

Identification of Mammalian E2F Regulatory Networks Using DNA Microarray Hybridization Analyses

Matthew J. Oberley and Peggy J. Farnham

Abstract

The E2F family of transcription factors has been intensely studied since the discovery that E2F1 can interact with the retinoblastoma tumor suppressor protein. Candidate gene approaches have led to the identification of E2F targets that play a role in cell cycle regulation, DNA replication, and apoptosis. The advent of new, microarray-based approaches has greatly expanded the base of knowledge about E2F and has revealed that E2F plays a role in far more cellular processes than originally envisioned. Here, we review the recent literature in which microarray technology has been used to gain insight into E2F activity.

Introduction

The E2F family of transcription factors plays a key role in cell cycle progression and the deregulation of E2F activity is linked to the development of human cancer. For example, certain E2Fs bind to and are repressed by the retinoblastoma (Rb) tumor suppressor protein, which is frequently mutated in several types of cancer.¹ The E2F family (which to date consists of E2Fs 1-7 and the heterodimerization partners DP1 and DP2) has been intensely studied since the identification in 1987 of E2F as a factor that recognizes the viral E1A promoter;² for a recent review see Trimarchi et al.³ Initial discoveries included the characterization of a consensus E2F binding site (TTTSSCGC) and the finding that E2Fs bind to and regulate promoters of genes that control processes such as cell cycle progression, DNA replication, and apoptosis.⁴ The pace of discovery has accelerated with the recent sequencing of the human genome, the increased use of bioinformatics, and the development of microarray technologies, the combination of which has allowed the identification and analysis of E2F targets in vivo in a high-throughput manner. To date, several E2F microarray analyses have been performed, identifying sets of genes regulated by individual E2Fs. Although one might criticize such studies as “fishing expeditions” that simply provide a large list of genes, carefully designed microarray experiments provide the opportunity to move beyond reductionist science by employing massively parallel analyses to identify metapatterns that can give novel insights into whether a factor regulates a specific process. In particular, such techniques can ultimately lead to the delineation of transcriptional regulatory networks in various normal tissues, assist in the char-

acterization of the deregulation of these networks in pathologies such as human cancers, and spur the development of therapeutics that can revert the regulatory network back to the normal state. Obviously, this is not a trivial goal given that there are estimated to be more than 2000 transcription factors in the human genome. However, large steps have been taken in the laboratory of R. Young who has begun to identify most of the transcriptional regulatory networks in yeast using a combination of gene expression analysis, chromatin immunoprecipitation coupled to microarray analysis (ChIP-chip), and bioinformatic algorithms.⁵ Although such comprehensive studies have not yet been performed in mammalian cells, investigators have taken two types of microarray-based approaches to high-throughput experimental identification of E2F targets: mRNA expression analysis following modulation of E2F activity and direct analysis of E2F binding by coupling ChIP with genomic microarray analysis. In this review, we summarize how these studies have provided novel insights concerning biological processes in which E2F participates and propose future experiments designed to more fully understand the complex regulatory networks in which E2F functions.

Identification of E2F Target Genes Using Oligonucleotide-Based Microarrays

E2F was initially identified as a cellular factor that regulates transcription from the adenoviral E2 promoter.⁶ The adenoviral system provided some knowledge about the function of E2F and allowed the sequence of a high affinity E2F binding site to be identified. However, E2F's role in regulating transcription of cellular genes continued to be unclear for several years. The first cellular gene determined to be regulated by the E2F family was dihydrofolate reductase (*dhfr*), a gene required for de novo purine biosynthesis (and thus important for the production of nucleotides required for DNA replication). The *dhfr* gene displays cell cycle-specific transcriptional regulation, with the highest promoter activity occurring at the G1/S phase boundary. Deletional analysis of the *dhfr* promoter revealed that an E2F site is critical for the increase in promoter activity at the G1/S phase boundary.⁷ Although this was the first indication that the E2F family is a critical regulator of S phase events, other genes involved in cell cycle progression or DNA replication were soon shown to be regulated by E2F.^{8,9}

These early studies of the E2F family relied upon candidate gene approaches to identify E2F-regulated genes. Such studies were time-consuming and biased towards identifying E2F target genes that controlled processes already known to be regulated by E2F. Therefore, it was not clear if the E2F family would be limited to controlling the G1/S phase transition or if it was more broadly involved in other cellular functions. Fortunately, the development of DNA-based microarrays soon allowed a less biased examination of the global cellular role of E2F. DNA microarrays have revolutionized the study of gene expression and are produced by the utilization of robotics that allows precise spatial control over the synthesis (or deposition) of DNA-based oligonucleotide sequences or PCR fragments onto solid platforms (usually glass slides) to use as a massively parallel probe of unknown samples of DNA or RNA.

To date, five independent groups have used high-density oligonucleotide microarrays to study the effects of overexpression of E2F1 protein levels on gene expression. Although each group employed a different system to introduce exogenous E2F1 and also used different mammalian cell lines, the studies resulted in complementary results that have contributed fundamentally to our knowledge of E2F activity in vivo. Interestingly, each group used an inducible system to express E2F1, which may have been required given that it has been demonstrated that E2F1 overexpression causes growth inhibition, both in normal and transformed cells.^{10,11}

In one of the first studies to emerge in this crowded field, Kalma et al studied the effects of E2F1 overexpression in a Rat1a cell line that had been stably transfected with a zinc-inducible E2F1 cassette.¹² They used this system to prepare RNA from cells in which expression of E2F1

had been induced for 12 or 16 hours. The RNA was then used to probe both an Atlas membrane containing several hundred rat cDNAs, as well as an Affymetrix rat-specific oligonucleotide array that represented ~8700 individual mRNAs. They identified >35 genes from the Affymetrix screen and three genes on the Atlas membrane screen that were reproducibly upregulated by E2F1 overexpression. Their results were similar to what was already known about the E2F family in that at least nine putative targets identified in their screens had previously been shown to be involved in DNA replication. One of the genes they found, RPA2, had not previously been identified as an E2F target, and they used several different types of follow-up analyses to conclude that RPA2 was indeed a bona fide E2F1 target *in vivo*. Thus, this study was one of the first to demonstrate that high-throughput microarray expression analysis could allow the identification of novel targets of E2F. However, these initial studies did not reveal any new processes in which E2F participates. The Ginsberg group next expanded their analysis and found 58 up- and 28 downregulated genes upon induction of E2F3 in Rat-1a fibroblasts. Clustering analysis showed that many of the putative E2F targets fell into three categories: DNA replication, DNA repair, and mitosis.¹³

Because of the already known link between certain E2F target genes and cell cycle regulation, the laboratory of J. Nevins used Affymetrix arrays to compare global gene expression changes as mouse embryo fibroblasts (MEFs) move through the cell cycle to changes caused by introduction of exogenous E2Fs.¹⁴ Using Affymetrix microarrays containing ~6200 sequences, they first identified changes in gene expression that occurred as cells progressed from serum starvation-induced quiescence into S phase and secondly, as cells were released from a block at the G1/S-phase boundary that was induced by treatment with hydroxyurea (HU). The results obtained were clustered based on expression patterns. Interestingly, examination of the gene expression clusters between cells brought out of quiescence and cells released from the HU block demonstrated that differences between the two different types of cycles exist, and these differences were most evident at the G1/S-phase boundary and during the G2 phase of the cell cycle. To compare these two types of cell cycle-induced changes in gene expression with E2F-induced gene expression, they infected quiescent MEFs with adenoviruses containing either E2F1 or E2F2 and collected RNA after 18 hours. They identified genes involved in transcriptional regulation and DNA repair as likely E2F target genes and, not surprisingly, found many common genes between those regulated by introduction of E2F and those that changed expression at the G1/S phase of the cell cycle. However, an additional set of putative E2F target genes were identified that changed expression in the G2 phase. Many genes involved in mitotic functions are activated in G2, and this study was the first to suggest a role for E2F in regulating mitosis.

A fairly comprehensive study of the effects of overexpression of human E2F1, -2, and -3 has also been performed.¹⁵ The Helin group used human U2OS cells to create stable cell lines that expressed estrogen receptor ligand binding domain (ER)-E2F fusion genes. They showed that upon addition of an estrogen receptor agonist, the ER-E2F protein moved into the nucleus and activated known E2F-regulated genes, such as cyclin E1. For novel target gene identification, they induced expression of ER-E2F1, ER-E2F2, or ER-E2F3 for eight hours. RNA from these cells was used for hybridization to Affymetrix oligonucleotide arrays that represented ~35,000 different mRNAs. Using the McNemar test to identify genes that were significantly altered in expression following E2F activation, they found 1240 mRNAs were regulated by at least one of the E2Fs. Although it was not practical to confirm each of these putative targets, independent validations of a subset of the mRNAs (via Northern analysis) revealed a very low false positive rate of 2%, suggesting that most of the identified genes are in fact responsive to changes in the levels of the E2Fs. Interestingly, they found that overexpression of E2Fs could lead to both increases and decreases in mRNA levels. They found that activation of genes by

ER-E2F fusion proteins did not require de novo protein synthesis, suggesting a direct involvement of E2F in the transcription of those genes. However, repression by ER-E2Fs did require de novo protein synthesis, suggesting that the repression required production of another transcription factor and that E2F indirectly regulates the repressed genes. To categorize large sets of E2F target genes into different functional groups, they performed a target gene bias analysis. The basis behind such an analysis is that if a disproportionate number of genes that are involved in specific cellular functions are regulated by E2F activation, then that cellular function must be regulated by E2F. Not surprisingly, they too found that E2F targets significantly overlapped mRNAs that are regulated by serum stimulation of quiescent fibroblasts.¹⁶ However, their analysis also showed a correlation between E2F target genes and genes involved in the TGF β signaling pathway as well as targets of other transcription factors such as homeobox family members, p53, and c-Myc. These data suggested that E2F plays a much broader physiological role than the previously characterized roles in DNA synthesis, apoptosis, and cell-cycle regulation.⁴

The laboratory of D. Cress used a mouse model system to examine the effects of expression of exogenous E2F1 in quiescent 3T3 cells. Thirty hours after infection with an adenovirus expressing E2F1, total RNA was isolated and hybridized to an Affymetrix array containing ~6500 sequences. This study produced several novel findings when compared to the other E2F1 overexpression studies.^{12,14,15,17} For example, Ma et al¹⁸ found that infection of quiescent fibroblasts with E2F1 affected the regulation of genes involved in signal transduction and cell membrane biology. Similar to the other studies, they found numerous genes that were repressed by exogenous E2F1 introduction. A detailed study, using the chromatin immunoprecipitation (ChIP) assay, of one of the putative E2F1 target genes revealed that the observed repression was a consequence of direct binding of E2F1 to the Mcl1 promoter.¹⁹ E2F1 is traditionally thought of as an activating transcription factor, yet this study and others²⁰ have clearly demonstrated that E2F1 can directly repress specific target genes by binding to their promoters. As indicated above, Muller et al relied upon results of cyclohexamide experiments to conclude that all of the genes that were downregulated were indirectly regulated by E2F. It is not clear why directly repressed targets were not identified in the Muller et al study, but perhaps the effects of cyclohexamide should be interpreted with caution.

Because it was known that overexpression of E2F1 can induce apoptosis, Stanelle et al introduced an ER-E2F1 fusion protein into the p53-negative Saos2 cell line to identify E2F1 targets that were not identified as a consequence of p53-mediated apoptosis.²¹ After treating the cells with tamoxifen for 8 hours, RNA was collected and hybridized to a Human Lifegrid 1.0 cDNA microarray containing 8400 cDNAs and ESTs. This study identified 470 upregulated genes, including a number of novel E2F targets that appeared to be both directly and indirectly regulated by E2F1. Similar to other studies using overexpressed E2F1, the identified putative targets play a role in cell-cycle control, DNA replication, and apoptosis.

Although these expression studies have successfully identified large sets of E2F1 target genes, it is surprising that more studies comparing the different E2Fs have not been performed. For example, microarrays have not been used to identify genes whose expression is altered upon overexpression of E2F4-7. However, a few studies have compared targets of E2F1, -2, and -3. In addition to the studies by Muller et al described above,^{15,22} the Nevins group infected quiescent MEFs with expression cassettes for Ras, Myc, or E2F1,-2, or -3 and then performed a metagene pattern analysis to identify a unique collection of genes that are regulated by a particular factor.²² The E2F metagenes included many previously known E2F targets, but also included many genes that had not previously been identified as E2F targets. Interestingly, the metagene patterns for E2F1, -2, and -3 were clearly different. One goal of identifying a metagene pattern is to allow a prediction as to whether a particular pathway has been deregulated in a

tumor sample. The authors tested the predictive value of their Myc and Ras metagene patterns using tumors derived from transgenic mice overexpressing Myc or Ras as well as other murine mammary tumors. The Myc metagene pattern was indeed able to predict whether a mammary tumor contained elevated Myc levels. Unfortunately, the E2F metagene patterns were not compared to tumor mRNA expression profiles; therefore it is not yet clear if an E2F metagene pattern will have predictive value when studying human tumors.

Clearly, combining the modulation of E2F protein levels with gene expression analysis has been a powerful tool to delineate targets and functions of this family of transcription factors. However, there are limitations with this approach to target identification. Most importantly, all of the studies described above modulate E2F protein to non-physiological levels, which may lead to artifactual results. For example, high levels of E2F may lead to non-specific transcriptional responses due to binding and sequestration of other transcription factors. An alternative approach to the overexpression studies was performed by Wells et al.,²³ in which oligonucleotide arrays were used to compare gene expression in the livers of wild-type and E2F1 nullizygous mice. In contrast to the overexpression studies that identified scores of target genes, loss of E2F1 had minor effects on the global transcriptional profiles. Such results suggest that perhaps many of the previously identified targets may only be regulated by E2F1 under superphysiological conditions. It is also possible that functional redundancy between the E2Fs may have limited the identification of targets. However, there were 11 mRNAs that were significantly decreased and 6 mRNAs that were significantly increased in the absence of E2F1. Interestingly, many of the E2F1-specific targets were involved in drug detoxification, as opposed to the more common functions of E2F target gene that are related to cell proliferation. To examine whether the genes that were up- or downregulated upon loss of E2F1 were directly bound by E2F1 in vivo, ChIP analysis was performed using livers from wild-type and nullizygous animals. Of the four genes examined whose expression profile changed in the null mice, three were directly bound by E2F1, but not by other E2Fs, providing the first evidence of in vivo target gene specificity among the E2F family members. Clearly this approach could be a powerful tool to identify targets specific to the other E2F family members as knockouts have been made of each individual E2F. Such studies may reveal unique cis elements and/or promoter structures that allow recruitment of one, but not all, E2F family members. Interestingly, the E2F1-specific promoters did not possess an identifiable E2F binding site, suggesting either that E2F can bind to DNA sequences other than the canonical TTTSSCGC binding site or that E2F1 is recruited to these loci via protein-protein interactions.

In summary, although both overexpression and knockout microarray studies can identify genes whose expression is altered by E2Fs, it is difficult to know if all of the identified genes are actually direct targets of E2F family members. One technique that was used to determine whether E2F targets were direct or indirect was treatment with cyclohexamide. However, the use of de novo protein synthesis inhibitors such as cyclohexamide cannot be used to formally conclude that gene expression changes are the direct result of E2F binding. For example, supraphysiologic levels of E2F may lead to the sequestration of coactivators and/or corepressors, which could result in artifactual gene expression in the absence of new protein synthesis. Although promoter-specific ChIP assays can provide information as to whether a gene is directly or an indirectly regulated by E2F, these follow-up assays were performed only in the studies by Wells et al.²³ It is not clear why the other laboratories did not attempt to validate a subset of their putative targets using ChIP assays. However, it is obvious that checking one promoter at a time would be time-consuming and laborious. In the next section, we will discuss the development and use of the ChIP-chip (chromatin immunoprecipitation coupled to microarrays) assay that has allowed a high throughput, relatively unbiased analysis of in vivo binding of E2F proteins expressed at their physiological levels.

Identification of Direct Targets of E2F Family Members Using a ChIP-Chip Approach

The chromatin immunoprecipitation assay has revolutionized the study of mechanisms of transcription because it allows protein-DNA interactions to be directly examined *in vivo*, under normal physiological conditions. Briefly, the ChIP assay involves creating protein-DNA and protein-protein crosslinks using formaldehyde, followed by sonication to break up the chromatin into small pieces. The sonicated chromatin can then be immunoprecipitated with antibodies specific to a protein of interest. Following extensive washing, and reversal of the crosslinks, one is left with a pool of DNA that represents all of the binding sites in the genome of the factor of interest. Using prior knowledge about the potential location of these factors, one can design primers specific to a locus of interest and assess whether the DNA has been enriched relative to a control IP.

Most studies employing the ChIP assay to examine promoter occupancy by E2Fs have relied on a candidate gene approach or the selection of targets identified by expression arrays (e.g., see Wells et al²³). Also, ChIP assays have been used to confirm E2F targets identified using a bioinformatics approach.²⁴ Although predictive techniques such as gene expression arrays and *in silico* analysis have provided a set of potential targets to be tested for direct recruitment of E2Fs, it is clearly not practical to test each and every potential E2F target gene by promoter-specific PCR using ChIP samples. Rather, it is desirable to have a system that will efficiently identify all of the sequences in the genome that are bound by E2F. Young and colleagues established such a technique with a genome-wide binding analysis in yeast that incorporates ChIP and spotted cDNA microarrays.²⁵ Because of the relatively small size of the yeast genome, it was possible to spot all intergenic regions of the yeast genome onto one microarray and to identify all loci that were enriched in the ChIP samples. These loci represented the *in vivo* binding sites of the factors they chose to study. Because the human genome has three orders of magnitude more sequence than yeast, a similar intergenic microarray has not been created and alternative microarrays have had to be utilized. The use of several different types of microarrays to identify large sets of direct E2F targets is described below.

One microarray-based approach to identify promoters bound by E2F utilizes CpG island microarrays. CpG islands are cytosine- and guanine-rich sequences greater than 200 basepairs in length that are found in the promoters and 5' ends of almost 70% of human genes.²⁶ Weinmann et al utilized these relatively unbiased promoter-enriched arrays to identify binding sites for the E2F4 transcription factor.²⁷ To avoid potential bias introduced by exponential amplification techniques such as LMPCR, they prepared E2F4-enriched chromatin by pooling a large number of ChIP samples from HeLa cells, labeling the DNA with Cy5 and hybridizing to a CpG island microarray containing 7776 CpG islands. As a reference, a similar number of samples of chromatin were collected in the absence of primary antibody, labeled with Cy3, and hybridized to the array. After normalization using repeat sequences that were invariably present in both the antibody and no antibody control samples, 68 independent loci were enriched at least two-fold in the E2F4-precipitated sample. 36 of the loci were promoters for characterized genes having known functions. Similar to the expression analyses described above, many of the identified E2F target genes were involved with cellular functions such as DNA replication, repair, and recombination. In support of the observations that E2F plays a role in G2 phase,¹⁴ two of the promoters (*TTK* and *CDC25C*) identified on the CpG arrays were known to be activated in G2 phase. In addition, a number of the E2F target genes were involved in chromatin structure and function (e.g., histones and histone binding proteins). Interestingly, many of the identified E2F targets were bidirectional promoters (15%), suggesting that many genes in the human genome may be co-regulated. Also, similar to the findings that E2F1 can bind to promoters that lack E2F consensus sites,²³ Weinmann et al found that around

25% of the promoters identified on the CpG array lacked sequences that resembled an E2F binding site within 1 Kb of the transcription start. Given that most validated E2F sites are found within 200 bp of the start of transcription,²⁴ it seems likely that many of the E2F4-bound promoters that lacked E2F sites either bound through a novel E2F recognition sequence or were recruited to the promoters via protein-protein interactions. The latter scenario has been demonstrated to occur through interactions of E2F1 with Sp1,²⁸ E2F2 with the YY1 transcription factor,²⁹ and E2F3 with TFE3.³⁰

Wells et al used a similar strategy to identify targets of E2F1 and Rb using the CpG-island microarray described above.³¹ This analysis employed large parallel preparations of chromatin using antibodies to E2F1 or Rb from human Raji cells synchronized in either G0/G1 or S phase using DMSO or thymidine plus aphidocolin, respectively. E2F1- or Rb-enriched chromatin was labeled with Cy5 and applied to the CpG microarray. As a reference, Raji cells were also subjected to an equal number of ChIPs with normal rabbit anti-serum. Unlike the E2F4 CpG studies, in this study the control DNA was also labeled with Cy5 but was applied to a second CpG microarray. The data from the two arrays were then normalized using repeat sequences known to be immunoprecipitated in an antibody-independent manner. Using chromatin from S phase cells, the E2F1 CpG arrays identified 124 loci and the Rb CpG arrays identified 32 loci. E2F targets encoded genes involved in DNA repair, chromatin assembly, development, and metabolism. Interestingly, in contrast to the commonly accepted model in which E2F-Rb interactions are disrupted during S phase to allow for gene activation, they found that Rb and E2F1 almost invariably colocalized (23/24 loci tested). In addition, Wells et al identified 42 loci bound by Rb using chromatin from G0/G1 phase cells. Again, all of the tested loci were bound by Rb and E2F1. Thus, they determined that E2F1 and Rb colocalize in both G0/G1 and S phase Raji cells.

Recently Oberley et al utilized the human 293 cell line to examine mechanisms of E2F6-mediated transcriptional activity.³² In these studies, a ChIP-chip analysis was performed to identify E2F6 targets in 293 cells. Although a similar CpG island microarray was used as in the previous studies, the number of individual ChIP samples used for the arrays was greatly reduced due to the incorporation of an LMPCR step (a detailed protocol is described in Oberley et al³³). The E2F6-precipitated chromatin was amplified by LMPCR and labeled with Cy5. As a reference, a sample of the starting input DNA was also subjected to LMPCR and labeled with Cy3. The Cy5 and Cy3 samples were combined and hybridized to a single CpG-island array. After the data was normalized by setting the median of ratios (Cy5/Cy3) to unity and low intensity signals were discarded, CpG islands were selected as putative positives if the sequences were enriched over the starting input population by at least two-fold in two independent assays. As a control, an identical ChIP-chip analysis was performed with rabbit IgG-enriched chromatin labeled with Cy5 and input chromatin labeled with Cy3. Any sequences that were non-specifically enriched were discarded from the list of target genes. The E2F6 targets identified included several previously known E2F targets such as HP1 α , as well as a cluster of genes involved with DNA damage sensing and repair, such as BRCA1 and CtIP. Independent ChIP analyses were performed to determine a false positive rate of 27% for the putative target genes. The newly identified E2F6 targets were further studied using RNA interference technology to specifically knockdown E2F6 protein. As E2F6 has been characterized as a repressive transcription factor,³⁴ it was not surprising that the activity of the E2F6-bound promoters (e.g., BRCA1) increased following E2F6 knockdown. By combining ChIP with RNAi, Oberley et al determined that the targets identified using the CpG arrays were negatively regulated as a result of E2F6 occluding the binding of activating E2Fs, such as E2F1. This led to the hypothesis that E2F6 overexpression may provide a proliferative advantage to tumorigenic cells, a model that is currently being tested.

Ren et al used an alternative approach to identify novel targets of E2F1 and E2F4 in primary quiescent human WI-38 fibroblasts.³⁵ Their arrays were somewhat more limited in scope than the CpG arrays described above in that they only contained promoters of 1444 human genes, 1200 of which were selected because they showed regulated expression as cells progressed through the cell cycle. Although these studies were biased in that specific promoters were selected for the arrays, the promoters did not have to fall within a CpG island and thus that form of experimental bias was removed. Their experimental approach was identical to that used to study genome-wide binding of transcription factors in yeast.²⁵ Briefly, ChIP was performed with antibodies specific to E2F1 or E2F4 and LMPCR was used to amplify the resultant pools of DNA. The E2F1 or E2F4 DNA samples were labeled with Cy5 and, as a reference, an equal amount of total input DNA was labeled with Cy3. The relative enrichment of each sequence over total input DNA (Cy5/Cy3) was determined using an error model²⁵ and a p-value was assigned to each ratio based on replicated experiments; enrichment of any particular sequence was deemed to be significant if the averaged p-value was less than 0.002. Using these criteria, 17 of 20 previously described E2F targets that were present on the array were positive in their analysis of E2F binding. As with the Weinmann et al³⁶ and Wells et al³¹ studies, Ren et al identified a significant number of novel E2F1 and E2F4 targets which they grouped into 7 distinct functional categories: cell-cycle regulation, DNA replication, DNA repair, DNA damage and G2/M checkpoint, chromosome transactions, and mitotic regulation. As expected, many of the identified targets were induced at least 2.5 fold as cells progressed through the cell cycle. However, there were a large number of putative targets whose expression did not change during the cell cycle, suggesting that many E2F target genes are not cell cycle regulated. Interestingly, Ren et al³⁵ noted that many of their putative targets did not contain a consensus E2F binding site, a conclusion also reached by other studies.^{27,31}

Summary

Gene expression microarrays and ChIP-chip analysis offer the opportunity to move beyond reductionist science and identify metapatterns of regulated genes that will give novel insight into the biological processes regulated by a particular DNA binding factor. Utilization of both types of microarrays has led to a remarkable advancement of knowledge of the multitude of ways in which the E2F family controls normal cellular physiology. Even though different cell systems and experimental designs were employed in the various microarray analyses, similar types of target genes were identified in the different studies. Although one should expect that many of the genes whose expression can be altered by over- or underexpression of E2Fs will be indirect targets that are responding to alterations in the cell cycle or signaling pathways, direct regulation by E2F, as measured by ChIP analysis, has now been shown to control genes involved in DNA synthesis, DNA replication, cell-cycle progression, DNA damage sensing and repair, chromatin assembly, transcriptional regulation, mitosis, and cellular metabolism (Fig. 1). Considering the large number of processes E2F is involved in, it is perhaps not surprising that the E2F family contains so many members (seven have been identified to date), perhaps providing a functional redundancy that may be required to maintain cell viability. Accordingly, Wells et al²³ found that only a handful of genes were deregulated in E2F1 nullizygous mice, suggesting that the other E2Fs compensate for loss of E2F1. Also in support of the concept of functional redundancy is the fact that, for the most part, loss of any single E2F does not result in embryonic lethality in murine models,³⁷⁻⁴⁴ although each knockout created a distinct phenotype. However, when multiple E2Fs are ablated, the resultant embryos are often not viable, as in the case of a E2F4 and E2F5 double-knockout.⁴⁵ In addition, mouse embryo fibroblasts are unable to proliferate when E2Fs 1,2, and 3 are all functionally inactivated.⁴⁶

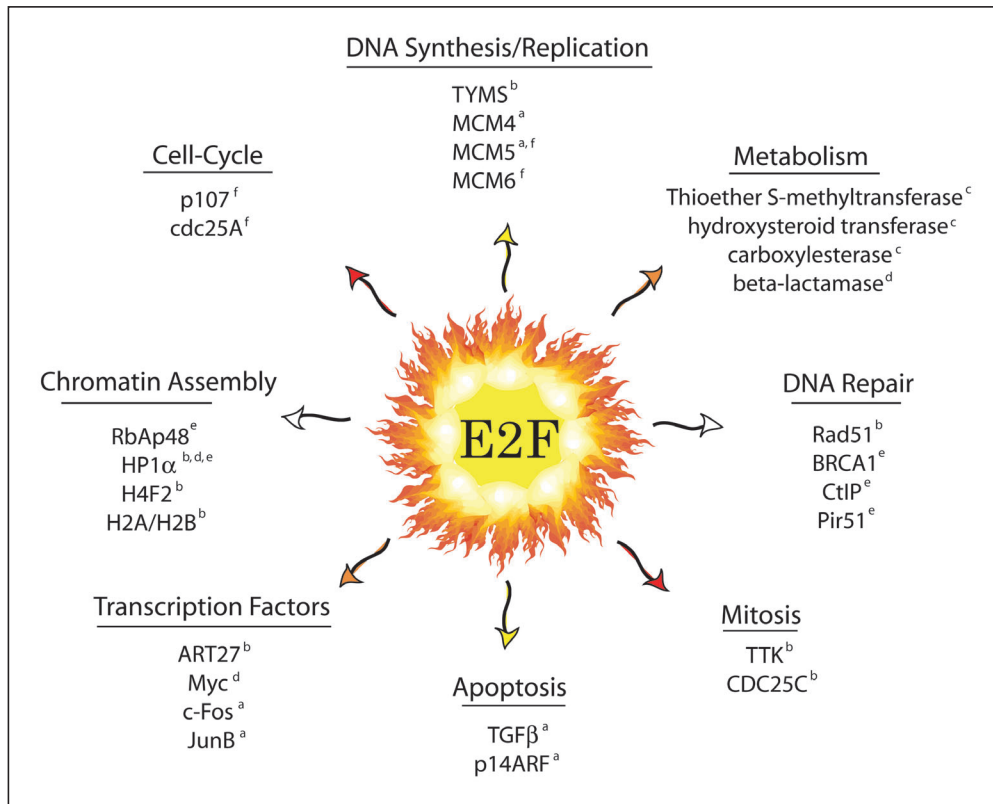


Figure 1. E2F Regulates Cellular Processes Through Direct Activation. Shown are examples of genes encoding proteins of diverse biological functions that have been identified as direct E2F targets from the ChIP-chip assays described in the text and then validated using independent ChIP assays. The listed genes are not comprehensive; other genes in each category have been identified. Although we recognize that many other investigators have performed ChIP assays on individual E2F target genes, these studies are not discussed due to space limitations. A) Kel et al., 2001. B) Weinmann et al., 2002. C) Wells et al., 2002. D) Wells et al., 2003. E) Oberley et al., 2003. F) Ren et al. 2002.

In addition to providing a list of E2F target genes, the microarray studies have also provided insight into the mechanisms by which E2Fs regulate transcription. Several studies found that a significant portion of the promoters bound directly by E2F1 or E2F4 contain no identifiable E2F binding sites.^{27,31,35} The Farnham group has validated that many of these promoter regions do indeed bind E2F and pocket proteins (Rb, p107, and p130) near the transcriptional start site using scanning ChIP analysis (M.J. Oberley, P.J. Farnham, unpublished data) and transient ChIP assays (J. Lavrrar, P. J. Farnham, unpublished data). Future studies are required to determine if the binding is mediated via protein-protein interactions and/or alternative DNA sequence elements.

Future Directions

According to Muller et al,¹⁵ as many as 7% of the known human genes may be regulated by E2F. Although many of these genes could be indirectly regulated by E2Fs, it is interesting that a similar percentage of mammalian promoters were shown to contain consensus E2F sequence elements using a bioinformatics approach to identify E2F targets.²⁴ Unfortunately, to date no one study has employed both expression arrays and ChIP-chip assays to analyze E2F

targets. Obviously, a direct comparison of the mRNAs whose regulation can be altered by an E2F with the list of promoters directly bound by that E2F should be the next goal of the E2F field. Also, to ensure that the analysis is complete, both the expression arrays and the promoter arrays should represent all of the known and predicted genes. Michael Zhang and colleagues have used *in silico* techniques to predict all promoters and first exons in the human genome using a program they developed called FirstEF;²⁶ this analysis has predicted ~38,000 promoters and first exons. Although the comparison of genomic arrays containing all 38,000 promoters to corresponding cDNA arrays containing regions transcribed from these promoters would allow the identification of the entire set of target genes directly regulated by E2F, such arrays unfortunately do not yet exist due to technological limitations. Gene expression analysis has a technological advantage presently in that a very large number of mRNAs can be analyzed using high-density oligonucleotide arrays. The genomic arrays used in the ChIP-chip analyses have been limited to a small subset of the human promoters because only relatively low density spotted arrays have been employed to date. Although high density oligonucleotide arrays have been made that span entire chromosomes;^{47,48} they have not yet been used for a ChIP-chip analysis. Arrays are now in development at NimbleGen Systems Inc. in which each of the 38,000 known or predicted promoters are represented by specific oligonucleotides, using a novel microarray fabrication technique.⁴⁹ With these arrays, location analysis can now be performed in mammalian systems in a manner similar to that used for the study of yeast transcription factors.²⁵ Similar arrays are also being designed to contain oligonucleotides representing the transcribed sequences corresponding to all known and predicted genes. Once investigators begin using these comprehensive arrays, bioinformatics analyses of the intersection of the expression and binding data sets can be used to develop transcriptional regulatory networks of the E2F family.

Bioinformatics and data mining has recently been used to illuminate the relationship between the E2F targets identified in the gene-expression analysis by Ishida et al,¹⁴ and the deregulation of upstream effectors, cyclin D1 and cyclin D3. Lamb et al⁵⁰ utilized a database called the Global Cancer Map (GCM) to examine expression profiles from 190 independent tumors and concluded that cyclin D1 overexpression in tumors did not correlate with changes in E2F target expression. However, there was a high correlation between Cyclin D3 expression in the tumors and upregulation of the previously identified E2F targets identified by Ishida et al ($p=0.002$). This study showed that bioinformatic data mining will allow functional differences between highly related effectors to be dissected *in silico*.

A long-standing issue in the E2F field is that of family member specificity. Several studies have found direct evidence of E2F family member-specific genes. For example, three promoters have been identified that are bound by E2F1, but not by other E2Fs²³ and one promoter has been shown to be specifically bound by E2F6.³² Other studies, have found evidence that certain promoters may be bound by only a subset of E2Fs.³⁵ However, to date a comprehensive comparison of the genomic binding pattern of all the E2Fs in a given cell type has not yet been performed. Due to the major technological advances allowed by the development of high density microarrays, such comparisons are now possible. Analysis of the promoters bound by multiple vs. specific E2Fs may reveal common promoter elements and/or structures preferred by specific E2F family members.

Finally, genomic microarrays can be used to study the mechanisms by which E2F family members regulate transcription. For example, a comparison of the binding pattern of an E2F with a co-activator or co-repressor (e.g., see Wells et al³¹) can provide insight into the biological significance of protein-protein interactions identified using techniques such as *in vitro* binding of purified components and/or yeast two hybrid assays. Also, a comparison of the binding pattern of an E2F with the location of acetylated or methylated histones can help to test the

hypothesis that E2Fs regulate transcription by recruiting histone modifying enzymes to promoter regions.

In summary, it is likely that new insights into the biological function of specific (or subsets of) E2Fs will be revealed by comprehensive gene expression and genomic array studies. Because E2F plays a fundamental role in cellular proliferation, a thorough knowledge base of E2F activities in normal physiology will likely lead to novel therapeutics to treat proliferative disorders such as cancer.

Acknowledgments

The authors would like to thank Antonis Kirmizis for a critical reading of this manuscript. This work was supported in part by Public Health Service grant CA45240, CA09135, CA14520, and the NIEHS training grant ES 07015-24.

References

1. Classon M, Harlow E. The retinoblastoma tumour suppressor in development and cancer. *Nat Rev Cancer* 2002; 2(12):910-917.
2. Reichel R, Kovetski I, Nevins JR. Developmental control of a promoter-specific factor that is also regulated by the E1A gene product. *Cell* 1987; 48(3):501-506.
3. Trimarchi JM, Lees JA. Sibling rivalry in the E2F family. *Nat Rev Mol Cell Biol* 2002; 3:11-20.
4. Dyson N. The regulation of E2F by pRB-family proteins. *Genes Dev* 1998; 12:2245-2262.
5. Lee TI, Rinaldi NJ, Robert F et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002; 298:799-804.
6. Kovetski I, Reichel R, Nevins JR. Role of an adenovirus E2 promoter binding factor in E1A-mediated coordinate gene control. *Proc Natl Acad Sci USA* 1987; 84(8):2180-2184.
7. Slansky JE, Li Y, Kaelin WG et al. A protein synthesis-dependent increase in E2F1 mRNA correlates with growth regulation of the dihydrofolate reductase promoter. *Mol Cell Biol* 1993; 13:1610-1618 [author's correction: 1613:7201].
8. Slansky JE, Farnham PJ. Introduction to the E2F family: protein structure and gene regulation. In: Farnham PJ, ed. *Transcriptional Control of Cell Growth: the E2F gene family*. Vol 208. New York, New York: Springer-Verlag; 1996:1-30.
9. Helin K. Regulation of cell proliferation by the E2F transcription factors. *Curr Opin Genet Devel* 1998; 8:28-35.
10. Lee TA, Farnham PJ. Exogenous E2F1 is growth inhibitory before, during, and after neoplastic transformation. *Oncogene* 2000; 19:2257-2268.
11. Ginsberg D. E2F1 pathways to apoptosis. *FEBS Lett* 2002; 529:122-125.
12. Kalma Y, Marash L, Lamed Y et al. Expression analysis using DNA microarrays demonstrates that E2F-1 up-regulates expression of DNA replication genes including replication protein A2. *Oncogene* 2001; 20(11):1379-1387.
13. Polager S, Kalma Y, Berkovich E et al. E2Fs up-regulate expression of genes involved in DNA replication, repair and mitosis. *Oncogene* 2002; 21(3):437-446.
14. Ishida S, Huang E, Zuzan H et al. Role for E2F in control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. *Mol Cell Biol* 2001; 21:4684-4699.
15. Muller H, Bracken AP, Vernell R et al. E2Fs regulate the expression of genes involved in differentiation, development, proliferation, and apoptosis. *Genes and Dev* 2001; 15:267-285.
16. Iyer VR, Eisen MB, Ross DT et al. The transcriptional program in the response of human fibroblasts to serum. *Science* 1999; 283(5398):83-87.
17. Stanelle J, Stiewe T, Theseling CC et al. Gene expression changes in response to E2F1 activation. *Nucleic Acids Res* 2002; 30(8):1859-1867.
18. Ma Y, Croxton R, Moorer RLJ et al. Identification of Novel E2F-1 Regulated Genes by Microarray. *Archives of Biochemistry and Biophysics* 2002; 399(2):212-224.
19. Croxton R, Ma Y, Song L et al. Direct repression of the Mcl-1 promoter by E2F1. *Oncogene* 2002; 21(9):1359-1369.

20. Crowe DL, Nguyen DC, Tsang KJ et al. E2F-1 represses transcription of the human telomerase reverse transcriptase gene. *Nucleic Acids Res* 2001; 29(13):2789-2794.
21. Stanelle J, Stiewe T, Theseling CC et al. Gene expression changes in response to E2F1 activation. *Nucleic Acids Res* 2002; 30:1859-1867.
22. Huang E, Ishida S, Pittman J et al. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat Genet* 2003; 34(2):226-230.
23. Wells J, Graveel CR, Bartley SM, Madore SJ, Farnham PJ. The Identification of E2F1-specific genes. *Proc Nat Acad Sci USA* 2002; 99:3890-3895.
24. Kel AE, Kel-Margoulis OV, Farnham PJ et al. Computer-assisted identification of cell cycle-related genes—new targets for E2F transcription factors. *J Mol Biol* 2001; 309:99-120.
25. Ren B, Robert F, Wyrick JJ et al. Genome-wide location and function of DNA binding proteins. *Science* 2000; 290:2306-2309.
26. Davuluri RV, Grosse I, Zhang MQ. Computational identification of promoters and first exons in the human genome. *Nature Genetics* 2001; 29:412-417.
27. Weinmann AS, Yan PS, Oberley MJ et al. Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. *Genes & Dev* 2002; 16:235-244.
28. Lin SY, Black AR, Kostic D et al. Cell cycle-regulated association of E2F1 and Sp1 is related to their functional interaction. *Mol Cell Biol* 1996; 16:1668-1675.
29. Schlisio S, Halperin T, Vidal M et al. Interaction of YY1 with E2Fs, mediated by RYBP, provides a mechanism for specificity of E2F function. *EMBO J* 2002; 21:5775-5786.
30. Giangrande PH, Hallstrom TC, Tunyaplin C et al. Identification of E-box factor TFE3 as a functional partner for the E2F3 transcription factor. *Mol Cell Biol* 2003; 23:3707-3720.
31. Wells J, Yan PS, Cechvala M et al. Identification of novel pRb binding sites using CpG microarrays suggests that E2F recruits pRb to specific genomic sites during S phase. *Oncogene* 2003; 22:1445-1460.
32. Oberley MJ, Inman D, Farnham PJ. E2F6 negatively regulates BRCA1 in human cancer cells without methylation of histone H3 on lysine 9. *J Biol Chem* 2003; 278(43):42466-42476.
33. Oberley MJ, Tsao J, Yau P et al. High throughput screening of chromatin immunoprecipitates using CpG island microarrays. *Methods in Enzymol* 2003; 376:316-335.
34. Morkel M, Wenkel J, Bannister AJ et al. An E2F-like repressor of transcription. *Nature* 1997; 390:567-568.
35. Ren B, Cam H, Takahashi Y et al. E2F integrates cell cycle progression with DNA repair, replication, and G2/M checkpoints. *Genes & Dev* 2002; 16:245-256.
36. Weinmann AS, Farnham PJ. Identification of unknown target genes of human transcription factors through the use of chromatin immunoprecipitation. *Methods* 2002; 26:37-47.
37. Yamasaki L, Jacks T, Bronson R et al. Tumor induction and tissue atrophy in mice lacking E2F-1. *Cell* 1996; 85:537-548.
38. Field SJ, Tsai F-Y, Kuo F et al. E2F-1 functions in mice to promote apoptosis and suppress proliferation. *Cell* 1996; 85:549-561.
39. Murga M, Fernandez-Capetillo O, Field SJ et al. Mutation of E2F2 in mice causes enhanced T lymphocyte proliferation, leading to the development of autoimmunity. *Immunity* 2001; 15(6):959-970.
40. Cloud JE, Rogers C, Reza TL et al. Mutant mouse models reveal the relative roles of E2F1 and E2F3 in vivo. *Molec Cell Biol* 2002; 22:2663-2672.
41. Humbert PO, Verona R, Trimarchi JM et al. E2F3 is critical for normal cellular proliferation. *Genes & Devel* 2000; 14:690-703.
42. Humbert PO, Rogers C, Ganiatsas S et al. E2F4 is essential for normal erythrocyte maturation and neonatal viability. *Mol Cell* 2000; 6(2):281-291.
43. Lindeman GJ, Dagnino L, Gaubatz S et al. A specific, nonproliferative role for E2F-5 in choroid plexus function revealed by gene targeting. *Genes Dev* 1998; 12:1092-1098.
44. Storre J, Elsasser HP, Fuchs M et al. Homeotic transformations of the axial skeleton that accompany a targeted deletion of E2F6. *EMBO Rep* 2002; 3(7):695-700.

-
45. Gaubatz S, Lindeman GJ, Ishida S et al. E2F4 and E2F5 play an essential role in pocket protein-mediated G1 control. *Mol Cell* 2000; 6(3):729-735.
 46. Wu L, Timmers C, Maiti B et al. The E2F1-3 transcription factors are essential for cellular proliferation. *Nature* 2001; 414(6862):457-462.
 47. Kapranov P, Cawley SE, Drenkow J et al. Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 2002; 296:916-919.
 48. Rinn JL, Euskirchen G, Bertone P et al. The transcriptional activity of human chromosome 22. *Genes & Devel* 2003; 17:529-540.
 49. Singh-Gasson S, Green RD, Yue Y et al. Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnol* 1999; 17:974-978.
 50. Lamb J, Ramaswamy S, Ford HL et al. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* 2003; 114:323-334.