

# Characterization of the c-MYC-regulated transcriptome by SAGE: Identification and analysis of c-MYC target genes

Antje Menssen and Heiko Hermeking\*

Max-Planck-Institute of Biochemistry, Molecular Oncology, Independent Junior Research Group, Am Klopferspitz 18A, D-82152 Martinsried/Munich, Germany

Edited by Bert Vogelstein, Johns Hopkins Oncology Center, Baltimore, MD, and approved February 21, 2002 (received for review January 4, 2002)

To identify target genes of the oncogenic transcription factor c-MYC, serial analysis of gene expression (SAGE) was performed after adenoviral expression of c-MYC in primary human umbilical vein endothelial cells: 216 different SAGE tags, corresponding to unique mRNAs, were induced, whereas 260 tags were repressed after c-MYC expression ( $P < 0.05$ ). The induction of 53 genes was confirmed by using microarray analysis and quantitative real-time PCR: among these genes was *MetAP2/p67*, which encodes an activator of translational initiation and represents a validated target for inhibition of neovascularization. Furthermore, c-MYC induced the cell cycle regulatory genes *CDC2-L1*, *Cyclin E binding protein 1*, and *Cyclin B1*. The DNA repair genes *BRCA1*, *MSH2*, and *APEX* were induced by c-MYC, suggesting that c-MYC couples DNA replication to processes preserving the integrity of the genome. *MNT*, a MAX-binding antagonist of c-MYC function, was up-regulated, implying a negative feedback loop. *In vivo* promoter occupancy by c-MYC was detected by chromatin immunoprecipitation for *CDK4*, *Prohibitin*, *MNT*, *Cyclin B1*, and *Cyclin E binding protein 1*, showing that these genes are direct c-MYC targets. The c-MYC-regulated genes/tags identified here will help to define the set of bona fide c-MYC targets and may have potential therapeutic value for inhibition of cancer cell proliferation, tumor-vascularization, and restenosis.

The proto-oncogene c-MYC is at the center of a transcription factor network that regulates cellular proliferation, replicative potential, growth, differentiation, and apoptosis (1–4). Expression of c-MYC is rapidly induced by a diverse catalog of mitogens, and down-regulated during differentiation (1). Deregulation of c-MYC expression has been implicated in the genesis of diverse human cancers (1). The c-MYC gene encodes a nuclear transcription factor with a C-terminal basic/helix-loop-helix/leucine-zipper (bHLHzip) domain and an N-terminal transactivation domain. The HLHzip motif allows dimerization with the bHLHzip-protein MAX, which is a prerequisite for specific binding to DNA at E-box sequences (5'-CA(C/T)G(T/C)G-3') in the vicinity of target gene promoters (5). The MAX protein has alternative dimerization partners (*MAD-1*, -3, and -4, *MXI1*, and *MNT*), which represent antagonists of c-MYC function (1, 3). Dimerization with MAX and specific binding to DNA are required for induction of cell cycle progression, apoptosis, and transformation by c-MYC (1, 6), suggesting that c-MYC exerts its oncogenic effects by transactivation of target genes via E-boxes. However, transcriptional repression has also been implicated in transformation by c-MYC (1), although the mechanisms involved are less understood.

The nature of c-MYC target genes is expected to ultimately elucidate why c-MYC is a potent activator of carcinogenesis. Intriguingly, other mitogenic transcription factors (e.g., E2F-1) do not display significant transforming activity, although they have similar effects on cell cycle progression. Furthermore, bona fide c-MYC target genes will be instrumental for understanding the molecular mechanisms of c-MYC-mediated gene regulation,

as exemplified by recent studies on c-MYC-mediated changes in histone acetylation (7, 8).

A number of c-MYC-regulated genes have been identified by using microarray analysis of cell lines with conditional c-MYC alleles (9–12) or tumor cells expressing various c-MYC levels (13). However, identification of c-MYC targets in established cell lines may be obscured by genetic and epigenetic changes, which are selected for during passaging or immortalization and affect c-MYC target gene expression. To identify c-MYC target genes, we therefore decided to analyze global gene expression shortly after adenoviral transfer of an ectopic c-MYC allele into primary human cells.

## Materials and Methods

**Tissue Culture.** Human umbilical vein endothelial cells (HUVEC) and their respective media were obtained from Clonetics (San Diego). For serum starvation, HUVEC were kept in media containing 0.25% FBS for 24 h. P493-6 cells, RAT1A fibroblast (subclone TGR-1), and c-Myc<sup>-/-</sup>-RAT1A (subclone HO15.19) were maintained as described previously (12, 14).

## Serial Analysis of Gene Expression (SAGE) and Microarray Analysis.

Total RNA was prepared by using CsCl-gradient ultracentrifugation of guanidinium isothiocyanate-lysed cells. Poly(A) mRNA was obtained by using the MessageMaker kit (GIBCO/BRL). SAGE was performed as described previously (15, 16). For data and statistical analysis, the SAGE-2000 software was used. For microarray analysis, 600 ng poly(A) mRNA was converted to cDNA with incorporation of Cy3- or Cy5-labeled dNTPs. Hybridization to arrays coated on glass, quality control, and normalization were performed by IncyteGenomics (Palo Alto, CA). The "Human Unigene1" microarray contained probes for 8,392 annotated genes/expressed sequence tag (EST) clusters and 74 nonannotated genes/ESTs (probes are derived from the 5' cDNA region), whereas the "Human Drug Target" array represented 8,031 unique annotated genes/EST clusters and 207 nonannotated genes/EST clusters (probes are derived from the 3' cDNA region). Because of an overlap of 1,261 probes, the two arrays represented 15,443 unique human mRNA species.

**Northern Blot Analysis.** RNA was isolated by using the RNagents-kit (Promega). Probes directed against the 3'-untranslated region of the respective mRNAs were generated by PCR using ESTs as templates and subsequent gel purification. Hybridiza-

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: SAGE, serial analysis of gene expression; HUVEC, human umbilical vein endothelial cell; EST, expressed sequence tag; qPCR, quantitative real-time PCR; TSS, transcription start site.

See commentary on page 5757.

\*To whom reprint requests should be addressed. E-mail: herme@biochem.mpg.de.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

tions were performed in QuickHyb, following the manufacturer's instructions (Stratagene).

**Quantitative Real-Time PCR (qPCR).** qPCR was performed by using the LightCycler and the FastStart DNA Master SYBR Green 1 kit (Roche Applied Science). For qPCR of cDNA, primer pairs were designed to generate intron-spanning products of 110–335 bp (except 408 bp for *PHB*). Primer sequences are available as Table 4 (which is published as supporting information on the PNAS web site, www.pnas.org). cDNA was generated by using the RevertAid First Strand cDNA synthesis kit (MBI Fermentas, St. Leonrod, Germany). The generation of specific PCR products was confirmed by melting curve analysis and gel electrophoresis. Each primer pair was tested with a logarithmic dilution of a cDNA mix to generate a linear standard curve (crossing point CP plotted vs. log of template concentration), which was used to calculate the primer pair efficiency ( $E = 10^{(-1/\text{slope})}$ ). Elongation factor 1 $\alpha$  (*ELF1 $\alpha$* ) mRNA was used as an external standard because its expression was not altered significantly by c-MYC as detected by SAGE and Northern blot analysis (data not shown). For data analysis, the second derivative maximum method was applied, and induction of a cDNA species (geneX) was calculated according to Pfaffl (17):

$$\frac{(E^{\Delta\text{CP}(\text{GFP cDNA} - \text{MYC cDNA})\text{geneX}})}{(E^{\Delta\text{CP}(\text{GFP cDNA} - \text{MYC cDNA})\text{ELF1}\alpha})} = \text{fold induction.} \quad [1]$$

**Chromatin Immunoprecipitation.** HeLa cells ( $4 \times 10^8$ ) were fixed in 1% formaldehyde. Chromatin was sheared to an average size of 500 bp by sonication in the presence of glass beads [3 times for 30 s at continuous maximum power setting; Bandelin Sonopuls HD 70 w. MS73 Sonotrode (3 mm), Berlin, Germany]. Lysates corresponding to  $1.33 \times 10^8$  HeLa cells were rotated at 4°C for 12 h with 5  $\mu\text{g}$  of polyclonal antibodies specific for c-MYC (sc-764, Santa Cruz Biotechnology), CDC25C (as an irrelevant antibody, sc-327), or no antibody. Washing and reversal of cross-links was performed as described (18). Precipitated DNA fragments were quantified by using qPCR, with primer pairs flanking E-boxes in the respective promoters (PCR-I to -VI; see Fig. 2A) or primer pairs generating fragments located in more than 1.5 kbp distance 5' or 3' to the binding sites (PCR-A to -F; Fig. 2A). A genomic fragment corresponding to *ELF1 $\alpha$*  was used as an external standard. Enrichment of a DNA fragment putatively bound to c-MYC (gene X) was calculated as follows:

$$\frac{(E^{\Delta\text{CP}(\text{irr. AB-IP} - \text{MYC-IP})\text{gene X}})}{(E^{\Delta\text{CP}(\text{irr. AB-IP} - \text{MYC-IP})\text{ELF1}\alpha\text{ genomic}})} = \text{fold enrichment.} \quad [2]$$

## Results

**SAGE After Ectopic Expression of c-MYC.** SAGE libraries were generated from RNA isolated 12 h after viral infection of serum-starved HUVEC with Ad-MYC or as a control, Ad-GFP. At this time point, levels of ectopic c-MYC protein had reached maximum levels (data not shown). With SAGE, unique 10-bp tags are isolated from cDNA, concatenated, and sequenced (16). The abundance of a given tag in the SAGE library corresponds to the expression level of the corresponding gene. Combined, the two SAGE libraries (Ad-MYC-lib.: 55,430 tags; Ad-GFP-lib.: 37,048 tags) represented 27,283 unique tags, which corresponded to  $\approx 8,500$  different mRNAs, when only tags that occurred at least twice were taken into account. Because c-MYC is not sufficient to induce cell cycle reentry in HUVEC (15), the detected changes in gene expression are not secondary to c-MYC-induced cell cycle progression. According to a statistical analysis using a Monte Carlo simulation, 476 tags were significantly, differentially regulated ( $P < 0.05$ ), with 216 tags induced

and 260 tags repressed by c-MYC. Examples of previously described c-MYC target genes corresponding to induced tags include *HSP70* (43:0 tags), *ODC* (14:4 tags), *CAD* (9:2 tags), *LDH-A* (9:2 tags), *eIF4E* (11:2 tags), and  $\alpha$ -*prothymosin* (36:12 tags). The complete set of detected SAGE tags is provided online at <http://www.biochem.mpg.de/hermeking/myc-sage.html>.

**Validation of SAGE Results.** To confirm the SAGE results, two microarray analyzes were performed (see *Materials and Methods*). Of 153 annotated cDNAs, which were significantly induced by c-MYC according to SAGE, 102 were represented by probes on the microarrays. Forty-two of these transcripts showed a significant induction as detected by microarray analysis (Table 1). This result most likely represents an underestimation of correct tag-to-gene assignments, which may be due to a lower sensitivity of microarray analysis vs. the SAGE method. For a subset of interesting genes, the expression data obtained by SAGE were confirmed by using qPCR, because the sensitivity of qPCR exceeds that of microarray hybridizations (Fig. 1A, Table 1). In the case of *MSH2* and *BRC1*, qPCR analysis demonstrated that a negative result by microarray analysis does not exclude regulation by c-MYC.

**Classification of c-MYC Target Genes.** The 53 confirmed tag-to-gene assignments were grouped into functional classes (Table 1). The complete set of 476 tags significantly, differentially regulated by c-MYC is listed in Tables 2 and 3 (which are published as supporting information on the PNAS web site), with assignment to cDNAs and functional classes. Most of the functional classes are in agreement with previous studies aimed to identify c-MYC target genes (9–13). However, the genes listed in Table 1–3 significantly increase the number of known c-MYC-regulated genes involved in these processes and also indicate pathways not previously linked to c-MYC: e.g., the induction of a number of genes involved in DNA repair (*BRC1*, *MSH2*, and *APEX*) was unexpected and points to a role of c-MYC in the coordination of synthesis and repair of DNA.

**Requirement of c-MYC for Serum-Induced Expression.** c-MYC is an immediate early growth response gene and mediates changes in gene expression observed after serum stimulation. Therefore, we asked whether the c-MYC-regulated genes identified here would be induced in a c-MYC-dependent manner after restimulation of mitogen-deprived HUVEC. As shown in Fig. 1B, restimulation of serum-starved cells led to induction of all genes tested as detected by qPCR. Furthermore, a dominant-negative mutant of c-MYC (MADMYC), in which the transactivation domain was replaced by a histone-deacetylase recruiting Sin3-domain (according to ref. 19), was sufficient to block the serum induction of these genes, implying that the c-MYC gene mediates the effect of serum on the expression of these genes (Fig. 1B).

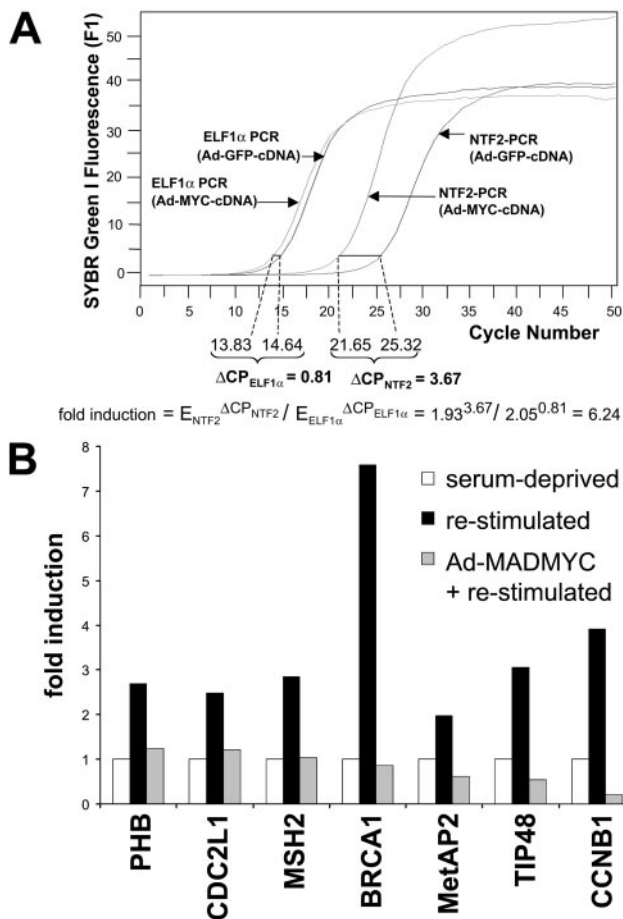
**Detection of *in Vivo* Promoter Occupation by c-MYC.** For selected, c-MYC-induced genes, analysis of the genomic sequence revealed several E-boxes upstream of their respective transcription start sites (TSS; Fig. 2A). To determine whether the endogenous c-MYC protein binds to these E-boxes and therefore directly regulates these genes, *in vivo* chromatin immunoprecipitation (ChIP) assays were performed (Fig. 2B). The previously described occupancy of the *CAD* promoter by c-MYC (18) served as a positive control: a DNA fragment spanning two E-boxes was enriched after coimmunoprecipitation of chromatin with a c-MYC-specific antibody in three independent ChIP assays, whereas a fragment located 9,595 bp upstream of the *CAD* TSS was not enriched (Fig. 2A and B). Cyclin dependent kinase (CDK4) has been identified and characterized as a c-MYC-regulated gene previously (15). However, *in vivo* promoter

Table 1. Identification and confirmation of c-MYC-induced genes

Symbol	SAGE		Micro Array		qPCR	Gene (Functional Category)	Unigene Acc. no.	
	Tag sequence	MYC	GFP	U1	DT			fold (st.dev.)
		no.	no.	fold	fold			
<b>CCNB1*</b>	TGCCATCTGT	7	0	1.1	1.1	2.3 (0.37)	Cell cycle regulation Cyclin B1	Hs. 23960
<b>CDC2L1</b>	TTTTGTTTT	6	0	1.1	/	4.1 (1.20)	Cell division cycle 2-like 1 (PITSLRE protein)	214291
<b>CDK4*</b>	GAAGGAAGAA	6	0	/	/		Cyclin dependent kinase 4	95577
<b>CEB1</b>	GTTTTTCATT	12	2	/	1.2	3.19 (1)	Cyclin E binding protein 1	26663
							<b>Transcription</b>	
<b>POLR2F</b>	CCCAATTTTC	8	2	1.8	/		RNA polymerase II subunit F	46405
<b>PCAF</b>	CTTTATGTGA	6	0	1.0	/	5.54 (3.4)	p300/CBP-associated factor	199061
<b>PIAS1*</b>	TATAAGGTGG	6	0	2.5	/		DEAD/H box polypeptide 21	169531
<b>MNT</b>	TAATAAGCAC	5	0	1.0	/	2.5 (0.98)	MAX binding protein	25497
<b>TIP48</b>	GACACCTCCT	9	1	2.1	/	2.8 (0.89)	RuvB (E. coli homolog)-like 2	6455
<b>NME1*</b>	GGCAGAGGAC	19	5	2.7	2.2		NM23A	118638
							<b>RNA processing / splicing</b>	
<b>U5-116KD</b>	GTTTGCAAGT	9	0	1.9	/		U5 snRNP-specific protein, 116 kDa	151787
<b>U3-55KD</b>	GTGTCGGCG	6	0	2.5	/		U3 snoRNP-associated 55-kDa protein	153768
<b>HNRPD</b>	TTAAACCTCA	20	2	1.8	1.7		Heterogeneous nuclear ribonucleoprotein D	170311
							<b>Protein synthesis</b>	
<b>RPS19*</b>	CAGCAGCAAA	8	1	2.3	/		Ribosomal protein S19	126701
<b>RPL3*</b>	TCATCTTTGT	5	1	2.6	2.7		Ribosomal protein L3, mitochondrial	79086
<b>RPL32*</b>	TGCACGTTTT	179	95	/	1.8		Ribosomal protein L32	169793
<b>EPRS</b>	AACTAAAAA	156	56	1.9	/		Glutamyl-prolyl-tRNA synthetase	55921
<b>MetAP2</b>	GAAACTGAAC	6	0	1.6	/	2.6 (0.44)	eIF-2-associated p67	78935
<b>EIF3S9</b>	GAATATAAAG	20	4	2.0	/		Translation initiation factor 3, s.u. 9 (eta)	57783
<b>IF3X</b>	ATGGCTGCTG	7	0	2.0	2.0		Translation initiation factor 3, s.u., putative	22616
<b>EIF5*</b>	GGCTTTACCC	49	19	1.2	2.1		Translation initiation factor 5A	119140
<b>NPM1*</b>	TGAAATAAAA	111	5	2.2	2.4		Nucleophosmin (B23, numatrin)	173205
<b>NCL*</b>	TACAAAAACCA	17	0	3.7	1.4		Nucleolin	79110
<b>NOP56</b>	AAGGAAGCAA	13	1	2.7	2.5		Nucleolar protein (KKE/D repeat)	296585
							<b>Chaperones / Protein folding</b>	
<b>HSP105B</b>	TGAAAGTGTG	7	1	2.9	/		Heat shock 105kD	36927
<b>HSPCB*</b>	GTGAGCCCAT	14	1	2.1	/		Heat shock 90kD protein 1, beta	74335
<b>HSPCA</b>	GAAGCTTTGC	31	5	1.0	3.3		Heat shock 90kD protein 1, alpha	289088
<b>HSP75</b>	GTGGGCGGCT	12	3	3.6	/		Heat shock protein 75	182366
<b>HSPA1*</b>	CAGAGATGAA	43	0	9.1	2.4		Heat shock 70kD protein 1A	8997
<b>HSPA8</b>	CCAGGAGGAA	72	28	2.4	/		Heat shock 70kD protein 8	180414
<b>HSPA9B</b>	TTTGTAGATG	6	1	2.9	/		Heat shock 70kD protein 9B (mortalin-2)	3069
<b>HSPD1*</b>	TACCAGTGTA	7	2	5.0	3.6		Heat shock 60kD protein 1	79037
<b>HSPF1*</b>	AGACCAAGT	10	1	/	2.8		Heat shock 40kD protein 1	82646
<b>HSPF1*</b>	TAATAAATTT	28	3	/	3.6		Heat shock 10kD protein 1 (chaperonin 10)	1197
<b>CCT3</b>	TAATCTTCT	18	4	1.8	/		TCP1, subunit 3 (gamma)	1708
							<b>DNA repair / metabolism</b>	
<b>BRCA1</b>	GTGGTGGACA	10	1	1.0	1.1	2.7 (0.85)	Breast cancer 1	194143
<b>MSH2</b>	TAGTTTGTGG	6	0	/	1.1	2.9 (0.22)	MSH2	78934
<b>APEX*</b>	AAAATAAAGA	11	1	1.9	2.0		Apurinic/aprimidinic exonuclease 1 (Ref-1)	73722
<b>CAD*</b>	GGGGTGGGGC	9	2	/	2.5		Carbamoylphosphate synthetase	6859
							<b>Signalling</b>	
<b>YWHAE</b>	CTTTTCAGCA	9	0	/	/	1.8 (0.25)	14-3-3 epsilon	79474
<b>FKBP52*</b>	CTTGAGCAAT	10	0	2.5	/		FK506-binding protein 4 (59kD)	848
<b>C1QBP</b>	ATAGACATAA	10	1	2.1	/		Complement comp. 1, q subcomp. bdg. prot.	78614
							<b>Transport</b>	
<b>NTF2</b>	ATACAATAAA	16	0	1.4	1.4	7.3 (2.10)	Nuclear transport factor 2	151734
							<b>Metabolism</b>	
<b>LDHB</b>	GATCACAGTT	11	1	/	2.2		Lactate dehydrogenase B	234489
<b>LDHA*</b>	TCTTGTGCAT	9	2	2.2	/		Lactate dehydrogenase A	2795
<b>SRM*</b>	CTGTGCAGCA	27	3	2.3	/		Spermidine synthase	76244
<b>PFKM</b>	ACTTCTGCCC	6	0	1.7	2		Phosphofructokinase	75160
<b>ENO1</b>	AGAAAAAAA	79	34	2.1	2.2		Enolase 1, alpha	254105
<b>CLA1</b>	CAGGACGGGC	12	1	2.3	1.0		CD36-antigen-like 1, HDL uptake receptor	180616
<b>MGST1</b>	AAATAAAGAA	9	0	2.1	1.1		Microsomal glutathione S-transferase 1	790
							<b>Apoptosis / Mitochondria</b>	
<b>BZRP</b>	GAATTTTATA	23	1	/	1.1	3.2 (0.14)	Benzodiazepine receptor (peripheral)	202
<b>PIG8</b>	AGCTGGTTTC	12	1	/	1.8	4.3 (0.73)	p53 induced gene 8	286027
<b>PHB</b>	ATTGGCTTAA	14	2	1.8	2.2	2.4 (0.53)	Prohibitin	75323
<b>PP1F*</b>	GATTTGTGTT	9	1	2.0	1.7		Cyclophilin F	173125

Assignment of 53 SAGE tags induced by c-MYC to the indicated genes was confirmed by microarray and/or qPCR analysis. The occurrence of specific SAGE tags in the respective SAGE libraries is depicted in columns 3 (MYC, Ad-MYC-infected HUVEC) and 4 (GFP, Ad-GFP-infected HUVEC). The results of two different microarray hybridizations are given as balanced differential expression as provided by IncyteGenomics (U1, human Unigene 1 array; DT, human drug target array; /, probe for cDNA not present on the array). Gray shading indicates fold induction values >1.7, which is the cut-off for reproducible differential expression according to IncyteGenomics. qPCR analysis was performed as described in Fig. 1. Induction of cdk4 by c-MYC has been confirmed by Hermeking *et al.* previously (15).

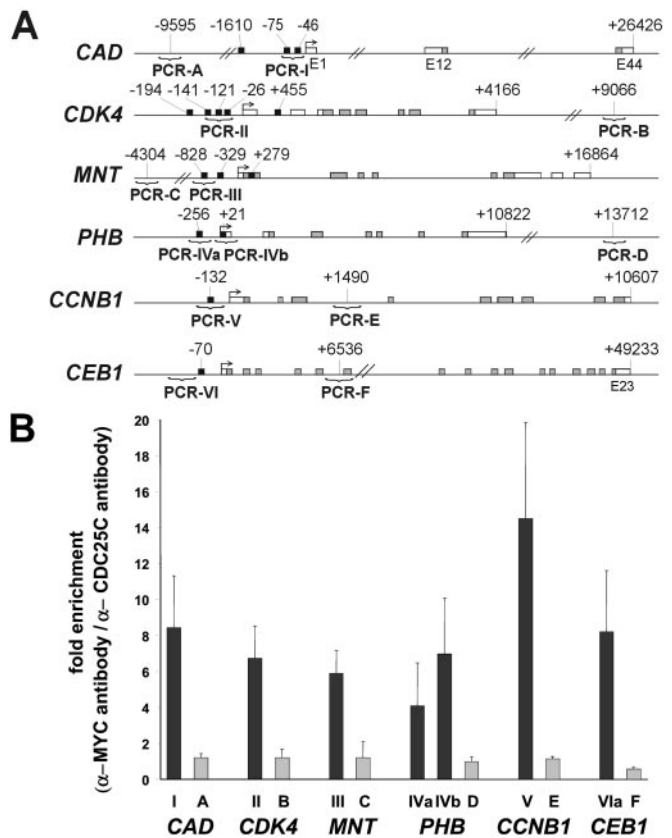
\*Previously identified as regulated by c- or N-MYC in a nonendothelial cell type.



**Fig. 1.** Analysis of c-MYC-induced genes by qPCR. Serum-starved HUVEC were infected with Ad-MYC, Ad-MADMYC, or Ad-GFP adenovirus. For re-stimulation, serum was added 4 h after infection. RNA was isolated 12 h after adenoviral infection or re-stimulation with serum. For details, see *Material and Methods*. (A) qPCR determination of fold induction of *NTF2* mRNA after c-MYC expression (primer efficiency: *NTF2*,  $E = 1.93$ ; *ELF1α*,  $E = 2.05$ ). (B) c-MYC dependence of serum-induced gene activation. The dominant-negative c-MYC mutant MADMYC was used to inactivate endogenous c-MYC function.

occupancy by c-MYC has not been shown for *CDK4*. The precipitation with a c-MYC-specific antibody led to an enrichment of a *CDK4* promoter fragment spanning three E-boxes, whereas sequences 9,066 bp downstream of the TSS were not enriched (Fig. 2*A* and *B*). These results demonstrate that c-MYC is present at the *CDK4* promoter *in vivo* and directly regulates *CDK4* expression. In the *MNT* gene, three candidate c-MYC binding sites are present (Fig. 2*A*). For the E-box at  $-828$  bp, occupancy with c-MYC was detected, implying that c-MYC directly regulates the expression of *MNT* (Fig. 2*B*). Consistent with a direct regulation by c-MYC the *PHB*, *CEB1*, and *Cyclin B1* genes harbor E-boxes in the vicinity of their promoters, which are occupied by c-MYC *in vivo* (Fig. 2*A* and *B*).

**Regulation by c-MYC in Heterologous Systems.** To prove that putative c-MYC targets are bona fide, global target genes of c-MYC, it is necessary to show that regulation by c-MYC occurs in different cell types and species. Therefore, we tested whether some of the identified genes were regulated in the human B cell line P493-6, which harbors a tetracycline-regulated c-MYC allele (described in ref. 12). As shown in Fig. 3*A*, removal of tetracycline led to induction of c-MYC mRNA. The induction of *B23*



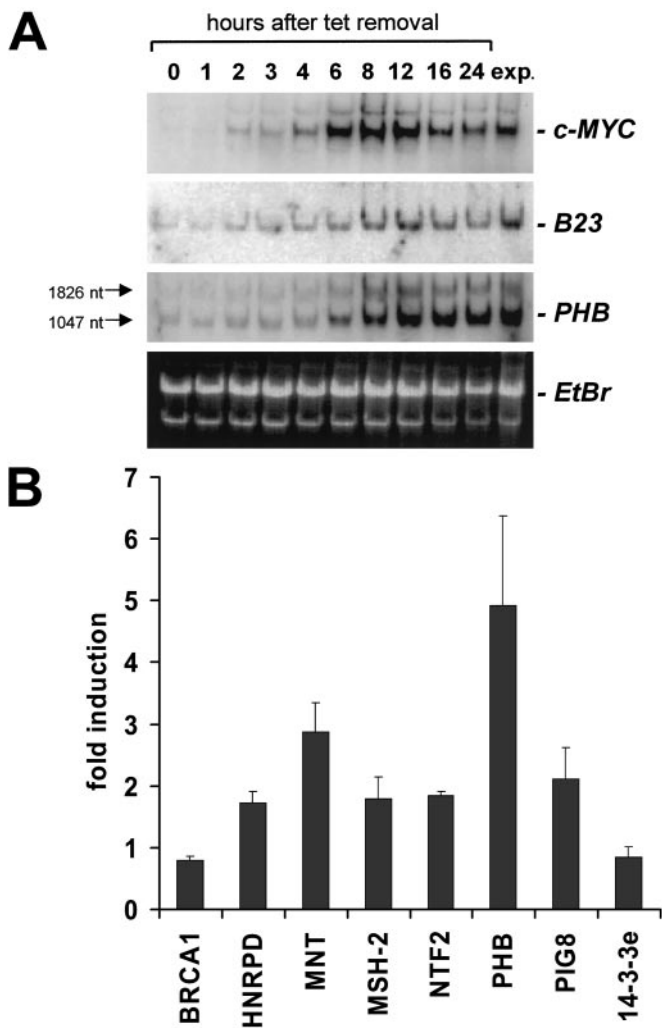
**Fig. 2.** *In vivo* binding of c-MYC to promoter sequences. (A) Maps of c-MYC-induced genes indicating the positions of E-boxes (black rectangles). Exons are represented with ORFs (gray rectangles). Arrows indicate the transcription start site (TSS), which corresponds to base pair +1. Horizontal brackets represent qPCR amplicons. (B) Chromatin immunoprecipitation assay of c-MYC-regulated genes. Each result is the average of at least three independent experiments. For details see *Material and Methods*.

and *PHB* mRNA lagged behind by 2–4 h, which is consistent with direct regulation by the newly synthesized c-MYC protein. Eight c-MYC-regulated genes identified in HUVEC were analyzed in the conditional B cell system by using qPCR (Fig. 3*B*). Only *BRCA1* and *14-3-3epsilon* showed no further induction by c-MYC (Fig. 3*B*). One possible explanation may be that these genes are already up-regulated by other factors in the P493-6 cells.

Another system that allows the validation of regulation by c-MYC are RAT1A fibroblasts in which c-Myc has been removed by homologous recombination (14, 20). In the parental RAT1A cells, induction of *PHB*, *MSH2*, and *BRCA1* mRNA was detectable 6–9 h after re-stimulation whereas, in c-Myc-deficient RAT1A cells, no significant increase in mRNA levels was detectable up to 24 h after re-stimulation (Fig. 4). Therefore, c-Myc is required for normal induction of *PHB*, *MSH2*, and *BRCA1* by serum. In summary, c-MYC regulation of most genes analyzed was conserved between species and cell-types, suggesting that the majority of the genes identified here are general targets of c-MYC.

## Discussion

The detection of a large number of genes regulated by c-MYC by using SAGE supports the notion that c-MYC is a central regulator that activates diverse cellular processes associated with progression through the cell cycle and synthesis of cellular components in preparation of cell division. In this study, we

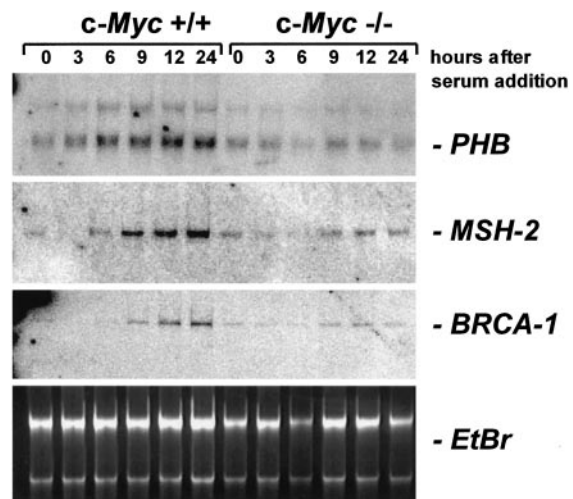


**Fig. 3.** Analysis of *c-MYC*-regulated genes identified in HUVEC in a human B cell line. P493-6 cells, which harbor a *c-MYC* gene under control of a tetracycline-responsive element (12), were cultured in the presence of tetracycline (0.1  $\mu\text{g}/\text{ml}$ ). After 72 h, tetracycline was removed for the indicated periods to induce ectopic *c-MYC* and total RNA was isolated. exp., cells asynchronously growing in absence of tetracycline for 3 days. (A) Northern blot analysis. Total RNA (4  $\mu\text{g}$ ) was loaded per lane. After transfer, the membrane was hybridized with the indicated probes. As a loading control, the ethidium bromide (*EtdBr*) stained gel is shown. (B) qPCR analysis. Total RNA was isolated 12 h after removal of tetracycline. The average of two determinations is presented.

focused on genes whose induction by *c-MYC* could be confirmed by microarray and/or qPCR analysis. However, many of the additional genes detected by SAGE, which are depicted in Tables 2 and 3, may be directly regulated by *c-MYC*.

Because target gene expression in established cell lines may be obscured by genetic and epigenetic changes acquired during *ex vivo* passaging, primary human cells were used for the SAGE screen. An example justifying this rationale may be the *CDK4* gene, which is expressed at high levels in NIH 3T3 cells and cannot be induced further after activation of a conditional *MycER* allele in this cell line (15). However, *CDK4* is induced by *c-MYC* in primary HUVECs, conditionally transformed B cells (15), and RAT1A cells (9, 15). Moreover, endogenous *c-MYC* binds to the E-boxes present in the *CDK4* promoter (this study), showing that *CDK4* is a direct target gene of *c-MYC*.

The comparison of SAGE with microarray analysis revealed a limited overlap in the results. This overlap may be explained by



**Fig. 4.** Requirement of *c-Myc* for normal induction of *BRCA1*, *MSH2*, and *PHB* after serum stimulation. *c-Myc*<sup>+/+</sup> and *c-Myc*<sup>-/-</sup> RAT1A cells were maintained in DMEM containing 0.25% calf serum for 48 h. After stimulation with 8% calf serum, total RNA was isolated at the indicated time points. Northern blot analysis was performed with the indicated probes. Total RNA was stained with ethidium bromide (*EtdBr*).

the inherent constraints of microarray hybridizations: e.g., non-specific cross-hybridization of homologous gene family members. Presumably, the higher sensitivity of SAGE further increased the difference between the assays: in general, tags with low numbers could be verified by qPCR and rarely showed induction in the microarray analysis (e.g., *Cyclin B1*, *CDC2-L1*, *PCAF*, and *MNT*).

The genes identified by SAGE point to a number of interesting new biological functions of *c-MYC* and further substantiate the role of *c-MYC* in previously identified pathways regulated by *c-MYC*: direct induction of *CDC2-L1* and *Cyclin B1* by *c-MYC* is consistent with an activating role of *c-MYC* not only in  $G_1/S$  progression, but also in the  $G_2$  phase of the cell cycle. Consistent with such a role, *c-Myc*-deficient RAT1A cells show a prolonged  $G_2$  phase (14). Although the function of the *c-MYC*-induced *CEB1* gene is still obscure, the reported association between *CEB1* and *Cyclin E* suggests a role during cell cycle progression (21). Consistent with the direct induction of *MNT* by *c-MYC* reported here, it has been shown previously that expression of *MNT* correlates with *c-MYC* expression (22). Because *MNT* is a functional antagonist of *c-MYC*, the induction of *MNT* may constitute a negative feedback loop that serves to restrict the transactivation activity of *c-MYC*. Another negative feedback mechanism may be represented by the induction of *hnRNP-D*, which mediates the degradation of mRNAs encoding cytokines and proto-oncogenes (e.g., *c-MYC*) through AU-rich elements in their 3'-untranslated regions (23). Conversely, induction of *TIP48* by *c-MYC* suggests the existence of a feed-forward loop, because binding of the RNA-helicase *TIP48* to a conserved region in the *c-MYC* transactivation domain (*MYC-BOXII*) is necessary for *c-MYC*-mediated transformation and transactivation (24). In addition, *TIP48* may be a cofactor for other transcription factors that are activated during cell cycle progression. *c-MYC* presumably mediates the induction of transcriptional cofactors and components of DNA-Polymerase II (e.g., PolII subunits F and B) to facilitate the overall increase in transcriptional activity during  $G_1/S$  transition. Induction of p300/CBP-associated factor (*PCAF*), a transcriptional cofactor with histone acetyltransferase activity that binds to p300/CBP (25), may also contribute to an overall increase in transcriptional activity. *Lactate dehydrogenase A* (*LDHA*) is a known *c-MYC*

target gene (Table 1) that links c-MYC to the activation of glycolysis (26). Shim *et al.* (26) suggested that deregulated c-MYC expression is central to the activation of glycolysis, which is found in most tumor cells (also known as the Warburg effect). Here, three additional c-MYC-induced genes encoding enzymes of the glycolytic pathway were identified: LDHB, which forms the tetrameric LDH-complex together with two LDHA molecules, phosphofructokinase and glucose phosphate isomerase.

The regulation of *BRCA1* and *MSH2* by c-MYC is consistent with the following observations: *BRCA1* is up-regulated during G<sub>1</sub>/S transition of restimulated cells (27) and expression of *MSH2* in epithelia is limited to the proliferating compartment (28). *BRCA1* and *MSH2* proteins are components of a large DNA damage repair complex, BASC (*BRCA1*-associated genome surveillance complex), that contains several other tumor-suppressor-gene products (29). For p53, which is a major sensor of DNA damage and mediator of responses to genotoxic insults, transcriptional and posttranscriptional effects mediated by c-MYC have been described previously (30, 31). Taken together, c-MYC seems to couple DNA synthesis, which presumably elevates the frequency of endogenous DNA damage, to the induction of genes involved in DNA repair, some of which function as tumor suppressor genes (*BRCA1*, *MSH2*, and *P53*). This finding is in agreement with observations from *Saccharomyces cerevisiae*, where DNA repair genes are required during DNA replication to preserve the integrity of the genome (32). Genetic inactivation of genes involved in DNA repair in the presence of an activated c-MYC gene may therefore lead to the bypass of DNA damage checkpoints, accumulation of mutations, and emergence of tumor cells.

c-MYC induces genes encoding components of the permeability transition pore complex (PTPC): up-regulation of the *benzodiazepin-receptor gene* was detected in this study, whereas induction of *VDAC* and *cyclophilin F* by c-MYC was identified previously (9, 11) and in this study (Table 1). The coordinated up-regulation of components of the PTPC may lower the threshold for cytochrome *c* release from mitochondria and

presumably contributes to the sensitization toward apoptotic stimuli, which is generally observed after c-MYC expression (reviewed in refs. 1 and 4). In agreement with this scenario, studies with c-Myc-deficient cells provide evidence that c-MYC affects the competence for apoptosis at the mitochondrial level (33).

*Prohibitin (PHB)* has been implicated in tumor suppression and senescence (34). PHB protein seems to function as a mitochondrial chaperone (35). An antiproliferative capacity has been assigned to the 3'-nontranslated region of the 1,826-nt *PHB* transcript (36). Interestingly, PHB is consistently expressed at higher levels in tumor cells (34), which may be due to the transactivation by c-MYC observed in this study. The up-regulation of the *nuclear transport factor 2 (NTF2)* gene suggests that nuclear import of nuclear localization signal-containing proteins is augmented after activation of c-MYC. c-MYC has been implicated in the control of translation previously because it directly induces *eIF4E* and *eIF-2 $\alpha$*  (37). Our results imply, that c-MYC mediates the serum induction of *MetAP2/p67*. After mitogenic stimulation, *MetAP2/p67* is induced and binds to eIF-2, thereby protecting it from inhibitory phosphorylation (38). Interestingly, the anti-angiogenic substance fumagillin and its derivative TNP-470 both inhibit neovascularization through irreversibly binding to *MetAP2* (39). Because TNP-470 has proven to be effective in animal model studies, phase III antitumor clinical trials have been initiated. This example shows that certain target genes of c-MYC may represent valuable therapeutic targets for interfering with endothelial cell proliferation during tumor vascularization and restenosis. Because of the widespread activation of c-MYC in cancer, c-MYC target genes may be useful candidates for inhibition of tumor cell proliferation in general.

We thank Bert Vogelstein, Ken Kinzler, and Axel Ullrich for their support, and Carlo Rago for expert technical assistance. We also thank Dirk Eick and John Sedivy for providing cell lines, Peggy Farnham and Julie Wells for technical advice, and Gregory J. Cost and members of the lab for comments.

- Oster, S. K., Ho, C. S. W., Soucie, E. L. & Penn, L. Z. (2002) *Adv. Cancer Res.* **84**, 81–154.
- Dang, C. V. (1999) *Mol. Cell. Biol.* **19**, 1–11.
- Eisenman, R. N. (2001) *Genes Dev.* **15**, 2023–2030.
- Grandori, C., Cowley, S. M., James, L. P. & Eisenman, R. N. (2000) *Annu. Rev. Cell. Dev. Biol.* **16**, 653–699.
- Grandori, C. & Eisenman, R. N. (1997) *Trends Biochem. Sci.* **22**, 177–181.
- Amati, B., Brooks, M. W., Levy, N., Littlewood, T. D., Evan, G. I. & Land, H. (1993) *Cell* **72**, 233–245.
- Frank, S. R., Schroeder, M., Fernandez, P., Taubert, S. & Amati, B. (2001) *Genes Dev.* **15**, 2069–2082.
- Bouchard, C., Dittrich, O., Kiermaier, A., Dohmann, K., Menkel, A., Eilers, M. & Lüscher, B. (2001) *Genes Dev.* **15**, 2042–2047.
- Guo, Q. M., Malek, R. L., Kim, S., Chiao, C., He, M., Ruffly, M., Sanka, K., Lee, N. H., Dang, C. V. & Liu, E. T. (2000) *Cancer Res.* **60**, 5922–5928.
- O'Hagan, R. C., Schreiber-Agus, N., Chen, K., David, G., Engelman, J. A., Schwab, R., Alland, L., Thomson, C., Ronning, D. R., Sacchetti, J. C., Meltzer, P. & DePinho, R. A. (2000) *Nat. Genet.* **24**, 113–119.
- Coller, H. A., Grandori, C., Tamayo, P., Colbert, T., Lander, E. S., Eisenman, R. N. & Golub, T. R. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 3260–3265.
- Schuhmacher, M., Kohlhuber, F., Holzel, M., Kaiser, C., Burtscher, H., Jarsch, M., Bornkamm, G. W., Laux, G., Polack, A., Weidle, U. H. & Eick, D. (2001) *Nucleic Acids Res.* **29**, 397–406.
- Schuldiner, O. & Benvenisty, N. (2001) *Oncogene* **20**, 4984–4994.
- Mateyak, M. K., Obaya, A. J., Adachi, S. & Sedivy, J. (1997) *Cell Growth Differ.* **8**, 1039–1048.
- Hermeking, H., Rago, C., Schuhmacher, M., Li, Q., Barrett, J. F., Obaya, A. J., O'Connell, B. C., Mateyak, M. K., Tam, W., Kohlhuber, F., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97**, 2229–2234.
- Veclulescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. W. (1995) *Science* **270**, 484–487.
- Pfaffl, M. W. (2001) *Nucleic Acids Res.* **29**, 2003–2007.
- Boyd, K., Wells, J., Gutman, J., Bartley, S. M. & Farnham, P. J. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 13887–13892.
- Berns, K., Hijmans, E. M. & Bernards, R. (1997) *Oncogene* **11**, 1347–1356.
- Bush, A., Mateyak, M., Dugan, K., Obaya, A., Adachi, S., Sedivy, J. & Cole, M. (1998) *Genes Dev.* **12**, 3797–3802.
- Mitsui, K., Nakanishi, M., Ohtsuka, S., Norwood, T. H., Okabayashi, K., Miyamoto, C., Tanaka, K., Yoshimura, A. & Ohtsubo, M. (1999) *Biochem. Biophys. Res. Commun.* **261**, 115–122.
- Hurlin, P. J., Queva, C. & Eisenman, R. N. (1997) *Genes Dev.* **11**, 44–58.
- Laroya, G., Cuesta, R., Brewer, G. & Schneider, R. J. (1999) *Science* **284**, 499–502.
- Wood, M. A., McMahon, S. B. & Cole, M. D. (2000) *Mol. Cell* **5**, 321–330.
- Yang, X.-J., Ogrzyzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. (1996) *Nature (London)* **382**, 319–324.
- Shim, H., Chun, Y. S., Lewis, B. C. & Dang, C. V. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 1511–1516.
- Vaughn, J. P., Davis, P. L., Jarboe, M. D., Huper, G., Evans, A. C., Wiseman, R. W., Berchuck, A., Iglehart, J. D., Futreal, P. A. & Marks, J. R. (1996) *Cell Growth Differ.* **7**, 711–715.
- Leach, F. S., Polyak, K., Burrell, M., Johnson, K. A., Hill, D., Dunlop, M. G., Wyllie, A. H., Peltomaki, P., de la Chapelle, A., Hamilton, S. R., Kinzler, K. W. & Vogelstein, B. (1996) *Cancer Res.* **56**, 235–240.
- Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J. & Qin, J. (2000) *Genes Dev.* **14**, 927–939.
- Reisman, D. & Loding, W. T. (1998) *Semin. Cancer Biol.* **8**, 317–324.
- Hermeking, H. & Eick, D. (1994) *Science* **265**, 2091–2093.
- Myung, K., Datta, A. & Kolodner, R. D. (2001) *Cell* **104**, 397–408.
- Soucie, E. L., Annis, M. G., Sedivy, J., Filmus, J., Leber, B., Andrews, D. W. & Penn, L. Z. (2001) *Mol. Cell. Biol.* **21**, 4725–4736.
- McClung, J. K., Jupe, E. R., Liu, X. T. & Dell'Orco, R. T. (1995) *Exp. Gerontol.* **30**, 99–124.
- Nijtmans, L. G., de Jong, L., Artal Sanz, M., Coates, P. J., Berden, J. A., Back, J. W., Muijsers, A. O., van der Spek, H. & Grivell, L. A. (2000) *EMBO J.* **19**, 2444–2451.
- Jupe, E. R., Liu, X. T., Kiehlauch, J. L., McClung, J. K. & Dell'Orco, R. T. (1996) *Exp. Cell Res.* **224**, 128–135.
- Rosenwald, I. B., Rhoads, D. B., Callanan, L. D., Isselbacher, K. J. & Schmidt, E. V. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6175–6178.
- Wu, S., Gupta, S., Chatterjee, N., Hileman, R. E., Kinzy, T. G., Denslow, N. D., Merrick, W. C., Chakrabarti, D., Osterman, J. C. & Gupta, N. K. (1993) *J. Biol. Chem.* **268**, 10796–10801.
- Sin, N., Meng, L., Wang, M. Q. W., Wen, J. J., Bornmann, W. G. & Crews, C. M. (1997) *Proc. Natl. Acad. Sci.* **94**, 6099–6103.