The CCAAT box is a prototypical promoter element, almost invariably found between −60 and −100 upstream of the major transcription start site. It is bound and activated by the histone fold trimer NF-Y. We performed chromatin immunoprecipitation (ChIP) on chip experiments on two different CpG islands arrays using chromatin from hepatic HepG2 and pre-B cell leukemia NALM-6 cell lines, with different protocols of probe preparation and labeling. We analyzed and classified 239 known or predicted targets; we validated several by conventional ChIPs with anti-YB and anti-YC antibodies, in vitro EMSAs, and ChIP scanning. The importance of NF-Y binding for gene expression was verified by the use of a dominant negative NF-YA mutant. All but four genes are new NF-Y targets, falling into different functional categories. This analysis reinforces the notion that NF-Y is an important regulator of cell growth, and novel unexpected findings emerged from this unbiased approach. (i) A remarkable proportion of NF-Y targets, 40%, are complex transcriptional units composed of divergent, convergent, and tandem promoters. (ii) 40–50% of NF-Y sites are not in core promoters but are in introns or at distant 3′ or 5′ locations. The abundance of “unorthodox” CCAAT positions highlights an unexpected complexity of the NF-Y-mediated transcriptional network.

The CCAAT box is a DNA element that controls transcriptional initiation in eukaryotic promoters; recent bioinformatic studies unambiguously identify it as one of the most widespread. The analysis on 1031 human promoters isolated through unbiased determination of mRNA start sites suggested that the CCAAT box or its reverse ATTGG is present in as many as 67% of promoters (1). A statistical, unbiased analysis of neural octanucleotides on a large 13,000-promoter data set confirmed that the CCAAT is second only to the Sp1-binding GC box in terms of abundance, despite the fact that the percentage of CCAAT promoters was inferior, 7.5% (2). Furthermore, analysis of cell cycle-regulated genes identified the CCAAT box as specifically present in promoters of G2/M genes (3). Most importantly, specific flanking nucleotides emerging from these studies matched specifically the consensus of the NF-Y transcription factor. A combination of EMSAs and transfections with highly diagnostic dominant negative vectors implicated NF-Y as the CCAAT activator (4). It is composed of three subunits, NF-YA, NF-YB, and NF-YC, all necessary for sequence-specific binding to a G/A, G/A, C, C, A, A, T, C/G, A/G, G/C consensus. NF-YB and NF-YC contain evolutionarily conserved histone fold motifs common to all core histones, mediating dimerization, a feature strictly required for NF-YA association and sequence-specific DNA binding (5, 6). In essentially all cases described so far, the binding of the trimer is important or essential for transcriptional regulation (7).

NF-Y is considered as a general promoter organizer: thanks to its histone-like nature, it presets chromatin structure locally (8), interfacing well with nucleosomes (9), it helps the binding of neighboring factors (reviewed in Refs. 4 and 5) and attracts coactivators, such as p300/CREB-binding protein (8, 10). The location of the CCAAT box is far from random, being positioned between −60 and −100 in the vast majority of the promoters analyzed. In general, our knowledge of the anatomy of NF-Y-binding sites in terms of flanking sequences, position with respect to transcriptional start sites, and promoter context (6, 11, 12) enables us to make predictions as to whether a gene will be regulated by NF-Y.

Chromatin Immunoprecipitation (ChIP)1 experiments determined that NF-Y is bound in vivo before gene activation (10–13); NF-Y is bound to a transcribing cyclin B1 promoter during mitosis in HeLa cells (14). Indeed, binding to cell cycle-regulated promoters is not constitutive but is time-regulated, being found before activation and displaced when promoters are repressed (10). Furthermore, conditional knock-out experiments of CBP-B (NF-YA) unambiguously determined that the protein is required for cell proliferation of mouse embryo fibroblasts and mouse development (15).

The analysis of 130 mammalian CCAAT-containing promoters suggests a prevalence in genes that are active in a tissue- or

---

1 The abbreviations used are: ChIP, chromatin immunoprecipitation; Pipes, 1,4-piperazinedinedithanesulfonic acid; PMSF, phenylmethylsulfonyl fluoride; IP, immunoprecipitation; CTU, complex transcriptional unit; EMSA, electrophoretic mobility shift assay; CREB, cAMP-responsive element-binding protein.
Identification of NF-Y Targets

The generation of amplicons from individual ChiPs was performed following the protocols of LM-PCR described in Refs. 20 and 21. Briefly, two unidirectional linkers (oligonucleotide JW102, 5′-GGGTTGACGCCGGGATAGTATC-3′; oligonucleotide JW103, 5′-GAATTCA-GATC-3′) were annealed and ligated to the chromatin IPs, previously brought to 0°C by Direct polymerase. The first amplifications were generated by PCR (one cycle at 55 °C for 2 min, 72 °C for 5 min, and 95 °C for 2 min, followed by 15 cycles at 95 °C for 30 min, 55 °C for min, 72 °C for 1 min, and a final extension of 4 min at 72 °C). The reaction was purified using the Qiaquick PCR purification kit (Qiagen) or the GFX PCR purification kit (Amersham Biosciences) according to the manufacturer’s instructions. One-tenth of these initial reactions were used to generate more amplicons, using the same PCR program for a subsequent 30 cycles. After each iteration of the latter, the quantity of the amplicons was quantified and examined by gene-specific PCR to ensure that the initial enrichment was maintained. 5 µg of amplicons for α-NF-YB, α-FLAG, and input DNA (subjected to the same number of PCR manipulations as the IPs) were labeled using the LabelIt Cy5/Cy3 nucleic acid labeling kit (Mirus), following the manufacturer’s instructions, with a reagent/DNA ratio of 2.5 for Cy5 (IPs) and 1.5 for Cy3 (input).

CpG Microarray Hybridization

7776 CpG Array—The development of the 7776 CpG island array was described previously (21–23). Prior to hybridization, spotted CpG island slides were incubated with a solution of 3× SSC, 0.25% SDS, and 1.5 µg/µl salmon testis DNA under a glass coverslip at 37 °C for 30 min to block nonspecific binding. Slides were washed twice with water and dried for 5 min at 600 rpm in a centrifuge. Labeled DNA was added to hybridization buffer (0.25 M NaPO4, 4.5% SDS, 1 mM EDTA, and 1× SSC), denatured at 95 °C for 2 min, cooled to 60 °C, and dropped onto slides placed in prewarmed hybridization chambers. Incubation was performed at 60 °C overnight. After hybridization, the slides were washed successively at 50 °C with 1× SSC, 0.1% SDS at room temperature with 1× SSC (0.1%) and at room temperature with 0.2 SSC for 5 min each and then dried. Hybridized slides were scanned on the GenePix 4000A scanner (Axon), and the acquired images were analyzed with the software GenePix Pro, Version 3.0. A global normalization factor was determined for each replica, evaluating the anti-NF-YB ChIP Cy5/control ChIP Cy5 ratio relative to control repetitive elements. Data were normalized prior to comparison. After normalization, positive loci were defined by hybridization intensities at least 2 times greater than that of control.

12K Array—The Cy5- and Cy3-labeled DNA were each resuspended in 10 µl of 1 µg/µl Cot-1 DNA (Invitrogen) and mixed together in order to have the same amount of input Cy3-labeled DNA for each IP Cy5-labeled DNA. The hybridization solution was then added to a final composition of 43% formamide, 4.3× SSPE, 0.42% SDS, 42 µg of salmon sperm DNA, 0.2 µg of tRNA, heated for 2 min at 95 °C and cooled down to 37 °C over 30 min. 95 µL of each mixture solution was applied to two human CpG 12K slides (University Health Network, The Microarray Center, Toronto, Canada) and hybridized at 37 °C for 18 h. The slides were prehybridized for 1 h at 42 °C with 25% formamide, 5× SSC, 0.1% SDS, and 10 µg/µl bovine serum albumin.

The slides were washed at room temperature for 5 min twice in 2× SSC, 0.1% SDS, once in 1× SSC, 0.1% SDS, and one final time in 0.1× SSC; dried; and immediately scanned using a ScanArray4000 scanner (Packard). The hybridized microarrays were analyzed using the Quantarray microarray analysis software (Packard). Features of poor intensity (<500) and those that did not meet the quality control criteria (visual inspection, spot circularity, spot uniformity, and background uniformity for both channels) were discarded. After the background subtraction for each spot, the data were normalized to median (i.e. the median value of all spots in the Cy5 channel (IP DNA) was normalized to the ratio of the median value of the control channel (Cy3 = input)). From a direct comparison of the arrays hybridized with the DNA of α-NF-YB IP and the α-FLAG IP, only the spots that showed an enrichment >2-fold in the YB samples were further analyzed. Two independent experiments were performed, each consisting of one α-NF-YB IP and one control α-FLAG IP slide, normalized to the same input DNA, and the commonly enriched spots were considered.

EXPERIMENTAL PROCEDURES

Chromatin Immunoprecipitation Assay

The procedure for ChiP was essentially as described previously (10) with some modifications. Rabbit polyclonal anti-YB and anti-YC antibodies were derived by purification of the corresponding sera on affinity columns containing purified recombinant NF-YB or NF-YC linked to CNBr-Sepharose (Sigma). Nalm-6 and HepG2 cells, grown in RPMI, supplemented with 10% fetal calf serum, 2 mM L-glutamine, 50 µg/ml β-mercaptoethanol, were treated by adding formaldehyde directly to tissue culture media to a final concentration of 1% and incubated for 10 min at room temperature. Approximately 5 × 10⁶ cells were used for each immunoprecipitation. Cross-linking reactions were stopped by the addition of phosphate-buffered saline-glycine to a final concentration of 0.125 M. Cells were washed twice with ice-cold phosphate-buffered saline, scraped, and centrifuged at 2000 rpm for 2 min. Cells were then resuspended in cell lysis buffer (5 mM Pipes, pH 8.0, 85 mM KCl, and 0.5% Nonidet P-40) containing protease inhibitors (100 ng/ml aprotinin and 100 ng/ml leupeptin) and 0.5 mM PMSF and kept on ice for 15 min. Cells were homogenized using a Dounce homogenizer (B pestle) several times, and the resultant homogenates were centrifuged at 5000 rpm for 5 min at 4 °C to pellet the nuclei. The pellets were resuspended in nuclei lysis buffer (50 mM Tris-HCl, pH 8.1, 10 mM EDTA, 0.1% SDS, and 0.5% deoxycholic acid) containing protease inhibitors and PMSF and kept on ice for 20 min. The nuclear lysates were sonicated on ice to an average chromatin length of 2.5 kb and then centrifuged at 12,000 rpm for 10 min at 4 °C. The supernatants were incubated in IP buffer (50 mM Tris-HCl, pH 8.1, 10 mM EDTA, 0.1% SDS, 0.5% deoxycholic acid, and 500 mM LiCl) containing protease inhibitors and PMSF, with Protein G-agarose, the precleared lysates were used as soluble chromatin for immunoprecipitation. The generated immunoprecipitated chromatin was used as template for random priming reactions in the presence of 10 mM amino allyl-UTP (Sigma catalog no. A-0410) using the BioPrime DNA labeling system (Invitrogen). The DNA were desalted and concentrated with a Microcon YM30 filter column (Millipore Corp.) and then lyophilized. After resuspension in water, amino allyl-dUTP-labeled chromatin was coupled with Cy5 dye (Amersham Biosciences) solubilized in 0.1 M sodium bicarbonate, pH 9.0, for 1 h in the dark. After the addition of 0.1 M sodium acetate, pH 5.2, DNAs were purified with QiAquick columns (Qiagen) and lyophilized.
Identification of NF-Y Targets

Data Analysis

Positive clones were sequenced and mapped with BLAT. The presence of CCAAT sequences were searched for 2 kb on the flanking of the 7776 Cpg island array and 500 bp on the 12K array, annotated in individual files corresponding to the genomic loci identified. The criteria for classifications are described below. Mouse orthologs were retrieved using BLAT. The annotated genes were classified according to functional categories, and the classification was compared with those performed on the MYC and E2F4 targets.

Expression Analysis of NF-Y-targeted Genes

HepG2 cells were infected with control green fluorescent protein, wild type NF-YA, or dominant negative YAm29 adenovirus. Adenovirus vectors to express NF-YA or the YAm29 dominant negative mutant were generated using AdEasy, using HindIII and XbaI from the corresponding pcDNA3-based vectors, and introduced into the same sites of the shuttle vector pAdTrack-CMV. This plasmid was recombined with the vector pAdEasy1, followed by treatment with PacI and transfection into an E1-complementing cell line. We infected exponentially growing cells for 7 h in the absence of serum. Fetal calf serum was then added, and cells were incubated for 48 h. RNA was extracted using an RNA-Easy kit (Qiagen), according to the manufacturer’s protocol. For cDNA synthesis, 4 μg of RNA were used with the M-MLV-RT kit (Invitrogen).

Electrophoretic Mobility Shift Analysis of NF-Y Binding

EMSA analyses of Fig. 3 were performed under standard NF-Y conditions (6, 11, 22, 23), with anti-YB supershift antibodies and recombinant NF-Y and the indicated oligonucleotides. 32P-Labeled oligonucleotides were incubated in 20 mM Tris-HCl, pH 7.8, 50 mM NaCl, 1 mM dithiothreitol, 3% glycerol, 5 mM MgCl2 for 30 min at 20 °C with 5 ng of recombinant NF-Y trimer or with 5 μg of HepG2 nuclear extracts together with 200 ng of poly(dI-dC) (Sigma). The samples were loaded on a 4.5% polyacrylamide gel, run for 2 h, dried, and exposed. To produce recombinant NF-Y, E. coli cells; BL21 DE3LysS was induced at an $A_{260}$ value of 0.6 by the addition of isopropyl-$\beta$-D-thiogalacto-

RESULTS

Our goal was to identify novel targets of NF-Y in an unbiased way. The combination of chromatin immunoprecipitation with microarray analysis was performed in yeast (24–26) and humans (20, 27–30). In particular, DNA microarrays containing genomic fragments with CpG islands, often corresponding to regulatory regions, were probed with DNA recovered from chromatin immunoprecipitated using MYC, E2F4/6, and methyl CpG binding domain protein antibodies. We decided to take the same route and used two different reagents: a 7776 array and a 12K array from UHN (Toronto, Canada). We also tested two different ways of preparing probes for hybridization. For the 7776 array, we prepared several sequential ChIPs with chromatin from the liver HepG2 cell line, with a highly specific anti-NF-Y antibody, and, in parallel, control ChIPs with a commercially available anti-FLAG control. The chromatin used in this procedure was larger (~2–2.5 kb) than the one used in conventional ChIPs (0.5–1 kb). Because of the modifications of our routine ChIPs with extended chromatin, we first verified whether immunoprecipitated DNAs were indeed enriched in NF-Y-targeted fragments. We used oligonucleotides amplifying several CCAAT-containing promoters in semiquantitative PCR analyses. In the upper panel, genes predicted to be targeted, according to the CCAAT promoter data set (F. Romani and R. Mantovani, unpublished observations). In the lower panel, bona fide CCAAT-less promoters were assayed. B, the NF-YA gene promoter was PCR-amplified from NALM-6 chromatin ChIP (upper panel) and after the LM-PCR amplification procedure for probe preparation (see “Experimental Procedures”).

in this procedure was larger (~2–2.5 kb) than the one used in conventional ChIPs (0.5–1 kb). Because of the modifications of our routine ChIPs with extended chromatin, we first verified whether immunoprecipitated DNAs were indeed enriched in NF-Y-targeted fragments. We used oligonucleotides amplifying several CCAAT-containing promoters in semiquantitative PCR analyses. In the upper panel, genes predicted to be targeted, according to the CCAAT promoter data set (F. Romani and R. Mantovani, unpublished observations). In the lower panel, bona fide CCAAT-less promoters were assayed. B, the NF-YA gene promoter was PCR-amplified from NALM-6 chromatin ChIP (upper panel) and after the LM-PCR amplification procedure for probe preparation (see “Experimental Procedures”).
NF-Y targets in the amplicons. Indeed, Fig. 1B shows that the NF-YA promoter amplicon is no less, and in fact probably more, enriched in the final LM-PCR chromatin compared with the initial starting material. Therefore, we conclude that both of these procedures yield sufficiently enriched DNA for further genomic analysis.

**Results of the 7776 Array**—We used DNAs from 20–30 individual ChIPs to generate probes for the 7776 array screening. We identified at least 230 spots, in which the corrected signal obtained with the NF-YB chromatin was at least 2-fold higher than the anti-FLAG signal. We sequenced all positive clones and derived their chromosomal localizations. A positive clone will indicate that a bound NF-Y site lies somewhere within 2.5 kb of the CpG island. The genomic sequences surrounding the CpG island were therefore scrutinized for the presence of CCAAT sequences for a length of 2 kb on either sides. Table I shows a list of the positive clones. Several criteria helped us to classify them as follows.

Flanking sequences are essential for high affinity NF-Y binding both at the 5′ and 3′ of the pentanucleotide, with a variation of 2 logs in \( K_d \) in vitro, between high and low affinity sites (for details, see Refs. 6, 7, 11, and 23). In essence, functional low affinity CCAAT boxes are rare and mostly found in proximity of high affinity ones. We classified as high affinity those NF-Y sites having optimal sequences both at the 5′ and at the 3′ of the pentanucleotide (+ + + in Table I); medium affinity those with optimal nucleotides at the 5′ or 3′ end (+ + in Table I); low affinity those only harboring the CCAAT pentanucleotide (+ in Table I). In most clones, multiple CCAAT boxes were identified, with various degrees of consensus match; in these cases, we referred only to the highest affinity ones.

We singled out the clones with a location appropriate for a “promoter” definition (i.e. whenever a mapped known gene or multiple clustered expressed sequence tags generated from a localized area were nearby). This is because the CCAAT position is quite constant, 60–100 bp from the transcriptional start site within the promoters analyzed (7), and exceptions to this rule are sporadic (32–34). In all cases in which multiple CCAAT boxes were detected throughout the locus, the clone was classified as “canonical” if one of them was present in the promoter, within 200 bp from the transcriptional start sites.

We further separated the promoters into two categories, based upon the type of transcriptional unit. CpG islands are abundant not only in simple promoters but also in divergent, convergent, and tandemly linked promoters as well (18, 35); we collectively classified them as complex transcriptional units (CTUs).

Species conservation of TF target sites or regulatory regions in general (and of CCAAT boxes in particular) is a hallmark of functional importance, as detailed in transfection experiments and phylogenetic footprints. We thus retrieved information of the mouse orthologous genes and analyzed them for the presence of CCAAT sequences at the corresponding position. This could only be possible, with a good degree of confidence, for the promoter (canonical and CTUs) data set, by taking the transcriptional start site as the pivotal point. The sequences of all the loci are individually provided as Supplemental Table III.

In all clones retrieved, at least one CCAAT pentanucleotide could be found. This is well expected, given the 4–5 kb of DNA analyzed on both sides of the CpG islands and the average frequency of the core CCAAT (or ATTGG) pentanucleotide, one every 0.5 kb. However, a consensus high affinity NF-Y site (+ + + in Table I) is theoretically present every 16 kb (7). Given the overall length of DNA analyzed in all of the loci (750 kb), the total number of CCAAT boxes expected would be 1500, with 46 high affinity ones. Indeed, 1135 CCAAT were scored, with 252 of these matching the NF-Y consensus; thus, although there is a slight negative skewing for the pentanucleotide around the CpG island regions analyzed, the NF-Y optimal sites were 6-fold overrepresented.

To validate our analysis, we performed conventional ChIPs, with 1 kb of chromatin. Selections of the identified targets in each of the different classes were probed with anti-NF-YB and NF-YC antibodies. Furthermore, we also performed sequential immunoprecipitations of chromatin with both antibodies (re-ChIP). The results of these experiments are shown in Fig. 2. All targets tested scored positive, further confirming that clones emerging from the ChIP on chip analysis are indeed positive for NF-Y binding in vivo.

**Results of the 12K Array**—Table II contains the genes that emerged from the 12K array screenings with the anti-YB probes. The criteria mentioned above for the classification were also applied here, except that the flanking DNA considered was shorter (1 kb) due to the restricted length of the probe. Clones showing >2-fold higher signals with respect to the FLAG control were 1205 and 783 on 5121 and 4371 spots analyzed, respectively, corresponding to 23 and 18% of positivity. 119 clones were in common; of these, 65 clones were mappable based on the sequences retrievable from the Sanger Centre. Core promoters were 10%, and noncanonical CCAAT were nearly 50%. Several CTUs were also present. Overall, the distribution was highly reminiscent of the 7776 array. Here, again, we validated selected clones by conventional ChIP; all showed a substantial enrichment with respect to the FLAG control (Fig. 2A, right panels).

**Analysis of NF-Y Targets**—To pinpoint specifically the sites of interactions, we performed in vitro EMSAs with HepG2 nuclear extracts and oligonucleotides corresponding to selected CCAAT boxes found in the CpG island regions of the HepG2 targets. Fig. 3 shows that essentially all CCAAT boxes are able to interact with a binding activity in nuclear extracts. This is identified as NF-Y by (i) supershift with the diagnostic anti-YB antibody and (ii) association with recombinant NF-Y. Therefore, we conclude that the identified targeted genes do contain NF-Y binding CCAAT boxes.

To further check whether the predicted CCAAT boxes were correctly evaluated, we performed ChIP scanning experiments on three loci. We immunoprecipitated chromatin from HepG2 with anti-YB and anti-YC antibodies and amplified three different regions of the CDKN2A-MTAP and EMX2-EMX2OS CTUs and the canonical PMSC6 gene. Results shown in Fig. 4 indicate that only one amplicon of the CDKN2A-MTAP locus was positive with both NF-Y antibodies, corresponding to the + + + CCAAT box indicated in Table I, despite the presence of other CCAAT elements in the proximity of the negative amplicons. In the case of the EMX2-EMX2OS locus, amplicons 2 and 3 were positive, corresponding to the core promoter regions of both genes, whereas an amplicon in the proximity of two high affinity sites in an intronic region of EMX2OS was not enriched compared with the control. In the PMSC6 locus, only the high affinity core promoter CCAAT was bound in vivo. Collectively, these experiments support the classification of Tables I and II and suggest that the genes are indeed under NF-Y control.

**Function of NF-Y on Selected Targets**—To verify the role of NF-Y binding in the expression of the newly discovered genes, we used adenoviral vectors expressing wild type NF-YA and the well characterized dominant negative YAm29 mutant, capable of associating to the histone fold motif dimer but crippled in the DNA-binding subdomain and hence incapable of binding to the CCAAT box. HepG2 cells were used for the infections, and mRNA analysis of a number of target genes retrieved from the 7776 array was performed by semiquantitative reverse
NF-Y-targeted genes are classified according to three categories (canonical, noncanonical, and complex transcriptional units) and to the relative match to a consensus NF-Y-binding site (+1/-1/+2/matches at the 5' or 3' ends, or no match). The presence of a CCAAT sequence in the mouse orthologs is indicated, as well as the locus link or accession numbers. In the noncanonical CCAAT cohort, we indicated whether a CCAAT was present in introns (in), or at the far 5' or 3' ends of the gene.

### TABLE 1

**Classification of NF-Y-targeted genes**

| NF-Y-targeted genes | Canonical | Noncanonical | Complex |
|---------------------|-----------|--------------|---------|
| Identification of NF-Y Targets | 13610 | | |

### facebook

**Identification of NF-Y Targets**

| Gene | Description | Function | Location | Accession |
|------|-------------|----------|----------|-----------|
| CCAAT | | | | |
transcription-PCR. The results are shown in Fig. 5. The control YBL1 transcript generated by a CCAAT-less promoter was unchanged (Fig. 5, bottom panel). All other loci were variably affected; in some cases (CDC10, BET1, and EMX2), the effect of the YAm29 was relatively modest with respect to the controls. For other genes, reduction was quite severe; expression of TIMP3, MTAP, TIP-1, and SHOX2 was nearly abolished. We also analyzed two complex loci. In the TYMS-s.FLJ147447 divergent units, both mRNA were affected, albeit modestly; in the TLP19 locus, in which the CpG island is located between exons 2 and 3, we analyzed three transcripts: in addition to TLP19, the convergent SBBI18, generated just upstream from the CpG island, and the divergent FLJ14844, which starts far upstream. Interestingly, the TLP19 and SBBI18 were severely affected by YAm29, whereas FLJ14844 was not. Taken collectively, these data confirm that the targeted genes are indeed affected by inhibition of NF-Y activity and that CTUs can be regulated simultaneously.

**DISCUSSION**

The results of the ChIP on chips analysis presented here represent a major and unexpected advance in our understanding of NF-Y genomic strategy in two specific directions: the identification of a high proportion of CTUs and of NF-Y sites away from promoters found in introns or at distant 3' or 5' locations.

The number of NF-Y-regulated genes found in our analysis is more in line with the 7.6% figure recently obtained by Fitzgerald et al. (2); in fact, were NF-Y indeed involved in the majority (67%) of promoters, as suggested by Suzuki et al. (1), we would expect a much larger number of positive clones. However, several considerations can be put forward to explain the relative paucity of isolated targets. (i) In the 7776 CpG experiments, we applied a stringent cut-off by normalizing for the higher signals observed with anti-YB DNA in clones containing repetitive sequences; recent reports, however, suggested that CCAAT boxes are present and conserved in some families of repetitive DNA of retroviral origin (36). This finding matches the well known importance of NF-Y sites in many (actually most) retroviral long terminal repeats (reviewed in Ref. 7). Thus, our normalization is likely to have obscured a larger set of targets. (ii) It is likely that only a minority of genes are expressed at high levels in all cells and hence activated by NF-Y at all times. Many of the ubiquitous genes, in fact, are only active under specific circumstances (stress, apoptotic signals, a specific cell cycle phase, or environmental stimulus). Cell cycle promoters, for example, to which NF-Y association fluctuates considerably (10), are potentially underrepresented; indeed, other anti-YB positives, such as cyclin B1, that scored between 1.5 and 2 in fluorescence intensity above the FLAG control in the 7776 array are bona fide NF-Y targets. (iii) In similar ChIP on chip experiments, an equivalent number of clones were retrieved for MYC (28) and fewer for E2F4, E2F6, and methyl CpG binding domain proteins (20, 27, 30). Alternative approaches indicate that MYC high affinity sites are only a part of the overall binding strategy (37). Thus, it is likely that our data constitute a fraction of all of the potential NF-Y targets. (iv) Most importantly, only clones that showed positivity for multiple hybridizations were considered. In the case of the 12K array, positivity was scored in 15–20% of clones of the individual loci.
experiments, yet only 119 of them overlapped. We believe that suboptimal hybridization conditions prevent the successive and reproducible identification of the same set of targets, precluding the possibility of calculating the exact number of NF-Y-targeted loci. These shortcomings notwithstanding, our data lead to several interesting considerations.

**Conservation of NF-Y Sites**—Among the identified genes, only four were previously established through mutagenesis of CCAAT, but not by ChIP analysis: (i) the UNG2-UNG1 tandemly linked genes were functionally dissected, and CCAAT boxes were found to be of importance for both genes (38); (ii) TIMP2 (and the related TIMP1/3) are clearly under NF-Y control (39–42); (iii) proliferating cell nuclear antigen, a CTU in which the CCAAT box is found in the first intron (43); and (iv) MTAP, a gene in which two separated suboptimal CCAAT boxes are important (44). For others, NF-Y-binding was more than suspected. (i) The divergent promoters of the H2B-H3 and of H2A-H4 loci belong to the wide family of histone genes; detailed mutational analysis of other histone promoters clearly evidenced the importance of NF-Y (45, 46). (ii) Functional analysis of the NKX6.1 promoter pointed to a double CCAAT region as essential (47). (iii) PAX2 and TLP19 belong to gene families for which formal proof of NF-Y involvement was obtained with other members: PAX3/7/8 (48–50) and other endoplasmic reticulum stress-inducible genes, respectively (51–53). The analysis of the conservation between human and mouse promoters represents a good example of phylogenetic footprint, since 52 of 72 (74%) mouse orthologous promoters do contain CCAAT at the expected position; this percentage increases to 86% if we consider the optimal NF-Y sites (Tables I and II). Thus, the notion that conservation of the CCAAT is an integral part of the expression strategy within gene families and across species is reinstated.

**The CCAAT Box and Complex Transcriptional Units**—It is somewhat surprising to find a high frequency of CTUs in our analysis; 24% of the loci analyzed contained bidirectional promoters, and 15–17% contained tandem promoters. Most bidirectional promoters are divergent (60%), and the rest are convergent, generating partially overlapping transcripts. This result was not anticipated; previous data identified only a minute number (essentially histones, UNG1-UNG2, and AIRC-GPAT) actually containing such units (7). An unexpected abundance of bidirectional promoters in the human genome has recently been documented; as many as 11% of the total are divergent, either with overlapping or nonoverlapping transcripts (35, 54). Furthermore, the bidirectional arrangement is often conserved among mouse orthologs and important for expression of both transcripts. We analyzed the ChIP on chip experiments previously performed on the 7776 array, obtaining figures of 15% for E2F4 and 23% for MYC targets as bidirectional promoters (27, 29). This suggests that (i) MYC and NF-Y sites are enriched in bidirectional promoters and/or (ii) that CpG islands are indeed specifically abundant in such units. CCAAT-less bidirectional promoters, do exist (e.g. the YBL1 promoter analyzed in Figs. 1 and 5); NF-Y, therefore, cannot be considered as a hallmark for such units. Nevertheless, in all systems tested so far, centrally located CCAAT boxes are important for the expression of divergent genes (45); the data obtained with the dominant negative NF-Yam29 presented in Fig. 5 on the TYMS and TLP19 loci confirm this assumption. The biological significance of the higher frequency of complex CTUs regulated by NF-Y as well as the molecular details of divergent co-regulation require further dissection.

**CCAAT at Distant Locations**—A second unanticipated result is the abundance (40–50%) of sites away from promoters, with almost half of them located in introns. This clearly means that NF-Y is not a promoter-specific factor. It is im-

---

6 R. Mantovani, unpublished results.
important to emphasize that this finding would have been completely obscured had we used a promoter array chip, as available information on CCAAT locations would have suggested. Of course, the assumption that the CCAAT box was almost exclusively a promoter element was based upon standard promoter-driven analysis, thus merely reflecting the fact that
far greater information had been gathered from such sequences. Only a handful of cases of distant locations were previously described. (i) In the major histocompatibility complex class II genes, upstream enhancers were shown to be dependent upon Y-boxes and neighboring RFX-binding sites (34). (ii) Sequences were found in the HOXB4 gene, that contain a highly conserved NF-Y site in a crucial intronic enhancer (in fact, it is not even a perfect pentanucleotide, CCATT or GCAAT), and similar deviant sequences were noticed at corresponding locations in other introns of HOX gene clusters (33). Interestingly, CCAAT boxes exist in HOX gene promoters as well (55–57), one of which (HOXB13) was identified here; they are perfect CCAAT, whereas the intronic ones are modified, most likely to accommodate the binding of additional cooperating factors, as shown for YY1 in the case of HOXB4 (33). This suggests that there might be a plethora of specialized CCAAT versions, slightly deviating from optimal sites. It is even possible that we are largely underestimating the number of binding sites by focusing on the perfect pentanucleotide. NF-Y binding has been so far invariably associated with regulatory regions, which is confirmed by the expression analysis with the dominant negative NF-Yam29 shown here. An important implication of our data, therefore, is that new enhancers or regulatory regions could be uncovered.

Fig. 5. Functional importance of NF-Y binding for target genes expression. Shown is semiquantitative reverse transcription-PCR analysis of mRNAs extracted from HepG2 cells uninfected or infected with control green fluorescent protein (GFP) adenovirus or with adenovirus producing wild type or the DN YAm29 mutant, respectively. The unaffected CCAAT-less YBL1 transcript was used for mRNA normalization. Details on the oligonucleotides used are shown in Supplemental Table II. The positions of the CCAAT boxes are marked by circles (+ (light gray), ++ (gray), and +++ (dark gray)), positions of exons are marked by black boxes, positions of the hybridizing CpG island clones are marked by white boxes, and positions of transcripts are marked by arrows.

Fig. 6. Functional classification of NF-Y target genes. The genes targeted by NF-Y whose function is known or inferred were classified according to the function of the encoded protein.
er via this strategy. In vivo functional dissection of the distant regions isolated here with enhancer-based assays is necessary to establish this point.

Functional Classification of NF-Y Targets—Fig. 6 shows the functional classification of the annotated genes. In both HeP2 and Nalm-6, prominent classes are (i) DNA-binding and transcription factors in general, which represent >25% of the total, and (ii) membrane/extracellular matrix proteins coding genes and signal transduction genes. Far fewer genes code for structural proteins, proteins involved in mRNA processing and in vesicular and nuclear trafficking. This could be due to particular skewing of the Cpg library (16), but we note that many of the genes identified in both HeP2 and Nalm-6 are indeed important for cell growth.

An important corollary to this analysis is the identification of genes that are targeted by NF-Y and MYC/E2F4. LALP1, MRS3/4, GTP2H2, and EIF3SS8 are shared with MYC; CSDA, RAD51, TYMS, and D1S155E are shared with MYC and E2F4. Thus, new transcriptional networks can be constructed; this is particularly relevant, since NF-Y is known to be cooperating with E2Fs in many systems (58) and to be controlled by MYC through direct protein-protein interactions (59). A precise mapping of the E2F and MYC regulatory areas, possibly adjacent to CCAAT boxes, as well as functional co-expression experiments, should provide a clue for their coregulation.

In conclusion, although we are still far from having a complete map of NF-Y targets on hand, the criteria employed here reveal new twists in the genomic strategy, mainly concerning its role in complex units and at nonpromoter locations. To build a complete understanding of the transcriptional networks in which the trimer takes part, it will be important to widen the analysis to lower affinity sites, in different cell types, under which the trimer takes part, it will be important to widen the analysis to lower affinity sites, in different cell types, under various growth conditions and with various partner activators.

Acknowledgment—We thank L. Penn for many helpful discussions.

REFERENCES
1. Suzuki, Y., Teunoda, T., Sese, J., Taira, H., Mizushima-Sugano, J., Hata, H., Ota, T., Isogoi, T., Tanaka, T., Nakamura, Y., Sugama, A., Sasaki, Y., Morishita, S., Okubo, K., and Sugano, S. (2001) Genome Res. 11, 677–684
2. FitzGerald, P. C., Shlyakhtenko, A., Mir, A. A., and Vinson, C. (2004) Mol. Cell. Biol. 24, 1652–1658
3. Elkon, R., Linnhart, C., Sharan, R., Shamir, R., and Sholil, Y. (2003) Genome Res. 13, 773–780
4. Mantovani, R. (1999) Gene 239, 15–27
5. Maiti, S. N., and de Crombrugghe, B. (1998) Trends Biochem. Sci. 23, 174–178
6. Romier, C., Cocchiarella, F., Mantovani, R., and Moras, D. (2003) J. Biol. Chem. 278, 1336–1345
7. Mantovani, R. (1998) Nucleic Acids Res. 26, 1135–1143
8. Li, G., Herry, M., Landsberger, N., Kaluder, N., Ogryzko, V., Nakatani, Y., and Wolfle, A. P. (1998) EMBO J. 17, 6300–6315
9. Country, F., Hu, Q., de Crombrugghe, B., and Maiti, S. N. (2001) J. Biol. Chem. 276, 40621–40630
10. Caretti, G., Salsi, V., Vecchi, C., Imbriano, C., and Mantovani, R. (2003) J. Biol. Chem. 278, 30435–30440
11. Dorn, A., Bollekes, J., Staub, A., Benoist, C., and Mathis, D. (1987) Cell 50, 863–872
12. Liang, S. G., and Maiti, S. N. (1998) J. Biol. Chem. 273, 31590–31598
13. Bennett, M. K., and Osborne, T. F. (1999) Proc. Natl. Acad. Sci. U. S. A. 97, 6340–6344
14. Sciortino, S., Gurtner, A., Manni, I., Fontemaggi, G., Dey, A., Sacchi, A., Ozato, K., and Wolffe, A. P. (1998) Mol. Cell. Biol. 18, 6389–6399
15. Rapp, R., Doenecke, D., and Albig, W. (1999) Biochim. Biophys. Acta 1446, 341–351
16. Meergans, T., Albig, W., and Doenecke, D. (1998) Eur. J. Biochem. 256, 436–446
17. Watada, H., Mirmira, R. G., Leung, J., and German, M. S. (2000) J. Biol. Chem. 275, 34224–34230
18. Okladnova, O., Poleev, A., Fantes, J., Lee, M., Plachov, D., and Horst, J. (1997) Genomics 42, 452–461
19. Okladnova, O., Syagailo, V. Y., Transit, M., Riederer, P., Stober, G., Mussen, R., and Lesch, K. P. (1999) Genomics 57, 110–119
20. Syagailo, V. Y., Okladnova, O., Reimer, E., Grassle, M., Mussen, R., Gattenloher, S., Marx, A., Meyer, J., and Lesch, K. P. (2002) Gene 294, 259–268
21. Tasanen, K., Oikarinen, J., Kivirikko, K. I., and Pihlajaniemi, T. (1992) J. Biol. Chem. 267, 15113–15119
22. Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., and Mori, K. (2001) Mol. Cell. Biol. 21, 1239–1248
23. Alenen, H. I., Williamson, R. A., Howard, M. J., Lappi, A. K., Jantti, H. P., Rautio, S., Miettinen, S., and Ruddle, L. W. (2003) J. Biol. Chem. 278, 28912–28920
24. Trinklein N. D., Aldred, S. J., Saldanha A. J., and Myers, R. M. (2003) Genome Res. 13, 308–312
25. Zwartkruis, F., Hoenmachers, T., Deschamps, J., and Meijlink, F. (1992) Nucleic Acids Res. 20, 1599–1606
26. Zhu, J., Giannoli, D. M., Zhang, Y., Rivera, A. J., and Emerson, S. G. (2003) Blood 102, 2420–2427
27. Meccia, E., Bottero, L., Felicetti, F., Peschle, C., Della, M., and Care, A. (2003) Biochim. Biophys. Acta 1626, 1–9
28. van Ginkel, P. R., Hsiao, H. C., Schievren, H., and Farnham, P. J. (1997) J. Biol. Chem. 272, 18372–18376
29. Izumi, H., Molander, C., Penn, L. Z., Ishiakai, A., Kohno, K., and Funa, K. (2001) J. Cell Sci. 114, 1533–1544