2586–2592 *Nucleic Acids Research, 1998, Vol. 26, No. 11*

A method for analyzing the qualitative and quantitative aspects of gene expression: a transcriptional profile revealed for HeLa cells

Akihide Ryo, Nobuo Kondoh[1], Toru Wakatsuki[1], Akiyuki Hada[1], Naoki Yamamoto and Mikio Yamamoto[1,*]

Department of Microbiology and Molecular Virology, Tokyo Medical and Dental University, Tokyo 113, Japan and [1]Department of Biochemistry II, National Defense Medical College, 3-2 Namiki, Tokorozawa, Saitama 359, Japan

Received February 24, 1998; Revised and Accepted April 7, 1998

ABSTRACT

A number of strategies have been devised by which differentially expressed genes in different cell types or tissues can be identified. We here report an efficient method to analyze the qualitative and quantitative aspects of transcripts and to construct an extensive gene expression profile in any kind of cell or tissue of interest. This method enables us to analyze the composition of mRNA species, reflecting gene activities, by measuring the frequency of appearance of concatamerized 17mer cDNA mini-fragments, which are proportional to the abundance of mRNA. As compared with a related method previously described by others, we can analyze ~3–4 bp longer cDNA fragments derived from amounts of total RNA as small as 1 μg. Using this technique we examined 10 100 cDNA mini-fragments from HeLa cells and constructed a gene expression profile consisting of 3665 genes. This method should thus provide an overall indication of gene activities and a rational means for monitoring gene fluctuation in different cells or tissues at different stages of development, in normal and disease states.

INTRODUCTION

The Human Genome Project elucidating the structures of genes has entered its final phase. Research following the Human Genome Project will consist of investigation and analysis of the functions of genes and their transcripts. We are urged to expand our limited knowledge of human gene expression, systematically documenting expression patterns in various human organs, tissues and cell types (1). Okubo *et al*. have initiated the large scale sequencing of a 3′-directed cDNA library designed so as to faithfully represent the RNA population from several human cell lines and tissues in order to construct a body map, which is a set of gene expression profiles of the entire human body (2,3). Although this strategy has the potential to be a significant and valuable approach for evaluating the qualitative and quantitative aspects of gene transcripts, it is inefficient and time consuming, because it is based on one gene to one clone correspondence in sequencing. Serial analysis of gene expression (SAGE) (4,5) may also serve as an appropriate approach for surveying differential expression of many genes. SAGE is a highly efficient method that allows quantitative analysis of a large number of transcripts (6–8). However, the method has the following disadvantages: (i) the sequence tag obtained is very short (13 or 14 bp, including the restriction enzyme site); (ii) a large poly(A)$^+$ RNA sample (>5 μg) must be prepared; (iii) extraction of each sequence tag with a proper nucleotide length may be difficult because there is no spacer nucleotide within every di-tag, which is ligated directly in a tail-to-tail manner. To overcome these drawbacks and to simplify the procedure we devised a method termed ‘mini-fragment analysis of gene expression’ (MAGE). The method permits a cDNA library to be synthesized from a sample as small as 1 μg total RNA. The library is composed of concatamerized longer sequence tags (17 bp, including a common 3′-GCTGC tail, i.e. N_{12}GCTGC) flanked on both sides by defined spacer sequences.

As a demonstration of the usefulness of the method we examined 10 100 cDNA mini-fragments from HeLa cells and constructed a gene expression profile that included 3665 different genes.

MATERIALS AND METHODS

Cell culture

The HeLa cells were grown in a CO_2 incubator in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal calf serum, 100 U/ml penicillin and 100 μg/ml streptomycin. Cells were passaged twice a week and seeded at 80% confluency.

Construction of a mini-fragment cDNA library

Schematic representation of the procedure for mini-fragment analysis is shown in Figure 1. Total RNA was prepared from ~80% confluent HeLa cells using Isogen (Nippon Gene, Tokyo, Japan) and usually 50 μg total RNA was used to synthesize double-stranded cDNA with a cDNA synthesizing kit (TaKaRa, Tokyo, Japan) and Oligotex [oligo(dT_{30}) immobilized on latex beads; TaKaRa] as a primer for first strand synthesis. Double-stranded cDNA was digested with *Bbv*I (New England Biolabs, Beverly, MA), which recognizes GCTGC and cleaves 12 bp upstream from the recognition site. After capture of the 3′-portion of cDNA fragments by centrifugation, 3′ recessive ends were filled-in and an extra non-templated A residue was added at the

*To whom correspondence should be addressed. Tel: +81 429 95 1469; Fax: +81 429 96 5190

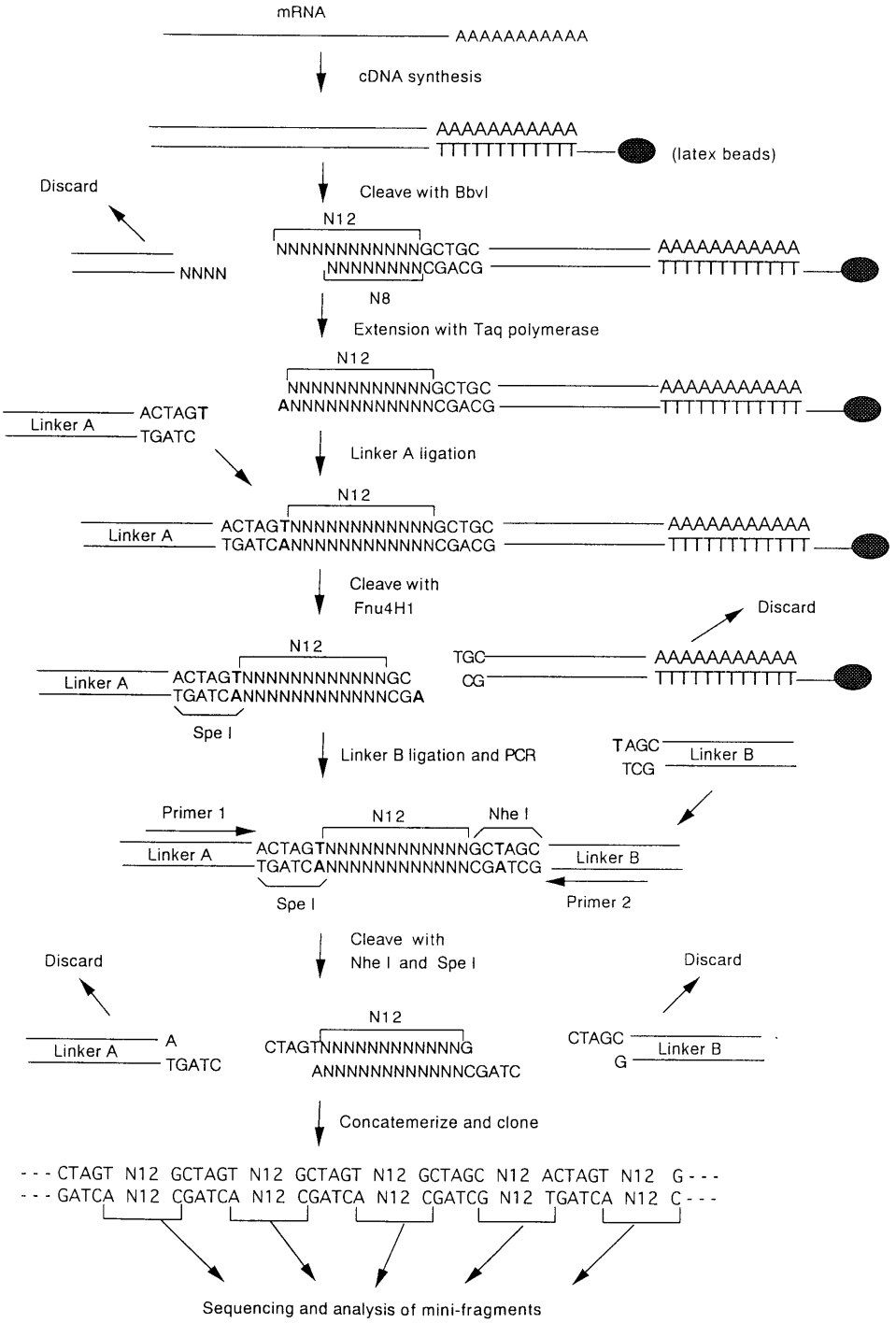

Figure 1. Schematic presentation of mini-fragment analysis of gene expression (MAGE). See the text for a detailed explanation.

3′-end with Taq DNA polymerase (Boehringer, Mannheim, Germany) in reaction buffer at 72°C for 2 h. Linker A, designed to have a 3′ single T overhang, was then ligated to the cDNA fragments similar to the A-T cloning procedure (9). After extensive washing to remove unligated linker fragments, cDNA immobilized on latex beads was digested with *Fnu*4HI, which recognizes GC/NGC. Since the 5′-most site should be the same as the *Bbv*I site used in the former step, the linker-contiguous cDNA fragments must have a 5′ single A overhang. After collecting the supernatant fraction, linker B, containing a 5′ single T overhang, was ligated as above. The cDNA fragments flanking both 5′- and 3′-sides with known oligonucleotide linkers were then amplified with the Expand High Fidelity PCR System (Boehringer). The PCR cycles consisted of an initial denaturation step at 95°C for 2 min followed by 25–28 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 30 s, with a final extension step at 72°C for 2 min, using 5′-TGTTACGTGGTTGACTAGT-3′ and 5′-TACGACGGCTGCGAGCCGCTAGC-3′ as primers 1 and 2

respectively. The PCR products were cleaved with *Nhe*I and *Spe*I simultaneously and separated by 12% PAGE to remove the major portions of the linkers. The band containing the DNA mini-fragments were excised, purified and concatamerized with T4 DNA ligase (0.5 U/μl) at 4°C for 8 h. The sample was subjected to 4% PAGE and products >400 bp were recovered. These products were cloned into the *Xba*I-digested and dephosphorylated pUC119 vector. Plasmid DNAs were isolated and screened for their insert size by *Eco*RI and *Hin*dIII double digestion. Recombinant plasmid DNAs from selected clones were sequenced with the TaqFS Dye terminater kits (Perkin Elmer) and analyzed using an ABI 373A automated sequencer.

Linker A and linker B were

```
5′-GCTGTTACGTGGTTGACTAGT-3′
3′-CGACAATGCACCAACTGATC-5′
```

and

```
5′-TAGCGGCTCGCAGCCGTCGTATG-3′
3′-TCGCCGAGCGTCGGCAGCATAC-5′
```

respectively.

Data analysis

Sequence files were manifested and analyzed by GENETYX-MAC v.8.0 software (Software Development Co. Ltd, Tokyo, Japan), which identified the four bases of sequences (CTAG) at the proper spacing and extracted 13 bp of sequence tags between them. Next, by adding CTGC at the 3′ portion of each tag as a part of the *Bbv*I recognition sequence, 17 bp mini-fragments were obtained. On average ~15–25 concatamerized mini-fragments were obtained from one bacterial clone. Sequence tags containing unidentified bases (N) and an illegitimate number of nucleotides were all discarded. Of all the representative mini-fragments those which occurred more than four times were searched for homologies using a non-redundant nucleic acid sequence database updated daily (nr-nt) at the Human Genome Center (Institute of Medical Science, The University of Tokyo) with the BLAST program (10). To select a match with nr-nt entries we down-loaded the full sequence showing 100% matches and examined whether each 17 bp sequence located at the proper position (i.e. the 3′-most *Bbv*I site in the mRNA sequence; Fig. 1). When there were multiple proper entries we adopted a characterized mRNA sequence entry arbitrarily. In cases of no match with characterized mRNA sequence entries we adopted one of the EST sequence entries meeting the above-described requirements. The sequences matching with none of the filed sequences were denoted as ‘no match’.

Oligonucleotide colony hybridization

To verify the quantitative nature of MAGE and to obtain novel gene sequences the oligo(dT)-primed HeLa cell cDNA library (Invitrogen Corp., Carlsbad, CA) was screened using the 17 bp of cDNA mini-fragments as radiolabeled probes. The filters, on which 4000–20 000 colonies were transferred, were prehybridized in a solution of 6× SET (0.9 M NaCl, 120 mM Tris–HCl, pH 7.8, 6 mM EDTA), 5× Denhardt’s solution, 10 μg/ml tRNA and 0.1% SDS for 1 h at 55°C. Hybridization proceeded in the prehybridization solution containing a ^{32}P-end-labeled oligonucleotide (3×10^5–1×10^6 c.p.m./ml) for 16 h at 45°C. The filters were washed twice in 6× SSC and 0.1% SDS at 52°C for 15 min, then autoradiographed using a Fuji Imaging Plate (Fuji Photo Co., Tokyo, Japan).

Northern blot analysis

A 10 μg amount of total RNA was separated on a 1.5% agarose gel containing 0.66 M formaldehyde and transferred to a Duralon-UV membrane (Stratagene, La Jolla, CA) according to the manufacturer’s protocol. The filters were baked at 80°C for 2 h. The cDNAs of ~300 bp were amplified by RT-PCR and labeled with [^{32}P]dCTP using a DNA labeling kit (Nippon Gene). Membrane filters were hybridized with labeled probes (1×10^6 c.p.m./ml) in buffer (40% deionized formamide, 4× SSC, 10% dextran sulfate, 1× Denhardt’s solution, 40 μg/ml sonicated and denatured salmon sperm DNA, 0.1% SDS, 20 mM Tris, pH 7.5) at 42°C for 16 h. The filters were washed twice for 15 min at room temperature and once for 30 min at 56°C with 2× SSC containing 0.1% SDS and exposed using a Fuji Imaging Plate.

RESULTS

Following the procedure schematized in Figure 1 we analyzed 10 100 mini-fragments of 17 bp cDNA sequences representing HeLa cell mRNAs. Of these transcripts those containing unidentified nucleotides were eliminated from further analysis. Consequently, 7858 transcripts were obtained in total for gene identification and quantification. These transcripts represented 3665 different genes that ranged in frequency of appearance from only once to as many as 552 times (Table 1 and 2). As the histogram in Figure 2 shows, frequencies of gene species were generally continuous from low to high abundance classes. Only five gene species appeared >100 times (1.27% of the total transcripts), 265 gene species 4–99 times (~0.05–1.27%) and 2814 gene species only once (Table 1). Two hundred and seventy species of mini-fragments occurred more than four times (47% of the total transcripts) and were matched against the nr-nt database using the BLAST program. Eighty seven percent of these sequence tags matched the nr-nt entries, of which 49.4% (116 species) corresponded to characterized mRNA sequence entries, whereas 50.6% (119 species) corresponded to uncharacterized expressed sequence tag (EST) entries. Table 2 shows the gene expression profile of the 33 most abundant species. One hundred and sixteen characterized genes were categorized arbitrarily into several groups based on their functions or intracellular location (Table 3). Among these the abundance of genes related to protein synthesis [e.g. elongation factor-1α (EF-1α) and ribosomal proteins] appeared with a remarkably high abundance, presumably reflecting the high activity of protein synthesis in this cell.

Table 1. Summary of the gene expression profile in HeLa cells

Fragment	Frequency in transcripts					Total
	>100	>10 ≤100	>4 ≤10	2/3	1	
Total fragments	1046	1625	1053	1320	2814	7858
Genes[a]	5	74	191	581	2814	3665
GenBank (EST)[b]	5 (0)	70 (31)	160 (88)	ND	ND	

[a]For genes, the number denotes the number of different genes represented in the indicated abundance class.
[b]For GenBank entries, the first number indicates the number of different genes that matched entries in GenBank in the indicated abundance class. Numbers in parentheses indicate the number of matched EST entries.
ND, not determined.

Table 2. Highly expressed genes in HeLa cells

Mini-fragment	*n	Percent	GenBank match (accession number)
GTGGACAAGAAGGCTGC	552	7.02	Elongation factor-1 alpha(X16869)
GGCAAGAAATGTGCTGC	133	1.69	Putative protein kinase C inhibitor (U51004)
GATCAAGTTCCCGCTGC	131	1.67	Ribosomal protein L18a (L05093)
GAAACTGATGAAGCTGC	126	1.6	Ferritin heavy chain (L20941)
GTGGGTCTCATTGCTGC	104	1.32	Ribosomal protein L8 (Z28407)
GGTGAAGATCATGCTGC	83	1.06	EST (T58602)
AGGTAAGCCCTGGCTGC	75	0.95	EST (T15957)
AGCGCAGGACAGGCTGC	66	0.84	Ribosomal protein L11 (L05092)
GGCCCATAGAGGGCTGC	57	0.73	Ribosomal protein S18 (X69150)
CCCCCTTCCCATGCTGC	50	0.64	Cofilin (D00682)
GGAATCATGGATGCTGC	50	0.64	E-B virus small RNAs associated protein (X59357)
GTGGAAGGAGTGGCTGC	41	0.52	Helix-loop-helix protein (X69111,S55813)
GTCCCACTATGCGCTGC	40	0.51	EST (T60734)
GATCCAAAAGGAGCTGC	34	0.43	L-lactate dehydrogenase-A (X02152)
CCTGTAATCCCAGCTGC	33	0.42	EST(AA112414)
GGCCATCTTGTTGCTGC	33	0.42	EST(AA043186)
GTGGACAAAAAAGCTGC	32	0.41	EST (Z14347)
TGGGAGGTTGAGGCTGC	31	0.39	Ornithine decarboxylase (M31061)
ATCGTGTTCGTGGCTGC	26	0.33	Calcium-binding protein S100P (X65614)
GCCGAGAAGAAGGCTGC	26	0.33	PLA-X (X06705)
TGGGAGGTGGAGGCTGC	25	0.32	EST (AA090560)
AAGGTCACCAAGGCTGC	24	0.31	Ribosomal protein S11 (X06617)
CCTGTAGTCCCAGCTGC	24	0.31	EST (T11857)
GCTACAGTACAAGCTGC	24	0.31	EST (D58897)
TATGGAGGAGGGGCTGC	24	0.31	KIAA0098 gene (D43950)
AGTAAAACGGTAGCTGC	23	0.29	Topoisomerase II beta isozyme (X68060)
ATGGTGGTTCCTGCTGC	23	0.29	EST (AA112825)
CCCACTGAAAGGGCTGC	23	0.29	Cystein rich peptide (M33146)
GTCCATCTTCCAGCTGC	23	0.29	EST (AA100991)
GTTCTTGACTCCGCTGC	23	0.29	Ribosomal protein L39 (D28397)
GCCACATTACAAGCTGC	21	0.27	Epithelin 1 and 2 (X62320)
GCTAACGATTGGGCTGC	21	0.27	Mitochondrial dodecanoyl-CoA delta-isomerase (Z25820)
TGCCATCACCTAGCTGC	21	0.27	Elongation factor-1 gamma (Z11531)

*n indicates the frequencies of appearance of each cDNA mini-fragment in the 7858 analyzed transcripts.
GenBank matches and accession nos are given for the most significant matches (see text).
Species are ordered in accordance with frequency in the profile.

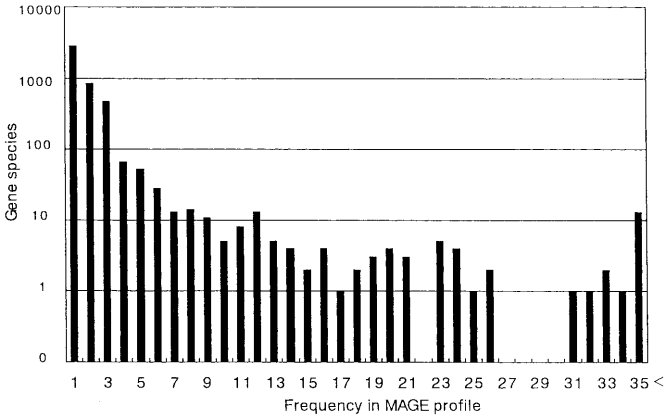

Figure 2. Gene expression profile in HeLa cells. The number of gene species is plotted on the abscissa. The frequency of each mini-fragment in the profile is plotted on the ordinate.

The quantitative nature of this method was evaluated by colony hybridization analysis using an oligo(dT)-primed HeLa cell cDNA library screened with 17 bp cDNA mini-fragments as probes. As shown in Figure 3, regression analysis between the results obtained by mini-fragment analysis and those obtained by colony hybridization analysis was performed. A regression equation was calculated as $y = 0.183 + 0.447x$, where y indicates the percent occurrence obtained by colony hybridization analysis and x represents those obtained by mini-fragment analysis. The correlation coefficient was 0.989 ($P < 0.01$, F-test), indicating that the abundance of mini-fragments faithfully represented the mRNA population.

To further analyze the composition of the MAGE library we performed Northern blot analysis of the total RNA from HeLa cells using radiolabeled cDNA probes for ferritin heavy chain, ribosomal protein L8, ribosomal protein S18, L-lactate dehydrogenase A, calcium binding protein S100P and cytochrome P450 reductase, which occurred 124, 104, 57, 34, 26 and 5 times in the MAGE profile (Fig. 4). Band intensities appeared to be quasi-proportional to the frequency levels observed in the mini-fragment analysis. These results suggest that the MAGE library is a relative representation of the mRNA population.

Along with data accumulation, the kinetics of six mini-fragment species which occurred abundantly in the final profile were examined. The occurrence of each mini-fragment increased in a near linear fashion as the total numbers of mini-fragments accumulated (data not shown). These results also confirmed the quantitative nature of the present method.

DISCUSSION

In this study we describe the development of a method providing an extensive gene expression profile in any kind of eukaryotic cell of interest. The profile is the list of active genes and their relative copy numbers of transcripts in a given cell. The profile gives us quantitative information as well as a chance to isolate novel expressed genes. Furthermore, by comparing two or more profiles constructed by the same procedure one can readily examine overall differences in gene activities among them. Such a comparison would readily uncover the genes specific to various cellular conditions, such as disease, development, differentiation, morphogenesis and many other physiological and pathological states, where the involvement of gene activity is considered.

Several strategies for surveying differential expression of many genes have recently been developed, including differential display (11), subtractive hybridization (12,13), microarray technologies (14) and the SAGE method. Both the differential display and subtractive hybridization methods are suitable for identifying genes expressed in relatively high abundance and depend on displaying substantial expression differences between the samples to be compared. Recently described microarray technologies may provide comprehensive transcriptional profiles in a very short period of time (15), but are currently restricted to known genes. The method also requires the development and the use of specialized and costly equipment for imaging.

Currently SAGE may be the best method, at least in essence, for quantitative analysis of gene expression, but there seems to be much room for improvement; for example, very short sequence tags, uncertainty of length of sequence tags due to di-tag formation by direct tail-to-tail ligation, the need for fairly large amounts of poly(A)$^+$ RNA, the difficulty of the overall procedure and so forth are current drawbacks.

Table 3. Categories of identified genes

Gene (Accession number)	Frequency	Gene (Accession number)	Frequency
protein synthesis			
elongation factor-1 α (X16869)	552	glutamine synthetase (U08626)	4
ribosomal protein L18a (L05093)	131	electron transfer flavoprotein alpha (J04058)	4
ribosomal protein L8(Z28407)	104	squalene synthetase (ERG9) (L06070)	4
ribosomal protein L11 (L05092)	66	MSS1(D11094)	4
ribosomal protein S18 (X69150)	57	cytosolic malate dehydrogenase (D55654)	4
Helix-loop-helix protein (X69111)	41	stathmin (X53305)	4
ribosomal protein S11(X06617)	24		
ribosomal protein L39 (D28397)	23		
elongation factor-1 γ (Z11531)	21	Secretory protein	
ribosomal protein S25(M64716)	20	ferritin heavy chain(L20941)	126
ribosomal protein L5 (U14966)	17	epithelin 1 and 2 (X62320)	21
ribosomal protein S26(HSS26)	15	anti-lectin antibody epitope(X58236)	18
ribosomal protein L27 (L19527)	14	lipocortin(X05908)	12
ribosomal protein S24 (M31520)	12	macrophage migration inhibitory factor(L19686)	10
ribosomal protein L19(X63527)	12	thymosin beta (S54005)	8
acidic ribosomal phosphoprotein P1(M17886)	12	IFN-gamma antagonist cytokine (A25270)	6
ribosomal proteinS16 (M60854)	11	ferritin light chain (M10119)	5
putative ribosomal protein (X53777)	7	macrophage capping protein (M94345)	4
ribosomal protein L32 (K03002)	5		
ribosomal protein S5 (U14970)	5		
histidyl-tRNA synthetase (Z11518)	5	Cytoskeleton	
ribophorin II (Y00282)	5	F-actin capping protein beta subunit(U03271)	16
		myosin light chain 3 non-muscle (M31212)	9
		profilin I (J03191)	7
Cytoplasm & organella		actin depolymerizing factor(S65738)	7
putative protein kinase C inhibitor (U51004)	133	beta-actin related pseudogene (V00481)	6
cofilin(D00682)	50	microfibril-associated glycoprotein (U19718)	4
E-B virus small RNAs associated protein (X59357)	50		
L-lactate dehydrogenase-A(X02152)	34		
ornithine decarboxylase (M31061)	31	Membrane protein	
calcium-binding protein S100P(X65614)	26	interleukin-6 receptor (M20566)	14
PLA-X (X06705)	26	guanine nucleotide-binding regulatory (G) protein (M16538)	12
plectin (U53204)	25	glucose transporter(K03195)	9
dodecanoyl-CoA delta-isomerase(Z25820)	21	integral membrane protein E16 (M80244)	8
glucose regulated protein (M19645)	20	surface antigen (M60922)	8
protein disulfide isomerase (J02783)	19	epithelial cell marker protein 1 (M93010)	7
retinoblastoma-binding protein (RbAp46)(U35143)	15	GTP-binding protein (X17644)	6
heat shock 27 KD protein (Z23090)	14	mannose 6-phosphate-specific receptor (M16985)	6
proline isomerase FK506-binding protein (L18980)	12	BST-2 (D28137)	6
cerebroside sulfate activator protein (M60257)	11	cell surface protein TAPA-1 (M33680)	5
cytochrome-c reductase core I protein (L16842)	10	voltage-dependent anion channel (L06132)	5
calmodulin(D45887)	9	membrane glycoprotein 4F2 antigen(j02939)	5
thiopurine methyltransferase(S62904)	8	laminin receptor (Human) (S37431)	5
glutathione peroxidase(Y00433)	8	cell surface glycoprotein gClq-R (X75913)	5
transketolase (U55017)	7	poliovirus receptor (X64123 X57516)	4
adrenodoxin (J03548)	7		
phenylalkylamine binding protein(Z37986)	6		
nicotinamide N-methyltransferase (U08021)	6	Nuclear protein	
casein kinase I delta (U29171)	6	topoisomerase II β isozyme(X68060)	23
semaphorin U (U28369)	6	hRlf beta subunit (D38073)	16
cystathionine beta-synthase subunit (L19501)	5	translocon-associated protein delta (Z69043)	12
ATP synthase (X83218)	5	Ran binding protein1 (D38076)	4
flavoprotein subunit of complex II (D30648)	5	U2 small nuclear RNA-associated B" antigen (M15841)	4
HIV-1 TAR RNA binding protein (L22453)	5	DNA replication licensing factor (D83986)	4
golgi antigen gcp372 (D25542)	5		
ERK activator kinase (MEK2) (L11285)	5		
casein kinase I epsilon (L37043)	5	Unknown functions	
heat shock protein 10 (X75821)	5	KIAA0098 gene(D43950)	24
cytochrome P450 reductase (S90469)	5	cystein rich peptide(M33146)	23
antioxidant enzyme AOE37-2 (U25182)	4	KIAA0002 gene(D13627)	18
aldolase A (M11560)	4	clone 23732 mRNA(U79258)	7
phosphoglycerate kinase (K03201)	4	14-3-3 epsilon mRNA (U54778)	6
17-beta-hydroxysteroid dehydrogenase (X87176)	4	large CLP mRNA (L54057)	6
cyclophilin (X52851)	4	enhancer of rudimentary homolog (U66871)	5
heat shock protein 90-alpha (X15183)	4	KIAA9001gene(D42040)	4
mitochondrial acetoacetyl-CoA thiolase (D10509)	4	radiation-inducible immediate-early gene (S81914)	4

The present method, on the other hand, gives 17 bp cDNA mini-fragments (~3–4 bp longer than SAGE tags), resulting in much higher reliability in identification of individual genes and usefulness for further gene cloning. The MAGE library can be made successfully from as little as 1 μg total RNA (unpublished results). This method should, therefore, be suitable for analyzing small amounts of valuable samples, such as biopsied or resected tissues from patients or aliquots. In addition, the biases derived from PCR should be significantly decreased because each mini-fragment is flanked on both sides by two defined linkers.

However, the profile constructed by our method does not reflect the entire mRNA population because endonuclease *Bbv*I cleaves every 1024 (4^5) bp on average, indicating that some mRNAs are devoid of a *Bbv*I site. This may be complemented by the use of other combinations of restriction enzymes (such as *Alu*I + *Bpm*I or *Msp*I +*Bsm*FI for 16 bp tags and so forth).

a

	Sequence	Gene	MAGE (%)	Colony hyb. (%)
a	GTGGACAAGAAGGCTGC	Elongation factor-1 α	7.02	3.27
b	GGCAAGAAATGTGCTGC	Putative protein kinase C inhibitor	1.69	1.2
c	GATCAAGTTCCCGCTGC	Ribosomal protein L18a	1.67	0.85
d	GTGGGTCTCATTGCTGC	Ribosomal protein L8	1.32	0.9
e	AGCGCAGGACAGGCTGC	Ribosomal protein L11	0.84	0.5
f	GTAGCGTATGGTGCTGC	Glucose regulated protein	0.25	0.1

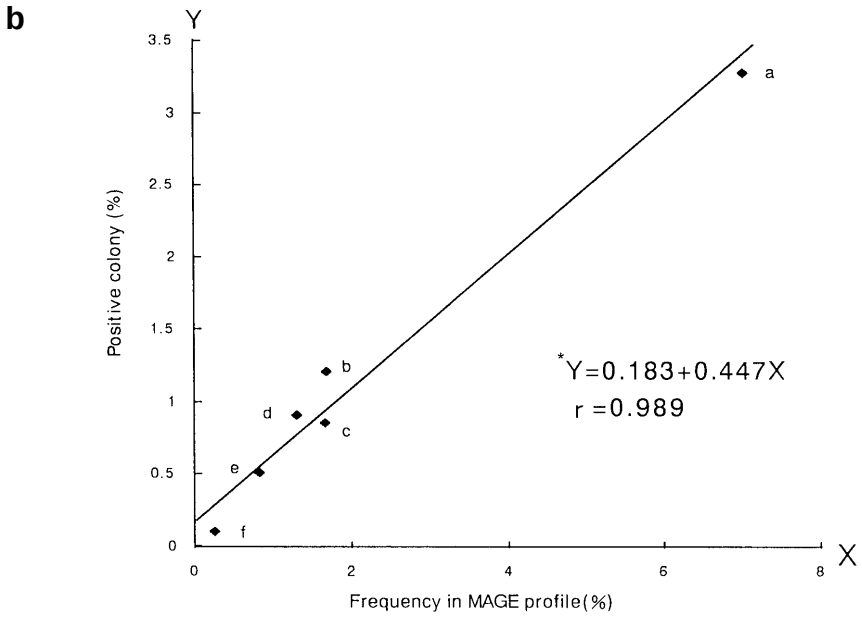

Figure 3. (**a**) Comparison of frequency of genes in the MAGE profile and those in a conventional cDNA library. MAGE indicates the percent abundance in the MAGE profile. Colony hyb. indicates the percent abundance of positive colonies in screening a cDNA library. (**b**) Regression analysis between MAGE and a conventional cDNA library. Percent abundance in screening a cDNA library is defined as an independent variable (*y*). Percent abundance in the MAGE profile is defined as a dependent variable (*x*). *Regression equation; *r*, correlation coefficient.

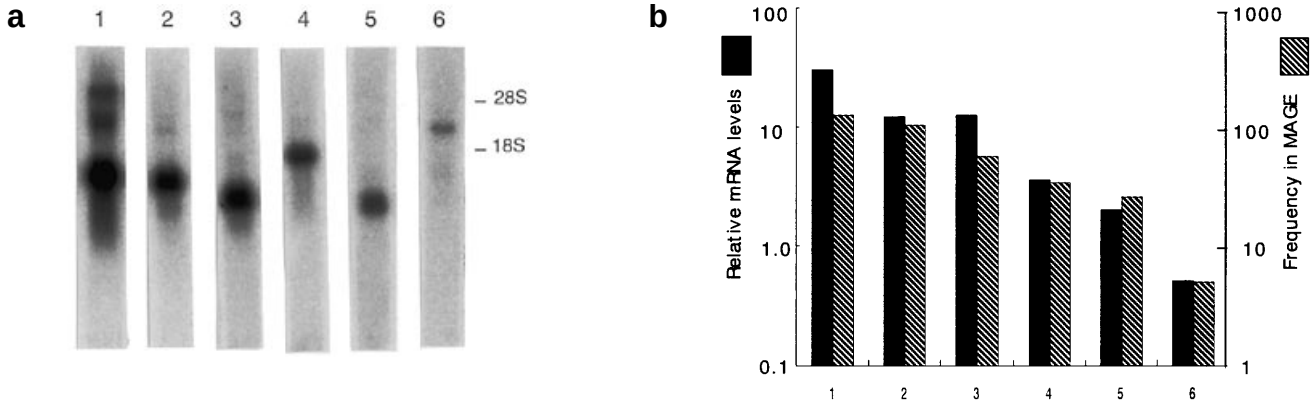

Figure 4. The relationship between mRNA levels and frequency in the MAGE profile. (**a**) The identical Northern blots using 10 µg total RNA were hybridized with the following ^{32}P-labeled cDNA fragments: lane 1, ferritin heavy chain; lane 2, ribosomal protein L8; lane 3, ribosomal protein S18; lane 4, L-lactate dehydrogenase A; lane 5, calcium binding protein S100P; lane 6, cytochrome P450 reductase. (**b**) Comparison of relative mRNA levels [relative band intensities in (a)] and frequency in the MAGE profile. The numbers on the ordinate represent the genes as in (a).

This is the first systematic profiling of expressed genes revealed for HeLa cells. Since only a limited number of non-biased large scale mRNA profiles are available at present (Expression Database of Human Genes, Osaka University, http://cookie.imcb.osaka-u.ac.jp/bodymap/), it is difficult to make detailed comparisons of the present profile with any others. However, it was obvious that the HeLa cell has its own expression characteristics, namely high activity of genes related to protein synthesis (1159 of 2340 total transcripts occurred more than four times; Table 3), as described in Results. In particular, the intense

expression of EF-1α was remarkable (552 times in 7868 fragments). Okubo *et al*. have indicated that this gene was one of the most active genes in HepG2 cells, showing 22 occurrences among 982 random cDNA clones (2). Although EF-1α is expressed in almost all kinds of mammalian cells and often is one of the most abundant proteins (16), it was not listed as a highly abundant gene in pancreas tissue (4). Even in liver and hepatoma tissues, which are akin to HepG2, EF-1α is ranked only in the moderately abundant class (our unpublished data). An enormously high transcription activity of EF-1α may therefore be unique to cultured cell lines having a high rate of protein synthesis (or growth properties).

An unexpected and rather embarrassing result was the differences in mRNA abundance among the ribosomal proteins, which are thought to be present in near equal amounts in mature ribosomal particles (Table 3). Analysis of HepG2 cells also revealed at least a 7-fold difference in ribosomal protein mRNAs (2) and at least a 3.2 times difference in normal colorectal epithelial cells (6). These observations may reflect the different turnover rate (synthesis and/or degradation) of each ribosomal subunit protein even in apparently highly stable ribosomal particles. Another explanation is that, taking into account the totally different compositions of the ribosomal components reported by Okubo *et al.* (2) compared with our findings (this study), eukaryotic ribosomal particles may be composed of a diverse set of heterogeneous complexes, not an assembly of homogeneous units. Many reports have shown previously that several ribosomal proteins were increased upon malignant transformation (17–21). Although we also noticed that genes for several ribosomal proteins were indeed up-regulated on analysis of hepatoma compared with normal liver tissue, some were instead down-regulated (unpublished observations). All these observations indicate that many, if not all, ribosomal proteins can be regulated independently according to changes in cellular physiology.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Masaaki Arai, Mizue Shichita (Fujiyakuhin Co. Ltd) and Kenji Tanaka for their excellent technical assistance. The authors are also grateful to the Human Genome Center of the Institute of Medical Science, The University of Tokyo, for providing valuable databases.

REFERENCES

1 Strachan,T., Abitbol,M., Davidson,D. and Beckmann,J.S. (1997) *Nature Genet*., **16**, 126–132.
2 Okubo,K., Hori,N., Matoba,R., Niiyama,T., Fukushima,A., Kojima,Y. and Matsubara,K. (1992) *Nature Genet*., **2**, 173–179.
3 Okubo,K., Itoh,K., Fukushima,A., Yoshii,J. and Matsubara,K. (1995) *Genomics*, **30**, 178–186.
4 Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) *Science*, **270**, 484–487.
5 Velculescu,V.E., Zhang,L., Zhou,W., Vogelstein,J., Basrai,M.A., Bassett,D.E.,Jr, Hieter,P., Vogelstein,B. and Kinzler,K.W. (1997) *Cell*, **88**, 243–251.
6 Zhang,L., Zhou,W., Velculescu,V.E., Kern,S.E., Hruban,R.H., Hamilton,S.R., Vogelstein,B. and Kinzler,K.W. (1997) *Science*, **276**, 1268–1272.
7 Polyak,K., Xia,Y., Zweier,J.L., Kinzler,K.W. and Vogelstein,B. (1997) *Nature*, **389**, 300–305.
8 Madden,S.L., Galella,E.A., Zhu,J., Bertelsen,A.H. and Beaudry,G.A. (1997) *Oncogene*, **15**, 1079–1085.
9 Marchuk,D., Drumm,M., Saulino,A. and Collins,F.S. (1991) *Nucleic Acids Res*., **19**, 1154.
10 Pearson,W.R. and Lipman,D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
11 Schena,M., Shalon,D., Davis,R.W. and Brown,P.O. (1995) *Science*, **270**, 467–470.
12 DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996) *Nature Genet*., **14**, 457–460.
13 Liang,P. and Pardee,A.B. (1992) *Science*, **257**, 967–971.
14 Hara,E., Kato,T., Nakada,S., Sekiya,S. and Oda,K. (1991) *Nucleic Acids Res*., **19**, 7097–7104.
15 Hampson,I.N., Pope,L., Cowling,G.J. and Dexter,T.M. (1992) *Nucleic Acids Res*., **20**, 2899.
16 Uetsuki,T., Naito,A., Nagata,S. and Kaziro,Y. (1989) *J. Biol. Chem*., **264**, 5791–5798.
17 Kondoh,N., Schweinfest,C.W., Henderson,K.W. and Papas,T.S. (1992) *Cancer Res*., **52**, 791–796.
18 Barnard,G.F., Staniunas,R.J., Bao,S., Mafune,K., Steele,G.D.,Jr, Gollan,J.L. and Chen,L.B. (1992) *Cancer Res*., **52**, 3067–3072.
19 Chiao,P.J., Shin,D.M., Sacks,P.G., Hong,W.K. and Tainsky,M.A. (1992) *Mol. Carcinogen*., **5**, 219–231.
20 Denis,M.G., Chadeneau,C., Lecabellec,M.T., LeMoullac,B., LeMevel,B., Meflah,K. and Lustenberger,P. (1993) *Int. J. Cancer*., **55**, 275–280.
21 Frigerio,J.M., Berthezene,P., Garrido,P., Ortiz,E., Barthellemy,S., Vasseur,S., Sastre,B., Seleznieff,I., Dagorn,J.C. and Iovanna,J.L. (1995) *Hum. Mol. Genet*., **4**, 37–43.