellular conditions. However, the labor-intensive nature of this technique, which requires extensive sequencing in the range of tens of thousands of sequencing reactions, and the advanced computer analysis required to interpret the sequencing results is likely to limit its application.

Several cDNA cloning approaches have been used successfully to discover a handful of c-Myc target genes. For example, Benvenisty *et al.* (77) identified *ECA39* as a Myc-regulated gene using a subtractive hybridization approach. This group reasoned that c-Myc target genes that are critical for tumorigenesis will be deregulated in a variety of c-Myc induced tumors. Accordingly, cDNA prepared from a brain tumor cell line derived from a c-Myc transgenic mouse was subtracted with normal brain mRNA to yield a library of potential Myc target genes. In a second step, the library was rescreened with cDNA probes from both c-Myc induced brain and lymphoma cell lines to identify common clones between the two tumor lines. The first subtraction refined the search from 25,000 phage down to 500 phage, and the second round of screening resulted in a total of 9 positive cDNA clones. One such clone, named *ECA39*, contained an E box element and was characterized as a direct target of c-Myc through promoter studies. Unfortunately, the function of *ECA39* in cell growth remains elusive. Additional c-Myc target genes have been identified through screening a subtracted library that was enriched for genes that exhibit mid-G1 serum response kinetics (78). Using labeled cDNAs derived from cells expressing a conditionally inducible form of c-Myc, clones corresponding to ornithine decarboxylase (*odc*), lactate dehydrogenase (*ldh-a*), and one novel sequence were defined as Myc-responsive genes. Another group studying Myc's role in tumorigenesis used the RDA approach as a screen for genes that contribute to the anchorage-independent growth phenotype of Ratla-Myc cells (79). RDA was performed with template mRNA prepared from both parental and Myc-transformed Ratla cells grown under nonadherent conditions. Following two or three rounds of RDA, differentially expressed products were verified by performing Southern slot blot analysis with the original Ratla and Ratla-Myc cDNA populations. Further analysis of the confirmed clones using Northern blots revealed that, of 23 positive clones examined, 20 displayed differential expression with a range from 2–20-fold. Revealed in this screen were several previously identified Myc target genes including *odc*,  $\alpha$ -*tubulin*, *ldh-a*, and two collagen genes, as well as five novel sequences. Importantly, one of the novel genes, called *rcl*, was shown to be sufficient to induce anchorage-independent growth and was shown to be a direct target of the Myc-ER fusion protein, suggesting *rcl* may be a critical link between c-Myc and tumorigenesis. In a follow-up study, Shim *et al.* (80) demonstrated that the *ldh-a* promoter is a direct target of Myc activity and that *ldh-a* contains two E box elements. It was

also established that *ldh-a* plays an important role in transformation since *ldh-a* antisense RNA reduces clonogenicity of c-Myc transformed lymphoblastoid cells.

cDNA cloning strategies have also been applied in studies investigating the downstream cellular target genes of the Ets family of transcription factors. In one such study, Robinson *et al.* (81) performed differential display analysis to identify Ets1 and Ets2 targets. Expression of either Ets1 or Ets2 has been shown to transform 3T3 cells. Therefore to screen for target genes related to this phenotype, differential display PCR was performed with RNA prepared from parental NIH 3T3 cells or cells transfected with either an Ets1 or Ets2 expression vector. From eight different primer sets, the authors identified 82 differentially expressed cDNA bands. Strikingly, many of the clones were false positives since only 16 of the clones showed reproducible differential expression by Northern analysis. Of these clones, only three were known genes: *cbf* (CArG box binding factor), *pla2p* (phospholipaseA2 activating protein), and *egr1*. Interestingly, *pla2p* was only differentially expressed in the presence of Ets2 overexpression and *egr1* was only differentially expressed in the presence of Ets1 overexpression. These results suggest that Ets1 and Ets2 may activate unique downstream targets. The authors established that *egr1* is a direct target of Ets1 activity through an Ets1 binding site screen (discussed in the next section) and mutational analysis of the promoter. Other groups interested in the role of Ets domain fusion transcription factors in neoplasia have also used cDNA cloning methods to search for downstream target genes. For example, Lawlor *et al.* (82) performed differential display analysis to identify target genes of the EWS-ETS fusion protein through a comparison of pNET (primitive neuroectodermal tumor) cell lines that are characterized by an EWS-ETS chromosomal translocation and other small round cell tumor lines. One differentially expressed band, confirmed by Northern analysis, corresponded to the Gastrin-releasing peptide (*ghp*) gene. Although elevated expression of *ghp* was observed in all pNET cell lines and primary tumors that were tested, *ghp* is not a direct transcriptional target of the EWS-ETS fusion protein in transfection assays. In another approach, Braun *et al.* (10) used RDA as a screen for EWS/FLI target genes. Overexpression of the EWS/FLI fusion protein, but not wild type FLI-1, can transform NIH 3T3 cells. Therefore, to identify unique EWS/FLI targets, RDA was performed on mRNA harvested from NIH 3T3 cells stably expressing either FLI-1 or the EWS/FLI fusion protein at both high constitutive levels and under conditional regulation. The authors characterized eight upregulated transcripts and two downregulated transcripts in the presence of EWS/FLI. Of these, a rapid increase in expression of stromelysin following conditional EWS/FLI expression suggested that it may be a direct transcriptional target. Other EWS/FLI activated transcripts identified in similar screens have been characterized as the human homolog of the *Drosophila* manic fringe gene, *mfng*, and the cyclin ubiquitin conjugating

enzyme related gene, *mE2-C* (83, 84). The observation that expression of *mfng* significantly increases the tumorigenicity of NIH 3T3 cells when injected into SCID mice provides a link between EWS/FLI and transformation. Upregulation of *mE2-C* by EWS/FLI also represents another pathway that could impact on the cell cycle. However, *mfng* and *mE2-C*, like many other genes identified in these screens are not direct downstream targets of EWS/FLI (C. Denny, personal communication).

In summary, investigators have used cDNA cloning methods to identify Myc and Ets target genes; similar methods are currently being used to identify E2F target genes (N. Heintz, personal communication). However, these techniques are not problem-free. Although in certain cases, a low rate of false positives has been obtained (79), cDNA cloning methods such as RDA often yield a high rate of false positives. Even though RDA and differential display can identify genes that display altered expression upon deregulation of a transcription factor, these techniques are more qualitative than quantitative. Therefore, for each of the Expression Profiling methods described above, it is important to use Northern analysis, RNase protection, or RT-PCR to confirm the differential expression of mRNAs corresponding to the clones. Also, although both Expression Profiling and cDNA cloning methods can provide a correlation between increased activity of a particular transcription factor and altered expression levels of various mRNA species, it is not possible to know *a priori* which identified genes are direct transcriptional targets and which display differential regulation simply due to being downstream from the transcription factor in a signal transduction cascade. Since both of these techniques are based on cDNA sequences, it is difficult to determine if the corresponding gene is directly regulated by the factor without the sequence of the promoter. Unfortunately, the current database, which is mainly composed of coding sequences, lacks promoter sequences for most genes.

**Cloning of Genomic Fragments Containing Binding Sites.** As noted above, a distinct disadvantage to the Expression Profiling and cDNA cloning methods is that the identified genes could be either directly or indirectly regulated by the transcription factor of interest. To circumvent this problem, investigators have developed methods that allow the cloning of DNA fragments that are bound directly by transcription factors (Fig. 6). The general principle behind this type of approach is that target genes can be isolated from a pool of genomic DNA or native chromatin through the direct association of these promoters with bound transcription factors. In these experiments, transcription factor-bound DNA or chromatin is immunoprecipitated using an antibody against the transcription factor. The immunoselected DNA is purified and either directly analyzed using Southern blotting or PCR methods (to examine known target genes) or inserted into a cloning vector (for the analysis of novel target genes). In addition to the fact that this method will identify direct targets of transcription factors,

chromatin immunoprecipitation has a further advantage; unlike gel shift experiments that use isolated binding sites, it allows for the analysis of factor binding within the context of the entire promoter region. Several direct cloning methods are currently being used to identify target genes of the Myc, Ets, and E2F families.

Ets1 target genes have been cloned by Robinson *et al.* (81) using a genomic DNA immunoprecipitation approach. The method used in this report, called whole genome PCR, was to digest genomic DNA, ligate the cut DNA to linkers for PCR amplification, and incubate the DNA/linker population with recombinant Ets1 protein. These *in vitro* Ets1-DNA formed complexes were isolated through immunoprecipitation using an Ets1 antibody. Immunoselected DNA was amplified by PCR and subcloned into a cloning vector. Sequences homologous to the promoters of human serglycin, preproapolipoprotein C II, and Egr1 were cloned; all of these promoters contain one or more Ets binding site. An additional 40 clones containing unique sequences were also obtained. One drawback of this particular approach is that it uses large amounts of recombinant protein and *in vitro* binding conditions to identify target genes. Therefore, preliminary targets must be further analyzed for physiologic relevance. Accordingly, the authors have also identified Egr1 as a cellular Ets1 target using a differential display approach (as mentioned in the previous section) and demonstrate that expression from the *egr1* promoter requires one of the two Ets binding sites. A similar *in vitro* genomic DNA approach has recently been used to identify candidate target genes of the Evi-1 zinc finger oncoprotein (85). In the first step of this technique, purified Evi-1 protein was incubated with a plasmid-based mouse genomic library. Protein-DNA complexes were isolated by nitrocellulose filtration, and selected plasmids were amplified in bacteria and rescreened for Evi-1 binding. A large library of genomic fragments containing Evi-1 binding sites was generated; however, no clones matched any known genes. Therefore, in the second step, the selected genomic clones were used to screen a cDNA library revealing that several clones represented known genes such as *itpr2* (inositol triphosphate receptor, type 2) which was identified by three individual clones in this screen.

Chromatin immunoprecipitation methods have been applied to the search for c-Myc target genes and used to analyze *in vivo* binding specificities of Myc family members to known target genes. Unlike the previous Ets1 and Evi-1 studies that used naked DNA, Myc binding was analyzed within the context of native chromatin structure in an intact cell. Grandori *et al.* (86) isolated targets of c-Myc from a human lymphocyte cell line that was transformed through the stable introduction of c-Myc and Max expression plasmids. The protocol used by this group was to isolate nuclei from the cells, solubilize the chromatin with nucleases, and immunoprecipitate the soluble chromatin in two successive rounds with antibodies against Max followed by antibodies against Myc. Immunoselected DNA

fragments were cloned, and those clones that bound the Myc/Max heterodimer in an *in vitro* gel shift assay were sequenced. This protocol yielded 20 clones that contained one or more Myc binding sites (E boxes). Interestingly, only 4 of the 20 DNA fragments cloned by immunoprecipitation contained the consensus E box sequence CACGTG. The remaining clones contained several noncanonical E box sequences (which had been previously shown by Blackwell *et al.* (47) to bind Myc/Max *in vitro*) demonstrating that *in vivo* the Myc/Max complex has a broad binding specificity. These experiments also demonstrated that the nucleotides that flank the E box influence which sites will be bound by Myc *in vivo*. Although the clones isolated by Grandori *et al.* (86) are direct targets of Myc binding, none of them directly matched the sequences of any previously identified gene or promoter region, as was the case for many of the clones identified in the Ets1 study. Therefore, library screens were performed to identify one clone obtained from the Myc immunoselection experiments, named *mrdb*, as a novel gene encoding a putative DEAD box RNA helicase. Although this study has the advantage of allowing promoter context to be taken into account when identifying target genes, the experiments were performed in the presence of levels of Myc and Max that were greatly increased over normal physiologic levels.

A similar approach was taken by Boyd *et al.* (45, 58) to examine the occupancy of Myc versus USF at the E boxes of potential target genes. In these experiments, cultures of cells were subjected to limited treatment with UV light or formaldehyde, which causes covalent crosslinks between DNA binding proteins and the sites at which they are bound at the time of the treatment (87, 88). Crosslinked chromatin was immunoprecipitated with antibodies to Myc or USF, and the immunoselected population of DNA was analyzed for the presence of a particular target gene through Southern blot analysis or PCR amplification. An important aspect of this technique is that transcription factor binding can be analyzed, not in an overexpression system, but in normal cells at any stage of the cell cycle. Using this technique the authors made several observations. First, in NIH 3T3 cells, Myc and USF binding can be detected at both the promoter of the putative Myc target gene, *cad*, which contains a consensus E box and at noncanonical E boxes within the promoters of other genes such as *dhfr*. Importantly, transient transfection experiments demonstrate that the transactivation domain of Myc activates transcription from *cad* but has little effect on expression of *dhfr*. Taken together these results suggest that target gene selectivity between Myc family members may occur at a level beyond DNA binding and that Myc binding may not always correlate with transcriptional activity. Based on these observations, one would predict that assays based on Myc function (such as RDA) may be better suited for initial Myc target gene screening. In subsequent steps, chromatin immunoprecipitation can then be used to identify whether Myc directly associates with the candidate target genes.

Unlike Myc and USF, members of the E2F family appear to have a more defined binding specificity *in vivo*. Using the formaldehyde crosslinking approach, Wells *et al.* have shown that subsets of the E2F proteins display binding to different subsets of target genes (unpublished data). Studies are now in progress to identify novel E2F target genes by cloning chromatin fragments that are immunoprecipitated using an E2F antibody. For the purposes of these experiments, only normal physiological levels of E2F protein are present, and the protein-bound DNA fragments are isolated in the context of natural chromatin to allow for promoter context specificity. Although these experiments are not yet completed, it is encouraging that an immunoselection experiment of this type was recently used successfully to identify a TAL-1 target gene in mouse erythro-leukemia cells (89).

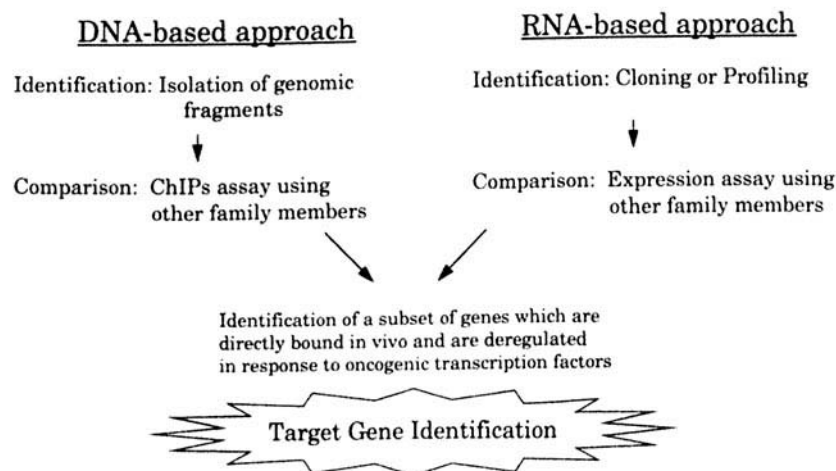
In summary, Ets, Myc, and E2F target genes can be identified based on the isolation of genomic fragments (either naked DNA or chromatin) that are directly bound by a member of these families of transcription factors. A powerful feature of chromatin immunoprecipitation is that it does not destroy any specificity in target gene selectivity that is dictated by promoter context (i.e., all of the other promoter-specific factors as well as histones and basal transcription factors are in their native context). However, since a vast number of mammalian sequences in the database are cDNA sequences rather than genomic sequences, it may not be easy to match cloned DNA fragments to previously cloned genes. Clearly the lack of genomic sequences in the current databases will impede our ability to rapidly confirm that the cloned DNAs are direct targets of DNA-binding factors. In many cases, additional cloning and characterization of larger genomic fragments will be required. Of course, as an intermediate step, confirmation that the cloned fragment is bound by the factor of interest can be achieved using PCR analysis of the fragment in a subsequent chromatin immunoprecipitation step. Such studies indicate that target genes of the Ets, Myc, and E2F families can be analyzed in this way (58; J. Wells, personal communication).

Another caveat in the interpretation of data obtained from these DNA and chromatin immunoprecipitation approaches is that transcription factor binding may not always correlate directly with function. For example, studies suggest that Myc family members do not regulate transcription from all binding sites (58).

## Conclusions

The main purpose of this review has been to provide a general approach that can be used to develop an understanding of how a particular member of a family of transcription factors mediates its unique biological response. One family of transcription factors can regulate the expression of many different cellular genes. Since it is possible that a subset of target genes of a given family of transcription factors may encode essential proteins or may even encode proteins that counteract certain neoplastic properties, providing the cell with an agent that inhibits all activity of these transcription factors might be harmful. If we wish to block the oncogenic functions of a particular transcription factor, we should target the specific target gene(s) that are responsible for conferring a neoplastic phenotype and not all target genes regulated by the entire family of proteins. This requires that we first obtain a large base of knowledge concerning the set of target genes that are specific for a particular member in a family of transcription factors. As described in Figure 7, this can be best accomplished by initially taking two parallel approaches.

The approach outlined in the right panel of Figure 7 will lead to the identification of a set of cellular genes whose expression is deregulated in response to increased abundance and/or activity of a transcription factor. These methods that are based on either subtractive hybridization and cloning of cDNAs or simply examining the expression level of a large number of previously cloned cDNAs will provide an initial subset of putative target genes. The response of the genes in this set to the transcription factor must be confirmed using more quantitative measures such as RNase protection and/or quantitative PCR assays. Those



**Figure 7.** General scheme for the identification of target genes of oncogenic transcription factors. See "Conclusions" for details.

candidates that are confirmed can then be examined for their response to a panel of different family members to determine which are uniquely responsive to some, but not all, factors. This step may provide key information for future studies. For example, it is unlikely that genes that are induced by both oncogenic and nononcogenic members of a family are critical determinants in conferring the oncogenic phenotype.

The approach outlined in the left panel of Figure 7 will lead to the identification of a set of cellular genes that are bound directly *in vivo* by the oncogenic transcription factor. This approach is based on the ability to select genomic fragments directly are bound by a transcription factor using an immunoprecipitation-based method. The sequence of the fragments obtained using this approach can be compared to a consensus sequence obtained using *in vitro* methods such as those that select high-affinity binding sites from a collection of random oligonucleotides. However, as noted above, the sites bound by protein *in vivo* may be different from the highest affinity site obtained using isolated oligonucleotides in an *in vitro* reaction. Once the genomic fragments are confirmed to bind the factor of interest, a separate immunoprecipitation experiment can be performed to determine if a fragment is bound specifically by that family member or by any or all of the other family members.

The final identification of the set of cellular genes that are direct targets of oncogenic transcription factors requires the combination of the knowledge obtained in the two parallel approaches described above. For example, many genes are bound by a factor *in vivo*, but that factor cannot activate the promoter (e.g., USF is bound to the *cad* promoter, but this factor cannot activate *cad* transcription see Ref. 58). However, if a particular gene can be shown to be bound by a factor *in vivo* and to be deregulated when that same factor is overexpressed, this will be a good candidate for a true target gene. If this same gene responds specifically to overexpression of the oncogenic factor and not to a very similar family member that lacks oncogenic potential, this provides a potential therapeutic target.

1. Struhl K. Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins. *TIBS* **14**:137-140, 1989.
2. Tenen DG, Hromas R, Licht JD, Zhang DE. Transcription factors, normal myeloid development, and leukemia. *Blood* **90**:489-519, 1997.
3. Gelmann EP, Psallidopoulos MC, Papas TS, Dalla Favera R. Identification of reciprocal translocation sites within the *c-myc* oncogene and immunoglobulin *mu* locus in a Burkitt lymphoma. *Nature* **306**:799-803, 1984.
4. Schwab M. Amplification of N-myc as a prognostic marker for patients with neuroblastoma. *Semin Cancer Biol* **4**:13-18, 1993.
5. Li W, Fan J, Hochhauser D, Banerjee D, Zielinski Z, Almasan A, Yin Y, Kelly R, Wahl GM, Bertino JR. Lack of functional retinoblastoma protein mediates increased resistance to antimetabolites in human sarcoma cell lines. *Proc Natl Acad Sci USA* **92**:10436-10440, 1995.
6. Dittmer J, Nordheim A. Ets transcription factors and human disease. *Biochim Biophys Acta* **1377**:F1-F11, 1998.
7. Chang CH, Scott GK, Kuo WL, Xiong X, Suzdaltseva Y, Park JW, Sayre P, Erny K, Collins C, Gray JW, Benz CC. ESX: A structurally

- unique Ets overexpressed early during human breast tumorigenesis. *Oncogene* **14**:1617-1622, 1997.
8. Golub TR, McLean T, Stegmaier K, Carroll M, Tomasson M, Gilliland DG. The *tel* gene and human leukemia. *Biochim Biophys Acta* **1288**:M7-M10, 1996.
9. Hiebert SW, Sun W, Davis JN, Golub T, Shurtleff S, Buijs A, Downing JR, Grosveld G, Roussel MF, Gilliland DG, Lenny N, Meyers S. The t(12;21) translocation converts AML-1B from an activator to a repressor of transcription. *Mol Cell Biol* **16**:1349-1355, 1996.
10. Braun BS, Frieden R, Lessnick SL, May WA, Denny CT. Identification of target genes for the Ewing's sarcoma EWS/FLI fusion protein by representational difference analysis. *Mol Cell Biol* **15**:4623-4630, 1995.
11. Facchini LM, Penn LZ. The molecular role of Myc in growth and transformation: Recent discoveries lead to new insights. *FASEB J* **12**:633-651, 1998.
12. Asker CE, Magnusson KP, Piccoli SP, Andersson K, Klein G, Cole MD, Wilman KG. Mouse and rat B-Myc share amino acid sequence homology with the c-Myc transcriptional activator domain and contain a B-Myc specific carboxy terminal region. *Oncogene* **11**:1963-1969, 1995.
13. Wenzel A, Schwab M. The mycN/max protein complex in neuroblastoma: Short review. *Eur J Cancer* **31**:516-519, 1995.
14. Barrett J, Birrer MJ, Kato GJ, Dosaka-Akita H, Dang CV. Activation domains of L-Myc and c-Myc determine their transforming potencies in rat embryo cells. *Mol Cell Biol* **12**:3130-3137, 1992.
15. McMahon SB, Van Buskirk HA, Dugan KA, Copeland TD, Cole MD. The novel ATM-related protein TRRAP is an essential cofactor for the c-Myc and E2F oncoproteins. *Cell* **94**:363-374, 1998.
16. Sakamuro D, Elliot KJ, Wechsler-Reya R, Prendergast GC. BIN1 is a novel Myc-interacting protein with features of a tumour suppressor. *Nat Genet* **14**:69-76, 1996.
17. Bush A, Mateyka M, Dugan K, Obaya A, Adachi S, Sedivy J, Cole M. c-Myc null cells misregulate *cad* and *godd45* but not other proposed c-Myc targets. *Genes Dev* **12**:3797-3802, 1998.
18. Li L-H, Nerlov C, Prendergast G, MacGregor D, Ziff EB. c-Myc represses transcription *in vivo* by a novel mechanism dependent on the initiator element and Myc box II. *EMBO J* **18**:4070-4079, 1994.
19. Roy AL, Carruthers C, Gutjahr T, Roeder RG. Direct role for myc in transcription initiation mediated by interaction with TFII-I. *Nature* **365**:359-361, 1993.
20. Desbarats L, Gaubatz S, Eilers M. Discrimination between different E-box-binding proteins at an endogenous target gene of c-Myc. *Genes Dev* **10**:447-460, 1996.
21. Xiao Q, Claassen G, Shi J, Adachi S, Sedivy J, Hann SR. Transactivation-defective c-MycS retains the ability to regulate proliferation and apoptosis. *Genes Dev* **12**:3803-3808, 1998.
22. Davis AC, Wims M, Spotts GD, Hann SR, Bradley A. A null *c-myc* mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. *Genes Dev* **7**:671-682, 1993.
23. Moens CB, Auerbach AB, Conlon RA, Joyner AL, Rossant J. A targeted mutation reveals a role for N-Myc in branching morphogenesis in the embryonic lung. *Genes Dev* **6**:691-704, 1992.
24. Charron J, Malynn BA, Fisher P, Stewart V, Jeannotte L, Goff SP, Robertson EJ, Alt FW. Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. *Genes Dev* **6**:2248-2257, 1992.
25. Sawai S, Shimono A, Wakamatsu Y, Palmes C, Hanaoka K, Kondoh H. Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* **117**:1445-1455, 1993.
26. Stanton BR, Perkins AS, Tessarollo L, Sassoon DA, Parada LF. Loss of N-myc function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes Dev* **6**:2235-2247, 1992.
27. Sirito M, Lin Q, Deng JM, Behringer RR, Sawadogo M. Overlapping

- roles and asymmetrical cross-regulation of the USF proteins in mice. *Proc Natl Acad Sci U S A* **95**:3758–3763, 1998.
28. Moroy T, Fisher P, Guidos C, Ma A, Zimmerman K, Tesfaye A, DePinho R, Weissman I, Alt FW. IgH enhancer deregulated expression of *L-myc*: Abnormal T-lymphocyte development and T-cell lymphomagenesis. *EMBO J* **9**:3659–3666, 1990.
  29. Dildrop R, Ma A, Zimmerman K, Hsu E, Tesfaye A, DePinho R, Alt FW. IgH enhancer-mediated deregulation of N-Myc gene expression in transgenic mice: Generation of lymphoid neoplasias that lack c-Myc expression. *EMBO J* **8**:1121–1128, 1989.
  30. Rosenbaum H, Webb E, Adams JM, Cory S, Harris AW. N-Myc transgene promotes B lymphoid proliferation, elicits lymphomas, and reveals cross-regulation with c-Myc. *EMBO J* **8**:749–755, 1989.
  31. Slansky JE, Farnham PJ. Introduction to the E2F family: Protein structure and gene regulation. In: Farnham PJ, Ed. New York: Springer-Verlag, pp1–30, 1996.
  32. Dyson N. The regulation of E2F by pRB-family proteins. *Genes Dev* **12**:2245–2262, 1998.
  33. Nevins JR. Toward an understanding of the functional complexity of the E2F and retinoblastoma families. *Cell Growth Differ* **9**:585–593, 1998.
  34. Bremner R, Cohen BL, Sopta M, Hamel PA, Ingles CJ, Gallie BL, Phillips RA. Direct transcriptional repression by pRB and its reversal by specific cyclins. *Mol Cell Biol* **15**:3256–3265, 1995.
  35. Ferreira R, Magnaghi-Jaulin L, Robin P, Harel-Bellan A, Trouche D. The three members of the pocket proteins family share the ability to repress E2F activity through recruitment of a histone deacetylase. *Proc Natl Acad Sci U S A* **95**:10493–10498, 1998.
  36. Parr MJ, Manome Y, Tanaka T, Wen P, Kufe DW, Kaelin WGJ, Fine HA. Tumor-selective transgene expression *in vivo* mediated by an E2F-responsive adenoviral vector. *Nat Med* **3**:1145–1149, 1997.
  37. Adams PD, Kaelin WG. Transcriptional control of cell growth: The E2F gene family. In: Farnham PJ, Ed. *Transcriptional Control of Cell Growth: The E2F Gene Family*. New York: Springer-Verlag, pp79–93, 1995.
  38. Pierce AM, Gimenez Conti IB, Schneider-Broussard R, Martinez LA, Conti CJ, Johnson DG. Increased E2F1 activity induces skin tumors in mice heterozygous and nullizygous for p53. *Proc Natl Acad Sci U S A* **95**:8858–8863, 1998.
  39. Pierce AM, Fisher SM, Conti DJ, Johnson DG. Deregulated expression of E2F1 induces hyperplasia and cooperates with *ras* in skin tumor development. *Oncogene* **16**:1267–1276, 1998.
  40. Yamasaki L, Jacks T, Bronson R, Goillot E, Harlow E, Dyson NJ. Tumor induction and tissue atrophy in mice lacking E2F1. *Cell* **85**:537–548, 1996.
  41. Field SJ, Tsai F-Y, Kuo F, Zubiaga AM, Kaelin WG, Jr., Livingston DM, Orkin SH, Greenberg ME. E2F1 functions in mice to promote apoptosis and suppress proliferation. *Cell* **85**:549–561, 1996.
  42. Lindeman GJ, Dagnino L, Gaubatz S, Xu Y, Bronson RT, Warren HB, Livingston DM. A specific, nonproliferative role for E2F5 in choroid plexus function revealed by gene targeting. *Genes Dev* **12**:1092–1098, 1998.
  43. Solomon DL, Amati B, Land H. Distinct DNA binding preferences for the c-Myc/Max and Max/Max dimers. *Nucleic Acids Res* **21**:5372–5376, 1993.
  44. Bendall AJ, Molloy PL. Base preferences for DNA binding by the bHLH-Zip protein USF: Effects of MgCl<sub>2</sub> on specificity and comparison with binding of Myc family members. *Nucleic Acids Res* **22**:2801–2810, 1994.
  45. Boyd KE, Farnham PJ. Myc versus USF: Discrimination at the *cad* gene is determined by core promoter elements. *Mol Cell Biol* **17**:2529–2537, 1997.
  46. Fisher F, Crouch DH, Javaraman PS, Clark W, Gillespie DA, Goding CR. Transcription activation by Myc and Max: Flanking sequences target activation to a subset of CACGTG motifs *in vivo*. *EMBO J* **12**:5075–5082, 1993.
  47. Blackwell TK, Huang J, Ma A, Krstzner L, Alt F, Eisenman R, Weintraub H. Binding of Myc proteins to canonical and noncanonical DNA sequences. *Mol Cell Biol* **13**:5216–5224, 1993.
  48. Hann SR, Dixit M, Sears RC, Sealy L. The alternatively initiated c-Myc proteins differentially regulate transcription through a noncanonical DNA binding site. *Genes Dev* **8**:2441–2452, 1994.
  49. Nye JA, Petersen JM, Gunther CV, Jonsen MD, Graves BJ. Interaction of murine Ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. *Genes Dev* **6**:975–990, 1992.
  50. Thompson CG, Wang C-Y, Ho I-C, Bohjanen PR, Petryniak B, June CH, Miesfeldt S, Zhang L, Nabel GJ, Karpinski B, Leiden JM. *cis*-Acting sequences required for inducible interleukin-2 enhancer function bind a novel Ets-related protein, Elf-1. *Mol Cell Biol* **12**:1043–1053, 1992.
  51. Shore P, Whitmarsh AJ, Bhaskaran R, Davis RJ, Waltho JP, Sharrocks AD. Determinants of DNA-binding specificity of ETS-domain transcription factors. *Mol Cell Biol* **16**:3338–3349, 1996.
  52. Magnaghi-Jaulin L, Masutani H, Robin P, Lipinski M, Harel-Bellan A. SRE elements are binding sites for the fusion protein EWS-FLI-1. *Nucleic Acids Res* **24**:1052–1058, 1996.
  53. Liu N, Lucibello FC, Engeland K, Muller R. A new model of cell cycle-regulated transcription: Repression of the *cyclin A* promoter by CDF-1 and antirepression by E2F. *Oncogene* **16**:2957–2963, 1998.
  54. DeGregori J, Leone G, Miron A, Jakoi L, Nevins JR. Distinct roles for E2F proteins in cell growth control and apoptosis. *Proc Natl Acad Sci U S A* **94**:7245–7250, 1997.
  55. Tao Y, Kassatly R, Cress WD, Horowitz JM. Subunit composition determines E2F DNA-binding site specificity. *Mol Cell Biol* **17**:6994–7007, 1997.
  56. Buttice G, Duterque-Coquillaud M, Basuyaux JP, Carrere S, Kurkinen M, Stehelin D. Erg, an Ets family member, differentially regulates human collagenase1 (MMP1) and stromelysin1 (MMP3) gene expression by physically interacting with the Fos/Jun complex. *Oncogene* **13**:2297–2306, 1996.
  57. Shaw PE, Schröter H, Nordheim A. The ability of a ternary complex to form over the serum response element correlates with serum inducibility of the human *c-fos* promoter. *Cell* **56**:563–572, 1989.
  58. Boyd KE, Wells J, Gutman J, Bartley SM, Farnham PJ. c-Myc target gene specificity is determined by a post-DNA-binding mechanism. *Proc Natl Acad Sci U S A* **95**:13887–13892, 1998.
  59. Luo X, Sawadogo M. Functional domains of the transcription factor USF2: Atypical nuclear localization signals and context-dependent transcriptional activation domains. *Mol Cell Biol* **16**:1367–1375, 1996.
  60. Kollmar R, Sukow KA, Sponagle SK, Farnham PJ. Start site selection at the TATA-less carbamoyl-phosphate synthase (glutamine-hydrolyzing)/aspartate carbamoyltransferase/dihydroorotase promoter. *J Biol Chem* **269**:2252–2257, 1994.
  61. Watanabe G, Albanese C, Lee RJ, Reutens A, Vairo G, Henglein B, Pestell RG. Inhibition of cyclin D1 kinase activity is associated with E2F-mediated inhibition of cyclin D1 promoter activity through E2F and Sp1. *Mol Cell Biol* **18**:3212–3222, 1998.
  62. Karlseder J, Rotheneder H, Wintersberger E. Interaction of Sp1 with the growth and cell cycle-regulated transcription factor E2F. *Mol Cell Biol* **16**:1659–1667, 1996.
  63. Slansky JE, Li Y, Kaelin WG, Farnham PJ. A protein synthesis-dependent increase in E2F1 mRNA correlates with growth regulation of the dihydrofolate reductase promoter. *Mol Cell Biol* **13**:1610–1618 [author's correction: 13:7201], 1993.
  64. Fry CJ, Slansky JE, Farnham PJ. Position-dependent transcriptional regulation of the murine dihydrofolate reductase promoter by the E2F transactivation domain. *Mol Cell Biol* **17**:1966–1976, 1997.
  65. Lin S-Y, Black AR, Kostic D, Pajovic S, Hoover CN, Azizkhan JC. Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. *Mol Cell Biol* **16**:1668–1675, 1996.
  66. De Francesco L. Taking the measure of the message. *The Scientist* **12**:20–21, 1998.

67. Marshall A, Hodgson J. DNA chips: An array of possibilities. *Nature Biotechnology* 16:27–31, 1998.
68. DeRisi J, Penland L, Brown PG, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JMG. Use of cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14:457–460, 1996.
69. Lockhart D, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* 14:1675–1680, 1996.
70. Bello-Fernandez C, Packham G, Cleveland JL. The ornithine decarboxylase gene is a transcriptional target of c-Myc. *Proc Natl Acad Sci U S A* 90:7804–7808, 1993.
71. Eilers M, Schirm S, Bishop JM. The Myc protein activates transcription of the  $\alpha$ -prothymosin gene. *EMBO J* 10:133–141, 1991.
72. Lutz W, Stohr M, Schurmann J, Wenzel A, Lohr A, Schwab M. Conditional expression of N-Myc in human neuroblastoma cells increases expression of  $\alpha$ -prothymosin and ornithine decarboxylase and accelerates progression into S-phase early after mitogenic stimulation of quiescent cells. *Oncogene* 13:809–812, 1996.
73. Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* 257:967–971, 1992.
74. Hubank M, Schatz DG. Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res* 22:5640–5648, 1994.
75. Lisitsyn N, Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 259:946–951, 1993.
76. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 270:484–487, 1995.
77. Benvenisty N, Leder A, Kuo A, Leder P. An embryonically expressed gene is a target for c-Myc regulation via the c-Myc-binding sequence. *Genes Dev* 6:2513–2523, 1992.
78. Tavtigian SV, Zabludoff S, Wold BJ. Cloning of mid-G<sub>1</sub> serum response genes and identification of a subset regulated by conditional myc expression. *Mol Biol Cell* 5:375–388, 1994.
79. Lewis BC, Shim H, Li Q, Wu CS, Lee LA, Maity A, Dang CV. Identification of putative c-Myc-responsive genes: Characterization of *rc1*, a novel growth-related gene. *Mol Cell Biol* 17:4967–4978, 1997.
80. Shim H, Dolde C, Lewis BC, Wu C-S, Dang G, Jungmann RA, Dalla-Favera R, Dang CV. c-Myc transactivation of *LDH-A*: Implications for tumor metabolism and growth. *Proc Natl Acad Sci USA* 94:6658–6663, 1997.
81. Robinson L, Panayiotakis A, Papas TS, Kola I, Seth A. Ets target genes: Identification of *Egr1* as a target by RNA differential display and whole genome PCR techniques. *Proc Natl Acad Sci U S A* 94:7170–7175, 1997.
82. Lawlor ER, Lim JF, Tao W, Poremba C, Chow CJ, Kalousek IV, Kovar H, MacDonald TJ, Sorensen PH. The Ewing tumor family of peripheral primitive neuroectodermal tumors expresses human gastrin-releasing peptide. *Cancer Res* 58:2469–2476, 1998.
83. May WA, Arvand A, Thompson AD, Braun BS, Wright M, Denny CT. EWS/FLI1-induced manic fringe renders NIH 3T3 cells tumorigenic. *Nat Genet* 17:495–497, 1997.
84. Arvand A, Bastians H, Welford SM, Thompson AD, Ruderman JV, Denny CT. EWS/FLI1 upregulates mE2-C, a cyclin-selective ubiquitin conjugating enzyme involved in cyclin B destruction. *Oncogene* 17:2039–2045, 1998.
85. Kim JH, Hui P, Yue D, Aycock J, Leclerc C, Björing AR, Perkins AS. Identification of candidate target genes for EVI-1, a zinc finger oncoprotein, using a novel selection strategy. *Oncogene* 17:1527–1538, 1998.
86. Grandori C, Mac J, Siëbelt F, Ayer DE, Eisenman R. Myc-Max heterodimers activate a DEAD box gene and interact with multiple E box-related sites *in vivo*. *EMBO J* 15:4344–4357, 1996.
87. Walter J, Biggin MD. Measurement of *in vivo* DNA binding by sequence-specific transcription factors using UV cross-linking. *Methods* 11:215–224, 1997.
88. Orlando V, Strutt H, Paro R. Analysis of chromatin structure by *in vivo* formaldehyde cross-linking. *Methods* 11:205–214, 1997.
89. Cohen-Kaminsky S, Maouche-Chretien L, Vitelli L, Vinit M-A, Blanchard I, Yamamoto M, Peschle C, Romeo P-H. Chromatin immunoselection defines a TAL-1 target. *EMBO J* 17:5151–5160, 1998.