DNA RESEARCH 3, 321–329 (1996)

Prediction of the Coding Sequences of Unidentified Human Genes. VI. The Coding Sequences of 80 New Genes (KIAA0201-KIAA0280) Deduced by Analysis of cDNA Clones from Cell Line KG-1 and Brain

Takahiro NAGASE, Naohiko SEKI, Ken-ichi ISHIKAWA, Miki OHIRA, Yutaka KAWARABAYASI, Osamu OHARA, Ayako TANAKA, Hirokazu KOTANI, Nobuyuki MIYAJIMA, and Nobuo NOMURA*

Kazusa DNA Research Institute, 1532-3 Yana, Kisarazu, Chiba 292, Japan

(Received 4 October 1996; revised 24 October 1996)

Abstract

In this series of projects of sequencing human cDNA clones which correspond to relatively long and nearly full-length transcripts, we newly determined the sequences of 80 clones, and predicted the coding sequences of the corresponding genes, named KIAA0201 to KIAA0280. Among the sequenced clones, 68 were obtained from human immature myeloid cell line KG-1 and 12 from human brain. The average size of the clones was 5.3 kb, and that of distinct ORFs in clones was 2.8 kb, corresponding to a protein of approximately 100 kDa. Computer search against the public databases indicated that the sequences of 22 genes were unrelated to any reported genes, while the remaining 58 genes carried sequences which show some similarities to known genes. Protein motifs that matched those in the PROSITE motif database were found in 25 genes and significant transmembrane domains were identified in 30 genes. Among the known genes to which significant similarity was shown, the genes that play key roles in regulation of developmental stages, apoptosis and cell-to-cell interaction were included. Taking into account of both the search data on sequence similarity and protein motifs, at least seven genes were considered to be related to transcriptional regulation and six genes to signal transduction. When the expression profiles of the cDNA clones were examined with different human tissues, about half of the clones from brain (5 of 11) showed significant tissue-specificity, while approximately 80% of the genes from KG-1 were expressed ubiquitously.

Key words: full-length cDNA sequence; mRNA expression; chromosomal location; myeloid cell line KG-1; brain.

1. Introduction

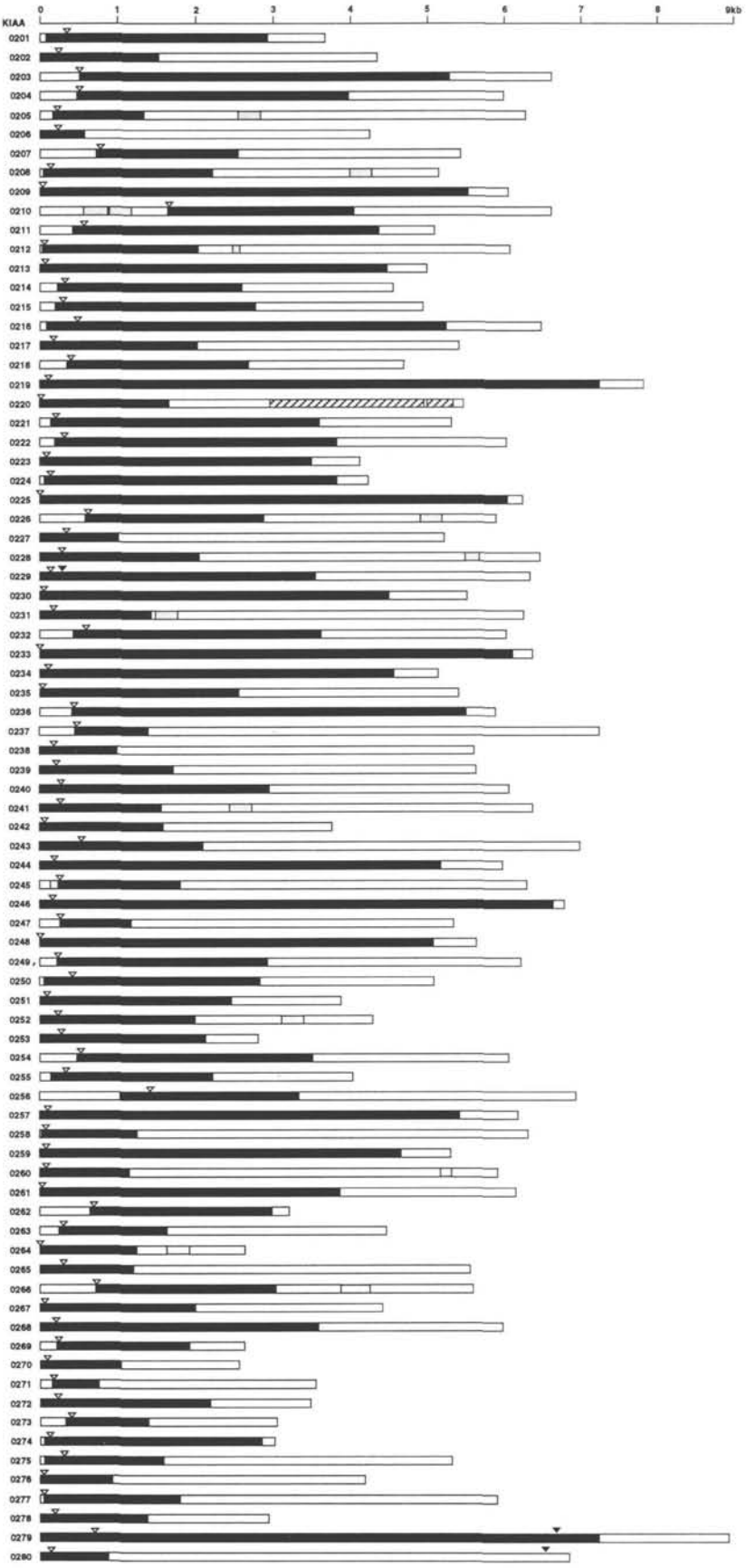
To accumulate information on the coding sequences of unidentified human genes, we have begun a project for sequencing the entire cDNA clones which correspond to relatively long and nearly full-length transcripts.[1,2] Whereas many genes of functional importance appear to be expressed in longer transcripts, little effort has been made to analyze such transcripts mostly due to technical difficulties. Although a large amount of expressed sequence-tags (ESTs) obtained by one path sequencing of cDNA libraries have been accumulated for comprehensive understanding of expression profiles, the information obtained is limited to relatively short cDNA species.[3–5] By our sequencing strategy, cDNA libraries enriched with clones corresponding to relatively long transcripts were constructed, and the clones that carry unreported terminal sequences are first selected. Then, the sizes of the mRNA corresponding to these clones are analyzed by Northern hybridization, and the entire nucleotide sequences of clones that comprised nearly full-length transcripts were determined. We have already predicted the coding sequences of 180 new genes from analysis of cDNA clones which were isolated from human immature myeloid cell line KG-1.[1,2] By computer search, 68 genes were found to be new, and most of the remaining genes (96 genes) were related to those with biologically important function. In this paper, we report the coding sequences of additional 80 genes and their sequence features as well as expression profiles.

2. Materials and Methods

The source of the KG-1 cDNA clones was identical to that used in the previous paper.[1] The brain cDNA clones were selected from a cDNA library which was con-

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +81-438-52-3930, Fax. +81-438-52-3931



Fifure 1. Physical maps of the 80 cDNA clones analyzed.

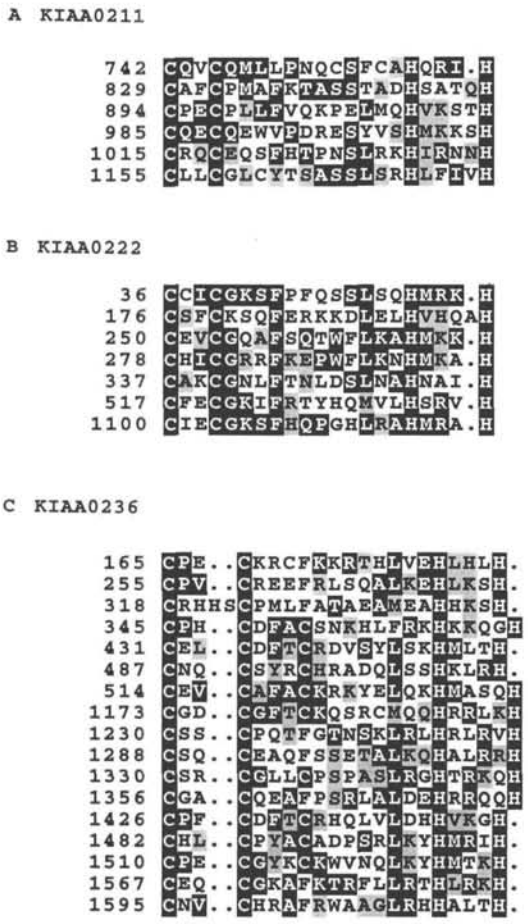


Figure 2. Distinct C2H2 type zinc finger repeats in products encoded in genes: **A**, KIAA0211; **B**, KIAA0222 ; **C**, KIAA0236.

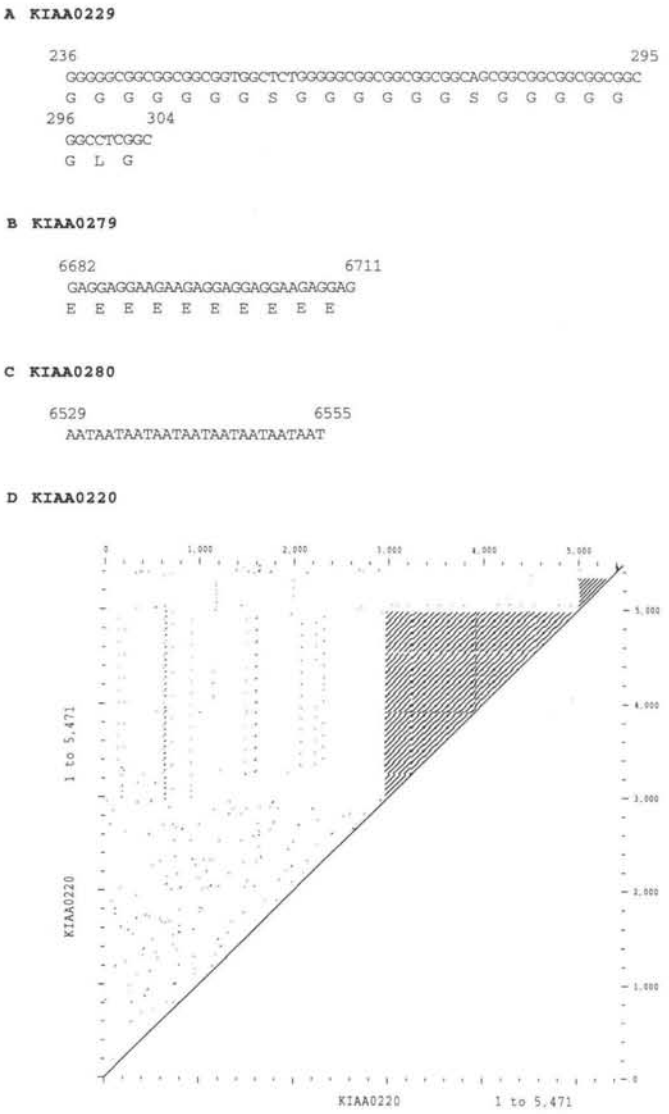


Figure 3. Typical repeats observed in cDNA clones from identified genes. **A**, GGC repeats in KIAA0229: **B**, GAG repeats in KIAA0279: **C**, AAT repeats in KIAA0280: **D**, reiteration of two types of 55 nucleotide repeats in KIAA0220. In **A** and **B**, translated amino acids are indicated below the DNA sequences. Numerals above the sequences are nucleotide positions in each clone.

structed from a whole brain mRNA fraction of CLONTECH (California, USA) by a method essentially identical to that used for the library of KG-1 cells. The methods used for selection of clones, Northern hybridization, sequence analysis, computer analysis of sequences and chromosomal mapping of cDNA clones were described previously.[1,6]

3. Results and Discussion

3.1. Sequence features of analyzed cDNA clones

As in the previous papers,[1,2] the cDNA clones carrying inserts longer than 2 kb were randomly selected from the libraries constructed from the medium-sized cDNA class, and both the terminal sequences were analyzed to select unidentified clones with poly(A) tails. The clones harboring inserts more than 90% of the length of the corresponding transcripts were further selected by northern hybridization, and their sequences were determined. Among 85 clones fully sequenced, 80 clones were

Figure 1. The horizontal scale represents the cDNA length in kb, and gene numbers are given on the left. Open reading frames (ORFs) within coding regions, untranslated regions, Alu sequences, and other repetitive sequences are indicated by solid, open, dotted, and hatched boxes, respectively. The details of repeats in KIAA0220 are shown in Fig. 3D. The positions of the first ATG codon in each ORF are represented by open triangles. The solid triangles show the positions of the triplet repeats listed in Figs. 3A, B and C. The nucleotide sequence data reported in this paper were deposited in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases under the accession numbers shown in Table 3.

Table 1. Genes with similarities to nucleotide and amino acid sequence database files.

Gene no. (KIAA)	Database files	Accession no.[a]	Identities (%)	Overlap[b] (amino acid residues)
0201	heat shock protein 105 kD alpha (M)	D67016	93.4	858
0202	KIAA0128 (H)	D50918	77.7	413
0203	coiled-coil protein CC1 (M)	X82318	84.4	141
0204	serine/threonine protein kinase Krs-1 (H)	U60206	30.5	483
0205	cosmid C01C10 (Ce)	U23526	28.0	344
0207	growth factor receptor-binding protein Grb10(M)	U18996	90.0	428
0208	dishevelled-3 (M)	U41285	98.7	681
0209	major CRK-binding protein DOCK180 (H)	D50857	62.3	1729
0210	cosmid B0393 (Ce)	Z37983	39.6	345
0211	finger protein 1 (H)	A32891[c]	23.2	343
0212	cosmid C47E12 (Ce)	Z68882	63.8	458
0214	cosmid ZK1248 (Ce)	U29244	35.2	739
0215	KIAA0239 (H)	D87076	60.4	312
0216	myosin heavy chain (R)	S21801[c]	22.5	900
0218	hypothetical 29.6 kD protein (E)	P27859[d]	30.1	204
0219	translational activator of GCN1 (Sc)	L12467	32.1	1888
0221	NAM7 (Sc)	P30771[d]	54.5	858
0222	finger protein MKR3 (M)	S03677[c]	23.9	323
0223	cosmid ZK669 (Ce)	Z37093	30.0	274
0224	putative ATP-dependent RNA helicase K03H1.2 (Ce)	P34498[d]	59.3	1064
0225	cosmid K12D12 (Ce)	Z49069	18.7	1301
0229	ankyrin 1 (H)	S08275[c]	21.7	652
0230	peroxidasin precursor (D)	U11052	38.0	1412
0231	adenylate cyclase (Sc)	S56776[c]	22.9	317
0233	cosmid T20D3 (Ce)	Z68220	33.9	427
0234	XE169 (H)	L25270	83.9	1363
0235	KIAA0099 (H)	D43951	81.5	850
0236	zinc finger protein ZNF142 (H)	U09849	98.8	170
0237	cosmid T10A3 (Ce)	U41035	40.4	228
0238	yolk sac permease-like molecule 1 (M)	U25739	38.7	223
0239	KIAA0215 (H)	D86969	60.4	311
0241	cosmid T26A5 (Ce)	U00043	45.8	164
0242	hypothetical 51.6kD protein ZK353.8 (Ce)	P34631[d]	25.3	351
0244	transcription elongation factor TFIIS (H)	X57198	29.6	161
0245	amino acid permease PRM1 (Sm)	L25068	40.9	480
0246	notch 3 (M)	X74760	27.5	299
0248	protein transport protein SEC7 (Sc)	P11075[d]	35.4	209
0249	KIAA0188 (H)	D80010	49.1	895
0251	cosmid C14H10 (Ce)	Z50863	24.0	407
0252	glutamic acid-rich protein precursor (Pf)	A54514[c]	18.3	503
0255	70kD endomembrane protein EMP70 (Sc)	P32802[d]	35.2	612
0257	cosmid C27F2 (Ce)	U40419	36.1	228
0259	rad4/cut5 protein (Sp)	P32372[d]	24.4	236
0260	cosmid C52E12 (Ce)	U50135	49.1	115
0261	parallel sister chromatids protein (D)	U40214	32.0	469
0263	hypothetical protein YD9335.01 (Sc)	S54638[c]	30.7	197
0266	hypothetical protein 5 (Sc)	S49634[c]	25.9	704
0267	Na+/H+ exchanger 2 (H)	A57644[c]	29.0	418
0268	C219-reactive peptide (H)	L34688	100[e]	136
0269	extensin-like protein (Zm)	S49915[c]	29.9	228
0271	transforming protein bcl-2 (H)	C37332[c]	45.1	164
0272	hypothetical C08B11.7 protein (Ce)	Q09444[d]	33.6	294
0274	hypothetical NO330 protein (Sc)	P42837[d]	35.1	655
0275	testican (H)	S33293[c]	49.0	358
0276	hypothetical protein L3111 (Sc)	S59316[c]	25.1	180
0277	CDC25 protein homolog (H)	L26584	26.4	235
0278	growth factor Arc (R)	U19866	92.4	396
0279	cadherin-related tumor suppressor hFat protein (H)	X87241	26.5	656

[a] EMBL/NCBI/GSDB/DDBJ database files are shown unless specified.

[b] The size of regions which show similarities.

[c] PIR database files

[d] SWISS-PROT database files

[e] A partial sequence spanning aa positions 592 -727 of KIAA0268 has been reported.

Ce, *Caenorhabditis elegance*; D, *Drosophila melanogaster*; E, *Escherichia* coli; H, human; M, mouse; Pf, *Plasmodium falciparum*; R, rat; Sc, *Saccharomyces cerevisiae*; Sm, *Schistosoma mansoni*; Sp, *Schizosaccharomyces pombe*; Zm, *Zea mays*.

Table 2. Genes with regions that matched motifs in the PROSITE database.

Motifs	Description	Gene number (KIAA)	References
HSP70 3	Heat shock hsp70 proteins family	0201	18
ATP GTP A	ATP/GTP-binding site motif A (P-loop)	0202, 0212, 0214, 0216 0221, 0222, 0224, 0250	19
PROTEIN KINASE ST	Protein kinases	0204, 0213	20
PROTEIN KINASE ATP	Protein kinases	0204, 0213	20
ZINC FINGER C2H2	Zinc finger, C2H2 type	0211, 0222, 0236	11
CYTOCHROME C	Cytochrome c family heme-binding site	0211, 0223	21
YBL055C-1	Hypothetical YBL055c/yjjV family	0218	22
YBL055C-2	Hypothetical YBL055c/yjjV family	0218	22
DEAH ATP HELICASE	DEAD and DEAH box families ATP-dependent helicases	0224	23
IG MHC	Immunoglobulins and major histocompatibility complex proteins	0233	24
GLYCOSYL HYDROL F1 1	Glycosyl hydrolases family 1	0237	25
CYTOCHROME P450	Cytochrome P450 cysteine heme-iron ligand	0246	26
EGF	EGF-like domain cysteine pattern	0246, 0279	27
AA TRNA LIGASE II 2	Aminoacyl-transfer RNA synthetases class-II	0248	28
ATPASE C	ATP synthase c subunit	0256	29
ATPASE ALPHA BETA	ATP synthase alpha and beta subunits	0257	29
ZINC FINGER C3HC4	Zinc finger, C3HC4 type	0262	12
PRENYLATION	Prenyl group binding site	0270	15
BCL2	Apoptosis regulator proteins, Bcl-2 family	0271	9
THYROGLOBULIN 1	Thyroglobulin type-1 repeat	0275	30
CADHERIN	Cadherins extracellular repeated domain	0279	10

found to contain distinct open reading frames (ORFs) and were subjected to further analysis. The ORFs and the first ATG codon in each ORF are shown in Fig. 1 by solid boxes and open triangles, respectively. In the figure, KIAA0201 to KIAA0268 represent the clones from the KG-1 cDNA library and KIAA0269 to KIAA0280 represent the clones from the brain cDNA library. In-frame termination codons upstream of the first ATG codon were identified in 38 clones, suggesting that at least half of the clones analyzed harbor the complete coding region.

The results of computer analysis with the GCG software package are shown in Tables 1 and 2, Figs. 2, 3 and 4 and in the figure in the Supplement section. Sequence features are summarized below.

1. Sequences of 22 genes were unrelated to any reported sequences in the database files, except for ESTs (GenBank release 96.0, August 1996). The remaining 58 genes carried sequences with at least some similarities to known genes (Table 1). The genes that we are particularly noted are as follows. KIAA0208 was a human homolog of mouse *dishevelled-3*[7] and KIAA0246 carried a sequence with considerable similarity to the *Notch* gene family,[8] both of which are known to mediate cell fate decisions during development. KIAA0271 showed significant similarity to the *bcl-2* gene family which plays important roles in apoptosis.[9] KIAA0279 was related to a gene involved in cell-to-cell interaction.[10]

2. Protein motifs that matched those in the PROSITE motif database were found in 25 genes (Table 2).

3. On the basis of the search data of similarity and protein motifs, at least seven genes were considered to be involved in transcriptional regulation. Those senen genes are KIAA0211, 0215, 0219, 0222, 0236, 0239 and 0262, in which KIAA0211, 0222 and 0236 carried the C2H2 type zinc finger[11] (Fig. 2) and KIAA0262 the C3HC4 type zinc finger.[12]

4. The search data of similarity and protein-motifs also suggested 6 genes, KIAA0204, 0209, 0213, 0231, 0277 and 0278, are relating to signal transduction: The product encoded in KIAA0231 harbors a leucine-rich domain with significant structural similarity to that of adenylate cyclase[13] (Fig. 4A), and that of KIAA0277 shows a high degree of sequence similarity to proteins encoded in the *CDC25, Sos1* and *Ste6* (Fig. 4B).[14]

5. Significant transmembrane domains were identified in 30 genes, 11 of which harbored multiple hydrophobic regions. It is also noted that KIAA0270 harbors a binding site of the prenyl group which is assumed to anchor to membranes.[15]

6. Three genes harbored triplet repeats, which were often correlated with genetic disorders:[16] GGC(Gly) repeats occurred 20 times within a 23-triplet stretch in KIAA0229 , GAG(Glu) repeats occurred 7 times in a 10-triplet stretch in KIAA0279, and 9 AAT repeats were detected in the 3′-untranslated region (UTR) of KIAA0280 (Figs. 3A, B and C).

7. Alu sequences were identified in the 5′-UTRs of 2 genes and in the 3′-UTRs of 12 genes, respectively. The presence of Alu in the 5′-UTR has already been

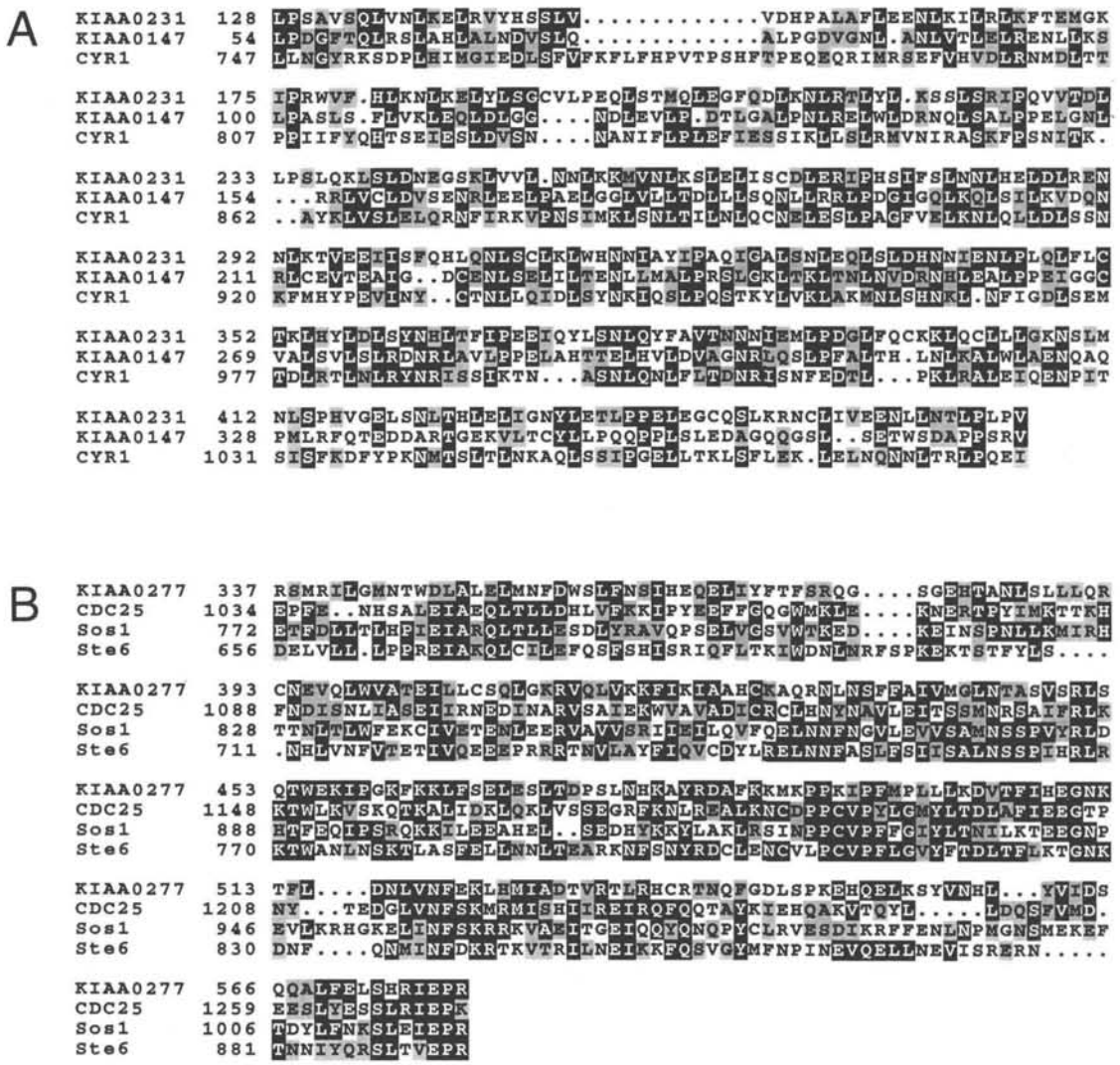


Figure 4. Sequence comparison of (A) a leucine-rich domain in adenylate cyclase family and (B) *CDC25* gene family. Identical and similar amino acids are indicated by black and grey shading, respectively. Numerals represent the number of amino acid residues from the start codon.

reported by Yulug et al.[17] Although the mechanism for Alu integration is not fully understood yet, the lower occurrence rate of Alu in the 5′-UTR than 3′-UTR in cDNAs may be due to differences in the target size for Alu retroposition. It should be noted that the Alu sequence at the 5′-UTR of KIAA0245 overlaps the ORF, but as the first ATG in the ORF is present downstream of Alu, the Alu sequence does not seem to be translated.

8. Two repeated sequences, both of which are 55 nucleotides long, were found in KIAA0220 (Fig. 3D). They were reiterated 32 and 8 times, respectively, in the 3′-UTR of KIAA0220.

3.2. *Expression profiles in tissues*

The expression profiles of the sequenced genes were examined with 16 different human tissues and 2 cell lines including the KG-1 cell, and clear patterns were obtained for all but two clones. The results are summarized in Table 3. Seventy-nine percent (53 out of 67) of clones from the KG-1 cells and 36% (4 out of 11) of those from brain were found to be expressed ubiquitously, albeit to varying degrees, in all the tissues examined. Five out of 11 clones (45%) from brain were expressed specifically, if not exclusively, in brain. The patterns of two representative clones are shown in Figs. 5A and B. This is in sharp contrast to the fact that most of KG-1 clones exhibited ubiquitous expression. It was also noted that 10

Table 3. Summary of cDNA sequence data and expression patterns of identified genes in human tissues and cell lines.

Gene number (KIAA)	Total length of cDNA (bp)[a]	Amino acid residues	Expression[b] KG-1	HeLa	He	Br	Pl	Lu	Li	Sk.m	Ki	Pa	Sp	Th	Pr	Te	Ov	Sm.i	Co	Pe.b	Chromosomal[c] location	Accession[d] number
0201[f,g]	3,614	858	+	+/−	+	++	+	++	+/−	+	+	+	+	−	−	++	+	+	+	+	13	D86956
0202[f,g]	4,344	508	+	+	+/−	++	++	++	+	+	+++	+	+	+	+	+/−	+	+	+	−	6	D86957
0203[e,f]	6,614	1,591	+	+	+	+	+	−	−	+	−	+	+	+	+	+	+	+	+	+	8,20,21	D86958
0204[f,g]	5,988	1,152	+	+	+	+	+	+	+	++	++	+	+	+	+	+	+	+	+	+	10	D86959
0205[e,f]	6,253	370	+	+	+	+	+	+	+	+	+	+/−	+	+	−	+	−	+	+/−	+	1,13	D86960
0206[e]	4,249	193	+	+	+/−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	−	5,11	D86961
0207[f]	5,431	588	+	++	−	−	+/−	+	+	+	+	+	+	+	+	+	+	+	−	−	7	D86962
0208[f]	5,146	693	++	++	+	+	++	++	+	+++	+++	+	++	++	++	++	+++	++	++	++	3	D86963
0209[f]	6,050	1,842	+	−	−	−	−	+	++	+	−	+	++	++	−	−	+	+	−	+++	6	D86964
0210[e,f]	6,611	795	+	+	+/−	+	+	++	+	+	++	+	+	+	+	+	+	+	+	+	3	D86965
0211[f,g]	5,086	1,267	+	+	+++	+	+	+	+	+++	+	+	+	+	+	+	+	+	+	+	15,18	D86966
0212e[f,g]	6,072	657	++	+	+	+	+	+	++	−	+	++	+	++	+	+	+	+	+	+	3	D86967
0213[g]	4,990	1,491	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	6	D86968
0214[e,f,g]	4,550	757	+	+	+++	+	+	+	+	+++	+	+	+	+	+	+	+	+	+	+	1	D86987
0215[f]	4,935	823	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	3,X	D86969
0216[f,g]	6,479	1,581	+	+	+	+	+	+	+	+++	+	+	+	+	+	+	+	+	+	+	17	D86970
0217	5,404	673	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	10	D86971
0218[f,g]	4,689	761	+	+	+	+	+	++	+	+	+	+	+	++	+	+	++	+	+	++	3	D86972
0219[e,f]	7,819	2,412	+	−	+	+/−	+	+	+	++	+	+	+	+	+	+	+	+	+	+	6,12	D86973
0220	5,471	553	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	14,16	D86974
0221[f,g]	5,311	1,129	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	19	D86988
0222[f,g]	6,033	1,163	+	−	+	+	+	+	+	+	+	+	+	+	+	++	+	+	−	+	18	D86975
0223[f,g]	4,121	1,165	++	−	++	+	+	+	+	+	+	+	++	++	+	+	+	+	+	++	19	D86976
0224[f,g]	4,226	1,227	+	−	+	+	+	+	+	+	+	−	+	+	+	++	+	+	+	+	16	D86977
0225[e,f]	6,237	2,013	+	++	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	7	D86978
0226[e]	5,891	751	+	+	+	+	+	+	+/−	+	+	+	+	+	+	+	+	+	+	+	3	D86979
0227	5,217	336	++	++	++	+	+	+	+	++	+	++	+	+	+	+	+	+	+	+	14	D86980
0228[e]	6,465	681	+	+	+	+	+	+	n.d.	+	+	+	+	+	+	+	+	+	+	+	17	D86981
0229[f]	6,335	1,180	+++	+++	+++	+	+	+	+	+++	+	+++	+	+	+	+	+	+	+	+	6	D86982
0230[e,f]	5,510	1,496	+	+	++	−	++	++	+	+	+	+	+	+	+	+	++	+	+	−	2	D86983
0231[f]	6,248	476	+	+	+/−	+	+/−	+/−	+/−	+/−	+/−	+/−	+	+	+	+	+	+	+	+	1	D86984
0232	6,025	1,008	+	+	++	+	+	−	+	++	+	+	+	+	+	+	+	+	+	+	4	D86985
0233[e,f,g]	6,368	2,035	+	+	+	+/−	+	+	+	+	+	+	+	+	+	+	+	+	+	++	16	D87071
0234[f]	5,134	1,482	+	+	+	+	+	+	+	+	−	+	+	++	+	+	+	+	+	+	Y	D87072
0235[f]	5,399	850	++	++	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	2	D87078
0236[f,g]	5,878	1,687	+	+	+	+	+	+	+	++	+	+	+	+	+	++	+	+	+	+	2*	D87073
0237[f,g]	7,239	308	+	−	+	++	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1*	D87074
0238[e,f]	5,608	330	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	20*	D87075
0239[f]	5,630	571	+	+	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	5	D87076
0240	6,060	983	+	+	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	6*	D87077
0241[f]	6,371	522	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	7*	D87682
0242[e,f]	3,760	529	+	+	+	+	+	+	++	++	+	+	+	+	+	+	+	+	+	+	2*	D87684
0243	6,984	699	+	+	+	+	+	+	−	+	+	−	+	+	+	+	+	+	+/−	+	9*	D87683
0244[f]	5,975	1,723	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	6*	D87685
0245[e,f]	6,296	515	+	+	+	+	+	+	−	+	+	+/−	+	++	+	+	+	+	+	+	16*	D87432
0246[e,f,g]	6,777	2,212	+	+	−	−	+	++	+	+	+	+/−	++	+/−	+/−	+/−	+	+	−	++	3*	D87433
0247[e]	5,338	303	+	+	+	+	+	+	+	+	+	+/−	+	+	+	+	+	+	+/−	+	14*	D87434
0248[e,f,g]	5,634	1,691	+	+	+	+	+	+	+	++	+	+	+	+	+	+	+	+	+	+	10*	D87435
0249[f]	6,219	896	+	+	+	+	+	+	++	+/−	++	+	+	+	+	+	+	+	+/−	+	18*	D87436
0250[e,g]	5,082	802	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	1*	D87437
0251[f]	3,875	820	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	16*	D87438
0252[f]	4,288	664	+	+	+	+	+	+	+	++	+	+	+	+	+	+	+	+	+	+	15*	D87440
0253[e]	2,805	708	+	+	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	++	1*	D87442
0254[e]	6,049	992	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	11*	D87443
0255[e,f]	4,028	625	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	20	D87444
0256[g]	6,935	635	+	+/−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+/−	+/−	15*	D87445
0257[e,f,g]	6,178	1,805	+	+	+	+	+	+	+	++	+	+	+	+	+	+	+	+	+	+	2*	D87446
0258	6,313	391	+	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	9*	D87447
0259[f]	5,298	1,550	+	+/−	+	+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	3*	D87448
0260[e,f]	5,918	383	+	+	+	+/−	+	+	++	+	+	+	+	+	+	+	+	+	+	+	1*	D87449
0261[f]	6,155	1,287	+	+	+/−	+/−	+	+	+	+	+	+	+	+	+	+	+	+	+/−	+	10*	D87450
0262[g]	3,205	761	++	++	+++	+	+	+	+	+++	+	++	+	+	+	++	+	+	+	+	12*	D87451
0263[f]	4,461	441	+	+	+	+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	3*	D87452
0264	2,635	415	++	++	+++	+	+	+	+	+++	+	+	+	+	+	+	+	+	+	+	5*	D87453
0265	5,551	401	+	+	+	+	+	+	n.d.	++	+	+	+	+	+	+	+	+	+	+	7*	D87454
0266[f]	5,585	766	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	13*	D87455
0267[e,f]	4,408	666	+	+	+	++	+	+	+	++	+	+	+	+	+	+	+	+	+	+	X*	D87743
0268[e,f]	5,976	1,193	+	+	+	+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	1*	D87742
0269[f]	2,625	559	−	−	−	+++	−	−	−	−	+/−	+/−	−	+/−	−	++	+	+/−	+	+/−	6*	D87459
0270[g]	2,552	345	+	+	+	++	+	+	+	+	+	+	+	+	+	+	++	+	+	−	19*	D87460
0271[e,f,g]	3,542	193	+	+	+++	+++	+	+	+	+	+	+++	+	+	++	++	+	+	++	+	14*	D87461
0272[f]	3,474	726	+	+	++	++	+	+	+	+++	+	+	+	+	+	+++	+	++	++	+	3*	D87462
0273	3,040	330	−	−	−	+++	−	−	−	−	−	−	−	−	+/−	+	+/−	+/−	+/−	−	8*	D87463
0274[e,f]	3,010	907	+	+	+	+	+	+	+	+	+	+	+	+	+	++	+	+	+	+	6*	D87464
0275[e,f,g]	5,316	424	+	+	+	++	+/−	+	−	+	+	−	+	++	+/−	+	+	+	−	++	10*	D87465
0276[f]	4,185	309	n.d.	n.d.	n.d.	+	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	4*	D87466

Table 3. Continued.

Gene number (KIAA)	Total length of cDNA (bp)[a]	Amino acid residues	Expression[b] KG-1	HeLa	He	Br	Pl	Lu	Li	Sk.m	Ki	Pa	Sp	Th	Pr	Te	Ov	Sm.i	Co	Pe.b	Chromosomal[c] location	Accession[d] number
0277[f]	5,900	580	−	−	+/−	+	+/−	+/−	−	−	−	+/−	+	−	−	+	−	+/−	−	−	7*	D87467
0278[f]	2,935	460	−	−	+	+	−	+	−	+	−	+	−	−	+	+	+/−	+/−	+/−	+/−	8*	D87468
0279[e,f,g]	8,924	2,408	−	−	+/−	++	−	−	−	+	+	+/−	−	−	+	++	−	−	+/−	+/−	1*	D87469
0280	6,837	291	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	11*	D87470

n.d., not determined; He, heart; Br, brain; Pl, placenta; Lu, lung; Li, liver; Sk.m., skeletal muscle; Ki, kidney; Pa, pancreas; Sp, spleen; Th, thymus; Pr, prostate; Te, testis; Ov, ovary; Sm.i, small intestine; Co, colon; Pe.b, peripheral blood leukocytes.

a) Values excluding poly(A) sequences.

b) Expression of mRNA in indicated cells and human tissues (Clontech, USA) was examined by northern hybridization, and the strength of the positive signals are indicated (+/−, +, ++, +++).

c) Asterisks indicate that choromosome localization has been determined only by radiation hybrid mapping. In the others, the panels of both radiation hybrid and human-rodent hybrid were used.

d) Accession number of GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases.

e) Putative transmembrane domains were contained (see Supplemental pages).

f) Similarities to known genes were identified (see **Table 1** and Supplemental pages).

g) Protein motifs were recognized (see **Table 2** and Supplemental pages).

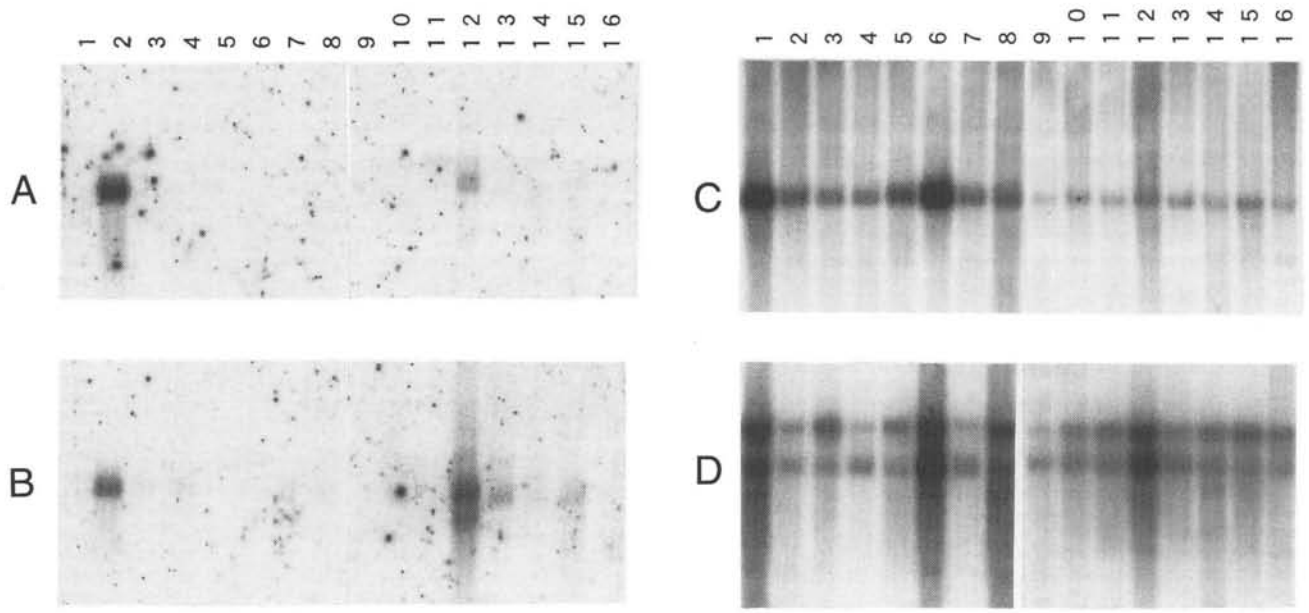


Figure 5. The typical expression patterns of representative genes. cDNA fragments were randomly labeled and hybridization was carried out as described previously. Human MTN blots were purchased from CLONTECH Laboratories, Inc.. **A**, KIAA0273; **B**, KIAA0269; **C**, KIAA0264; **D**, KIAA0262. Lane 1, heart; 2, brain; 3, placenta; 4, lung; 5, liver; 6, skeletal muscle; 7, kidney; 8, pancreas; 9, spleen; 10, thymus; 11, prostate; 12, testis; 13, ovary; 14, small intestine; 15, colon; 16, peripheral blood leukocyte.

clones yielded relatively strong signals in both heart and skeletal muscle (as shown in Figs. 5C and D, lanes 1 and 6).

The chromosomal location of these genes has been determined by the panels of radiation hybrid[6] and/or human-rodent hybrid cell lines (see Table3).[1]

Acknowledgments: This project was supported by grants from the Kazusa DNA Research Institute Foundation.

References

1. Nomura, N., Miyajima, N., Sazuka, T. et al. 1994, Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1, *DNA Res.*, **1**, 27–35, Supplement; **1**, 47–56.
2. Nagase, T., Seki, N., Ishikawa, K.-i., Tanaka, A., and Nomura, N. 1996, Prediction of the coding sequences of unidentified human genes. V. The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by analysis of cDNA clones from human cell line KG-1, *DNA Res.*, **3**,

17–24, Supplement; **3**, 43–53.
3. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D. et al. 1995, Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence, *Nature*, **377** (Supplement), 3–174.
4. Houlgatte, R., Mariage-Samson, R., Duprat, S. et al. 1995, The Genexpress Index: A resource for gene discovery and the genic map of the human genome, *Genome Res.*, **5**, 272–304.
5. Okubo, K., Hori, N., Matoba, R. et al. 1992, Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression, *Nature Genet.*, **2**, 173–179.
6. Schuler, G. D., Boguski, M. S., Stewart, E. A. et al. 1996, A gene map of the human genome, *Science*, **274**, 540–546.
7. Sussman, D. J., Klingensmith, J., Salinas, P., Adams, P. S., Nusse, R., and Perrimon, N. 1994, Isolation and characterization of a mouse homolog of the *Drosophila* segment polarity gene *dishevelled, Dev. Biol.*, **166**, 73–86.
8. Fortini, M. E. and Artavanis-Tsakonas, S. 1993, *Notch*: Neurogenesis is only part of the picture, *Cell*, **75**, 1245–1247.
9. Kiefer, M. C., Brauer, M. J., Powers, V. C. et al. 1995, Modulation of apoptosis by the widely distributed Bcl-2 homologue Bak, *Nature*, **374**, 736–739.
10. Takeichi, M. 1990, Cadherins: A molecular family important in selective cell adhesion, *Annu. Rev. Biochem.*, **59**, 237–252.
11. Rosenfeld, R. and Margalit, H. 1993, Zinc fingers: conserved properties that can distinguish between spurious and actual DNA-binding motifs, *J. Biomol. Struct. Dyn.*, **11**, 557–570.
12. Haupt, Y., Alexander, W. S., Barri, G., Klinken, S. P., and Adams, J. M. 1991, Novel zinc finger gene implicated as *myc* collaborator by retrovirally accelerated lymphomagenesis in Eμ-*myc* transgenic mice, *Cell*, **65**, 753–763.
13. Kataoka, T., Broek, D., and Wigler, M. 1985, DNA sequence and characterization of the S. cerevisiae gene encoding adenylate cyclase, *Cell*, **43**, 493–505.
14. Sadhu, K., Reed, S. I., Richardson, H., and Russell, P. 1990, Human homolog of fission yeast cdc25 mitotic inducer is predominantly expressed in G_2, *Proc. Natl. Acad. Sci. USA*, **87**, 5139–5143.
15. Lowy, D. R. and Willumsen, B. M. 1989, New clue to ras lipid glue, *Nature*, **341**, 384–385.
16. Mandel, J.-L. 1994, Trinucleotide diseases on the rise, *Nature Genet.*, **7**, 453–455.
17. Yulug, I. G., Yulug, A., and Fisher, E. M. C. 1995, The frequency and position of *Alu* repeats in cDNAs, as determined by database searching, *Genomics*, **27**, 544–548.
18. Pelham, H. R. B. 1986, Speculations on the functions of the major heat shock and glucose-regulated proteins, *Cell*, **46**, 959–961.
19. Walker, J. E., Saraste, M., Runswick, M. J., and Gay, N. J. 1982, Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold, *EMBO J.* **1**, 945–951.
20. Hunter, T. 1991, Protein Kinase Classification, In: Hunter, T., Sefton, B. M. (eds) Methods in Enzymology, Vol. 200. Academic Press, New York, pp. 3–37.
21. Mathews, F. S. 1985, The structure, function and evolution of cytochromes, *Prog. Biophys. Mol. Biol.*, **45**, 1–56.
22. Bairoch, A. and Rudd, K. E. 1995, unpublished observations.
23. Linder, P., Lasko, P., Ashburner, M. et al. 1989, Birth of the DEAD box, *Nature*, **337**, 121–122.
24. Beck, S. and Barrell, B. G. 1988, Human cytomegalovirus encodes a glycoprotein homologous to MHC class-I antigens, *Nature*, **331**, 269–272.
25. El Hassouni, M., Henrissat, B., Chippaux, M., and Barras, F., 1992, Nucleotide sequences of the *arb* genes, which control β-glucoside utilization in *Erwinia chrysanthemi, J. Bacteriol.*, **174**, 765–777.
26. Nelson, D. R., Kamataki, T., Waxman, D. J. et al. 1993, The P450 superfamily: Update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature, *DNA and Cell Biology*, **12**, 1–51.
27. Ziegler, S. F., Bird, T. A., Schneringer, J. A., Schooley, K. A., and Baum, P. R. 1993, Molecular cloning and characterization of a novel receptor protein tyrosine kinase from human placenta, *Oncogene*, **8**, 663–670.
28. Schimmel, P. 1987, Aminoacyl tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs, *Annu. Rev. Biochem.*, **56**, 125–158.
29. Futai, M., Noumi, T., and Maeda, M. 1989, ATP synthase (H^+-ATPase): Results by combined biochemical and molecular biological approaches, *Annu. Rev. Biochem.*, **58**, 111–136.
30. Koch, N., Lauer, W., Habicht, J., and Dobberstein, B. 1987, Primary structure of the gene for the murine Ia antigen-associated invariant chains (Ii). An alternatively spliced exon encodes a cysteine-rich domain highly homologous to a repetitive sequence of thyroglobulin, *The EMBO Journal*, **6**, 1677–1683.