J. Biochem. 119, 749-767 (1996)

A Catalogue of Genes in Mouse Embryonal Carcinoma F9 Cells Identified with Expressed Sequence Tags[1]

Seiji Nishiguchi,* Rui Sakuma,* Midori Nomura,* Zhihua Zou,* Juree Jearanaisilavong,*,[2] Tadashi Joh,* Teruo Yasunaga,† and Kazunori Shimada*,[3]

*Department of Medical Genetics, Division of Molecular Biomedicine, Research Institute for Microbial Diseases, and †Genome Information Research Center, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565

Received for publication, November 6, 1995

We used expressed sequence tags (ESTs) to identify genes expressed in mouse embryonal carcinoma F9 cells and prepared 2132 ESTs from undifferentiated F9 cDNA libraries: 1026 were prepared after randomly selecting clones from one of the libraries and the remaining 1106 ESTs were prepared after classifying 2896 clones of the libraries into four classes, according to the levels and patterns of expression. Among the former 1026 ESTs, 797 (78%) matched known genes, 61 (6%) matched database sequences of uncharacterized cDNAs, and 168 (16%) represented novel genes. The ESTs matching known genes were catalogued according to putative structural and cellular functions. As many as 53% were related to transcription and translation, and 19% were related to energy metabolism, including transcripts of mitochondrial DNA. These percentages were significantly higher in F9 cells than in the human heart and brain, and a human liver cell line, HepG2. We found that approximately 7% of the ESTs corresponding to low-abundance mRNAs are either related to retinoic acid-regulated genes or mammalian development- and/or differentiation-related genes. Cataloguing of the genes expressed in the F9 cells paves the way for isolating genes involved in early mammalian development.

Key words: embryonal carcinoma, expressed sequence tags, F9, low-abundance mRNAs.

Since mouse embryonal carcinoma F9 cells and embryonic stem cells of the mouse blastocyst show a close resemblance (*1*), a catalogue of genes expressed in F9 cells should be of significant interest, not only for investigators using this line to study early mammalian development (*2*), but also for those characterizing various gene functions using embryonic stem cells and gene targeting techniques (*3*).

Partial sequencing of clones from cDNA libraries of specific tissues or cell types to generate expressed sequence tags (ESTs) provides quantitative and qualitative information regarding gene expression in a variety of tissues and cells (*4-9*). We used this procedure and started the categorization of genes expressed in undifferentiated F9 cells, using ESTs matching genes of known sequence (*10*). To deduce the overall profile of the expressed genes in F9 cells and to identify genes involved in early mammalian development, we chose to sequence the 5′ portions of cDNAs isolated from cDNA libraries constructed from F9 cells; sequencing from the 5′ end is expected to reveal DNA sequences that encode amino acids (aa), and to provide hints as to the functional class to which each cDNA belongs.

Here, we report the sequencing of cDNA clones and the systematic classification of genes expressed in undifferentiated mouse embryonal carcinoma cells, using F9 cells.

MATERIALS AND METHODS

Cells—Mouse embryonal carcinoma F9 cells (*11*) were maintained in Dulbecco's modified Eagle's medium supplemented with 15% fetal calf serum, as described by Nishiguchi *et al.* (*10*).

RNA Preparation and RNA Analysis—Total cellular RNAs were extracted from undifferentiated F9 cells by the acid guanidinium thiocyanate-phenol-chloroform method (*12*). Poly(A)$^+$ RNAs were prepared using an mRNA purification kit (Pharmacia LKB) (*10*).

Construction of cDNA Libraries—Two kinds of cDNA libraries were constructed, as described previously (*10*); one was constructed using λZAPII as a vector, the titer being 1×10^6 pfu/ml. Most inserts ranged from 500 bp to 1.5 kb. This library was used only at the initial stage of the current work. The other cDNA library was constructed using λ uni-ZAP as a vector, the titer being 2×10^6 pfu/ml.

[1] This work was supported by a Grant-in-Aid for Creative Basic Research (06NP0401), and a Grant-in-Aid for Scientific Research (04454170) from the Ministry of Education, Science, Sports and Culture of Japan. A part of this work was also supported by a Grant for Aging from the Ministry of Health and Welfare of Japan, a Grant-in-Aid from the Japan Medical Association, and a Grant-in-Aid from the Uehara Memorial Foundation. The nucleotide sequence data reported in this paper have already been registered or will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence databases, under the following accession numbers: D21355 to D21795, D28603 to D28744, and D76447 to D77996.

[2] Present address: Department of Microbiology, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

[3] To whom all correspondence should be addressed. E-mail: shimada@biken.osaka-u.ac.jp

Abbreviations: aa, amino acid(s); ESTs, expressed sequence tags; nt, nucleotide(s); ORF, open reading frame; pfu, plaque-forming unit; RA, retinoic acid.

Most inserts ranged from 500 bp to 3 kb.

Isolation and Classification of cDNA Clones—Each of the phage plaques was stocked in a cluster tube (Costar), as described (*10*). To classify the phage clones, cDNA probes were synthesized on poly(A)$^+$ RNAs prepared from undifferentiated F9 cells and mouse fibroblast L cells (*13*), using an oligo(dT)$_{12-16}$ primer and SuperScript reverse transcriptase (GIBCO BRL) (*10*). Clones of the libraries were classified into four groups, according to the intensities of signals (*10*, see Table I).

Single Pass Sequencing—Templates for an automated DNA sequencer (373A Sequencer; Applied Biosystems, ABI) were prepared by PCR amplification of cDNA inserts (*10*). Sequencing reactions were performed using fluorescent dye-labeled primers and cycle sequencing kits, and were analyzed on 6% polyacrylamide gels using the 373A Sequencer (ABI) and accompanying software (*10*).

Database Search—We used the BLAST program mail server (*14*) at either the Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, or the National Center for Biotechnology Information, USA. The FASTA program mail server (*15*) at the Human Genome Center was also used. To search for similarities to sequenced cDNAs, the non-redundant nucleic acid sequence database constructed from gb (=GenBank nucleic acid sequence database), emb (=EMBL nucleic acid sequence database), and gbu (=cumulative daily updates of GenBank, GenBank-UPD) was used. Clones with sequences exhibiting similarities to database sequences with BLASTN or FASTA-opt scores higher than 200 were operationally classified as the identified group. Subsequently, the DNA sequences of clones other than those in the identified group were examined for similarities to a non-redundant protein sequence database constructed from sp (=SWISS-PROT protein sequence database), pir (=PIR protein sequence database), prf (=PRF protein sequence database), gp (=translated protein sequence from GenBank), and gpu (=translated protein sequence from GenBank-UPD), using the BLASTX program mail server (*14*) at the Human Genome Center. Clones with sequences showing similarities to database sequences with BLASTX scores higher than 100 were operationally added to the identified group. The final searches were made from September 1-6, 1995.

Clones with sequences showing significant similarities to database sequences of uncharacterized cDNAs were re-examined for similarities to a database of ESTs (=DBEST), using the BLAST program mail server at the National Center for Biotechnology Information. These searches were made from November 27-30, 1995, and provided information on the sources of the corresponding sequences.

RESULTS AND DISCUSSION

Analyses of cDNA Clones Sequenced—To obtain an overall profile of gene expression in F9 cells, we prepared 1,026 ESTs after randomly selecting clones from an F9 cDNA library (see Table I, class X). Approximately 84% of the partial sequences, or 858 ESTs, were identical or almost identical to known genes or ESTs (Table I). We reported that 41% of 582 ESTs prepared from low-abundance class IV clones correspond to identified ESTs (*10*). Recently, we re-examined these 582 ESTs in homology searches and found that 54% show similarities to sequences in the non-redundant nucleic acid or protein sequence databases, probably reflecting the rapid increases in the sizes of the databases (Table I). In the present work, we prepared 524 ESTs from three other classes and found that approximately 76% are identified ESTs (Table I). Altogether, we examined 4,428 clones of undifferentiated F9 cDNA libraries, prepared 2,132 ESTs, and found that 74% (1,570 ESTs) are identical or almost identical to known genes or ESTs (Table I), whereas 26% (562 ESTs) show no significant homology to any gene in any organism. These figures are comparable to data obtained in a similar study on 2,058 ESTs prepared from a human fetal-lung cDNA library (*9*), *i.e.*, about 69% were considered to be identified ESTs.

Catalogue of Identified ESTs—The identified ESTs were classified according to the distributions and functions of proteins deduced from their DNA sequences (Table II): 141 corresponded to transcripts of mitochondrial DNA; the remaining 1,429 ESTs corresponded to 633 different genes, and 436 of these 633 were unique ESTs, *i.e.*, they appeared only once in the identified ESTs. One-hundred and fifty-four of the 1,570 ESTs matched database sequences of uncharacterized cDNAs: 2 matched the *Caenorhabditis elegans* ESTs, 8 the mouse ESTs, and 13 the rat ESTs, whereas 131 matched the human ESTs (Table II).

Among the 1,570 identified ESTs, 695 (44%) were related to transcription/translation, and 252 (16%) to energy metabolism (Table II). One-hundred and forty-one of the latter 252 ESTs corresponded to transcripts of mitochondrial DNA, and most corresponded to discrete ORFs of mitochondrial DNA (Table II). Two ESTs (69D07

TABLE I. **Summary of database similarity search.**

Class		Examined cloned (%)[a]	Sequenced clone (%)[b]	Identified EST (%)[c]	Novel EST (%)[d]
1) ESTs prepared from unclassified library					
Class X		1,532	1,026 (67)	858 (84)	168 (16)
2) ESTs prepared from classified library					
Class I	F9+ L+	1,388 (48)	247 (18)	209 (85)	38 (15)
Class II	F9+ L−	381 (13)	209 (55)	142 (68)	67 (32)
Class III	F9− L+	248 (9)	68 (27)	48 (71)	20 (29)
Class IV	F9− L−	879 (30)	582 (66)	313 (54)[e]	269 (46)
	Sub-total	2,896	1,106 (38)	712 (64)	394 (36)
	Total numbers	4,428	2,132 (48)	1,570 (74)	562 (26)

[a]Numbers (%) of examined clones of each class. Class X corresponds to unclassified clones, and 2,896 clones of F9 cDNA libraries were classified into the four classes using F9 and L total cDNA probes (*10*). [b]Numbers (%) of ESTs prepared from each class. [c]Numbers (%) of identified ESTs among the total ESTs of each class. [d]Numbers (%) of novel ESTs among the total ESTs of each class. [e]When ESTs prepared from class IV by Nishiguchi *et al.* (*10*) were searched again, the ratio of identified ESTs increased from 41 to 54%.

TABLE II. **Mouse F9 ESTs matching known genes or proteins in the non-redundant nucleic acid or protein databases.**

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
Cytoskeletal and Contractile Elements							
72E11	IV	204	gb:BOVACT2	Actin 2	529	(0/ 2)	bo
95G04	X	105	gb:MMACTBR	Actin, beta, cytoskeletal	471		m
74H03	I	374	gb:MMGACTR	Actin, gamma, cytoplasmic	1381f	(5/ 6)	m
C1B04	X	225	gb:RNU06755	Calponin, acidic	243		r
29F08	IV	266	gb:RATCARHC	Clathrin heavy chain	597		r
95C08	X	376	gb:MUSCOF	Cofilin	1826	(1/ 2)	m
72E08	IV	209	gb:PIGDESTN	Destrin, actin-binding protein	350		p
A6C09	X	217	gp:U15303	Dynein, heavy chain subunit	165x		cr
B1D10	X	184	gb:MUSKIF4	Kif4	848		m
A2E04	X	109	gb:RATKINLB	Kinesin light chain B	455		r
A2G04	X	142	gb:HUMCLIP	Linker protein-170 alpha-2, cytoplasmic	503		h
93F01	X	299	gb:RNU05784	Microtubule-associated protein	1351		r
83E08	IV	333	gb:HUMMYLCB	Myosin alkali light chain, non-muscle	1110f	(0/ 2)	h
92F12	IV	164	gb:RNMYOLC1	Myosin light chain, muscle, mlc2 gene	522		r
B9B10	X	140	gb:S59342	Nuclear pore complex glycoprotein p62	646		m
96F11	X	108	gb:MUSPTAC97	Nuclear pore-targeting-complex component	441		m
71G11	IV	266	gb:MMNUCLEO	Nucleolin	759f	(1/ 3)	m
B6F08	X	271	gb:RNPLECT	Plectin	537		r
93H01	X	303	gb:MMTROPM5	Tropomyosin 5, non-muscle	816f	(1/ 2)	m
71G10	I	363	gb:MUSTUBA2M	Tubulin, M-alpha 2 or 6	1401f	(4/ 8)	m
72F04	II	326	gb:CGTUBB2	Tubulin, M-beta 2	1212f	(3/ 4)	ha
86F08	III	268	gb:MMTUBBM5	Tubulin, M-beta 5	1038f	(2/ 3)	ha
Extracellular Matrix							
B2F12	X	219	gb:RATAGR	Agrin	942		r
93H11	X	369	gb:MMSPARCR	Osteonectin (SPARC)	1388f		m
A6E01	X	266	gb:HUMDMP	Desmoplakin	1012		h
74E04	IV	242	gb:MMECADH	E-cadherin	921f		m
74C10	III	258	pi:S23127	Endozepine	128x		bo
B5F07	X	246	gb:RATFN3M2	Fibronectin	888f		r
71E02	IV	267	gb:HUMLAMBB	Lamin B2	428f		h
68D02	II	459	gb:MUSLAMP1	Lysosomal membrane glycoprotein-1	1699f	(3/ 6)	m
93D06	X	175	gb:HUMPA1V	Pro-alpha-1 (V) collagen	591		h
87D10	IV	278	gb:HSTEST	Testican	603		h
Energy Metabolism							
A7E05	X	223	gb:MMANTAP	Adenine nucleotide translocator	852f	(2/ 2)	m
81B02	IV	387	gb:RATKINASE2	Adenylate kinase 2	1021		r
67A01	IV	230	gb:RATPCOA	Alpha-propionyl-CoA carboxylase	392		r
71G08	II	475	gb:MSALEN	Alpha-enolase	1811f	(5/14)	m
84D03	IV	316	gb:MUSASPATC	Aspartate aminotransferase, cytosolic	972f	(0/ 2)	m
95H10	X	371	gb:MUSASPATM	Aspartate aminotransferase, mit.	1625		m
30B02	IV	177	gb:BTATP1	ATP synthase, mitochondrial	383f		bo
72B06	I	146	gb:MUSATPSYNX	ATP synthase alpha subunit	721	(1/ 2)	m
96B10	X	206	gb:HSU09813	ATP synthase subunit 9, mitochondrial	599f	(1/ 2)	h
A2G09	X	287	gb:MMU09874	Clp/HSP104 gene family of ATPases, SKD3	1408		m
92D09	IV	271	gb:M21197	Citrate synthase	567		p
82G12	IV	308	gb:HUMCOXCA	Cytochrome c oxidase subunit Vb	940		h
30A07	IV	114	gb:MMCOXVB	Cytochrome c oxidase subunit Vb	238f		m
B5H08	X	409	gb:MMCOX7C1	Cytochrome c oxidase subunit VI	1740		m
72H02	I	396	gb:MUSMTTIVMR	Cytochrome oxidase subunit IV	1678	(1/ 2)	m
68E03	IV	417	gb:HSCYP2D7A	Cytochrome P45..., pseudogene	437		h
83D09	IV	308	gb:HUMDIHYDH	Dihydroorotate dehydrogenase	289		h
B7D11	X	197	gb:RNDDOPTAU	D-dopachrome tautomerase	227f		r
93E07	X	373	gb:RATATPASEF	F1-ATPase beta subunit	1236f	(1/ 4)	r
90C02	IV	244	gb:MUSGPI	Glucose phosphate isomerase	941		m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
95A01	X	260	gb:RATGRP78	Glucose regulated protein 78-kDa	524f	(2/ 2)	m
B8F06	X	396	gb:MUSGAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1776	(5/11)	m
96D04	X	347	gb:MUSIMPD	IMP dehydrogenase	909	(2/ 3)	m
90C12	IV	288	gb:BTIDNADP	Isocitrate dehydrogenase, mitochondrial	1021		bo
B8F04	X	412	gb:MMLDHAP1	Lactate dehydrogenase-A pseudogene	1395f	(1/ 2)	m
86B05	IV	422	gb:MUSMDHA	Malate dehydrogenase	1680		m
71C06	IV	282	gb:DOGOST48A	Oligosaccharyl transferase	637f		do
91G01	II	432	gb:MUSPFK	Phosphofructokinase	1808	(1/ 3)	m
B9D09	x	156	gb:OA6PDGH	Phosphogluconate dehydrogenase	510		sh
77F10	I	382	gb:MUSPGK1PS2	Phosphoglycerate kinase-1	1402f	(5/11)	m
B1F04	X	317	gb:MUSPGK1PS1	Phosphoglycerate kinase, pseudogene	1491		m
77C06	I	320	gb:S63233	Phosphoglycerate mutase type B subunit	1091	(1/ 2)	r
92B11	IV	157	gb:HUMPYRUV	Pyruvate dehydrogenase E1-beta subunit	551	(1/ 2)	h
B6D08	X	252	gb:MUSSDHA1A	Pyruvate dehydrogenase, pdha-1	968		m
94D03	X	387	gb:MUSPKM	Pyruvate kinase, M	1283f	(3/ 7)	m
74G12	II	249	gb:RATRIP	Rieske iron-sulfur protein	309		r
89H06	II	278	gb:HUMSHMTB	Serine hydroxymethyltransferase, mitoch.	895f	(0/ 3)	h
40F03	IV	268	gb:HUMCOMIIP	Succinate-ubiquinone oxidoreductase	392		h
84C09	IV	486	gb:CHKTRANGLU	Transglutaminase	413		ch
94G09	X	387	gb:MMTPI1R	Triosephosphate isomerase	873	(2/ 3)	m
88A10	IV	264	gb:BTCYTBC1I	Ubiquinol-cytochrome-c reductase B	710f	(0/ 3)	bo
82E10	III	112	pi:S28238	Ubiquinone oxidoreductase, CI-13(IP)	111x		bo
77E11	I	334	gb:BTCI18IP	Ubiquinone oxidoreductase, CI-18(IP)	1256		bo
93F08	X	138	gb:BTCI19	Ubiquinone oxidoreductase, CI-19	420		bo
73F12	I	347	gb:BTCIB18	Ubiquinone oxidoreductase, CI-B18	766	(0/ 2)	bo
74E12	I	386	gb:BTCISGDH	Ubiquinone oxidoreductase, CI-SGDH	949	(1/ 4)	bo
(ESTs matching mitochondrial H-strand)[i]							
B9A05	X	325	gb: MITOMM	12S rRNA	1450	(2/ 2)	m
B1F10	X	365	gb: MITOMM	16S rRNA	1704	(3/ 3)	m
A5C04	X	299	gb: MITOMM	NADH dehydrogenase 1	1450	(7/10)	m
95A07	X	302	gb: MITOMM	NADH dehydrogenase 2	1465	(7/ 7)	m
B2E01	X	286	gb: MITOMM	Cytochrome oxidase I	1101	(17/24)	m
B9H01	X	349	gb: MITOMM	Cytochrome oxidase II	1390	(16/20)	m
B9A11	X	351	gb: MITOMM	ATPase 8 and/or 6	1674	(21/27)	m
B5F06	X	293	gb: MITOMM	Cytochrome oxidase III	1264	(15/16)	m
B6D03	X	293	gb: MITOMM	NADH dehydrogenase 3	1107f	(2/ 3)	m
C1C09	X	311	gb: MITOMM	NADH dehydrogenase 4L and/or 4	506	(10/14)	m
69F06	II	282	gb: MITOMM	NADH dehydrogenase 5	1360	(4/ 5)	m
94H06	X	467	gb: MITOMM	Cytochrome B	1693	(7/ 8)	m
(ESTs matching mitochondrial L-strand)[i]							
69D07	I	198	gb: MITOMM	tRNA-Ala, Asn and Cys	963		m
A6C08	X	343	gb: MITOMM	tRNA-Ser	1495		m
Hormones and Hormonal Regulation							
97B05	X	109	gb:RABEPIP	Endometrial progesterone-induced protein	273		rb
77D08	IV	295	gb:HUMBIGFII	Insulin-like growth fact. bind. protein 2	574f		h
93C01	X	110	em:S48643	Mammary-derived growth inhibitor	514		m
48B04	IV	283	gb:RNU05014	PHAS-I, target for insulin	880f	(0/ 4)	r
05E12	II	312	gb:MUSTHYB	Thymosin beta-4	1118f	(0/ 2)	m
B5C09	X	311	gb:MMTHYMOA	Prothymosin alpha	1323	(2/ 2)	m
Signal Transduction and Cell Regulation							
(Signal transduction)							
68G12	IV	382	gb:MME1433IS	14-3-3 protein epsilon subtype	1297f	(1/ 2)	m
59C01	IV	321	gb:RAT1433PG2	14-3-3 protein gamma subtype	828f	(0/ 2)	r
75H04	I	336	gb:RATTHETA	14-3-3 protein theta subtype	1593f	(3/ 5)	r
B9B04	X	361	gb:MMCALMOD	Calmodulin	1316f	(3/ 4)	m
81C11	IV	457	gb:MUSGB4A	G protein beta subunit	1747		m
67C08	II	477	gb:MUSGPBSL	G protein beta subunit homolog	1563f	(4/ 6)	m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
A4G03	X	171	gb:RATGAMMA5A	G protein gamma-5 subunit	727		r
89H11	IV	253	gb:HUMEGFGRBA	GRB2, EGF-receptor binding protein	205		h
75C06	I	314	gb:HSGST1	GST1-Hs, GTP-binding protein	1298		h
A0H06	X	216	gb:HSU02082	Guanine nucleotide regulatory protein	774		h
84B04	IV	342	gb:MUSGNDSA	Guanine nucleotide dissociation stimulator	728		m
A1E12	X	113	gb:HSIEF7442	IEF 7442, regulator of RAS-cAMP pathway	249		h
C0E07	X	413	gb:HSIEF9306	IEF 9306, regulator of RAS-cAMP pathway	957f		h
94B08	X	204	gb:HUMPHPLA2	Phospholipase A2	474f		h
60E11	IV	246	gb:MMRAB3A	Rab3A, GTP-binding protein	903f	(1/ 2)	m
86B09	II	183	gb:RNRABGDIB	Rab GDI beta	290		r
69D06	IV	307	gb:HSRHOC9	RhoC, Ras-related GTP-binding protein	357		h
A4D06	X	259	gb:HUMRHOGDI	Rho GDP-dissociation inhibitor	883		h
73C03	II	331	gb:MMU06923	Signal transducer and activator	1246		m
B9G02	X	315	gb:RATSTAT	Stathmin, neuron-enriched phosphoprotein	1203		r
69C05	IV	377	gb:HSU02609	Transducin-like protein	674f		h
(Kinases and phosphatases)							
70A06	I	248	gb:RATBCAK	Branched chain keto acid dehydrog. kinase	1015		r
A2D11	X	186	gb:HUMCNRA1	Calcineurin A1	826		h
82H06	IV	397	gb:MMCKIIB	Casein kinase II b subunit	595		m
84A02	IV	355	gb:MUSFAK	Focal adhesion kinase	1241		m
68A05	IV	106	gb:RATEIF2APK	Hemin-sensitive eIF-2a kinase	228		r
A2B09	X	104	gb:HSINPO5P	Inositol polyphosphate 5-phosphatase	261		h
92A01	IV	161	gb:RATMEK2	MAP kinase kinase-related protein	485f		r
88H01	II	360	gb:MUSMATRICI	Matricin, protein kinase	1279f	(2/ 4)	m
96F01	X	113	em:S59517	Microtubule-associated protein-2 kinase	502		m
83F11	II	367	gb:MMNDIPKB	Nucleoside diphosphate kinase B	1413f	(0/ 3)	m
75C08	IV	242	gb:HUMNPP	Nucleotide pyrophosphatase	275f		m
85B05	IV	265	gb:MUS2PHKG	Phosphorylase kinase g-subunit	481		m
39G09	IV	323	gb:S12091	Protein kinase crk-4	998f		m
88E05	IV	228	gb:MUSMEKPKI	Protein kinase MEK	776		m
A4G11	X	220	pr:1917290A	Protein kinase NPK1, yeast STE11 homolog	118x		to
66F10	IV	240	gb:RNP2A2	Protein phosphatase 2A subunit isotype a	781f		r
67G04	I	244	gb:HUMP2A	Protein phosphatase 2A regul. subunit b	820f	(0/ 3)	h
97D01	X	166	gb:MUSPP2C1	Protein phosphatase 2C beta	776		m
38F12	IV	352	gb:HSU02680	Protein tyrosine kinase	278		h
73D10	I	208	gb:MUSPTP36	Protein tyrosine phosphatase, PTP36	103x		m
96E01	X	254	gb:MUSS6KA	Ribosomal protein S6 kinase	969f	(1/ 2)	m
C0C03	X	324	gb:MMSEK2	Sek-2, receptor-protein tyrosine kinase	1117		m
(Developmental regulation)							
A2G08	X	258	gb:S68108	brg1 = brahma homolog, *hox* gene regulator	1095		m
90B11	IV	263	gb:S43105	Cyclin B1, cell cycle regulatory protein	1028f		m
71B09	IV	201	gb:MMCYB2	Cyclin B2	454		m
83E09	IV	337	gb:MMCYCLGMR	Cyclin G, target of p53	1273f		m
78C08	IV	253	gb:MMHOX31A	Hox-3.1 homeo gene 5' flank	205		m
B7H06	X	199	gb:MMHOX35	Hox-3.5	223		m
67A04	IV	377	gb:S73882	*rae-28, Drosophila polyhomeotic* gene-like	813		m
75A11	IV	271	gb:MMSIAHIB	siah-1B, seven in absentia gene-homolog	837f		m
83A06	IV	303	gb:MUSWNTVB	Wnt-5b protein prec.,Wnt-1/int-1 family	329f		m
(Oncogenes, tumor suppressors and tumor-related)							
74A01	IV	256	gb:S111271	b-Myb, proto-oncogene	335		m
72A05	IV	132	gb:HUMCRK	Crk-II, proto-oncogene	490		h
75F03	IV	247	gb:MMECA39	ECA39, target for c-Myc regulation	489		m
74C11	II	354	gb:MMP53	p53, transformation-associated protein	1321f	(1/ 2)	m
A1C07	X	404	gb:HUMAPE6ONC	Papillomavirus E6 oncogenic protein	781		h
B7B12	X	236	gb:MMRNAQM	QM, a putative tumor suppressor	893f	(2/ 3)	m
A2H12	X	305	gb:S66431	RBP2 = retinoblastoma binding protein 2	300f		h
66B01	IV	321	gb:HUMTYKRET	Ret, transforming gene	432f		h
66F07	IV	514	gb:HUMSET	Set, related to putative oncogene cat	1687f	(1/ 4)	h
71C04	I	225	gb:RATTUMASAN	Tumor-associated antigen	485		r

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
71D01	IV	281	gb:HUMIEF	Transformation sensitive protein	762f		h
86H08	IV	445	gb:HSTRE210	Tre, oncogene	359		h
Transcription and Translation							
(Transcription factors)							
C0B01	X	309	gb:XLAB21	AB21, promoter factor RPD3 homolog	965	(1/ 2)	xe
95B04	X	367	gb:S60998	Acute myeloid leukemia, mouse PFBP2 alp	382		h
78E12	III	273	gb:MMATF4	ATF-4 = activating transcription factor 4	792f	(0/ 2)	m
B7F09	X	139	gb:HUMHIP116A	ATPase, DNA-binding protein, SNF2/SWI2	542		h
82D06	IV	405	gb:HSBTF3B	BTF3b, transcription factor	805		h
B7C05	X	370	gb:HUMCBE	CBE = CCAAT-box-binding factor	1274	(1/ 2)	h
96F03	X	131	gb:MMDP1	DRTF-polypeptide-1, DP-1	352		m
94H07	X	307	gb:MUSEKLTF	Erythroid krueppel-like transcription fact.	450		m
B9E01	X	146	gb:MMEWS	EWS, putative RNA binding protein	1267		m
81A11	II	134	gb:HUMBTEB2	GC-box binding protein BTEB2	481	(0/ 2)	h
73F06	I	209	gb:RNU09551	HMG-box containing protein 1	717f		r
B4C12	X	271	gb:S69934	hRPB6 = RNA polymerase common subunit RPB6	904		ha
86C03	IV	458	gb:MUSID	Id, HLH-DNA binding protein regulator	407		m
95A12	X	368	gb:HUMTBP1	HIV Tat transactivator binding protein-1	1498		h
39B01	IV	356	gb:HUMMPSI	Metallopanstimulin, MPS1	1058f	(3/ 5)	h
B9C03	X	395	gb:MMNRF1	NRF1, leucine zipper transcription factor	1198		m
B9C06	X	367	gb:MUSOCT3	Oct-3, POU-domain transcription factor	1399f	(1/ 3)	m
77A04	I	259	gb:S64860	Octamer-binding protein	1096		h
A6G06	X	103	gb:S66656	p70 = transcription factor	488		h
74B03	II	269	gb:HSU02368	PAX3/forkhead transcription factor gene	494		h
73A07	I	186	gb:HUMPRGRN	Peregrin, leucine zipper dimer motif	225		h
84C12	IV	431	gb:RNPYB2	PYBP2, pyrimidine binding protein	344		r
B9F07	X	284	gb:MUSREX1	REX-1, zinc finger protein	1393	(3/ 4)	m
C0B11	X	329	gb:DROROX2Y	RRM-type RNA binding protein	319		dm
77D04	I	359	gb:HUMSP2A	Sp2 protein	1230		h
B9E10	X	314	gb:CGU22818	SREBP-2	294		ha
85G08	IV	399	gb:HSU14134	Transcription factor IIIA, finger protein	826f		H
66A04	IV	408	gb:HSUBF	Upstream binding factor, hUBF	209f		h
72H05	II	112	gb:HSUSFMR	Usf = late upstream transcription factor	479		h
94H09	X	304	gb:MUSZFP0	Zinc finger protein	541f	(2/ 2)	m
C0D06	X	233	gb:HUMMAZ	Zinc finger protein, MAZ	407		h
94A11	X	433	gb:HUMZFPJS	Zinc finger protein	1181f		h
84C05	IV	318	gb:HSZNF	Zinc finger transcription factor, ZNF6	589f		h
(Transcription and translation machinery)							
74C12	I	344	gb:MUSRGEETS	18S ribosomal RNA gene 5' end	1343		m
84D07	III	222	gb:MUSLASSB	Autoantigen La (SS-B)	1074		m
72D06	III	259	gb:MUSGBPA	Beta-galactoside binding protein	539		m
69E11	II	460	gb:HUMCSF	Cleavage stimulation factor	500		h
70H07	II	263	gb:DMCPO612G	Cpo 61.2 gene, RNA binding protein	174x		dm
73A09	I	367	gb:MMCYCM	Cyclophilin = cyclosporin binding protein	1652	(0/ 2)	m
99A09	X	299	sp:RPC9_YEAST	DNA-directed RNA polymerases I and III	160x		y
38E12	IV	362	gb:MUSSRPA	Docking protein, SRP14	783		m
72A09	IV	200	gb:MMSRP54	Docking protein, SRP54	660f	(0/ 2)	m
73C05	IV	184	gb:CFSRP68	Docking protein, SRP68	364		do
90H06	IV	283	gb:CFSRP72	Docking protein, SRP72	672		do
92A11	II	440	gb:HSDOCKP	Docking protein, SRP receptor protein	1384f	(1/ 3)	h
C0E12	X	386	gb:MMEF1A	Elongation factor 1-alpha	1455f	(27/42)	m
94H05	X	431	gb:HSEF1B	Elongation factor 1-beta	1127f	(5/ 6)	h
69F05	II	416	gb:HSEF1G	Elongation factor 1-gamma	1136f	(2/ 4)	h
69E06	II	456	gb:RNEF2R	Elongation factor 2 (EF-2)	1516f	(7/10)	r
90E05	III	270	sp:p19216	Elongation factor EF-TS	117x		sp
77E12	II	273	gb:RATEFCVFGM	Elongation factor G	702f		r
B8B02	X	148	gb:S63912	hnRNP, A1-gamm isoform	793f	(2/ 3)	h
73C08	I	273	gb:HUMRNPC2A	hnRNP, C2 protein	657f	(2/ 4)	h

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
91F01	II	322	gb:MUSPHOSPHO	hnRNP, K homolog	1274f	(2/ 3)	m
66F06	IV	524	gb:MUSNUABPRO	hnRNP X	1796f	(1/ 2)	m
96D10	X	291	gb:HSHTFIIAS	hTFIIAs smallest TFIIA subunit	783		h
68H05	IV	322	gb:RATEIF2B	Initiation factor, eukaryotic	874f	(0/ 2)	r
88F06	II	325	gb:T27975	Initiation factor 2	937f	(1/ 2)	h
66B02	IV	384	gb:MMEIF4AI	Initiation factor eIF-4AI	1500f	(2/ 4)	m
B4F07	X	210	gb:MUSEIF4AL	Initiation factor eIF4A, long form	402		m
85E03	II	333	gb:HSTIF4GP	Initiation factor eIF-4 gamma	880f	(0/ 2)	h
66A06	IV	300	gb:HSTIF	Initiation factor nuk 34	703f	(0/ 2)	h
90G11	IV	279	gb:HSANAC	NAC, alpha; polypeptide binding protein	612f		h
94C09	X	375	sp:NFS1YEAST	NIFS-like 54.5 kd protein	132x		y
72H01	II	306	gb:RATP49A	Nuclear protein p47	738f	(0/ 2)	r
73G07	II	269	gb:MMCNBPMRB	Nucleic acid binding protein, cellular	1022		m
A5H08	X	260	gb:RNNAP57	Nucleolar protein NAP	478		r
95A10	X	431	gb:MUSN038A	Nucleolar protein N038	1640f	(4/ 5)	m
94A12	X	408	gb:MMPOLYABP	Poly(A) binding protein	1149	(2/ 3)	m
98B11	X	136	gb:MUSPDIA	Protein disulfide isomerase (ERp59)	309		m
77B02	IV	249	pi:S42639	RNA helicase-like protein	136x		to
A2H02	X	256	gb:HSRPII145	RNA polymerase II 14.5 kD subunit	290		h
83D02	IV	219	gb:HUMRPOLAA	RNA polymerase subunit hRPB 33	679		h
77B09	II	271	gb:MMRDBPA	RNPS1 RNA/DNA-binding protein	770		m
B0D03	X	179	gb:HUMSEC61B	Sec61-complex b-subunit	556		h
83D07	II	155	gb:S65406	Signal-anchor protein, type II	640		r
88B08	III	415	gb:HSU08815	Splicesomal protein, SAP 61	1309f		h
A2B11	X	169	gb:HUM9G8SF	Splicing factor, 9G8	547		h
71G03	II	499	gb:HUMSUIISO	Suilisol, translation initiation factor	1836f	(1/ 4)	h
B6E10	X	364	gb:HUMCYSTRNA	tRNA-cys, partial	1120f	(2/ 3)	h
87E05	IV	297	gb:HUMGLYCYL	tRNA-gly synthetase	998f	(1/ 2)	h
84D04	IV	347	sp:SYMCYEAST	tRNA-met synthetase, cytoplasmic	176x	(0/ 2)	y
86A07	IV	360	sp:SYFC	tRNA-phe synthetase a chain	144x		y
74G07	II	263	gb:HUMTHRSYNT	tRNA-thr synthetase	793		h
95D06	X	272	gb:MMWRSA	tRNA-trp synthetase (WRS)	1229		m
B2D01	X	261	gb:MMX16MR	X16, with nucleic acid binding motifs	1047	(1/ 2)	m
(Ribosomal proteins)							
94D04	X	415	gb:MMARPPO	Ribosomal acidic phosphoprotein PO	1504f	(16/22)	m
B5H10	X	419	gb:RRRPP1	Ribosomal acicic phosphoprotein P1	1252f	(9/14)	r
B6A09	X	250	gb:RRRPP2	Ribosomal phosphoprotein P2	883f	(2/ 6)	r
94G11	X	388	gb:MUSKE3B	Ribosomal protein (Ke-3)	1464f	(10/14)	m
A1B08	X	427	gb:HUMRSP	Ribosomal protein L1	1370f	(14/19)	h
69D02	II	425	gb:MMJ1PRO	Ribosomal protein L3, yeast J1 homolog	1361f	(8/13)	m
B5H03	X	335	gb:RATRPL5	Ribosomal protein L5	1531	(3/ 3)	r
92E05	II	440	gb:HSRPL6AA	Ribosomal protein L6	1249f	(5/10)	h
95C12	X	444	gb:MUSRPL7R	Ribosomal protein L7	1589	(4/ 7)	m
66G08	II	444	gb:MUSSURFB	Ribosomal protein L7a	1414f	(6/ 9)	m
90F06	II	380	gb:RRRPL8	Ribosomal protein L8	1257f	(2/ 4)	r
66F04	III	345	gb:RRRPL9	Ribosomal protein L9	1198f	(4/ 6)	r
70G04	I	118	gb:RRRPL11	Ribosomal protein L11	404f	(1/ 2)	r
97G10	X	335	gb:MUSRPL12A	Ribosomal protein L12	1281f	(5/ 5)	m
71A03	IV	240	gb:RNSDRPL13	Ribosomal protein L13	287		r
B8F03	X	401	gb:RNRPL13A	Ribosomal protein L13a	1436f	(5/ 9)	r
C0B02	X	164	gb:RNSDRPL15	Ribosomal protein L15	592f	(2/ 2)	r
75F09	I	420	gb:RRRPL17	Ribosomal protein L17	1251f	(4/ 8)	r
B4A07	X	375	gb:MUSRPL18A	Ribosomal protein L18	1386f	(3/ 7)	m
94B09	X	411	gb:RRRPL18A	Ribosomal protein L18a	1411f	(12/17)	r
67H11	I	400	gb:MUSL19RP	Ribosomal protein L19	1495f	(4/ 8)	m
68E10	I	369	gb:RATL21A	Ribosomal protein L21	1280f	(4/ 8)	r
97C07	X	212	gb:RNRPL22SD	Ribosomal protein L22	731f	(3/ 3)	r
B8D03	X	153	gb:RRRPL23	Ribosomal protein L23	643		r
B9G04	X	313	gb:RRRPL23A	Ribosomal protein L23a	1136f	(7/ 8)	r
72E09	I	373	gb:RNRPL24	Ribosomal protein L24	1267f	(2/ 3)	r

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
68F03	II	406	gb:MML26MRN	Ribosomal protein L26	1503f	(4/ 6)	r
92D02	IV	276	gb:RRRPL27	Ribosomal protein L27	879f		r
73E02	I	385	gb:MMRNAL28	Ribosomal protein L28	1435f	(1/ 3)	m
B8G03	X	391	gb:MMCCL3R	Ribosomal protein L29, yeast homolog	1420f	(4/ 6)	m
66C04	IV	379	gb:MUSRPLPSA	Ribosomal protein L30-2 pseudogene	1453f	(1/ 3)	m
62E11	IV	327	gb:RATRPL31R	Ribosomal protein L31	1138f	(3/ 4)	r
66A05	IV	324	gb:RRRPL34	Ribosomal protein L34	1093f	(3/ 4)	r
94F10	X	362	gb:RATRPL35AA	Ribosomal protein L35, 60S subunit	1198f	(4/ 5)	r
90C05	IV	277	gb:RNRPL35A	Ribosomal protein L35A	564f	(1/ 2)	r
73E06	I	331	gb:RNRPL36	Ribosomal protein L36	1154f	(3/ 9)	r
48C08	II	404	gb:RATRPL36A	Ribosomal protein L36a	858f	(1/ 2)	r
72G05	I	337	gb:RNRPL37	Ribosomal protein L37	1462	(1/ 3)	r
69A08	I	226	gb:MMRP37A	Ribosomal protein L37a	1052	(2/ 4)	m
73G11	I	306	gb:RNRIPRL38	Ribosomal protein L38	1135	(1/ 2)	r
B6D01	X	327	gb:RNRPL39	Ribosomal protein L39	903	(3/ 4)	r
B6A06	X	150	gb:RNRPL41	Ribosomal protein L41	455f	(2/ 2)	r
94D06	X	318	gb:MUSRP75A	Ribosomal protein, large subunit	1467	(3/ 6)	m
72E05	I	345	gb:MUSRPL4A	Ribosomal protein pseudogene rpL32-4A	1706	(1/ 3)	m
B8A10	X	360	gb:MMRIBPS3	Ribosomal protein S3	1651	(6/ 7)	m
75C10	I	387	gb:RNRNAS3A	Ribosomal protein S3a	1275f	(8/13)	r
72D09	II	549	gb:MUSRSP4	Ribosomal protein S4	2089f	(10/17)	m
B8G05	X	343	gb:RNRPS5	Ribosomal protein S5	1370	(7/ 8)	r
95G03	X	392	gb:MMRPS6	Ribosomal protein S6	1482f	(6/10)	m
69B11	I	301	gb:RRRPS7	Ribosomal protein S7	1046f	(5/ 8)	r
73E04	I	368	gb:MMRPS8	Ribosomal protein S8	1387f	(7/ 8)	m
B9A03	X	298	gb:RNRPS9	Ribosomal protein S9	1253	(3/ 6)	r
90A10	II	383	gb:RNRPS10R	Ribosomal protein S10	1125f	(5/ 6)	r
B9A02	X	307	gb:RATRPS11	Ribosomal protein S11	1085f	(5/ 5)	r
B5B07	X	283	gb:MMRPS12	Ribosomal protein S12	882	(4/ 4)	m
71F11	I	363	gb:RRRSP13	Ribosomal protein S13	1253f	(9/14)	r
66D10	III	333	gb:RNRPS14	Ribosomal protein S14	1193f	(6/ 8)	r
74C06	I	325	gb:MUSRIGA	Ribosomal protein S15	1485	(2/ 6)	m
B1F09	X	335	gb:RRRPS16	Ribosomal protein S16	1167f	(6/ 8)	r
72G12	I	434	gb:MUSRPS17	Ribosomal protein S17	1617f	(3/ 5)	m
B6E06	X	325	gb:RRRPS19	Ribosomal protein S19	1176f	(7/13)	r
C1B11	X	209	gb:RRRPS20	Ribosomal protein S20	579	(3/ 3)	r
B7H09	X	329	gb:RNRPS21	Ribosomal protein S21	1070	(2/ 2)	r
73G02	I	431	gb:RNRPS23	Ribosomal protein S23	1392f	(4/ 7)	r
B8B05	X	328	gb:HUMRPS24A	Ribosomal protein S24	917f	(2/ 3)	h
78C04	I	304	gb:RRRPS25	Ribosomal protein S25	977		r
B6A01	X	144	gb:RRRPS26	Ribosomal protein S26	1200	(2/ 2)	r
B8E08	X	299	gb:MMU11248	Ribosomal protein S28	1074f	(2/ 2)	m
B7A04	X	259	gb:RNRPS29	Ribosomal protein S29	816		r
(Heat shock proteins)							
95G07	X	428	gb:MMCCTBE	Cctb CCT, chaperonin containing TCP-1	1617f	(2/ 3)	m
94A03	X	339	gb:MMCCTEP	Ccte CCT, chaperonin containing TCP-1	1300f	(2/ 4)	m
02H09	X	266	gb:MMCCTHE	Ccth CCT, chaperonin containing TCP-1	1276		m
A6G11	X	145	gb:MMCCTQG	Cctq CCT, chaperonin containing TCP-1	630f		m
A6E07	X	206	gb:MMCCTZE	Cctz CCT, chaperonin containing TCP-1	561		m
69B01	I	440	gb:MMU09659	Chaperonin 10	1585	(1/ 2)	m
88D07	III	433	gb:MMHSP65R	Heat shock protein 65	1575	(0/ 2)	m
95C06	X	397	gb:MUSHSPCA	Heat shock protein 70 cognate	1457f	(7/12)	m
B3C10	X	181	gb:MMU08215	Heat shock protein 70-related NST-1	833		m
75E05	II	264	gb:MUSHSP84B	Heat shock protein 84	473f		m
90G12	II	377	gb:MUSHSP86A	Heat shock protein 86	1378f	(18/29)	m
87C05	IV	349	em:HSDNAJHOM	Heat shock protein, DnaJ homologue	834f	(0/ 2)	h
B9F10	X	326	gb:MUSPBP	Peptide-binding protin 74	1304f		m
77A08	I	321	gb:MUSTCPAB	Tcp-1-b = t complex polypeptide 1	936f		m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
Membrane-Associated							
(Receptor and membrane-associated)							
46E06	IV	95	gb:MMBAP37	B-cell receptor-associated protein, BAP	405		m
A2H04	X	239	gb:RNCD37	CD37 antigen	1042		r
A7A04	X	128	gb:MUSMCD63	CD63 antigen	442f		m
86C04	IV	452	gb:MUSLINE	Glycine receptor beta-subunit	521		m
72A12	IV	155	gb:RATGCA	Guanylyl cyclase A/atrial natriuretic rec.	267f		r
B0A03	X	243	gb:HSBAT2	HLA-B associated transcript, Bat2	272		h
88H04	II	254	gb:HSPHAPII	HLA-DR associated protein	459		h
86G03	III	266	gb:S63813	HT7 antigen, member of IGM superfamily	1064f	(1/ 3)	m
B0F06	X	124	sp:E16_HUMAN	Integral membrane protein E16 (probable)	111x		h
72D07	I	419	gb:MUSLAMR	Laminin receptor	1535f	(5/ 7)	m
83B05	II	154	em:S94687	Mph, poliovirus receptor homolog	752		m
82E06	II	253	gb:MMN10R	N10, nuclear hormonal binding receptor	1020		m
B2D09	X	269	gb:R75073	Phosphatidylethanolamine-binding protein	882		m
85C03	II	326	gb:S85655	Prohibitin	445		h
75E04	IV	255	gb:RATYWKII	Sperm membrane protein	201f		r
A4H11	X	175	gb:MUSTRAB	Tumor rejection antigen P815A	669f	(1/ 2)	m
(Transporters)							
91F02	II	306	gb:SCMPRM1A_1	Amino acid permease	104x		sm
59G11	IV	411	gb:RATAP50A	Brain AP50, coat assembly protein	572		r
A9H03	X	129	gb:RNCHOT1	CHOT1, choline transporter	550		r
86E12	IV	414	gb:BTZCOP	Coatomer	990		bo
A4G05	X	281	pi:1811235A	Cystic fibrosis conductance regulator	139x		xe
93E05	X	358	gb:MUSGLUTRN	Glucose transport protein, facilitated	1609	(2/ 2)	m
A4G02	X	133	gb:MOUSGLUT10	Glucose transporter, GLUT3	337	(1/ 2)	m
85G11	IV	382	sp:AMT1_YEAST	Mitochondrial transporter ATM1 precursor	152x		y
A8G07	X	110	pi:A47131	Na(+)-dependent neutral amino acid transp.	120x		h
87B05	IV	404	gb:HSP63	P63, transmembrane protein	392		h
95G01	X	121	pi:1911369A	Perforin, HP-10 protein	117x		h
A3A08	X	113	gb:RATATPIF1	Plasma membrane Ca2+ ATPase-isoform 1	430		r
C1A10	X	334	gb:BTPLASPR	Plasmalemmal porin	1233		bo
B9D02	X	334	gb:RATPRPU	Proton pump polypeptide	539		m
A4G09	X	334	gb:MUSRCA1	Reticulocalbin, calcium binding protein	1399		m
82F01	IV	412	gb:MUSANNVII	Synexin, membrane-binding protein	1149		m
A0F12	X	225	gb:BOVHATPA	Vacuolar H+ ATPase 31kDa subunit	900		bo
B2F11	X	415	gb:BOVATPX	Vacuolar H+-ATPase	1446		bo
92C05	II	358	gb:MUSMVP	VMP, vacuolar H (+)-ATPase	1461	(0/ 3)	m
Other-metabolism							
(DNA metabolism)							
66D08	IV	406	gb:HUMACT1A	Activator 1, replication factor	477		h
82A09	IV	453	gb:MUSAPX	APEX nuclease, DNA repair enzyme	573		m
A5C06	X	194	gb:HUMAUTANT	Autoantigen, nuclear localization signal	428		h
74E05	I	269	em:HSBM28	BM28, nuclear early S-phase protein	1101		h
86E08	IV	451	gb:MMU04674	DNA ligase I	1617f	(0/ 2)	m
77G09	IV	263	gb:MUSTOPI	DNA topoisomerase I	529f		m
97C11	X	366	gb:MUSTOP2	DNA topoisomerase II	1385f	(2/ 2)	m
67B08	IV	494	gb:RATHIS2AZ	Histone H2A.Z	888f	(1/ 5)	h
67C04	II	449	gb:MMH33REP	Histone H3.3, replacement variant	1208f	(0/ 4)	m
86A03	IV	425	gb:RNHI0SH1	Histone H10	404		r
66A10	IV	452	pi:JU0038	HMG = high mobility group protein	113x		te
94G03	X	329	gb:MMHMG2	HMG 2	972f	(1/ 4)	m
84H01	IV	240	gb:MUSHMGIY	HMG-I(Y)	1339f	(0/ 5)	m
A1B07	X	290	gb:S50213	HMG1-related DNA-binding protein	1134f	(1/ 2)	m
82D12	IV	296	gb:MUSNAP1	Nucleosome assembly protein 1	239f		m
A0F02	X	142	gb:MMP1M	P1 protein, DNA polymerase alpha assoc.	528f		m
85H01	IV	386	gb:HSP1MCM3	P1-Mcm3, replication protein	988f		h
C0B10	X	399	gb:MMPCNAG	Proliferating cell nuclear antigen	1328f	(2/ 3)	m
38F06	IV	357	gb:MMU12270	Pendulin	1393		m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
94B03	X	379	gb:CIURCC1	RCC1, involved in chromosome condensation	1553		ha
75H02	IV	249	gb:HUMPOLACCB	Replicative polymerase accessory protein	450f		h
B6F02	X	285	gb:MUSBPP9	Single stranded DNA binding protein p9	1102f	(2/ 4)	m
94H02	X	311	gb:MUSSMBPG	Sm-B protein, nuclear antigen	1519		m
75C04	IV	219	gb:MUSTC4PROT	TC4, GTP-binding protein	521f		m
93F03	X	136	gb:MUSTHYS1	Thymidylate synthase	635		m
88E11	III	354	gb:HUMUMPS	UMP synthase	545		h
66F01	IV	485	gb:AGMUVDAMBP	UV-damaged DNA-binding protein	1432f		m
(Repetitive DNA and virus-related sequences)							
60F06	IV	152	gb:MUSB2REB	B2 repetitive sequence	535		m
88B03	IV	296	gb:HUMREPSTS	DNA repeat regions	287		h
B7B05	X	329	gb:HSAG1	HSAG-1 middle repetitive element	340f	(5/ 7)	h
A2F11	X	219	gb:MMIAP103	Integrase, intracisternal A-particle	850f		m
84G11	IV	298	gb:HSU04847	Ini1, integrase for incoming viral DNA	595f	(0/ 2)	h
93G02	X	281	gb:MMIAPIL3	Intracisternal A-particle IAP-IL3	1369		m
B6D02	X	242	gb:MUSRSA	LLRep3 protein from a repetitive element	900f	(1/ 2)	m
30B10	IV	274	gb:S69706	Ms6-hm, minisatellite locus	378f		m
B1C04	X	366	gb:MMBVL1	Virus-like (VL30) retrotransposon BVL-1	603		m
09F04	IV	138	gb:MUSRSRGO	VrDNA, repeated element	580		m
(Proteases and protease inhibitors)							
72F01	I	384	gb:HUM26SPSIV	26S protease (S4) regulatory subunit	1356		h
C0D09	X	223	gb:RNU10861	Calpain small subunit	953		r
86E10	IV	438	gb:MMFAU	FAU, ubiquitin-like-S30 fusion protein	1462f	(3/ 7)	m
84H05	IV	368	gb:HSHE4MR	HE4, major epididymis-specific protein	430		h
96E07	X	404	gb:S65367	Leucine aminopeptidase 2	1097		bo
B1A04	X	343	gb:RATMPP	Mitochondrial matrix processing protease	1391		r
77F09	I	366	gb:S69034	Preprocathepsin B	1553		m
93H09	X	423	gb:RATPRORD	Proteasome subunit R-DELTA	1133	(1/ 2)	r
96B04	X	263	gb:RATPROC2A	Proteasome C2 component	1045		r
A2A10	X	118	gb:RNPTSC9	Proteasome subunit RC9	271		r
A2C05	X	200	gb:RATPSRC7I	Proteasome subunit RC7-I	736		r
72G04	I	307	gb:MMUBIQU	Ubiquitin	1173f	(8/13)	m
70A08	II	175	gb:MUSUBA1	Ubiquitin activating enzyme E1	536	(0/ 2)	m
83A04	II	287	gb:HUME2EPI	Ubiquitin carrier protein (E2-EPF)	1129		h
B8D06	X	316	sp:UBC5_DROME	Ubiquitin-conjugating enzyme E2-17 kd	974		h
(Other metabolism)							
54C03	IV	308	gb:ECOMALS	1,4-alpha-D-glucanohydrolase (= *malS*)	505		ec
A4H08	X	407	gb:T10017	20b-hydroxysteroid dehydrogenase	872f		h
B9H10	X	257	gb:HSADE2H1	ADE2H1, homologous to SAICAR synthetase	988	(1/ 2)	h
68B12	IV	393	gb:RATALFUC	Alpha-L-fucosidase	208		r
28H08	IV	252	gb:MUSASSB	Argininosuccinate synthetase	540		m
82B03	IV	403	gb:LAASP	Asparaginase	321f		lu
77E04	IV	289	gb:RNU07202	Asparagine synthetase	916f	(2/ 3)	r
72G09	II	276	gb:MUSNB14G1	Beta 1-4 galactosyltransferase	1066		r
A0A11	X	184	gb:MUSBGAL45	Beta-galactosidase, BGAL	509		m
71E08	IV	267	gb:RATBGASTRB	Beta-galactoside-a 2,6-sialyltransferase	257f		r
83F08	IV	433	gb:RATC1H4SY	C-1-tetrahydrofolate synthase, cytoplasm	1458	(1/ 2)	r
93E04	X	145	gb:MUSCALNEXI	Calnexin	689		m
B7F11	X	260	gb:MUSCPP	Coproporphyringogen oxidase	1019		m
A1B05	X	240	gb:RATCGL	Cystathionine gamma-lyase	946	(2/ 2)	r
83C04	IV	345	gb:RATAKGE2	Dihydrolipoamide succinyltransferase	301		r
95F09	X	325	gb:RATFTBS	Farnesyltransferase beta subunit	1517		r
B7H06	X	300	gb:MMGSHPX	Glutathione peroxidase	1269	(1/ 2)	m
93C03	X	258	gb:MMMRNAGR	Glutathione reductase	1236		m
A1D07	X	149	gb:MMGSTII	Glutathione S-transferase II	709		m
B8D12	X	362	gb:COQ1_YEAST	Hexaprenyl pyrophosphate synthetase pr	123x		y
67D10	IV	217	gb:HUMGLYI	Lactoyl glutathione lyase	296	(0/ 2)	h
A5G12	X	113	gb:MML14L	Lectin, L14	457		m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
B4D09	X	363	gb:RATMETI	MT-I = metallothionein I	1150f	(6/ 9)	r
75A02	IV	250	gb:HUMBHAB	N-acetyl-beta-glucosaminidase	202		h
82A02	IV	394	gb:S70932	N-acetylgalactosamine-6-sulfate sulfatase	701		h
96A02	X	274	gb:RATODCB	Ornithine decarboxylase antizyme	1035f	(3/ 3)	r
96E10	X	337	gb:S56546S1	Paraoxonase/arylesterase	219		h
77B05	I	325	gb:HUMPDEC	Phosphodiesterase	2627		h
85D09	IV	468	gb:RATPHGP	Phospholipid hydroperoxide glutath. perox	1371f	(1/ 4)	r
93A05	IV	308	gb:T08379	Phosphoserine aminotransferase	514f		h
B9E11	X	258	gb:RNBINP	Resiniferatoxin binding protein, cytosolic	861f	(1/ 2)	r
86A11	IV	428	gb:MUSSAHH	S-adenosyl-L-homocysteine hydrolase	1542f	(1/ 2)	m
A4D12	X	297	gb:S36200	Saposin = sphingolipid activator protein	1382		m
89A09	IV	273	gb:MUSSAT	Spermidine/spermine N1-acetyltransferase	1255		m
B8H11	X	328	gb:MMSOD1	Superoxide dismutase, Cu-Zn	1232	(1/ 3)	m
77E01	II	276	gb:MMNSODR	Superoxide dismutase, manganese	278		m
91B07	IV	256	gb:PIGTTF	Thioltransferase	277		p
B6D09	X	265	gb:RNTR	Thioredoxin	957f	(1/ 3)	m
87D11	III	333	gb:HSPM5	Thyroid peroxidase = TPO, M5 protein	1212		h
(Not classified)							
86F02	IV	471	gb:RAT5E5ANTE	5E5 antigen, adult brain cDNA	487f		r
29D10	IV	234	gb:HUMANONYMO	Anonymous gene, 2/meningioma gene	416f		h
93E12	X	405	gb:MML21KD1	21 kd polypeptide	1884	(3/ 3)	m
93C04	X	279	gb:HUMCG1X	CG1 protein, containing heptad repeat	667		h
94C06	X	414	gb:MMU12403	Csa-19	1388f	(2/ 4)	m
96D03	X	461	gb:HUMRNAE	CTG-B33, triplet repeat gene in brain	1229f	(3/ 7)	h
B2D11	X	217	gb:MMEEP	E46 protein, brain specific	670		m
A2G03	X	202	gb:HSDAP1	DAP-1, cell death-associated cDNA	740		h
81H12	IV	395	gb:MMGAS5MR	gas5, growth arrest specific gene	712		m
B3B01	X	129	gb:MMH5RNA	H5, brain-specific expression	298f	(1/ 2)	m
77F01	IV	228	gb:T03846	Huntington disease region, clone gt123	348f		h
B6H10	X	347	sp:YM68_CAEEL	Hypothetical 208.3 kd protein K12h4.8	232x		ce
74G08	I	329	pi:S46802	Hypothetical protein, YHR004c	101x		y
85F11	IV	473	gb:DOGSNVD17A	Inserted sequence in a virus vector	422		do
78A11	I	316	gb:RNU21718	Intestinal epithelium proliferating cell	965		r
96B08	X	100	gb:MUSNEDD4	NEDD-4, developmentally down-regulated	446		m
A0E06	X	207	gb:CCOCP2	OCP-2, auditory organ-specific protein	568f		gp
B6B04	X	258	gb:MUSOSF3	OSF-3	947f	(1/ 3)	m
72H08	I	349	gb:MMPH34MRA	pH 34, down-regulated in EC cells	1354f	(4/ 8)	m
A3B02	X	120	gb:HSPP15	PP15, placental protein 15	492		h
90H02	III	258	gb:U00995	TA1, shows onco-fetal expression pattern	396		r
78A10	I	367	gb:MUSTCTEX	tctex-1 mRNA, t complex sterility locus	1480		m
B4F01	X	331	gb:RNSSRDSUB	TRAP-complex delta	929		r
04A10	II	260	gb:HUMOGC	Unknown protein from clone pHGR74 mRNA	217f		h
(ESTs matching database sequences of uncharacterized cDNAs)							
B2C09	X	230	gb: CEU19615	LET-858	223f		ce
93A08	X	217	gb: M89360	Mixed stage hermaphrodite cDNA library	242f		ce
A4D04	X	335	de:46606 Z33595	cDNA Atrium	1304		h
78C09	IV	247	de:81299 T31266	cDNA Brain	472		h
95H03	X	306	de:85403 T35370	cDNA Brain	1161		h
A1C11	X	313	de:281183 H21245	cDNA Brain, adult	1144		h
81G04	IV	382	de:351653 H50443	cDNA Brain, adult, N2b5HB55Y	348		h
93F09	X	269	gb: T10047	cDNA Brain, 3-month old neonate	405f		h
A2A05	X	224	de:29239 T10033	cDNA Brain, 3-month old neonate	421		h
97A04	X	255	de:29527 T10321	cDNA Brain, 3-month old neonate	386		h
89D09	IV	275	gb: T07272	cDNA Brain, fetal	265f		h
89B06	IV	271	de:329711 D53483	cDNA Brain, fetal	289		h
B7C10	X	300	de:342681 D56188	cDNA Brain, fetal	907		h
69C04	I	221	de:1399 M77859	cDNA Brain, fetal	809		h
94H11	X	302	de:2095 M85644	cDNA Brain, fetal	748		h

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
62B11	IV	305	de:2357	M85901 cDNA Brain, fetal	845		h
73F04	I	373	de:25245	T06191 cDNA Brain, fetal	1256		h
39G08	IV	323	de:25308	T06254 cDNA Brain, fetal	234		h
46G08	IV	276	gb:	T03819 cDNA Brain, infant	267f		h
C1A08	X	232	gb:	T08425 cDNA Brain, infant	298f		h
69G01	II	362	gb:	T08516 cDNA Brain, infant	476f		h
03H03	IV	177	gb:	T08743 cDNA Brain, infant	306		h
93D12	X	132	de:27103	T08049 cDNA Brain, infant	265		h
89D01	IV	278	de:27232	T08178 cDNA Brain, infant	482		h
83F07	IV	343	de:27421	T08367 cDNA Brain, infant	435		h
83B01	IV	293	de:28268	T09214 cDNA Brain, infant	319		h
62H03	IV	225	de:41453	T15733 cDNA Brain, infant	255		h
B2E12	X	169	gb:	R43141 cDNA Brain, infant, 1NIB	259f		h
B1A07	X	373	de:266856	H06976 cDNA Brain, infant, 1NIB	959		h
A9G09	X	138	de:270328	H10448 cDNA Brain, infant, 1NIB	367		h
B9E05	X	276	de:275288	H15350 cDNA Brain, infant, 1NIB	1022		h
B1G08	X	276	de:276362	H16424 cDNA Brain, infant, 1NIB	601		h
86F04	IV	498	de:182458	R11879 cDNA Brain, infant, 1NIB	593		h
B8H04	X	344	de:184246	R13667 cDNA Brain, infant, 1NIB	236		h
85E02	IV	449	de:184346	R13767 cDNA Brain, infant, 1NIB	305		h
C0B04	X	228	de:185355	R14671 cDNA Brain, infant, 1NIB	707		h
67D01	IV	425	de:188457	R18663 cDNA Brain, infant, 1NIB	719		h
89A11	IV	270	de:194605	R24560 cDNA Brain, infant, 1NIB	541		h
28C12	IV	336	de:207532	R37470 cDNA Brain, infant, 1NIB	229		h
B2G02	X	330	de:232608	R55871 cDNA Brain, infant, 1NIB	548		h
73H12	II	336	de:235919	R59164 cDNA Brain, infant, 1NIB	1258		h
81H03	IV	288	de:125172	T66121 cDNA Brain, infant, 1NIB	351		h
69H03	IV	256	de:137715	T75352 cDNA Brain, infant, 1NIB	212		h
86A12	IV	481	de:141026	T78663 cDNA Brain, infant, 1NIB	271		h
74B10	IV	245	de:142627	T80259 cDNA Brain, infant, 1NIB	235		h
04G11	IV	253	de:131952	F09837 cDNA Brain, normalized infant	221		h
75D04	IV	229	de:122553	F06892 cDNA Brain, normalized infant	529		h
68G06	IV	253	de:132291	F10164 cDNA Brain, normalized infant	651		h
75C01	III	397	de:148188	F11161 cDNA Brain, normalized infant	209		h
72D10	IV	265	de:148543	F11505 cDNA Brain, normalized infant	352		h
94F05	X	252	de:151746	F12834 cDNA Brain, normalized infant	612		h
B9H02	X	299	de:151992	F13077 cDNA Brain, normalized infant	379		h
66D02	IV	390	de:152004	F13089 cDNA Brain, normalized infant	479		h
B7D07	X	255	de:64618	Z39529 cDNA Brain, normalized infant	545		h
41B03	IV	313	de:66130	Z40005 cDNA Brain, normalized infant	204		h
84C03	IV	431	de:66608	Z41923 cDNA Brain, normalized infant	334		h
77F04	IV	243	de:69598	Z42861 cDNA Brain, normalized infant	1008		h
90D01	IV	288	de:70401	Z43511 cDNA Brain, normalized infant	589		h
69C03	IV	318	de:73576	Z44392 cDNA Brain, normalized infant	457		h
87F12	IV	326	de:73642	Z44456 cDNA Brain, normalized infant	306		h
A9E10	X	210	de:73757	Z44559 cDNA Brain, normalized infant	712		h
74C08	IV	245	de:73902	Z44689 cDNA Brain, normalized infant	504		h
83G02	IV	176	de:73931	Z44717 cDNA Brain, normalized infant	313		h
B2D10	X	217	de:74905	Z45775 cDNA Brain, normalized infant	334		h
96A01	X	300	de:42872	T17152 cDNA Brain, normalized infant	527		h
81B04	IV	396	de:286631	H26120 cDNA Breast 3NbHBst	300		h
87H04	IV	252	de:289436	H28877 cDNA Breast 3NbHBst	423		h
84A11	IV	333	de:305596	H44545 cDNA Breast 3NbHBst	225		h
72C05	IV	230	de:306547	H45496 cDNA Breast 2NbHBst	484		h
56G09	IV	350	de:248534	R71676 cDNA Breast 2NbHBst	1319		h
72D01	IV	251	de:252909	R76001 cDNA Breast 2NbHBst	452		h
98B06	X	181	de:61671	T24841 cDNA Colorectal cancer	758		h
B4F08	X	147	de:61612	T24782 cDNA Colorectal cancer	297		h
87F11	IV	285	de:61603	T24773 cDNA Colorectal cancer	670		h
84C06	IV	470	de:33004	T12405 cDNA Heart	772		h
85H11	IV	373	de:21833	Z22148 cDNA Heart	253		h

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
68E12	IV	421	de:153 M62106	cDNA Hippocampus	405		h
86B01	IV	479	de:137156 T74799	cDNA Liver	518		h
09A01	IV	358	de:17930 D12310	cDNA Liver HepG2 cell line	313		h
71G09	I	141	de:349857 H48648	cDNA Liver + spleen, fetal	553		h
82E12	IV	372	de:365087 H63637	cDNA Liver + spleen, fetal	768		h
85D12	IV	334	de:367140 H65552	cDNA Liver + spleen, fetal	878		h
77H02	IV	248	de:369512 H67924	cDNA Liver + spleen, fetal	289		h
B7D01	X	156	de:177990 R07694	cDNA Liver + spleen, fetal	239		h
94D05	X	378	de:178990 R08694	cDNA Liver + spleen, fetal	858		h
94C10	X	285	de:53401 T23879	cDNA Liver + spleen, fetal	347		h
40C05	II	312	de:131266 T70346	cDNA Liver + spleen, fetal	756		h
28A03	IV	316	de:170110 T99954	cDNA Liver + spleen, fetal	425		h
A6E10	X	243	de:19458 Z20973	cDNA Liver tissue	516		h
B6C11	X	283	gb: T82352	cDNA Lung	306f		h
85H12	IV	452	de:91286 T40482	cDNA Lung	206		h
55B01	IV	323	de:159834 T89713	cDNA Lung	253		h
B8F07	X	446	de:161789 T91668	cDNA Lung	1061		h
95B11	X	205	de:85603 T35570	cDNA Lung	762		h
A5H02	X	141	gb: HUMORF2	cDNA Myeloblast cell line, KG-1	543		h
73C12	II	244	gb: HUMORFT	cDNA Myeloblast cell line, KG-1	976		h
A2E12	X	110	gb: HUMKIAAP	cDNA Myeloblast cell Line, KG-1	275		h
73G01	I	248	gb: HUMRSC508	cDNA Myeloblast cell line, KG-1	677		h
A0B07	X	240	gb: HUMRSC765	cDNA Myeloblast cell line, KG-1	389		h
70F12	III	266	gb: T10491	cDNA Pancreatic islet	261f		h
B5F04	X	257	de:76638 T27396	cDNA Pancreatic islet	383		h
A6D06	X	322	de:76687 T27445	cDNA Pancreatic islet	1103		h
30A09	IV	200	de:354683 H53457	cDNA Pineal gland N3HPG	486		h
95A05	X	274	de:201172 R31127	cDNA Placenta Nb2HP	877		h
93F04	X	252	de:206411 R36350	cDNA Placenta Nb2HP	823		h
88H08	III	411	de:239444 R62689	cDNA Placenta Nb2HP	996		h
B7G09	X	313	de:241163 R64408	cDNA Placenta Nb2HP	212		h
93E09	X	377	de:253160 R76252	cDNA Placenta Nb2HP	967		h
73B09	I	393	de:257284 R80348	cDNA Placenta Nb2HP	1220		h
55H12	IV	315	de:340878 D58859	cDNA Placenta	369		h
85A11	IV	446	de:101951 T48669	cDNA Placenta	1183		h
A2G11	X	240	de:105927 T52645	cDNA Placenta	793		h
55D07	IV	246	de:110004 T56721	cDNA Placenta	384		h
A0H07	X	185	de:85222 T35189	cDNA Prostate gland	714		h
67A05	IV	431	de:341406 R93867	cDNA Retina N2b4HR	638		h
87H12	IV	268	gb: HUMMG44A	cDNA Retina MG44	399		h
71A05	IV	267	de:30104 Z24885	cDNA Skeletal muscle	302		h
B4B11	X	349	de:30457 Z25237	cDNA Skeletal muscle	592		h
83H01	IV	297	de:38616 Z28510	cDNA Skeletal muscle	464		h
93B12	X	258	de:38936 Z28866	cDNA Skeletal muscle	237		h
30B05	IV	202	de:99896 T48083	cDNA Spleen, fetal	248		h
B5C02	X	246	de:108345 T55062	cDNA Spleen, fetal	677		h
90B02	IV	294	de:109368 T56085	cDNA Spleen, fetal	574		h
B2E02	X	261	de:113268 T59201	cDNA Spleen, fetal	215		h
94E10	X	116	de:86120 T36087	cDNA Testis	468		h
95B02	X	275	gb: HSAAACYBP	cDNA UK-HGMP sequence	289f		h
83B04	II	163	de:81959 T31926	cDNA Uterus	234		h
78B06	I	254	de:82031 T31998	cDNA Uterus	851		h
66D03	IV	466	de:84175 T34142	cDNA White blood cells	1257	(0/ 2)	h
A5D11	X	167	de:84276 T34243	cDNA White blood cells	611		h
B5D05	X	173	de:84541 T34508	cDNA White blood cells	488		h
84D06	IV	438	de:84581 T34548	cDNA White blood cells	298		h
59A12	IV	339	de:252108 R74679	cDNA Brain	644		m
A2H05	X	222	de:252032 R75562	cDNA Brain	616		m
78F03	II	243	de:251491 R75592	cDNA Brain	265		m
42E06	IV	85	gb: MUSKA10X	cDNA Erythroleukemia	200f		m
88B04	IV	293	de:40925 L26824	cDNA Testis, lambda unizap	796		m

TABLE II (continued)

ESTs[a]	Class[b]	Seq[c]	DB:ID[d]	Putative identification[e]	Score[f]	Freq[g]	Sp[h]
B8H06	X	311	de:92462 Z31116	cDNA Testis T-ZAP	364		m
66G12	IV	246	de:92582 Z31236	cDNA Testis T-ZAP	201		m
B2C02	X	236	de:92586 Z31240	cDNA Testis T-ZAP	673		m
95C03	X	336	de:217233 R47168	cDNA Incisor	1555		r
B7E01	X	276	de:292012 H31388	cDNA PC-12 cells, untreated R...	967		r
91E02	IV	147	de:293030 H32406	cDNA PC-12 cells, untreated R.	339		r
59H07	IV	368	de:293430 H32806	cDNA PC-12 cells, untreated R...	1276		r
04A06	IV	105	de:293815 H33191	cDNA PC-12 cells, NGF-treated...	491		r
73G03	I	313	de:294015 H33391	cDNA PC-12 cells, NGF-treated...	720		r
93H02	X	245	de:294257 H33633	cDNA PC-12 cells, NGF-treated...	479		r
C0B05	X	295	de:294648 H34024	cDNA PC-12 cells, NGF-treated...	483		r
87F08	IV	295	de:294965 H34341	cDNA PC-12 cells, NGF-treated...	893		r
87A03	IV	437	de:295787 H35163	cDNA PC-12 cells, untreated R...	602		r
29A09	IV	247	de:295804 H35180	cDNA PC-12 cells, untreated R...	410		r
90E09	IV	311	de:295906 H35282	cDNA PC-12 cells, NGF-treated...	605		r
67C12	II	395	de:296398 H35774	cDNA PC-12 cells, untreated R...	285		r
Other-secreted proteins							
(Secreted proteins)							
86E11	II	273	gb:MUSAPOE	Apolipoprotein E	873f	(0/ 2)	m
B4B12	X	353	gb:MUSFERLA	Ferritin light chain	1582	(20/25)	m
A2H01	X	316	gb:RATREBT	Renin-binding protein	562		r
(Cytokines and growth factors)							
B6A10	X	313	gb:MUSAG	Acrogranin = acrosomal glycoprotein	1205f	(2/ 3)	m
81D09	IV	405	gb:MMU10118	Endothelial-monocyte activating protein	1274f		m
88H12	IV	367	gb:MUSCRIPTO	Cripto, Heart development-related protein	1321f		m
86A09	IV	493	gb:MUSEGFPRE	Heparin-binding EGF-like growth factor	1258f		m
40E01	IV	249	gb:HUMHDGF	Hepatoma-derived growth factor	412f		h
03B11	IV	229	gb:MUSIFNBB	Interferon-beta	569f		m
86H04	IV	271	gb:HUMINTERFN	Interferon-gamma	381f		h
29B05	IV	267	gb:MUSMKPG	Midkine, growth/differentiation factor	845		m
75B03	I	313	gb:MMMMIHFA	MIF, pituitary-derived cytokine	1029f	(1/ 3)	m
03C06	IV	225	gb:MMVALPRO	Valosin-containing protein	419f	(1/ 2)	m

TABLE II. Classification of identified ESTs was performed as described (*7, 8*), but with minor modifications. [a]Name of identified EST, when redundantly prepared, one of the ESTs is shown as a representative. As cloning vectors, λZAPII was used for ESTs from 00D08 to 62H03, and λuni-ZAP for all the other ESTs. [b]I, II, III, IV, and X represent classes I, II, III, IV, and X, respectively (see Table I) (*10*). [c]Sequenced lengths are shown in nt numbers. [d]Databases are indicated by the following symbols: de, DBEST; dj, DDBJ; em, EMBL; gb, GenBank or GenBank-updated; pi, PIR; pr, PRF; sw, Swiss-Prot. ID indicates the accession number of the matched sequence. [e]Names of identified genes, proteins, DNA sequences or sources of the cDNAs. [f]Scores indicate BLASTN scores, those marked with f indicate the FASTA opt scores, and those with x, the BLASTX scores. [g]Numbers in parentheses indicate the numbers of redundantly prepared ESTs; the first numbers indicate those prepared from class X, and the second, the total numbers prepared in this work. [h]Symbols are as follows: bo, bovine; ce, *Caenorhabditis elegans*; ch, chicken; cr, *Chlamydomonas reinharditi*; do, dog; dm, *Drosophila melanogaster*; ec, *E. coli*; go, goat; gp, guinea pig; h, human; ha, hamster; lu, *Lupinus arboreus*; m, mouse; p, pig; r, rat; rb, rabbit; sh, sheep; sm, *Schistosoma mansoni*; sp, *Spiroplasma citri*; te, *Tetrahymena*; to, tobacco; xe, *Xenopus laevis*; y, yeast. [i]ESTs matching mitochondrial DNA are further classified as to coding regions (*16*). Since only one of the redundantly prepared ESTs is shown in this table, we prepared a table showing the classification of all identified ESTs. On request, we will supply this table.

and A6C08) seemed to correspond to transcripts of the L-strand, while all the others corresponded to ones of the H-strand (Table II). Although the number of identified ESTs was small, we found ESTs corresponding to various ORFs present in mitochondrial DNA, other than those corresponding to the NADH dehydrogenase 6 gene, which is transcribed from the L-strand into a part of a large RNA (*16*). Since L-strand coded polyadenylated RNAs were reported to be short-lived RNAs (*16*), the profile of mitochondrial DNA expression obtained on analyses of ESTs seems to reflect the expression patterns and the properties of transcripts of mitochondrial DNA.

Homologies to repetitive and/or provirus-related sequences were found in only 18 ESTs, accounting for approximately 1% of the identified ESTs (Table II); this frequency is relatively low compared with data in the literature (*7, 9*). For example, a human fetal-brain cDNA library and a human fetal-lung cDNA library were reported to contain 11.9 and 2.7% of *Alu*-like sequences, respectively (*7, 9*). It was also reported that transcripts from the B1 and B2 repeat sequences are abundantly expressed in early mouse embryos (*17-19*), being estimated to account for approximately 5 to 6% of the total number of poly(A)$^+$ RNAs in early blastocysts (*19*). This dissimilarity may

TABLE III. **ESTs matching known genes.**

a) ESTs matching known genes in lower organisms

Query = A4G11, Sbjct = Protein kinase NPK1 (STE11 homolog), tobacco
BLASTX score: 118, Identities = 19/38 (50%), Positives = 27/38 (71%)

```
Query:   105 TGIRSVTGTPYWMSPXVISGEXYGRKADVWSLXCTVVE 218
             TG +S+ GTPYWM+P VI    +   AD+WS+ CT++E
Sbjct:   256 TGAKSMKGTPYWMAPEVILQTGHSFSADIWSVGCTIIE 293

                  Identities = 18/27 (66%), Positives = 23/27 (85%)

Query:     3 RNIKGANILRDSPGNVKLGDFGASKRL 83
             R+IKGANIL D+ G +KL DFGASK++
Sbjct:   223 RDIKGANILVDNKGCIKLADFGASKKV 249
```

Query = A6C09, Sbjct = Dynein, *Chlamydomonas reinharditi*
BLASTX score: 165, Identities = 29/72 (40%), Positives = 45/72 (62%)

```
Query:     1 VSKSCENYFQRYRRRAHVTPKSYLSFINGYKSIYTDKVKYINXQAERMNIGLDKLMEASXSLAKLSQDLAVK 216
             V+ +C+ YF++YRR  +VTPKSYLSF+ GYK +Y  K  +    A ++ +   K+ E    + K+  +LAVK
Sbjct:  2999 VTAACKEYFEKYRRYVYVTPKSYLSFLQGYKELYAKKWSFTKELAYQIEVACQKMFEPKADVNKMKAELAVK 3070
```

Query = A4G05, Sbjct = Cystic fibrosis conductance regulator, *Xenopus laevis*
BLASTX score: 139, Identities = 23/68 (33%), Positives = 43/68 (63%)

```
Query:    13 EPNPLQDANLXSPVSFWWLNPLFKTGHKRRLEEDDMFSVLPEDRSKHLGEELQRYWDKELLXAKKDSR 216
             +  PL+ A++ S + F W  P+   G+++RLE  D++ + P D + +L E L+R WD+E+  +KK+ +
Sbjct:     2 QKTPLEKASIFSQIFFSWTKPILWKGYRQRLELSDIYQIHPGDSADNLSERLEREWDREVATSKKNPK 69
```

Query = 77B02, Sbjct = RNA helicase-like protein, tobacco
BLASTX score: 136, Identities = 26/49 (53%), Positives = 34/49 (69%)

```
Query:    86 LKTVINYDVARDIDTHTHRIGRTGRAGEKGVAYTLLTPKDSNFAGDLVR 232
             ++ VINYD    I+ + HRIGRTGRAG  G+AYT  + +DS  A DLV+
Sbjct:   458 IRVVINYDFPTGIEDYVHRIGRTGRAGASGLAYTFFSDQDSKHALDLVK 506

                  Identities = 10/10 (100%), Positives = 10/10 (100%)

Query:    35 VLVATDVAAR 64
             VLVATDVAAR
Sbjct:   442 VLVATDVAAR 451
```

b) ESTs showing relatively low similarities to known genes in mammals

Query = 75E04, Sbjct = Sperm membrane protein (YWK-II), rat
BLASTN score: 201, Identities = 42/46 (91%)

```
Query:   127 GATGTTAAGGAAATCATTTTCAATGCTGAGAGAGTTGGAGGCCTTG 172
             |||||||||||||| ||||||||||||||||||||| || || |
Sbjct:   262 GATGTTAAGGAAATGATTTTCAATGCTGAGAGAGTTGGGGGTCTGG 307

                  Identities = 33/39 (84%)

Query:    14 AAGGCTCTGGAATGGCAGAAGGAGACGGGGGACTGATTG 52
             |||| |||||||||||||   ||| ||||| |||||||
Sbjct:   155 AAGGATCTGGAATGGCAGAGCAAGATGGGGGCCTGATTG 193

                  Identities = 27/29 (93%)

Query:   227 GCAATGCCCTTATTGGCTTGCTGGTTATC 255
             ||| ||||||||||||||||||||| |||
Sbjct:   359 GCAGTGCCCTTATTGGCTTGCTGGTCATC 387
```

reflect differing amounts of premature, unprocessed heterogeneous nuclear RNAs present in the respective mRNA molecule populations (*9*).

Redundantly Prepared ESTs—The frequency of each of the redundantly prepared, identified ESTs are also shown in Table II. The most repeatedly prepared ESTs corresponded to elongation factor 1-alpha, as for the human liver cell line, HepG2 (*6*), and represented 27 of the 1,026 ESTs in class X or 42 of the total 2,132 ESTs (see 27/42 in Table II, Transcription and translation machinery, C0E12, elongation factor 1-alpha). Other than ESTs related to ribosomal proteins and to mitochondrial DNA transcripts, those encoding heat-shock protein 86 (18/29), ferritin light chain (20/25), alpha-enolase (5/14), ubiquitin (8/13), heat shock protein 70 cognate (7/12), glyceraldehyde-3-phosphate dehydrogenase (5/11), phosphoglycerate kinase-1 (5/11), elongation factor 2 (7/10), and metallothionein I (6/9) were included in the top ten of the most redundantly prepared ESTs (Table II). Consistent with previous reports (*20, 21*), the levels of expression of heat-shock protein 86 and heat-shock protein 70 cognate were high in F9 cells.

ESTs Showing Higher Sequence Similarities to Known Proteins in Lower Organisms than to Ones in Mammals—Several ESTs showed significantly higher similarities to known genes or proteins from tobacco, yeast, *Chlamydomonas reinharditi, Xenopus laevis, Caenorhabditis elegans*, and *Flavobacterium*, than to ones from mammals (Table II). For example, we found on a homology search (BLASTX program) that the protein sequences deduced from the following ESTs show significant similarities to those of known proteins from lower organisms (Table IIIa): one block of 38 aa deduced from A4G11 showed 50% identity and another block of 27 aa deduced from the same EST showed 66% identity to tobacco protein kinase NPK1 (=a homologue of yeast STE11); A6C09 showed 40% identity in 72 aa to the *Chlamydomonas reinharditi* gamma heavy

TABLE III (continued)

```
Query = 73D10, Sbjct = Protein tyrosine phosphatase (PRP36), mouse
BLASTX score: 103, Identities = 22/49 (44%), Positives = 29/49 (59%)

Query:     19 QEQTRHIFSLHIKESLLAGHLXFSPEQAVELIALLVXTKFGDYNQNTAQ 165
              QE TR+ + L +K+ +L G L  S EQ + L  L V   FGDYNQ  +Q
Sbjct:    111 QEATRYQYYLQVKKDVLEGRLRCSLEQVIRLAGLAVQADFGDYNQFDSQ 159

Query = 74C10, Sbjct = Endozepine, bovine
BLASTX score: 128, Identities = 23/50 (46%), Positives = 32/50 (64%)

Query:     38 FKQVKVGNCNTPKPNFFDFEGKQKWEAWKALGDSSPSQAMQEYIAAVKKL 187
              +KQ  VG+ NT +P   DF+GK KW+AW  L  +S  AM+ YI  V++L
Sbjct:     31 YKQATVGDINTERPGMLDFKGKAKWDAWNELKGTSKEDAMKAYIDKVEEL 80
```

TABLE III. First, we used the BLASTN program to search for similarities to the non-redundant nucleic acid sequence database. ESTs exhibiting no significant similarities to the database sequences were examined for similarities to a non-redundant protein sequence database, using the BLASTX program (see "MATERIALS AND METHODS," *Database Search*). BLASTN compares a "Query" nt sequence against a "Sbjct" nt sequence database by identifying ungapped alinged segments, and BLASTX, the six-frame conceptual translation products of a "Query" nt sequence (both strands) against a protein sequence database. The BLASTN and BLASTX scores reflect the similarity between the "Query" sequence and a "Sbjct" sequence, however, there is no direct relationship between scores and percent similarity (for detailed explanations, see Refs. *14* and *26*). "Identities" indicate numbers (%) of identical aa or nt, and "positives," numbers (%) of identical plus conservatively substituted aa.

TABLE IV. **Distribution of mouse F9, human heart and brain, and liver HepG2 cell line ESTs with database matches as to functional categories.**

Category	F9[a]					Heart	Brain	Liver
	X	I	II	III	IV			
Cytoskeletal/structural/contractile	3.4	2.5	2.2	4.4	3.0	31.6	26.3	3.1
Extracellular matrix	1.0	0.5	0.7	2.2	1.7	8.1	0.0	0.0
Energy metabolism	19.4	19.8	17.0	8.9	12.7	13.5	6.1	9.2
Hormones/hormonal regulation	0.5	0.5	3.0	0.0	0.8	4.2	0.7	1.0
Signal transduction/cell regulation	4.4	3.5	6.7	2.2	16.9	7.9	19.9	8.2
Transcription/translation	52.8	58.9	49.6	60.0	27.0	18.7	18.8	31.8
Membrane-associated	3.1	1.5	4.4	6.7	5.1	8.8	15.3	2.1
Other-metabolism	12.2	10.9	13.3	15.6	29.1	6.2	11.7	14.4
Other-secreted protein	3.1	2.0	3.0	0.0	3.8	1.2	1.3	30.3
Total	99.9	100.1	99.9	100.0	100.1	100.0	100.0	100.0
(n=)	(797)	(202)	(135)	(45)	(237)	(520)	(864)	(195)

[a]The distribution of F9 ESTs is summarized according to their classes (see Table I). Presented are the percentage of ESTs in each category, with the actual number of ESTs given in parentheses. The figures for human heart (*8*) and brain (*7*), and a liver cell line, HepG2 (*6*), were cited as summarized in Liew *et al.* (*8*).

chain subunit of outer-arm dynein; A4G05 showed 33% identity in 68 aa to African clawed frog cystic fibrosis conductance regulator; one block of 49 aa deduced from 77B02 showed 53% identity and another block of 10 aa deduced from the same EST showed 100% identity to tobacco RNA helicase-like protein. These results suggest that these code for mouse homologues of the respective proteins and that they could be putative candidates for functional analysis.

ESTs Identified as Known Genes or Proteins in Mammals with Relatively Low BLASTN or BLASTX Scores— Some of the ESTs were identified as known genes or

TABLE V. **ESTs reportedly related to development and/or differentiation.**

ESTs[a)]	Class	DB:ID	Putative identification	Freq	Species
Down-regulated in embryonal carcinoma cells by retinoic acid treatment					
60F06	IV	gb:MUSB2REB	B2 repetitive sequence		m
75F03	IV	gb:MMECA39	ECA39, target for c-Myc regulation		m
95C06	X	gb:MUSHSPCA	Heat shock protein 70 cognate	(7/12)	m
84H01	IV	gb:MUSHMGIY	HMG-I(Y)	(0/ 5)	m
86C03	IV	gb:MUSID	Id, helix-loop-helix DNA binding protein regulator		m
78H05	IV	gb:MUSOCT3	Oct-3, POU-domain transcription factor	(1/ 3)	m
72H08	II	gb:MUSPH34MRA	pH 34, down-regulated in EC cells	(4/ 8)	m
A4H07	X	gb:MUSREX1	REX-1, zinc finger protein	(3/ 4)	m
72F04	II	gb:CGTUBB2	Tubulin, beta 2	(3/ 4)	m
A2G06	X	gb:MUSTUBA6M	Tubulin, M-alpha 2 or 6	(4/ 8)	m
Up-regulated in embryonal carcinoma cells by retinoic acid treatment					
68G12	IV	gb:MME1433IS	14-3-3 protein epsilon-subtype	(1/ 2)	m*
59C01	IV	gb:RAT1433PG2	14-3-3 protein gamma-subtype	(0/ 2)	r*
A1C06	X	gb:RATTHETA	14-3-3 protein theta-subtype	(3/ 5)	r*
A7A04	X	gb:MUSMCD63	CD63		m*
90G12	II	gb:MUSHSP86A	Heat shock protein 86	(18/29)	m
03B11	IV	gb:MUSIFNBB	Interferon-beta		m
68D02	II	gb:MUSLAMP1	Lysosomal membrane glycoprotein-1	(3/ 6)	m*
29B05	IV	gb:MUSMKPG	Midkine, growth/differentiation factor		m
93H11	X	gb:MMSPARCR	Osteonectin (SPARC)		m*
98B11	X	gb:MUSPDIA	Protein disulfide isomerase (ERp59)		m*,
67A04	IV	gb:S73882	*rae-28, Drosophila polyhomeotic* gene-like		m*
96F11	X	gb:MUSPTAC97	Nuclear pore-targeting-complex component of 97 kD		m*
Other development and/or differentiation-related					
B2F12	X	gb:RATAGR	Agrin		r
A2G08	X	gb:S68108	brg1, brahma homolog (Hox gene regulator)		m
70H07	II	gb:DMCPO612G	Cpo 61.2 gene, RNA binding protein		dm
88H12	IV	gb:MUSCRIPTO	Cripto, related to heart development		m
90B11	IV	gb:S43105	Cyclin B1, cell cycle regulatory protein		m
71B09	IV	gb:MMCYB2	Cyclin B2		m
83E09	IV	gb:MMCYCLGMR	Cyclin G		m
A4G05	X	pi:1811235A	Cystic fibrosis conductance regulator		xe
74E04	IV	gb:MMECADH	E-cadherin		m
84A02	IV	gb:MUSFAK	Focal adhesion kinase		m
81H12	IV	gb:MMGAS5MR	gas5, growth arrest specific gene		m
B7H06	X	gb:MMHOX35	Hox-3.5		m
B1D10	X	gb:MUSKIF4	kif4		m
94H07	X	gb:MUSEKLTF	Erythroid krueppel-like transcrption factor		m
93C01	X	em:S48643	Mammary-derived growth inhibitor		m
88E05	IV	gb:MUSMEKPKI	MEK, protein kinase		m
96B08	X	gb:MUSNEDD4	NEDD-4, developmentally down-regulated		m
74B03	II	gb:HSU02368	PAX3/forkhead transcription factor gene		h
A7H06	X	gb:MMPCNAG	PCNA, proliferating cell nuclear antigen	(2/ 3)	m
C0B11	X	gb:DROROX2Y	RRM-type RNA binding protein		dm
C0C03	X	gb:MMSEK2	Sek-2, receptor-protein tyrosine kinase		m
75A11	IV	gb:NMSIAHIB	siah-1B, seven in absentia gene-homologue		m
83A06	IV	gb:MUSWNTVB	Wnt-5b protein precursor, Wnt-1/int-1 family		m

[a)]ESTs, Class, DB:ID, Putative identification, Freq and Species are the same as in the footnote to Table II. * indicates that a homologous EST was isolated as a RA-inducible clone in our laboratory (*22* and unpublished data).

proteins in mammals with relatively low scores, such as BLASTN scores of around 200, or BLASTX scores of around 100 (Table II); they probably do not correspond to identical genes or proteins, but rather to novel genes or proteins sharing significant sequence similarities with previously identified ones. For example, 75E04 was identified as rat sperm membrane protein with a BLASTN score of 201, *i.e.*, three fragments of 46, 39, and 29 nucleotides (nt) showed 91, 84, and 93% identities to corresponding regions of the cDNA coding for rat sperm membrane protein (Table IIIb). 73D10 was identified as mouse protein tyrosine phosphatase with a BLASTX score of 103 (44% identity in 49 aa residues), and 74C10 to bovine endozepine with a score of 128 (46% identity in 50 aa residues) (Table IIIb). These results suggest that 74E04 corresponds to a novel mouse mRNA encoding a protein sharing partial homology with rat sperm membrane protein, but probably different from this protein. Similar possibilities can be considered for the proteins coded by 73D10 and 74C10. Among ESTs belonging to this group, we can expect to find ESTs corresponding not only to novel genes, but also biologically important genes.

Distributions of Identified ESTs—Among the 1,026 ESTs prepared from class X clones, 858 were identified: 797 corresponded to known genes and 61 to previously registered ESTs matching no known genes. The distribution of these 797 identified ESTs in the functional categories was compared with similar distributions of ESTs prepared from other classes of F9 clones, and from human heart (*8*) and brain (*7*), and HepG2 cells (*6*) (Table IV). The distribution of ESTs from class IV was apparently different from those prepared from the other classes: ESTs related to "Transcription/translation" were less abundant, but those related to "Other-metabolism" and "Signal transduction/cell regulation" were relatively abundant in this class (Table IV). Cytoskeletal proteins, which are abundant in heart and brain (31.6 and 26.3% of the total ESTs, respectively), were apparently less abundant in F9 and HepG2 cells. Signal transduction/cell regulation proteins, which were abundant in brain (19.9% of the total ESTs), were less abundant in F9, heart and HepG2 cells. Proteins related to transcription/translation were most abundant in F9 cells (52.9% of the ESTs from class X) and were also abundant in HepG2 cells (31.8%). Proteins related to energy metabolism were abundant in F9 cells (19.4% of the ESTs from class X) and heart (13.5%). On the other hand, proteins related to "Other-secreted proteins" were significantly abundant in HepG2 cells, but not in F9 cells, heart or brain (Table IV). These data indicate that mRNAs coding for proteins related to transcription/translation and energy metabolism are most abundant in F9 cells, and suggest that this should be a characteristic feature of metabolism in undifferentiated F9 cells, which resemble embryonic stem cells of the mouse blastocyst (*1*).

ESTs Possibly Related to Early Mammalian Development—We isolated 17 different cDNA clones corresponding to RA-inducible mRNAs in F9 cells by the differential hybridization method from about 2×10^5 cDNA clones (*22*). Interestingly, we found that the following ESTs correspond to 5 of the 17 RA-inducible cDNAs: 93H11 was identified as osteonectin; 67A04 as RAE-28 (*Drosophila polyhomeotic* gene-like); 68D02 and the other 5 ESTs as lysosomal membrane glycoprotein-1; A7A04 as CD63 antigen; and 98B11 as protein disulfide isomerase (Table V) (*22*). Other than these 17 clones, we isolated 2 more RA-inducible cDNA clones by the differential hybridization method: one named Rae-12 showed high sequence similarities to 14-3-3 family proteins, and the other, named Rae-25, corresponded to the nuclear pore-targeting-complex component of 97 kDa (*23*) (unpublished observations). Interestingly, we found in this study that the following 5 ESTs, 94C02, B9E11, 75H04, 75A01, and 68G12, match Rae-12, and one EST, named 96F11, matches Rae-25 (Table V). Altogether, we found at least 45 different ESTs which had previously been documented either to be regulated by RA in embryonal carcinoma cells or to be involved in mammalian development and/or differentiation (Table V). Thirty-two of the 45 ESTs listed in Table V were unique ESTs, appearing only once in the identified ESTs, which suggested that they belong to the low-abundance mRNA group. Thus, approximately 7% of the low-abundance mRNAs, *i.e.* 32 of the 436 unique ESTs, correspond to either RA-regulated genes or development- and/or differentiation-related genes. Our results strongly suggest that one can isolate various new development- and/or differentiation-related genes from novel ESTs corresponding to low-abundance mRNAs.

There are several potential applications of ESTs obtained from F9 cells. For example, we can use these ESTs as probes and examine the levels of the corresponding mRNAs, not only during the differentiation of F9 cells, but also in various developmental stages of mouse embryos. Through such experiments, we can screen genes potentially involved in early mammalian development (*22*). The roles of such genes in early development may be directly examined using F9 cells, because we can generate F9 cell lines, in which both alleles of a certain gene are disrupted by homologous recombination (*3, 24, 25*), and because F9 cells can be induced to differentiate into the parietal endoderm or into more complex structures called embryoid bodies (*1, 2*). Studies along these lines are in progress in our laboratory.

We wish to thank M. Ohara for the many helpful comments on the manuscript, and M. Miyazaki, T. Sakaura, and K. Sumitomo for their expert technical assistance. We also thank the National Center for Biotechnology Information, USA, for providing access to the network BLAST server, and the Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, for that to the network BLAST and FASTA servers.

REFERENCES

1. Martin, G.R. (1980) Teratocarcinomas and mammalian embryogenesis. *Science* **209**, 768-776
2. Gudas, L.J. (1991) Retinoic acid and teratocarcinoma stem cells. *Semin. Dev. Biol.* **2**, 171-179
3. Capecchi, M.R. (1989) Altering the genome by homologous recombination. *Science* **244**, 1288-1292
4. Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R., and Venter, J.C. (1991) Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* **252**, 1651-1656
5. Khan, A.S., Wilcox, A.S., Polymeropoulos, M.H., Hopkins, J.A., Stevens, T.J., Robinson, M., Orpana, A.K., and Sikela, J.M. (1992) Single pass sequencing and physical and genetic mapping of human brain cDNAs. *Nature Genet.* **2**, 180-185
6. Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A.,

Kojima, Y., and Matsubara, K. (1992) Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression. *Nature Genet.* **2**, 173-179
7. Adams, M.D., Kerlavage, A.R., Fields, C., and Venter, J.C. (1993) 3,400 new expressed sequence tags identify diversity of transcripts in human brain. *Nature Genet.* **4**, 256-267
8. Liew, C.C., Hwang, D.M., Fung, Y.W., Laurenssen, C., Cukerman, E., Tsui, S., and Lee, C.Y. (1994) A catalogue of genes in the cardiovascular system as identified by expressed sequence tags. *Proc. Natl. Acad. Sci. USA* **91**, 10645-10649
9. Sudo, K., Chinen, K., and Nakamura, Y. (1994) 2058 expressed sequence tags (ESTs) from a human fetal lung cDNA library. *Genomics* **24**, 276-279
10. Nishiguchi, S., Joh, T., Horie, K., Zou, Z., Yasunaga, T., and Shimada, K. (1994) A survey of genes expressed in undifferentiated mouse embryonal carcinoma F9 cells: Characterization of low-abundance mRNAs. *J. Biochem.* **116**, 128-139
11. Bernstine, E.G., Hooper, M.L., Grandchamp, S., and Ephrussi, B. (1973) Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci. USA* **70**, 3899-3903
12. Meyuhas, O. and Perry, R.P. (1979) Relationship between size, stability and abundance of the messenger RNA of mouse L cells. *Cell* **16**, 139-148
13. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
14. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
15. Pearson, W.R. and Lipman, D.J. (1988) Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**, 2444-2448
16. Attardi, G. (1985) Animal mitochondrial DNA: An extreme example of genetic economy. *Int. Rev. Cytol.* **93**, 93-144
17. Vasseur, M., Condamine, H., and Duprey, P. (1985) RNAs containing B2 repeated sequences are transcribed in the early stages of mouse embryogenesis. *EMBO J.* **4**, 1749-1753
18. Taylor, K.D. and Piko, L. (1987) Patterns of mRNA prevalence and expression of B1 and B2 transcripts in early mouse embryos. *Development* **101**, 877-892
19. Bachvarova, R. (1988) Small B2 RNAs in mouse oocytes, embryos, and somatic tissues. *Dev. Biol.* **130**, 513-523
20. Bensaude, O. and Morange, M. (1983) Spontaneous high expression of heat shock proteins in mouse embryonal carcinoma cells and ectoderm from day 8 mouse embryo. *EMBO J.* **2**, 173-177
21. Legagneux, V., Mezger, V., Quelard, C., Barnier, J.V., Bensaude, O., and Morange, M. (1989) High constitutive transcription of HSP86 gene in murine embryonal carcinoma cells. *Differentiation* **41**, 42-48
22. Nomura, M., Takihara, Y., and Shimada, K. (1994) Isolation and characterization of retinoic acid-inducible cDNA clones in F9 cells: One of the early inducible clones encodes a novel protein sharing several highly homologous regions with a *Drosophila polyhomeotic* protein. *Differentiation* **57**, 39-50
23. Radu, A., Blobel, G., and Moore, M.S. (1995) Identification of a protein complex that is required for nuclear protein import and mediates docking of import substrate to distinct nucleoporins. *Proc. Natl. Acad. Sci. USA* **92**, 1769-1773
24. Mortensen, R.M., Conner, D.A., Chao, S., Geisterfer-Lowrance, A.T.T., and Seidman, J.G. (1992) Production of homozygous mutant ES cells with a single targeting construct. *Mol. Cell. Biol.* **12**, 2391-2395
25. Coll, J.-L., Ben-Ze'ev, A., Ezzell, R.M., Fernandex, J.J.R., Baribault, H., Shima, R.G., and Adamson, E.D. (1995) Targeted disruption of vinculin genes in F9 and embryonic stem cells changes cell morphology, adhesion, and locomotion. *Proc. Natl. Acad. Sci. USA* **92**, 9161-9165
26. Altschul, S.F., Boguski, M.S., Gish, W., and Wootton, J.C. (1994) Issues in searching molecular sequence databases. *Nature Genet.* **6**, 119-129