Oncogene (2001) 20, 2225–2234

www.nature.com/onc

DNA microarrays identification of primary and secondary target genes regulated by p53

Karuppiah Kannan[1], Ninette Amariglio[2], Gideon Rechavi[2], Jasmine Jakob-Hirsch[2], Itai Kela[3], Naftali Kaminski[4], Gad Getz[3], Eytan Domany[3] and David Givol*[,1]

[1]*Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel;* [2]*Department of Pediatric Hematology-Oncology, The Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Israel;* [3]*Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel;* [4]*Functional Genomics Unit, The Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Israel*

The transcriptional program regulated by the tumor suppressor p53 was analysed using oligonucleotide microarrays. A human lung cancer cell line that expresses the temperature sensitive murine p53 was utilized to quantitate mRNA levels of various genes at different time points after shifting the temperature to 32°C. Inhibition of protein synthesis by cycloheximide (CHX) was used to distinguish between primary and secondary target genes regulated by p53. In the absence of CHX, 259 and 125 genes were up or down-regulated respectively; only 38 and 24 of these genes were up and down-regulated by p53 also in the presence of CHX and are considered primary targets in this cell line. Cluster analysis of these data using the super paramagnetic clustering (SPC) algorithm demonstrate that the primary genes can be distinguished as a single cluster among a large pool of p53 regulated genes. This procedure identified additional genes that co-cluster with the primary targets and can also be classified as such genes. In addition to cell cycle (e.g. *p21*, *TGF-β*, *Cyclin E*) and apoptosis (e.g. *Fas*, *Bak*, *IAP*) related genes, the primary targets of p53 include genes involved in many aspects of cell function, including cell adhesion (e.g. *Thymosin*, *Smoothelin*), signaling (e.g. *H-Ras*, *Diacylglycerol kinase*), transcription (e.g. *ATF3*, *LISCH7*), neuronal growth (e.g. *Ninjurin*, *NSCL2*) and DNA repair (e.g. *BTG2*, *DDB2*). The results suggest that p53 activates concerted opposing signals and exerts its effect through a diverse network of transcriptional changes that collectively alter the cell phenotype in response to stress. *Oncogene* (2001) **20,** 2225–2234.

Keywords: DNA microarray; cycloheximide; ts-p53; target genes; expression profile; clustering

Introduction

The function of p53 as tumor suppressor is mainly due to its activity as a transcription factor that activates many genes in response to various types of stress (El-Deiry *et al*., 1992). This may be the basis for p53 protection of cells against DNA damage and various stress conditions that lead usually to growth arrest or apoptosis (Levine, 1997). Mice with p53 which was rendered transcriptionally inactive by mutations at codon 25 and 26 are predisposed to tumors similar to p53 deficient mice (Jimenez *et al*., 2000), demonstrating the importance of transcriptional activity for p53 function. The heterogeneity of gene transcription profile in response to p53 was demonstrated recently using transcriptionally regulated p53 expressed in colon cancer cell line (Zhao *et al*., 2000; Yu *et al*., 1999). This heterogeneity could result partly from secondary effect and from dependence on other transcription factors. In order to overcome part of this heterogeneity and to distinguish primary targets of p53, we used the temperature sensitive p53 (denoted Val135) (Michalovitz *et al*., 1990) expressed in the human lung cancer cell line H1299 to analyse the transcriptional programs induced by p53. Shifting the temperature to 32°C causes Val135 to assume wild-type p53 conformation and induce target genes. This conformational change does not require protein synthesis and allows for the analysis of p53 induced genes in the presence of protein synthesis inhibitor to prevent secondary effects brought about by the activated genes.

Here we analysed the profile of gene expression regulated by p53 at 32°C in the presence and absence of cycloheximide (CHX) using DNA microarrays containing ~7000 probes for human genes (Affymetrix, Santa Clara, USA). Less than 20% of the genes regulated by p53 were also regulated in the presence of CHX and therefore are defined as primary targets of p53. We were able to distinguish these genes as a cluster of primary targets among a large pool of p53 regulated genes by subjecting the expression data to clustering analysis using Super Paramagnetic Clustering (SPC) algorithm (Blatt *et al*., 1996). The results indicate that in addition to genes involved in cell cycle

*Correspondence: D Givol
Received 7 November 2000; revised 21 January 2001; accepted 29 January 2001

control and apoptosis, p53 regulates directly a plethora of genes involved in many aspects of cellular functions such as DNA repair, cytoskeleton, extracellular matrix and signal transduction. This suggests that in addition to its effect on cell cycle, p53 may exert some of its tumor suppressive effects through a diverse network of transcriptional changes that alter cellular phenotype and behavior in response to stress.

Results

Effect of cycloheximide on p53 target genes

Analysis of mRNA from H1299Val135 cells and H1299 control cells by hybridization to DNA microarrays showed that after 12 h at 32°C, 212 and 77 genes were up and down-regulated respectively by more than 2.5-fold. It is likely that some of these genes may be regulated because of secondary effects induced by the primary targets of p53. To identify the primary targets, we analysed the effect of p53 in the presence and absence of cycloheximide (CHX) at 32°C in both H1299Val135 and H1299 control cells. The inhibition of protein synthesis by CHX presumably prevents most of the seconday gene regulation (O'Hagan *et al*., 2000; Coller *et al*., 2000) that is not transactivated directly by p53. The results showed that in the presence of CHX, p53 remained stable for at least 12 h (Figure 1A) and induce significantly the mRNA of *p21waf*, a major target for p53 (Figure 1B). Evidently protein synthesis was indeed shut down as no p21waf or hDM2 (human MDM2) proteins were detected, in contrast to their presence in the experiment without CHX (Figure 1C,D). It was shown previously that at 32°C p53Val135 is relocalized to the nucleus (Ginsberg *et al*., 1991) and that it may therefore be protected from degradation in the presence of CHX due to a lack of nuclear exclusion and the absence of hDM2 synthesis (Haupt *et al*., 1997; Kubbutat *et al*., 1997). Hence, direct target genes of p53 will be transcriptionally activated, but because of inhibition of protein synthesis, the p53-induced proteins will not be synthesized and will not induce secondary or indirect p53 target genes.

p53 primary target genes identified in the presence of cycloheximide

RNA from various time points (between 2 to 24 h) was isolated, labeled and hybridized to oligonucleotide microarrays (Genechip Hugene FL array, Affymetrix, Santa Clara, USA) which contains probes for $\sim$7000 human genes. Those genes that are transcriptionally regulated by p53 under both conditions, i.e. with and without CHX, are likely to be primary targets for p53-mediated transcription. In order to eliminate noisy data in the analysis of the hybridization experiments, we applied a very stringent filter; we selected genes that showed more than 2.5-fold induction or repression (over their controls) at three or more time points in the presence or absence of CHX. Two hundred and fifty-nine upregulated genes passed this filter for the experiment in the absence of CHX whereas 131 genes passed this filter for the experiment with CHX; 38 genes were found to be common to these two groups (see the Venn diagram in Figure 2A and Table 1). Denote by *G(38)* the group of 38 genes that have been identified by our analysis as possible primary upregulated targets of p53. It is important to note that $n=38$, the number of these genes, exceeds significantly the number n_r, that would have been obtained had we applied the same filtering procedure to random data. To estimate n_r, assume that for every gene in any single experiment, the over-expression level can exceed 2.5 with probability P, and that such over-expressions are independent random events. All together, nine measurements were taken for 7070 genes; by counting the number of occurrences of 2.5-fold over-expression, we estimate $P\approx 0.053$. The probability that a particular gene will be over-expressed at least three times at a level above 2.5 in the experiments without CHX is $5.7\cdot10^{-4}$; the same figure for the experiments with CHX is $1.4\cdot10^{-3}$; and for passing the filter in both

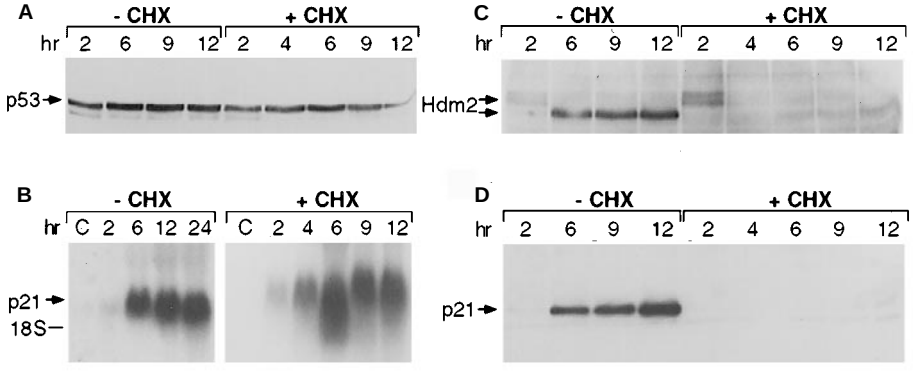

Figure 1 Effect of cycloheximide (CHX) on H1299Val135 cells, demonstrating synthesis of mRNA and lack of protein synthesis. (**A**) Western-blot analysis of p53Val135 in H1299 cells in the presence and absence of CHX. (**B**) Northern-blot analysis of *p21Waf* in lysates from H1299Val135 cells grown in the presence and absence of CHX. (**C**) Western-blot analysis of Hdm2, (**D**) Western-blot analysis of p21waf in lysates from H1299Val135 cells grown in the presence and absence of CHX. **C**, control cells of H1299 devoid of p53. hr, indicates time after shifting the temperature to 32°C

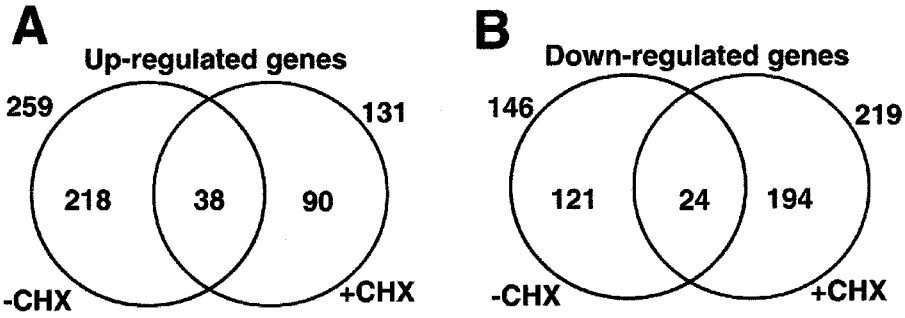

Figure 2 Venn diagram of the number of genes that were regulated by p53 in the presence and absence of CHX. (**A**) Only genes that showed at least 2.5-fold up or (**B**) down-regulation in at least three of the time points in each experiment i.e. with or without CHX, were listed in this analysis. Note that only 38 of the up-regulated and 24 of the down-regulated genes were unaffected by CHX

experiments $8{\cdot}10^{-7}$. Multiplying these probabilities by the total number of genes tested, we get estimates for the numbers of genes that would have passed the filter and would have been assigned to the three groups. Next to these estimates, which represent the expected number of genes for random expression data, i.e. for a randomized expression matrix, we placed in parentheses the actual numbers, obtained by our analysis for the real data: with CHX: 10 (vs 131); without CHX: 4 (vs 259); and the number of genes expected to be in the overlap; $n_r = 5{\cdot}10^{-3}$ (vs 38). The large disparity between the numbers obtained under the assumption of a random process and the actual measured numbers proves beyond doubt the statistical significance of our findings.

The down-regulated genes that passed this filter were 146 and 219 respectively in the experiments without and with CHX and 24 genes were common to these groups (Figure 2B). The common genes were regulated by p53 irrespective of the presence or absence of CHX (Tables 1 and 2). The remaining 218 upregulated and 121 downregulated genes (Figure 2) may be indirect or secondary targets of p53 (see Table at our web site: http://www.weizmann.ac.il/home/ligivol/primary.html). In this study, we focus mainly on the analysis of the upregulated genes. Apparently CHX by itself also induced many genes probably due to removal of inhibitory signals for gene expression or changes in gene expression associated with clonal variability. Had we performed the analysis of p53 induced genes only in the presence of CHX, we would have detected 131 genes (Figure 2A) as primary targets, most of which may be unrelated to direct p53 regulation.

In order to analyse genes regulated by CHX in the absence of p53, H1299 cells were incubated at 32°C in the presence of CHX and RNA was collected at 2 and 6 h and analysed by hybridization to the DNA microarray. Three genes from the 38 primary upregulated genes (Table 1) were also upregulated by CHX alone. These are the *ATF3* (activated transcription factor 3), *Histone 2A* like protein and *Bak*. In a similar way the effect of temperature on H1299 cells was analysed by incubating H1299 cells at 32°C and collecting RNA for hybridization at 2 and 12 h. Three genes from the primary genes: *Lysyl oxidase*, *Diacylglycerol kinase* and *LISCH7* were found to be upregulated also after 12 h at 32°C. We keep these genes in the primary target list although they were upregulated by the control conditions (CHX or temperature). It is possible that these genes are transactivated by various stress conditions as well as by p53. For example, $p21^{waf1}$, a major primary target of p53 is also induced by a variety of external stimuli and stress conditions independent of P53 (Michieli *et al*., 1994).

Identifying primary p53 target genes by cluster analysis

A major problem we must now face is that the identities of the genes of *G(38)*, that were designated above as possible primary upregulated targets of p53, are determined by our stringent filtering criteria. Had we set our threshold at, say, observing twofold increased expression (instead of 2.5) at two time points (instead of three), the number of genes that passed the filter would have been larger. The values of our filtering parameters (2.5-fold change at three or more times, in both experiments) were chosen in an arbitrary fashion; it is important to assess the extent to which relaxation of our filtering criteria will add primary p53 target genes beyond the 38 candidates that were already found. In order to reduce the dependence of our results on the precise values of the arbitrarily chosen filtering parameters, we performed cluster analysis on the data, using more relaxed filtering criteria to select the genes to be included in the analysis.

The first cluster analysis was done on the 259 genes that passed our filter for the experiment without CHX. That is, in our filtering we relaxed completely the restriction on the expression levels measured in the experiment with CHX (but did use the results of these experiments in the cluster analysis). Each of the 259 genes is represented by its expression levels taken at nine different time points (in the two experiments combined). The data were normalized as follows. Denote by A_{ij} the $\log_2$ ratio of the expression of gene i (where $i=1,2,...,259$), measured at experiment (and time) j (with $j=1,2,...,9$), with respect to the control. For $j=1,...,4$ we divided A_{ij} by $[\sum_{j=1}^{4} A_{ij}^2]^{1/2}$ and for

Table 1 Primary target genes upregulated by p53

Accession no.					*Ratio of gene expression at specific time points*								
					−CHX				*+CHX*				
				h	*2*	*6*	*12*	*24*	*2*	*4*	*6*	*9*	*12*
	Apoptosis	G(47)	G(38)	G(9)									
X63717	Fas/APO-1 cell surface antigen	+	+		2.2	5.6	7.9	8.3	3.2	6.8	3.3	5.7	9.8
U82987	Bcl-2 binding component 3 (bbc3)	+	+		6.2	16.5	16.8	8.4	7.6	13.4	28.5	32.5	30.5
U00115	Bcl-6	+	+		2.5	3.3	5.8	6.5	3.7	4.7	6.1	6.0	6.8
U16811	Bak	+	+		4.0	7.0	4.4	6.8	1.0	2.3	4.5	4.7	7.4
	Cell cycle												
U09579	p21 WAF1	+	+		4.3	18.0	16.9	13.0	3.5	6.7	9.2	8.2	7.4
D90070	ATL derived PMA responsive peptide	+	+		4.0	4.5	2.8	4.1	1.8	3.5	3.4	4.3	6.1
M60974	GADD45	+			2.3	3.2	3.4	3.6	1.6	1.6	2.1	2.1	2.0
	DNA repair/replication												
U72649	BTG2	+	+		3.0	4.9	4.8	5.2	11.0	15.9	20.1	17.9	16.5
U18300	Damage-specific DNA binding protein	+	+		1.3	4.0	5.4	7.8	1.1	2.6	8.6	6.1	9.4
U90551	Histone 2A-like protein	+	+		2.6	2.6	2.5	3.5	1.7	3.1	4.5	8.8	16.7
M15796	PCNA	+			2.3	3.9	5.3	5.3	1.7	1.4	1.7	1.9	3.1
	Receptors/ECM												
X72012	Endoglin		+		21.7	15.2	7.9	6.8	3.4	1.0	14.0	6.9	1.0
U16306	Versican	+	+		1.0	5.2	11.9	16.6	1.0	2.0	2.7	6.8	13.3
M21904	Heavy chain 4F2	+			7.7	6.5	8.2	6.1	0.9	1.0	1.6	1.3	1.4
AF010193	SMAD7			+	3.6	1.9	3.9	3.1	1.9	1.8	1.5	1.5	1.6
	Growth factors/inhibitors												
AB000584	TGF-Beta Superfamily protein	+	+		3.5	38.6	67.2	50.1	11.5	16.0	38.4	33.6	32.4
M62402	IGFBP6	+	+		17.5	17.7	18.5	17.6	22.3	13.9	27.8	25.4	28.9
L42379	Quiescin/QSCN6	+			2.7	2.4	2.5	2.8	1.3	2.2	2.3	2.3	1.7
X97324	Adipophilin	+			5.2	6.3	7.1	10.0	1.5	1.9	1.7	1.9	3.8
U72263	Multiple exostoses type II protein	+			2.9	2.7	3.1	6.2	1.5	2.0	1.4	1.8	3.2
	Cytoskeleton/cell adhesion												
X13839	Vascular smooth muscle alpha-actin	+	+		6.5	14.6	44.3	98.3	6.4	9.0	18.9	18.4	26.9
Z49989	Smoothelin	+	+		3.7	4.3	3.9	3.5	1.7	2.2	4.6	4.1	3.4
X05608	Neurofilament subunit NF-L	+	+		3.0	7.7	16.6	32.0	5.0	8.8	9.6	16.8	32.2
D82345	NB Thymosin beta	+	+		2.4	3.4	4.2	5.4	1.6	2.5	3.8	2.8	3.9
X93510	LIM domain protein		+		4.5	5.5	4.3	5.6	4.3	2.4	9.3	3.8	1.0
	Metabolism												
U24389	Lysyl oxidase-like protein	+	+		16.7	17.5	23.0	22.9	5.5	4.2	9.7	8.3	6.9
L38668	UDP-Galactose 4 epimerase (GALE)	+	+		4.2	3.6	3.2	4.4	3.2	3.7	5.1	3.4	2.8
Y12556	cAMP activated Protein Kinase B	+	+		4.0	1.0	19.7	22.2	1.0	3.1	3.1	7.1	37.5
U05572	Lysosomal Mannosidase alpha B		+		9.2	2.0	7.0	15.9	6.5	5.6	13.9	13.0	7.2
Y09616	Carboxylesterase (liver)	+	+		2.3	4.1	5.9	9.1	1.9	2.5	3.0	3.4	3.6
U78735	ABC3			+	4.5	3.0	2.8	3.6	1.9	2.0	1.9	1.4	1.1
M20902	Apolipoprotein C-I (VLDL)	+			3.1	3.9	2.9	4.1	2.1	2.9	2.3	2.2	1.7
U20325	CART	+			2.8	3.9	3.0	3.3	1.3	1.7	2.1	1.3	1.9
M12625	Lecithin-cholesterol acyltransferase	+			1.4	2.9	3.0	3.0	0.8	1.4	3.6	2.6	2.4
D87292	Rhodanese			+	7.7	4.9	5.4	10.5	1.6	1.6	1.7	1.3	1.2
	Neuronal growth												
U35139	NECDIN related protein		+		3.2	1.0	3.0	15.7	5.8	4.9	2.7	4.8	5.5
M96740	NSCL-2 gene	+	+		1.0	4.2	10.1	12.6	2.6	2.0	4.8	7.0	11.0
U60062	FEZI-T gene	+	+		1.0	2.7	2.7	4.0	1.6	2.3	2.6	5.6	8.7
U72661	Ninjurin 1	+	+		1.4	3.7	7.6	10.4	1.9	3.1	6.0	6.3	7.8
U48437	Amyloid precursor-like protein	+	+		1.9	2.7	3.7	4.7	1.0	1.5	3.6	2.7	3.8
	Signal transduction												
J00277	c-Ha-ras 1	+	+		2.3	3.1	4.7	4.8	1.2	1.5	4.1	5.6	5.7
X77777	Intestinal VIPR related protein		+		14.7	7.5	11.9	9.8	1.0	2.7	19.9	4.7	1.0
X62535	Diacylglycerol Kinase (alpha)	+	+		4.4	12.0	11.2	38.8	1.2	3.4	6.0	7.1	10.1
U56998	Putative ser/thr protein kinase	+	+		6.3	3.6	1.0	12.5	2.1	2.2	4.8	6.4	4.5
L08835	DM Kinase	+			2.5	3.4	3.3	3.7	0.9	1.2	1.8	1.5	1.5
L42176	DRAL-FHL2			+	5.4	3.8	3.8	6.5	2.0	2.0	1.8	1.5	1.7
	Transcription												
L19871	Activating transcription factor 3	+	+		1.4	4.4	5.9	5.2	4.2	7.2	18.9	8.8	9.9
U38315	ZNF127-Xp		+	+	4.4	5.1	9.2	9.5	6.5	6.4	3.9	7.2	3.1
AD000684	LISCH7	+	+		5.2	6.1	6.1	7.4	2.7	1.0	2.1	3.1	3.8
M29580	Zinc finger protein 7	+			1.2	2.6	3.6	4.7	1.2	1.2	1.2	1.3	1.2
U90913	Tip-1	+			3.4	2.7	3.1	3.7	1.5	2.1	2.3	1.9	2.3
HG3494	Nuclear factor NF-116			+	3.3	2.4	2.9	2.7	2.9	2.8	2.4	2.3	1.1
	Other												
U10099	POM-ZP3	+	+		3.2	4.2	4.7	4.8	1.8	2.0	4.0	3.5	3.3
D87434	KIAAA0247	+	+		1.0	4.5	4.7	6.6	3.5	7.5	10.3	15.4	18.9
U33147	Mammaglobin 1		+	+	3.5	1.0	4.5	4.1	4.5	4.2	2.9	4.6	1.0
J05016	Disulfide isomerase related protein	+			2.6	3.8	6.0	8.7	2.1	1.4	1.8	1.8	2.3

Continued

Table 1 (*Continued*)

Accession no.				Ratio of gene expression at specific time points								
				−CHX					+CHX			
				h	*2*	*6*	*12*	*24*	*2*	*4*	*6*	*9*	*12*
U81556	OS4		+	2.7	2.6	2.4	3.0	1.6	2.2	1.8	1.4	1.3
S58544	Infertility-related sperm protein	+		3.0	3.9	7.6	3.1	1.6	2.0	1.7	3.0	4.3
D63481	KIAA0147	+		1.8	2.6	3.1	2.6	1.2	1.2	1.5	2.1	1.7
Z35093	SURF-1		+	2.2	2.6	2.6	3.4	1.9	2.0	2.0	1.6	1.4
U94747	WD repeat protein HAN11	+		2.3	3.1	3.4	3.4	0.8	1.3	3.0	1.4	3.1

Primary targets are defined as genes that changed their expression over 2.5-fold in at least three time points in the presence and absence of cycloheximide (CHX). The ratio of gene expression was determined by the expression level at each time point divided by that of H1299 contol cells at 2 h at 32°C. The list denoted *G(38)* consists of the genes in the overlap in Figure 2. The list denoted *G(47)* constitutes the set of primary p53 targets generated from the cluster analysis in Figure 3 (see text). The list denoted *G(9)* contains the genes in cluster **a** in Figure 3A

Table 2 Primary target genes downregulated by p53

Accession no.		Ratio of gene expression at specific time points								
		−CHX				+CHX				
		h	*2*	*6*	*12*	*24*	*2*	*4*	*6*	*9*	*12*
	Apoptosis										
U45878	Inhibitor of Apoptosis protein 1 (MIHB)	−4.0	−8.0	−3.9	−3.5	−1.9	−1.8	−2.8	−3.1	−2.6
U37546	Inhibitor of Apoptosis protein 2(MIHC)	−4.2	−10.7	−14.9	−15.8	−2.8	−2.8	−3.0	−3.7	−3.2
	Angiogenesis									
L22548	Collagen type XVIII alpha 1	−8.4	−8.4	−8.4	−7.6	−2.9	−2.9	−2.9	−2.9	−2.9
	Cell Cycle									
L78833	BRCA-1	−2.9	−3.1	−3.1	−3.1	−4.1	−4.5	−4.5	−2.3	−4.5
M74093	Cyclin E	−0.8	−2.9	−4.0	−3.2	−1.4	−2.6	−6.9	−4.5	−3.7
U77949	Cdc6-related protein	−2.2	−3.3	−5.9	−6.4	−4.9	−4.4	−9.1	−1.6	−2.1
M72885	GOS2	−33.4	−137.0	−172.0	−143.0	−9.3	−15.1	−11.5	−21.0	−107.0
	DNA repair/replication									
M96684	Purine rich element binding protein A	−4.2	−9.3	−3.0	−2.5	−1.5	−3.5	−5.3	−5.3	−2.4
	Receptors/ECM									
U31201	Laminin gamma 2 chain	−8.5	−8.5	−8.5	−8.5	−2.0	−2.7	−7.1	−2.4	−3.9
U90716	Caxsackie virus and adenovirus receptor	−5.4	−3.2	−4.3	−2.5	−2.8	−3.0	−4.2	−2.3	−1.1
U60975	Receptor gp250 precursor	−2.9	−5.6	−1.9	−3.1	−1.5	−1.3	−6.3	−3.4	−2.8
J04970	Carboxypeptidase M	−6.2	−6.2	−5.8	−6.2	−6.9	−6.9	−6.9	−6.1	−6.9
U17566	Folate transporter	−0.8	−2.8	−10.3	−10.3	−3.0	−8.8	−10.3	−2.6	−2.5
	Metabolism									
U00238	Glutamine PRPP amidotransferase	−1.6	−3.0	−3.5	−3.9	−1.6	−2.0	−7.2	−9.2	−4.5
	Neuronal growth									
D11428	Peripheral myelin protein 22	−0.3	−2.7	−2.7	−2.7	−3.5	−2.9	−10.9	−9.2	−10.7
S78296	Neurofilament-66	−5.8	−6.9	−5.1	−6.1	−2.6	−2.6	−6.1	−2.8	−2.2
U79255	Amyloid B-precursor binding protein	−0.6	−8.3	−8.3	−8.3	−5.0	−5.0	−2.5	−5.0	−5.0
U73960	ADP-ribosylation factor-like protein 4	−4.7	−3.4	−6.4	−9.8	−1.4	−2.2	−5.9	−5.9	−4.2
	Transcription									
U80017	Basic transcription factor 2 p44	−2.3	−2.6	−2.7	−3.1	−1.5	−2.2	−8.0	−4.0	−3.3
X16706	Fra-2	−1.9	−5.4	−5.4	−5.4	−3.9	−2.8	−2.0	−1.9	−3.1
	Other									
HG2855	HSP 70	−1.9	−2.6	−2.9	−3.4	−1.0	−1.3	−3.1	−6.9	−20.2
L19183	MAC30	−3.2	−2.9	−3.3	−4.0	−2.0	−2.7	−3.8	−2.7	−3.4
AB000467	RES4-25	−2.4	−5.0	−5.0	−5.0	−2.1	−1.1	−6.3	−2.5	−6.3
U79273	Clone 23933	−4.9	−16.3	−2.5	−3.6	−4.2	−4.2	−4.2	−4.2	−2.7

The 24 genes that were downregulated in either the presence of absence of CHX

$j=5,...,9$ by $[\sum_{j=5}^{9} A_{ij}^2]^{1/2}$; the resulting 9-component vector represents gene *i*. The 259 genes were clustered by SPC (Blatt *et al*., 1996; Getz *et al*., 2000) (see below). Genes with similar expression profiles (over the time courses of both experiments) are represented by two nearby vectors and are placed in the same cluster. This cluster analysis answers directly two questions: (1) Do all, or a majority of the genes of *G(38)* cluster together?; and (2) What other genes cluster together with these possible primary targets? If the answer to the first question is positive, we can identify an expression profile which is characteristic of primaries, and identify those genes that share this profile and cluster together with the genes of *G(38)* as good candidates for being primaries (even though they did not pass the original stringent filtering process). Furthermore, if we find that some of the members of *G(38)* have significantly different expression kinetics than this characteristic profile, these genes should possibly be removed from the list of primaries.

Regarding the first question – the genes of *G(38)* are special in that their expression increased at least 2.5-fold at three or more time points of both

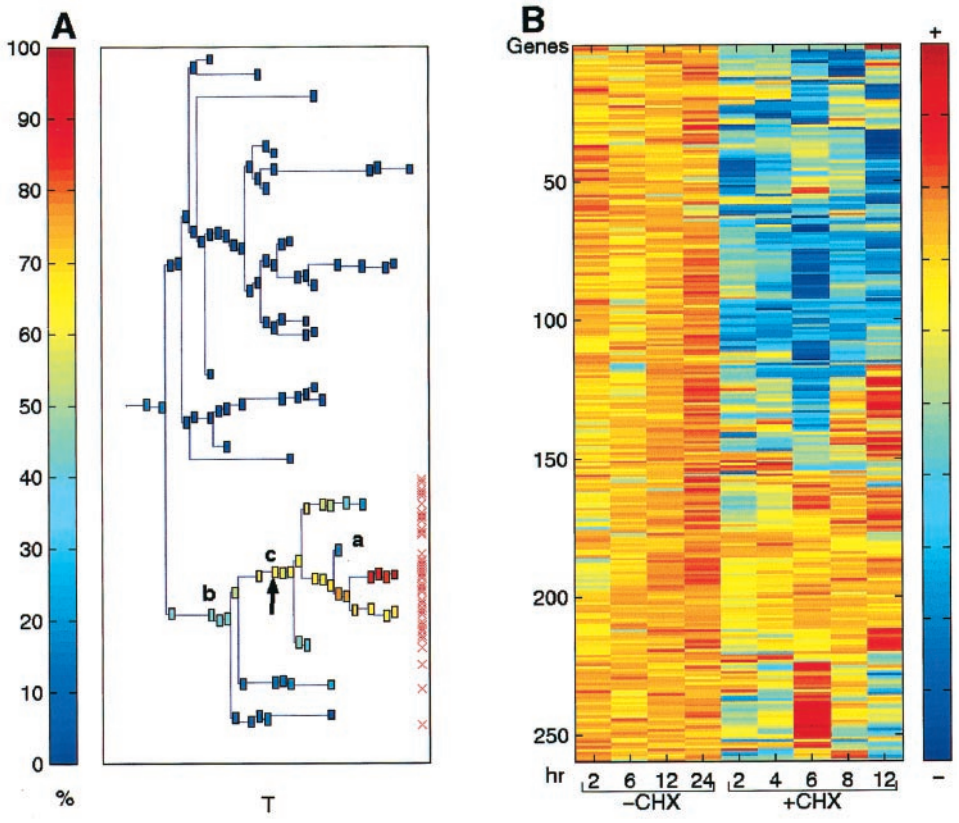

Figure 3 Clustering results using super-paramagnetic clustering (SPC) for the 259 genes that were upregulated at three time points or more upon activation of p53 in the experiment without cycloheximide (CHX). (**A**) The dendrogram of the genes that include clusters of size 4 and larger. Each cluster is represented by a box colored according to the per cent of primary target genes (38 genes, see Table 1) contained in the cluster. The distribution of the 38 primary target genes is marked by red crosses at the right. (**B**) The normalized log ratio of the nine experiments (four without CHX and five with CHX) are plotted. The genes are ordered according to the dendrogram on the left. The color represents induction (red) or repression (blue). T, a parameter of the SPC algorithm that controls the resolution at which the cluster is found. %, per cent of primary target genes in the cluster. The cluster marked by an arrow (**c**) contains 87% of the 38 primary genes. The cluster marked by **b** contains all the 38 primary genes and the cluster marked by **a** contains the nine genes that show different kinetics (Table 1 and Figure 4)

experiments. Hence their representative 9-component vectors are likely to be close – but some may also be uncorrelated. For example, a gene whose expression decreases with time will have low correlation with one that increases. This situation may happen even if both genes passed the filtering criteria.

The results of our cluster analysis are summarized in the dendrogram of Figure 3A. The parameter T on the horizontal axis controls the resolution at which the data are viewed. At $T=0$ all 259 genes are in a single cluster; as T increases, large groups split into smaller ones. The boxes indicate clusters that contain more than four genes. Each box is colored according to its 'purity' – the percentage of members of $G(38)$ among the genes contained in the corresponding cluster.

When we reorder the genes according to their position in the dendrogram, i.e. rearrange the rows of the expression data matrix according to the order imposed by the clustering process, the color-coded-matrix of Figure 3B is obtained.

Next, we marked the positions of the members of $G(38)$ by red crosses. All 38 are in the low-level cluster

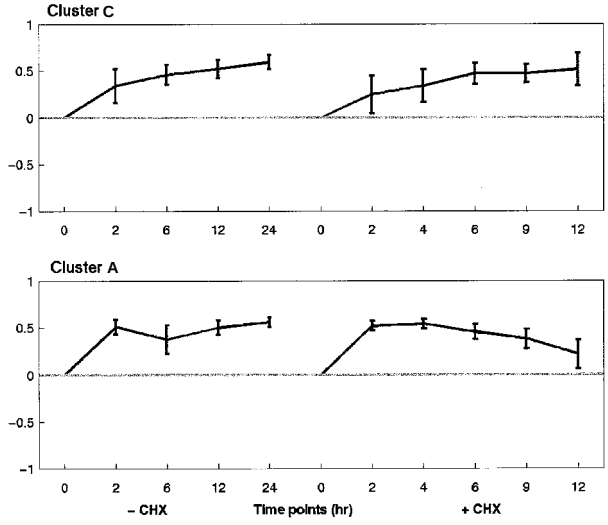

Figure 4 Average expression profiles of genes in clusters c and a of Figure 3A. The expression profile of each gene in the cluster was normalized as described in the text

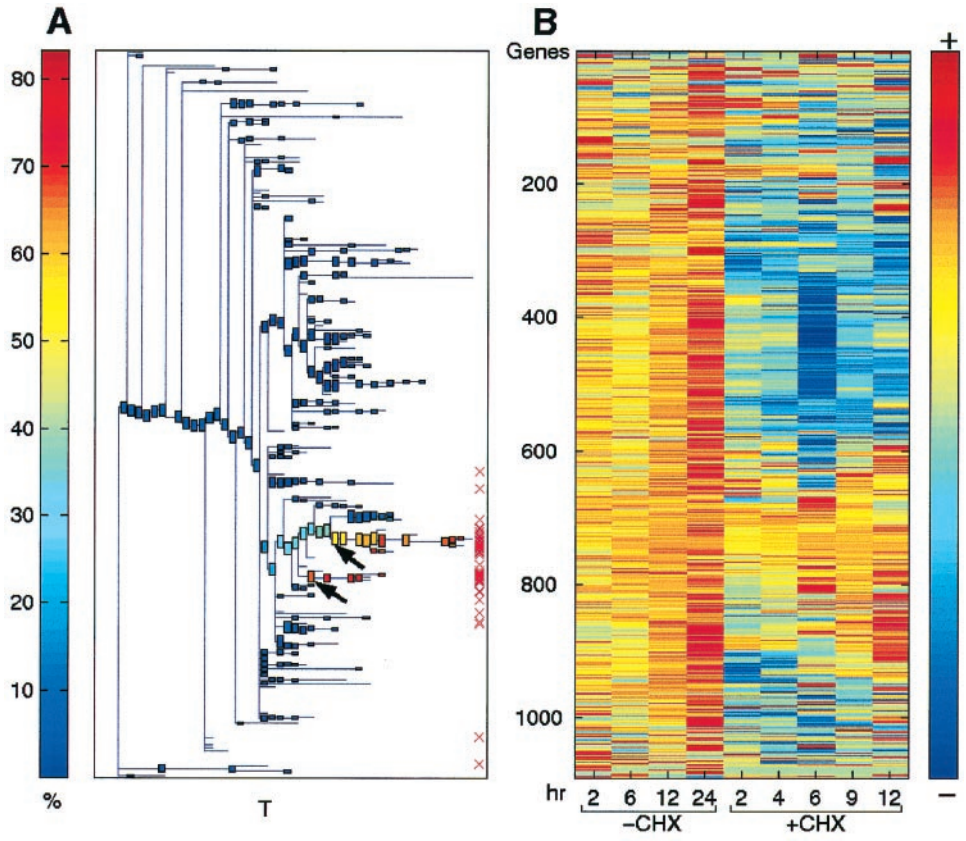

Figure 5 Clustering results using superparamagnetic clustering (SPC) for 1090 genes that were upregulated in the experiment without cycloheximide. In this analysis a relaxed filtering condition was used and all the genes that were upregulated 2.5-fold at least once (1090) in the experiment without CHX were included. (**A**) Dendrogram of the genes including clusters of size 5 and larger. The distribution of the 47 primary genes is marked by red crosses at the right. (**B**) Normalized log ratio of the nine experiments are plotted. The color represents induction (red) or repression (blue). Other details as in Figure 3. Note the primary gene-containing cluster resolves into two distinct clusters (marked by arrows) by splitting away the non-primary gene containing clusters

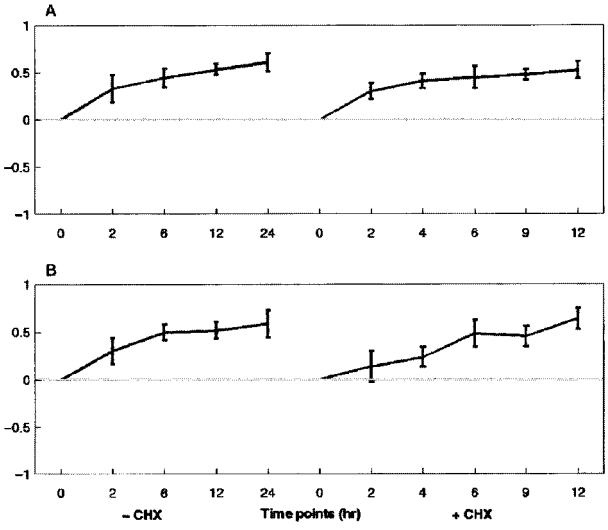

Figure 6 Average expression profiles of genes in clusters of Figure 5. (**A**) genes from the upper arrow cluster. (**B**) genes from the lower arrow cluster (Figure 5A). The expression profile of each gene was normalized as described in the text

denoted by *b* which, however, contains also 58 additional genes. This cluster branches and breaks into sub-clusters which have higher percentages of genes from *G(38)*. In order to identify a characteristic primary expression profile we want to work with a cluster which has many members of *G(38)* (for better statistics) and also has a high percentage of them. These two requirements conflict; as we move up on the dendrogram, the clusters become purer, but also decrease in size. Hence we decided to start with the cluster *c* marked by the arrow as our working point. It contains 33 of the *G(38)* genes and, in addition, 23 genes that did not pass the original filter, but have expression profiles that are similar to those of *G(38)*.

The average expression kinetics of the genes of this cluster in the two experiments is shown in Figure 4. Note the fairly similar kinetics with and without CHX, with the expression level increasing monotonously with time. All but nine genes of *c* share these features of the expression kinetics. The nine which differ appear on the dendrogram in the vicinity of the cluster denoted by *a*; two of these belong to *G(38)*. The average expression kinetics of these nine genes is shown in

Figure 4 (bottom panel); it clearly differs from that of Figure 4 (top panel), and their kinetics with and without CHX are different. Hence we decided to discard these nine genes from our list of designated primaries. The average expression kinetics of the remaining 47 genes is very similar to Figure 4, with reduced scatter (error bars). The group of these 47 genes, denoted *G(47)*, constitutes our final set of proposed primary p53 targets (Table 1).

This analysis identified a characteristic kinetic profile of primary p53 targets in either the presence or absence of CHX; the group of genes that share this profile contains 31 out of our 38 original candidates and 16 additional ones, that happened to fail our stringent filtering criteria. The various genes are listed and the groups to which they belong are properly identified in Table 1.

Some of the known p53 target genes such as *Gadd45* (Zhan *et al*., 1998) and *PCNA* (Shivakumar *et al*., 1995), were now included as primary targets in addition to the original 38 primary target genes, indicating that this is a sensible way to 'fish' for further potential primary targets. Most of these added genes exhibited $\sim$twofold induction at several of the time points both in the presence or absence of CHX experiments and were previously not known to be p53 target genes.

Next, we have put the stability of our identification of primary p53 targets against changing the procedure and parameters of selection to an extremely demanding test. We performed a similar cluster analysis on a much larger set of genes, including now all those that were upregulated at least twofold, at (at least) a single time point of the experiment without CHX. This very relaxed criterion selected 1090 genes to cluster; that is, as compared to the previous clustering analysis, the number of genes was increased fourfold, including now extremely noisy expression data. The results obtained when these 1090 genes were clustered are presented in Figure 5. To our satisfaction, we found that 3/4 of our 47 proposed primary p53 targets (that were identified above), belong to two very stable gene clusters, denoted by arrows on Figure 5A. The left of the two contains 24 of the genes of *G(47)* and 20 new genes, whereas the right one contains 11 from *G(47)* and seven new associated genes. The average expression kinetics of the two clusters is shown in Figure 6.

It is important to understand that adding all these extra noisy genes to the set of 259 that were analysed before could well have resulted in a total loss of the signal that was identified above. The fact that the two clusters that are rich in previously identified primaries indeed contain 35 out of the 47 is gratifying and indicates the stability of our method.

Finally, we turn to discuss our choice of the clustering algorithm to be used (SPC) (Blatt *et al*., 1996; Getz *et al*., 2000). The optimal algorithm for analysis of gene expression data should have the following properties: the number of clusters should be determined by the algorithm itself and not externally prescribed (as is done for SOM and K-means) (Sherlock, 2000; Young, 2000); stability against noise; generating a hierarchy (dendrogram) and providing a mechanism to identify in it robust, stable clusters; ability to identify a dense set of points, that form a cloud of an irregular, non-spherical shape, as a cluster. SPC, a hierarchical clustering method recently introduced by Blatt *et al*. (1996) is the algorithm that best fits these requirements.

Discussion

In this study we have used the temperature sensitive p53 expressed in the lung cancer cell line H1299 to analyse p53-regulated genes utilizing oligonucleotide microarray containing 7070 probes. We employed inhibition of protein synthesis by CHX in an effort to limit the p53-regulated genes to primary targets by exclusion of possible secondary effect by newly synthesized proteins. By measuring gene expression at four or five time points we clearly show that a group of genes show consistent p53-dependent regulation in either presence or absence of CHX. This group consists of less than 20% of the genes regulated by p53 in the absence of CHX, and defined here as primary targets. The criteria used to group these genes (2.5-fold change in at least three time points) is somewhat arbitrary and may discriminate against genes which show consistent changes but did not reach the 2.5-fold change in all the time points analysed. We therefore used a clustering method which allows the grouping of genes based on the kinetics of their expression and showed that most of the primary genes group together in one or two clusters. These clusters contain additional genes which show similar kinetic behaviors in their change of expression even though their level of expression did not search the stringent criteria. This group contains some of the established targets of p53 (e.g. *Gadd45* and *PCNA*) and demonstrate the usefulness of this clustering method, since it overcame small experimental variation in measurement of hybridization intensity and relies on the pattern of expression in the presence or absence of CHX.

In order to compare our results with known p53 targets we used the list compiled by El-Deiry (1998). Some of the genes in this list are not present on the chip (*Bax*, *Killer/DR5*, *PAG608*, *14-3-3σ* and *B99*). From the rest of the p53 targets on this list *p21waf*, *Fas/Apo*, *Gadd45* and *PCNA* were detected as primary target genes (Table 1). *MDM2* (a known p53 target) expression was increased more than 2.5-fold after 4 h in the presence of CHX but reached 2.67-fold expression only after 24 h without CHX. *CyclinG* expression level reached 2.4-fold only after 12 h in CHX and 2.1-fold after 12 h without CHX. *Cathepsin D*, *Thrombospondin*, *IGFBP5* and *PI3* kinase were not over pressed in this cell line in our experiment. This type of heterogeneity of expression indicates that not all p53-targets are activated in all conditions and a similar conclusion was pointed out also by other studies (Yu *et al*., 1999; Zhao *et al*., 2000). We also tried to analyse for the presence of p53 consensus

targets in several of the genes listed in Table 1. This was performed by searching for the sequence of the two consensus decamers RRRCWWGYYY, separated by a gap of 0–13 nucleotides, in the sequence which is 2 kbp upstream and downstream to the RNA start. In our list of primary targets, seven genes were shown previously to have consensus sequence and their promoters respond to p53; 15 genes were found to have consensus p53 targets with up to two mismatches and five genes did not have the genomic sequence available yet (data not shown). We feel that database search that is not supported by experimental analysis of individual genes is not very informative at present.

Our results from the analysis of p53 regulated primary target genes show that p53 directly upregulates proapoptotic and cell cycle inhibitors (e.g. *Bak* and *p21waf*) (see Table 1) and downregulates antiapoptotic and cell cycle genes (e.g. Inhibitor of Apoptosis and *Cyclin E*) (see Table 2). Thus p53 seems to activate concerted opposing signals to control cell proliferation and apoptosis. Although the identified primary genes cluster together, indicating the similarity in their expression kinetics, functional classification of these genes revealed that they are very different in their cellular activities (Table 1). The genes involved in cell cycle, growth arrest, apoptosis and perhaps DNA repair seem to be activated in most cell types analysed (e.g. *p21waf*, *Fas/ApoI*, *bcl2* binding protein) (Zhao *et al*., 2000; Komarova *et al*., 1998). On the other hand, genes involved in many other cellular functions like cytoskeleton, extracellular matrix (ECM), growth factors and their receptors and signal transduction may be more specific to the cell type analysed and this may be one of the causes for different lists of p53 regulated genes in various cells lines as was pointed out previously (Zhao *et al*., 2000; Yu *et al*., 1999). For example, in H1299 cells we found activation of several genes known to be involved in neuronal growth (Table 1) which may be unique to this cell type.

Several reasons may account for the heterogeneity of p53 target genes. It is possible that some of the targets require additional factors as coactivators. For example the transactivation of the mismatch repair gene *MSH* requires both p53 and c-*jun* or UV irradiation (Scherer *et al*., 2000). An additional factor may be the level of p53 activation by phosphorylation and other post-translation modification. Different transactivation properties may be related to a different level of p53 modification (Vousden, 2000; Oda *et al*., 2000).

We have noted before the induction of TGFβ by p53 (Kannan *et al*., 2000) and another study confirmed that *TGF-β* is a primary target of p53 and demonstrated its growth suppressor activity (Tan *et al*., 2000). This was also evident in our current experiments and may potentially suggest an additional paracrine mechanism by which p53 exerts its tumor suppressive effects on neighboring tumor cells. Other genes in our list of primary genes like endoglin (a coreceptor of TGF-β) and versican may also be part of the *TGF-β* activating system (Massague *et al*., 2000). Interestingly, the tumor suppressor activity of TGF-β may be related mainly to its function as an inducer of p15 (*ink4b*) expression (Hannon and Beach, 1994). The p15 is an inhibitor of cdk4 and cdk6 and its overexpression may lead to G1 arrest (Massague *et al*., 2000), thus TGF-β may behave as a tumor suppressor (Tang *et al*., 1998). The induction of *p21waf* by p53 and *p15* by TGF-β which itself is a target of p53 may provide an efficient control on the cell cycle. In the H1299 cell line, TGF-β, IGF binding protein 6, thymosin beta, diacylglycerol kinase alpha and neuronal growth related genes were substantially induced and were among the primary targets suggesting that these genes may also be relevant to the tumor suppressive phenotype induced by p53 as are the classical cell cycle related p53 inducible genes. Further analysis and cataloguing of p53 primary target genes from various cell lines may explain how p53 orchestrates diverse cellular signals as a tumor suppressor and could possibly lead to the identification of potential therapeutic molecular targets for therapy in different cellular targets. The approach presented here of defining primary targets of p53 is an important additional step in sorting out p53 targets at various cellular contexts.

Materials and methods

Cell lines, cycloheximide treatment and total RNA isolation

The human lung cancer cell line H1299 (lacking endogenous p53) expressing the mouse temperature sensitive mutant p53Val135 (a gift of Dr M Oren) which on temperature shift to 32°C will assume wild-type p53 conformation was used in this study. The control cell line used was the parental H1299 cells without ts-p53Val135. The cells were maintained at 37°C in RPMI medium containing 10% fetal bovine serum (FBS). Where specified, the cells were exposed to cycloheximide (10 μg/μl) 30 min prior to temperature shift. The cells were temperature shifted to 32°C, and after specific time intervals, harvested in TRIZOL solution (Gibco BRL, USA) and total RNA was isolated as per manufacturer's instructions.

Northern and Western-blot analyses

Equal amounts (10 μg) of total cellular RNA were used for Northern-blot analysis, using standard protocols (Ausubel *et al*., 1990). Autoradiography was done on Kodak X-ray films with intensifying screens at −70°C. For Western-blots, total cell lysates were prepared by lysing cells in NP40 lysis buffer (0.5% NP40, 150 mM NaCl, 50 mM Tris-Cl, pH 7.5, 1 mM DTT, 25 ng/ml aprotinin, 25 ng/ml leupeptin and 1 mM PMSF). Equal amounts (60 μg) of total protein were fractionated by SDS–PAGE on a 15% gel and transferred to nitrocellulose membranes (BA.85, Schleicher and Schuell). The primary antibodies used was a mixture of mAbs 421/248 for p53, Ab C19 (Santa Cruz) for p21 and a mixture of mAbs 4B2/2A9 for human MDM2 protein. The secondary antibodies used were HRP goat anti-mouse IgG (Sigma) and HRP goat anti-rabbit IgG (Jackson Laboratories).

Preparation of labeled cRNA and hybridization of microarrays

Total RNA was isolated from H1299val135 cells incubated at 32°C for 2, 6, 12, 24 h and H1299 control cells at 32°C for 2, 12 h. H1299val135 cells at 32°C in the presence of

cycloheximide were harvested only up to 12 h at 2, 4, 6, 9 and 12 h time points and H1299 control cells with cycloheximide at 32°C were harvested at 2 h time point and total RNA was isolated. Biotin labeled cRNA was synthesized and hybridized as described (Kaminski *et al*., 2000) to the Genechip HuGene FL array (Affymetrix, Santa Clara, USA) which contains probes for ~7000 human genes. Scanned output files were visually inspected for hybridization artifacts. Arrays lacking significant artifacts were analysed using Genechip 3.3 software (Affymetrix). Arrays were scaled to an average intensity of 1200 per gene and analysed independently. The expression value for each gene was determined by calculating the average of differences (perfect match intensity minus mismatch intensity) of the probe pairs in use for this gene. Ratios were determined by dividing the average difference of H1299Val135 cells for each time point to that of H1299 control cells at 2 h time point with or without cycloheximide in the respective experiments. A value of 100 was assigned to all measurements lower than 100. The expression data for all the hybridization experiments can be obtained from the corresponding author upon request.

Clustering analysis

Clustering analysis was performed twice; first to the 259 genes that were upregulated over 2.5-fold at least three times in the absence of CHX (in at least one of the three experiments, the gene had to be 'present' in the Present/Absent call provided by Affymetrix software). The second clustering analysis was applied to the 1090 genes that were upregulated over twofold at least once, in the absence of CHX experiment (also in that experiment it should be called 'present'). In both cases, each gene was represented by nine (four experiments without CHX and five experiments with CHX) values representing the log ratio of the average difference in the experiments to the average difference in their corresponding control. The log ratios were then normalized for each gene by dividing separately for the two experiments. The clustering algorithm measured the distance between the genes using the regular Euclidean distance between their normalized values. Note that this distance measure is more suitable than the Pearson correlation between the log ratios, since the latter may place induced and repressed genes in the same cluster.

Acknowledgments

We are grateful for the Arison Dorsman family's donation to the Center of DNA Chips in Pediatric Oncology. This work was supported in part by the Yad Abraham Research Center for Cancer Diagnosis and Therapy, the Rich Foundation for Leukemia Research and the Germany-Israel Science Foundation (GIF). G Rechavi holds the Gregorio and Dora Shapiro Chair for Hematologic Malignancies, Sackler School of Medicine, Tel Aviv University. We are thankful to the Crown Human Genome Center of the Weizmann Institute for the help received.

References

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA and Struhl K. (1990). *Current Protocols in Molecular Biology*. Green/Wiley Interscience, New York.

Blatt M, Wiseman S and Domany E. (1996). *Physical Rev. Lett.*, **76,** 3251–3254.

Coller H, Grandori C, Tamayo P, Colbert T, Lander ES, Eisenman RN and Golub TD. (2000). *Proc. Natl. Acad. Sci. USA,* **97,** 3260–3265.

El-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW and Vogelstein B. (1992). *Nat. Genet.*, **1,** 45–49.

El-Deiry WS. (1998). *Sem. Cancer Biol.*, **8,** 345–357.

Getz G, Levine E, Domany E and Zhang MQ. (2000). *Physica A.*, **279,** 457–464.

Ginsberg D, Michalovitz D, Ginsberg D and Oren M. (1991). *Mol. Cell Biol.*, **11,** 582–585.

Hannon GJ and Beach D. (1994). *Nature,* **371,** 257–261.

Haupt Y, Maya R, Kazaz A and Oren M. (1997). *Nature,* **387,** 296–299.

Jimenez GS, Nister M, Stommel JM. Beeche M, Barcarse EA, Zhang X-Q, O'Gorman S and Wahl GM. (2000). *Nat. Genet.*, **26,** 37–43.

Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJD, Morris D, Huang X, Sheppard D and Heller RA. (2000). *Proc. Natl. Acad. Sci. USA,* **97,** 1778–1783.

Kannan K, Amariglio N, Rechavi G and Givol D. (2000). *FEBS Lett.*, **470,** 77–82.

Komarova EA, Diatchenko L, Tokhlin OW, Hill JE, Wang ZJ, Krivokrysenko VI, Feinstein E and Gudkov AV. (1998). *Oncogene,* **17,** 1089–1096.

Kubbutat MH, Jones SN and Vousden KH. (1997). *Nature,* **387,** 299–303.

Levine AJ. (1997). *Cell,* **88,** 323–331.

Massague J, Blain SW and Lo SR. (2000). *Cell,* **103,** 295–309.

Michalovitz D, Halevy O and Oren M. (1990). *Cell,* **62,** 671–680.

Michieli P, Chedid M, Lin D, Pierce JH, Mercer WE and Givol D. (1994). *Cancer Res.*, **54,** 3391–3395.

Oda K, Arakawa H, Tanaka T, Matsuda K, Tanikawa C, Morit T, Nishimori H, Tamai K, Tokino T, Nakamura Y and Taya Y. (2000). *Cell,* **102,** 849–862.

O'Hagan RC, Schreiber-Agus N, Chen K, David G, Engelman JA, Schwab R, Alland L, Thomson C, Ronning DR, Sacchettini JC, Meltzer P and DePinho RA. (2000). *Nat. Genet.*, **24,** 113–119.

Scherer SJ, Maier SM, Seifert M, Hanselman RG, Zang KD, Muller-Hemerling HK, Angel P, Walter C and Schartl M. (2000). *J. Biol. Chem.*, **275,** 37469–37473.

Sherlock G. (2000). *Curr. Opin. Immunol.*, **12,** 201–205.

Shivakumar CV, Brown DR, Deb S and Deb SP. (1995). *Mol. Cell Biol.*, **15,** 6785–6793.

Tan M, Wang Y, Guan K and Sun Y. (2000). *Proc. Natl. Acad. Sci. USA,* **97,** 109–114.

Tang B, Bottinger EP, Jakowlew SB, Bagnall KM, Mariano J, Anver MR, Letterio JJ and Wakefield LM. (1998). *Nat. Med.*, **4,** 802–807.

Vousden KH. (2000). *Cell,* **103,** 691–694.

Young RA. (2000). *Cell,* **102,** 9–15.

Yu J, Zhang L, Hwang PM, Rago C, Kinzler KW and Vogelstein B. (1999). *Proc. Natl. Acad. Sci. USA,* **96,** 14517–14522.

Zhan Q, Chen IT, Antinore MJ and Fornace AJ. (1998). *Mol. Cell Biol.*, **18,** 2768–2778.

Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH and Levine AJ. (2000). *Genes Dev.*, **14,** 981–993.