Influence of the 3'-UTR-length of mKIAA cDNAs and their Sequence Features to the mRNA Expression Level in the Brain

Noriko OKAZAKI, Kazuhide IMAI, Reiko F. KIKUNO, Kazuharu MISAWA, Makoto KAWAI, Susumu INAMOTO, Reiko OHARA, Takahiro NAGASE, Osamu OHARA, and Hisashi KOGA

Kazusa DNA Research Institute, 2-6-7 Kazusa-Kamatari, Kisarazu, Chiba 292-0818, Chiba Industry Advancement Center, 2-6 Nakase, Mihama-ku, Chiba 261-7126, Institute of Research and Innovation, 1201 Takada, Kashiwa, Chiba 277-0861 and RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama City, Kanagawa, 230-0045, Japan

(Received 24 January 2005; revised 4 April 2005)

Abstract

We have previously described the sequence features of ~1500 mouse KIAA (mKIAA) genes in comparison with those of human KIAA genes (Okazaki, N., Kikuno, R., Inamoto, S., Hara, Y., Nagase, T., Ohara, O., and Koga, H. 2002, DNA Res., 9, 179–188; Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Aizawa, H., Yuasa, S., Nakajima, D., Nagase, T., Ohara, O., and Koga, H. 2003, DNA Res., 10, 35–48; Okazaki, N., Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S., Saga, Y., Nagase, T., Ohara, O., and Koga, H. 2003, DNA Res., 10, 167–180; and Okazaki, N., F-Kikuno, R., Ohara, R., Inamoto, S., Koseki, H., Hiraoka, S., Saga, Y., Seino, S., Nishimura, M., Kaisho, T., Hoshino, K., Kitamura, H., Nagase, T., Ohara, O., and Koga, H. 2004, DNA Res., 11, 205–218). To validate the orthologous relationship between mKIAA and KIAA genes in detail, we examined their chromosomal positions and evolutionary rate of synonymous substitutions and confirmed that >93% of the mKIAA/KIAA gene pairs are orthologous. During the sequence analysis of mKIAA genes, we found that 3'-untranslated region (3'-UTR) lengths of mKIAA and KIAA genes are extremely long. In the meanwhile, we have also examined the tissue-specific expression of ~1700 mKIAA genes using cDNA microarray and verified predominantly their expression in adult brain (Koga, H., Yuasa, S., Nagase, T., Shimada, K., Nagano, M., Imai, K., Ohara, R., Nakajima, D., Murakami, M., Kawai, M., Miki, F., Magae, J., Inamoto, S., Okazaki, N., Ohara, O. 2004, DNA Res., 11, 293–304). To connect these two evidences, we statistically analysed the relationship between them by using the mKIAA genes. Consequently, a positive correlation was observed between the 3'-UTR lengths and the relative expression intensities in adult brain. Furthermore, we searched sequence elements in the 3'-UTR possibly related with their expression and found some candidates regarding the brain-specific expression.

Key words: mKIAA; orthology; 3'-UTR; expression; brain

1. Introduction

After completion of the human genome project, numerous genes have been predicted by bioinformatic approaches. However, library construction still serves as a critical resource of actually existing genes and a platform for comprehensive analysis of the gene functions. With this in mind, we have constructed several cDNA libraries and initiated a human cDNA project to accumulate information regarding the long protein coding sequences (CDSs) of unidentified human genes since 1994.1 We have isolated and entirely sequenced long human cDNA clones (>4 kb) and have already registered >2000 genes (KIAA genes) to the public database to date. However, the function has been identified in some genes, half of the genes are still remained to be elucidated mainly depending on their difficulty for using human materials. Therefore, we decided to collect mouse KIAA cDNAs (mKIAA) for further functional analysis in a model organism.

We have previously reported the sequences of ~1500 mKIAA cDNAs and the tissue-specific expression...
of ~1700 mKIAA genes using cDNA microarray.\(^2\) The entire sequence of mKIAA cDNAs revealed that mKIAA (also KIAA) cDNAs have extremely long 3’-untranslated region (3’-UTR) sequences. Furthermore, cDNA microarray analysis elucidated that 30% of the genes are predominantly expressed in the brain. We thus assumed some correlations lie between the two observations. The 3’-UTR sequences are critical for determining mRNA stability, mRNA targeting and level of translation.\(^7\)\(^\text{13}\) The sequences are also known to be important for the pathogenesis of some disease through regulation of mRNA stability.\(^14\) Although molecular neuroscientists have empirically noted that mRNAs abundant in the brain have long 3’-UTR, statistical or comprehensive approaches on the correlation between the expression levels and the 3’-UTR lengths have not been performed.

In this paper, we first validate the orthology of mKIAA and KIAA cDNA pairs; subsequently, we report the statistical correlation between the relative expression levels in the brain and the 3’-UTR lengths and discuss the possible functional sequence elements in the 3’-UTRs found in mKIAA and KIAA cDNAs.

2. Materials and Methods

2.1. Nucleotide sequences

We obtained mKIAA and KIAA cDNA sequences from our ROUGE (http://www.kazusa.or.jp/rouge/index.html) and HUGE (http://www.kazusa.or.jp/huge/index.html) databases, respectively.

2.2. Estimation of orthology

We estimated the evolutionary relationship between KIAA and mKIAA by the following steps: (i) find chromosomal position of KIAA genes using the NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi/),\(^15\) (ii) find the mouse chromosomal region corresponding to the human chromosomal region encoding the KIAA genes using the NCBI Map Viewer; (iii) find the chromosomal position of mKIAA using BLAST search,\(^16\) and (iv) compare the mouse chromosomal region found in (ii) with the chromosomal position of mKIAA gene found in (iii). The number of synonymous substitutions per site (dS) was estimated by Nei and Gojobori’s method\(^17\) using SNAP program.\(^18\)\(^,19\)

2.3. A search for candidate elements effecting on the specific expression in brain

We extracted conserved 3’-UTR sequences between 50 mKIAA highly expressing in brain and their orthologous KIAA cDNAs. The MEME program (http://meme.sdsc.edu/meme/website/meme.html) was used to find 20–25 bp elements that appeared >50 times. The existence of newly identified sequence elements was searched with >60% nucleotide identity against the 3’-UTR sequences of 1031 mKIAA cDNAs, which were verified for the integrity of the 3’-UTR structure. The existence of the elements was also searched against 3’-UTRs of non-mKIAA genes in UTRdb (http://www.ba.itb.cnr.it/BIG/UTRHome/) entries using BLAST search. Differences of the relative expression levels in brain between mKIAA cDNAs with and without the elements were calculated by the Student’s t-test. The tissue specific expression of each gene was based on the descriptions in the literature or our original data (freely available through our database).\(^6\)

3. Results

3.1. Identification of mouse KIAA-homologous cDNAs

The mouse homologs of KIAA genes were isolated from size-fractionated cDNA libraries (Table 1). Approximately 55% of KIAA cDNAs (1116 out of

---

### Table 1. The details of cDNA libraries used in this study.

| Tissue                        | Age          | Strain | Number of clones sequenced from their 5’ or 3’ end | Number of clones entirely sequenced |
|-------------------------------|--------------|--------|---------------------------------------------------|-----------------------------------|
| Brain                         | Adult        | BALB/c | 23 290                                            | 712                               |
| Tail                          | Embryo       | ICR    | 31 537                                            | 404                               |
| Brain                         | Fetus (18 dpc)| ICR    | 20 929                                            | 152                               |
| Pancreas                      | Adult        | C57BL/6| 29 334                                            | 82                                |
| Thymus                        | Adult        | C57BL/6| 8149                                              | 54                                |
| Intestine                     | Embryo       | ICR    | 12 399                                            | 52                                |
| Spleen                        | Adult        | ICR    | 17 429                                            | 22                                |
| Natural killer T cells        |              | C57BL/6| 11 816                                            | 19                                |
| Bone marrow-derived dendritic cells |          | C57BL/6| 2113                                              | 1                                 |
| Splenocytes                   | Adult        | ICR    | 2917                                              | 1                                 |
| Total                         |              |        | 159 913                                           | 1499                              |

\(^a\) The brain library consists of ~3 × 10⁶ independent clones.
2038 cDNAs) were derived from adult human brain libraries; thus, we initially subjected 23,000 end sequences of adult mouse brain library for the screening of mKIAA genes and identified nearly half of mKIAA cDNAs. The remainder was isolated from other libraries and consequently 1499 mKIAA homologs were successfully isolated and entirely sequenced. By comparing the mKIAA cDNA sequences with human sequences, we designated those 1479 mouse cDNA clones as ‘mKIAA’ and the same four-digit number corresponding to human clones. Although 20 cDNA clones were eventually found not to be orthologous to any KIAA cDNAs, these cDNAs were conventionally designated as ‘mKIAA’ and a four-digit number that has not been allocated for human KIAA genes (mKIAA3000s). The average length of the cDNA inserts was 4.6 kb and deduced gene products was 830 amino acid residues. Multiple CDSs were found in 237 mKIAA cDNAs that are longer than 50 amino acid residues and have high amino acid identity (>50%) to corresponding KIAA by FASTA analysis. These multiple CDSs are thought to be produced by spurious CDS splits caused by retained intron(s), alternative splicing or splicing error, reverse transcription error(s) or other cloning artifacts. We evaluated the spurious CDS splits through the sequence comparison of mouse and human KIAA cDNA pairs and the genome/cDNA structures, and we found that 144 mKIAA cDNA clones appeared to retain intron(s). Taking the assumption into consideration, alternative prediction of the deduced gene products might increase to 873 amino acid residues.

3.2. Orthology between mouse and human KIAAs

We have only defined the cDNA sequences as mouse homologs of KIAA genes when they showed the highest homology against the corresponding human KIAA genes at the time of being sequenced. To support orthologous relationship of mKIAA and KIAA gene pairs, we assigned their chromosomal positions and examined whether they were derived from the common ancestral chromosomal location. We classified the positional relations of the gene pairs into the following three categories using the NCBI Map Viewer: (i) the 896 gene pairs located on the corresponding chromosomal regions between mouse and human; they would be judged to be orthologous; (ii) the 102 gene pairs not located on the corresponding chromosomal regions; they could not be judged to be orthologous; and (iii) the 481 gene pairs whose sequence or data on the chromosome location were not available; also, they could not be judged to be orthologous.

For further verification of the orthology, we also investigated the evolutionary rate of dS estimated by the Nei and Gojobori’s method. The rate of human/mouse orthologous pair is approximately constant, since they are diverged concurrently 80 million years ago and accumulated dS during the same period. Paralogous gene pairs usually show higher rate of dS than orthologous gene pairs. The dS of 1479 mKIAA/KIAA pairs (0.64 ± 0.22, max. = 4.86, min. = 0.01) is slightly higher but consistent with the previous study that calculated dS of orthologous gene pairs between human and mouse (0.55 ± 0.63). Therefore, most of the mKIAA/KIAA pair is thought to be orthologous.
Table 2. Orthology between mouse and human KIAAs.

| Distance from the mean<sup>a</sup> | Synonymous substitution<sup>b</sup> | Chromosomal localization<sup>c</sup> | Judgement by phylogenetic tree analysis<sup>d</sup> | mKIAA |
|----------------------------------|----------------------------------|-----------------------------------|-----------------------------------------------|-------|
| >Mean + 4 SD                     | NA                               | 2                                 | P                                             | mKIAA1032 |
|                                  | NA                               | 3                                 | O                                             | mKIAA1710 |
|                                  | 4.86                             | 1                                 | O                                             | mKIAA1561 |
|                                  | 2.62                             | 3                                 | O                                             | mKIAA0961 |
|                                  | 1.91                             | 2                                 | P                                             | mKIAA0420 |
| >Mean + 3 SD                     | 1.44                             | 1                                 | P                                             | mKIAA1811 |
|                                  | 1.36                             | 1                                 | O                                             | mKIAA0908 |
|                                  | 1.36                             | 1                                 | O                                             | mKIAA1023 |
| >Mean + 2 SD                     | 1.28                             | 3                                 | O                                             | mKIAA1881 |
|                                  | 1.28                             | 3                                 | O                                             | mKIAA1913 |
|                                  | 1.26                             | 3                                 | O                                             | mKIAA1273 |
|                                  | 1.25                             | 2                                 | P                                             | mKIAA1074 |
|                                  | 1.23                             | 1                                 | O                                             | mKIAA0883 |
|                                  | 1.23                             | 1                                 | O                                             | mKIAA1162 |
|                                  | 1.23                             | 3                                 | O                                             | mKIAA1483 |
|                                  | 1.22                             | 1                                 | O                                             | mKIAA0767 |
|                                  | 1.20                             | 3                                 | O                                             | mKIAA1600 |
|                                  | 1.19                             | 1                                 | O                                             | mKIAA1335 |
|                                  | 1.18                             | 3                                 | O                                             | mKIAA1323 |
|                                  | 1.17                             | 3                                 | O                                             | mKIAA1466 |
|                                  | 1.17                             | 3                                 | P                                             | mKIAA0088 |
|                                  | 1.16                             | 3                                 | O                                             | mKIAA0698 |
|                                  | 1.16                             | 1                                 | O                                             | mKIAA1878 |
|                                  | 1.15                             | 2                                 | O                                             | mKIAA1948 |
|                                  | 1.14                             | 1                                 | O                                             | mKIAA1015 |
|                                  | 1.10                             | 1                                 | O                                             | mKIAA0213 |
|                                  | 1.10                             | 1                                 | O                                             | mKIAA1676 |
|                                  | 1.10                             | 1                                 | O                                             | mKIAA0082 |
|                                  | 1.10                             | 3                                 | P                                             | mKIAA1120 |
|                                  | 1.10                             | 1                                 | O                                             | mKIAA0610 |
|                                  | 1.09                             | 1                                 | O                                             | mKIAA0439 |
|                                  | 1.08                             | 1                                 | O                                             | mKIAA1389 |
|                                  | 1.08                             | 1                                 | O                                             | mKIAA1716 |
|                                  | 1.08                             | 3                                 | O                                             | mKIAA0365 |
|                                  | 1.08                             | 1                                 | O                                             | mKIAA1729 |
|                                  | 1.08                             | 3                                 | O                                             | mKIAA0814 |
|                                  | 1.08                             | 1                                 | O                                             | mKIAA0803 |
|                                  | 1.08                             | 1                                 | O                                             | mKIAA0207 |
|                                  | 1.07                             | 3                                 | O                                             | mKIAA0637 |

The orthology of top 100 mKIAA/KIAA gene pairs that shows the high rate of dS were investigated in detail by phylogenetic tree analysis.

<sup>a</sup> The mean value of dS was 0.635 and the SD was 0.221.

<sup>b,c</sup> Value of dS and chromosomal localization were estimated as described in the legend of Fig. 1.

<sup>d</sup> Phylogenetic trees of KIAA and mKIAA genes were constructed and analysed as described in Fig. 1b. If the pair was judged as orthologous, it is indicated by ‘O’, otherwise, it is indicated by ‘P’. Because of occupancy of too much space the data for < mean + 2 SD were omitted (61 samples). In this range there is a paralog (mKIAA0345; ds = 1.06885; chromosomal localization = no data).
pairs were thought to be orthologous. We then compared the distribution of the dS values among the above-mentioned three categories based on the chromosomal positions (Fig. 1a). The distributions of dS values were quite similar, therefore, not many paralogs would be included if they exist in the gene pairs whose correspondence was not supported by the chromosomal locations. Furthermore, to confirm the orthology we examined in detail the orthology of the 100 gene pairs that showed unusually high rate of dS (mean + 1.31 SD) by the phylogenetic analysis (Table 2, Fig. 1b). Among them 93 mKIAA genes were confirmed to be orthologs, whereas the remaining seven mKIAA genes (mKIAA0345, mKIAA0420, mKIAA0588, mKIAA1032, mKIAA1074, mKIAA1120 and mKIAA1811) were determined to be paralogs of the corresponding KIAA genes. These results suggest that most of the KIAA/mKIAA gene pairs have orthologous relationship, and if the paralogous pairs existed it must be <7%.

3.3. Sequence comparison of mouse and human KIAA cDNA pairs

In our previous study, we reported the sequence identity of 3′-UTR and CDSs for the 100 mouse and human KIAA cDNA pairs. We extended the same analysis to the 1468 gene pairs (Table 3). Similar to our previous study, the average aligned length of CDS and 3′-UTR of mKIAA and KIAA cDNAs are 2–3 times longer than reported by the other group, while the observed sequence identities of KIAA and mKIAA cDNAs in CDS and 3′-UTR agreed with the values by the other group. In Saccharomyces cerevisiae, the UTR lengths are in a narrow range whereas the CDS lengths widely change in parallel with the mRNA lengths. Although this narrow range of the UTR lengths is applicable to other species, the considerably long 3′-UTR sequences of mKIAA/KIAA cDNAs would be exceptional. Thus, the longer length and high conservation of 3′-UTR sequences of mKIAA/KIAA cDNAs may imply the unverified important roles in the sequences.

3.4. Positive correlation between the brain-specific expression and the 3′-UTR length of the genes

Previously, we examined tissue-specific expression of mKIAA genes using mKIAA cDNA microarray. On the microarray, 1467 out of 1499 mKIAA cDNAs were spotted in this study and ~30% of the genes were predominantly expressed in the brain. Therefore, we focused on the long 3′-UTRs to examine the relationship between brain-specific expression and their sequence features. Statistical analysis of ~1031 clones using the Kendall rank correlation measurement revealed a significant correlation between 3′-UTR lengths and relative expression levels in the brain, exhibiting correlation coefficient (Tau) = 0.16, P = 1.7 × 10−14. However, there were no significant correlation between relative expression levels in other tissues and their 3′-UTR lengths (Table 4). To exclude the bias against the sources of cDNA libraries, the clones were subdivided into two samples to avoid this bias.
Influence of 3′-UTR of mKIAA cDNA

Table 5. The relative level of expression in brain was positively correlated with the length of 3′-UTR but not the length of CDS.

| Region\(^a\) | Tissue\(^b\) | Selection rule\(^c\) of clones | Sample number | Kendall rank correlation (Tau) | Probability value |
|-------------|--------------|-------------------------------|---------------|--------------------------------|-------------------|
| 3′-UTR length | Brain | 3′-terminal | 491 | 0.14 | 7.3 × 10\(^{-6}\) |
| 3′-UTR length | Brain | 5′- and 3′-terminal | 184 | 0.14 | 0.00048 |
| CDS length | Brain | 5′- and 3′-terminal | 184 | −0.056 | 0.26 |
| 5′-UTR length | Brain | 5′- and 3′-terminal | 184 | −0.029 | 0.55 |
| Total length | Brain | 5′- and 3′-terminal | 184 | 0.089 | 0.072 |
| 3′-UTR length | Other | 3′-terminal | 540 | 0.13 | 1.3 × 10\(^{-5}\) |
| 3′-UTR length | Other | 5′- and 3′-terminal | 217 | 0.12 | 0.0089 |
| CDS length | Other | 5′- and 3′-terminal | 217 | −0.064 | 0.16 |
| 5′-UTR length | Other | 5′- and 3′-terminal | 217 | −0.029 | 0.53 |
| Total length | Other | 5′- and 3′-terminal | 217 | −0.031 | 0.49 |

\(^{a,b,c}\) We further selected 401 clones (184 clones from brain-derived library, 217 clones from other tissues-derived library) from 1031 clones in Table 4, which were verified for the integrity of the CDS and the 5′-UTR structures. Our definition of the integrity is as follows, without warning for N-terminal truncation of the coding region by GeneMark analysis.\(^{37}\) Correlation between lengths of 3′-UTR, CDS and 5′-UTR and relative expression levels in each tissue (the expression value of each tissue/the mean expression value of 13 tissues) were calculated using the Kendall rank correlation measurement.\(^{b}\) To exclude the bias against the sources of cDNA libraries, the clones were subdivided into two groups (brain-derived and other tissues-derived).

4. Discussion

To discover novel functional aspect of UTRs, we prepared 1479 mKIAA/KIAA cDNA pairs and verified the orthology from their chromosomal location and the evolutionary rate of dS. Using these mouse cDNAs, we identified a positive correlation between the 3′-UTR length and the relative expression levels in the brain. Paying attention to the function of each gene, for instance, it appears that genes involved in G-protein signaling and vesicle trafficking dominantly expressing in the brain tended to have long 3′-UTR (data not shown). It was already assumed that 3′-UTR length increases with evolutionary age and organism complexity.\(^{26}\) Moreover, molecular neuroscientists have empirically noted that mRNAs, the most abundant in the brain, have long 3′-UTR. Therefore, it seems reasonable to propose that the genes predominantly expressed in mammalian brain have evolved to have long 3′-UTR. Growing evidences of RNA-binding proteins also demonstrates the importance of various 3′-UTR elements in regulation of mRNA turnover at the posttranscriptional level.\(^{37}\) These evidences and our result motivated us to find novel sequence elements in the 3′-UTR.

MEME is a web-based tool for discovering elements in a group of related DNA sequences, thus we applied this program to find the highly conserved sequence elements...
Table 6. Sequence elements found in the 3'-UTR mKIAA/KIAA gene pairs showing relatively high expression in the brain.

| Element | Consensus sequence | Relative level of expression in brain | t-test P-value | Number of mKIAA cDNA with the motif | Definition of non-mKIAA gene with the element | Species | High expression in brain | Reference |
|---------|--------------------|---------------------------------------|---------------|------------------------------------|-----------------------------------------------|---------|-------------------------|-----------|
|         |                    | with Mean ± SD                         | without Mean ± SD |                                    |                                               |         |                         |           |
| 1       | CCATCAACAGAGAAGGAGTT | 1.8 ± 1.2                             | 1.4 ± 0.9     | 3.7 × 10⁻¹⁰                        | FLI44606                                      | Human   | ?                       | No        |
| 2       | AAGGAAACCATGTAAGAGAAAGAG | 1.9 ± 1.2                             | 1.5 ± 1.0     | 2.7 × 10⁻¹⁰                        | Thioredoxin domain containing 5               | Human   | ?                       | No        |
| 3       | TCTTTTGTAAAAGAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Hypothetical protein                          | Human   | ?                       | Yes       | http://www.kazusa.or.jp |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | KIAA0269 (mKIAA0269 was not yet cloned)      | Human   | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | YB2                                           | Human   | Ubiquitous              | (38)      |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | vms-tm2 single-pass transmembrane protein     | Rat     | Yes                     | (39)      |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Unknown IMAGE:6530404                       | Rat     | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Highly similar to Homo sapiens Sadl           | Mouse   | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Protooncogene A-myb (MYBL2)                   | Mouse   | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | DKFZp451P2311                                 | Bovine  | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | DKFZp566G1424                                | Human   | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Glutathione S-transferase A1                  | Human   | ?                       | No        |
|         | TCTTTTGTAAAAGAGAAAT | 1.8 ± 1.1                             | 1.4 ± 0.9     | 2.9 × 10⁻⁷                         | Hypothetical protein LOC138846               | Human   | ?                       | No        |
| 4       | AATGCAGATGTAATGCAATTAA | 1.9 ± 1.2                             | 1.5 ± 1.0     | 3.3 × 10⁻⁶                         | Hypothetical protein                          | Human   | ?                       | No        |
|         | AATGCAGATGTAATGCAATTAA | 1.9 ± 1.2                             | 1.5 ± 1.0     | 3.3 × 10⁻⁶                         | Thioredoxin domain containing 5               | Human   | ?                       | No        |
| 5       | AAATATTACGTGTATCTTA | 1.7 ± 1.2                             | 1.4 ± 0.9     | 1.0 × 10⁻⁵                         | DKFZp668A05136                               | Human   | ?                       | No        |
| 6       | CAGTCAGAAAGCAGTGGGTCGC | 1.7 ± 1.2                             | 1.5 ± 0.9     | 2.1 × 10⁻⁵                         | Ring finger protein 130/goliath               | Human   | ?                       | No        |
| 7       | CCTTTGACTCTGTGAGAGT | 1.7 ± 1.1                             | 1.4 ± 1.0     | 1.4 × 10⁻⁴                         | RUSH-1beta                                   | Rabbit  | ?                       | No        |
in mKIAA/KIAA cDNAs revealed relatively high expression in the brain. We found 43 statistically significant elements in mKIAA/KIAA cDNAs and identified 7 out of 43 elements in the 3'-UTRs of 18 non-mKIAA genes in UTRdb. Among the several elements in 3'-UTRs that regulate the mRNA levels posttranscriptionally, the novel elements in this study might be one of such kind of elements. Biochemical approaches as well as accumulation of the expression profile of non-mKIAA genes containing the elements may help the verification of the novel elements on the predominant expression in brain. Among those elements, adenylate uridylate-rich elements (AU-rich elements, AREs) are the best characterized elements in 3'-UTR, and posttranscriptionally regulate the cytoplasmic half-life of the mRNAs encoding various proteins that regulate cellular proliferation/differentiation and response to inflammatory and environmental stimuli. Cytidine-rich 15-lipoxygenase differentiation control element (15-LOX DICE) is another well-characterized element in 3'-UTR and is a multifunctional cis-element found in numerous eukaryotic mRNAs. Although there were many AREs and 15-LOX DICEs in mKIAA cDNAs, the numbers of the elements simply correlated with the 3'-UTR length but not obviously with the relative expression level in the brain (data not shown).

The elements involved in tissue, stage or cell-type specific expression of the genes had also been reported; however, the specific expression of the genes is partly due to the specific expression of certain RNA-binding proteins. For example, mRNA of membrane-bound IL-1R accessory protein expressing in a tissue-specific manner has several elements in the 3'-UTR and the stability is thought to be parallel to the expression of some RNA-binding protein. Neurofilament-M (NF-M) expression is stage-specific and culminates at the most mature stages of axon development. This alteration is partly regulated by the NF-M mRNA stability, parallel with the binding of hnRNP to NF-M 3'-UTRs. Especially in the brain, several RNA-binding proteins expressing in neuronal cell-type or stage-specific manner have been already identified; for instance, three ELAV-like proteins (HuB, HuC and HuD), Musashi, Autoimmune antigens Nova, polypyrimidine tract-binding protein-like protein and Drl. In the brain, posttranscriptional regulation might be more popular than any other organ and multiple RNA-binding proteins might act through the long 3'-UTR of target genes and control them in a complex manner. Since numerous RNA-binding proteins have been reported in particularly matured neuron in which the proteins might play important roles in terminal neuronal differentiation, perpetual neurite outgrowth/retraction and synaptogenesis. Since the long 3'-UTR of mKIAA/KIAA mRNAs potentially have multiple functional elements, the mRNA levels might be regulated by the combination of multiple brain-specific RNA-binding proteins. Accumulating information about sequence elements in mRNA, neuron-specific RNA-binding proteins and their interactions is promising to solve these complicated regulations.

Acknowledgements: We thank Nobue Kashima for her technical assistance. This project was supported by grants from the Kazusa DNA Research Institute. This study was also supported by the CREATE Program (Collaboration of Regional Entities for the Advancement of Technological Excellent) from JST (Japan Science and Technology Corporation).

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.
2. Okazaki, N., Kikunou, R., Ohara, R., et al. 2002, Prediction of the coding sequences of mouse homologues of KIAA gene: I. The complete nucleotide sequences of 100 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries, DNA Res., 9, 179–188.
3. Okazaki, N., Kikunou, R., Ohara, R., et al. 2003, Prediction of the coding sequences of mouse homologues of KIAA gene: II. The complete nucleotide sequences of 400 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries, DNA Res., 10, 35–48.
4. Okazaki, N., Kikunou, R., Ohara, R., et al. 2003, Prediction of the coding sequences of mouse homologues of KIAA gene: III. The complete nucleotide sequences of 500 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries, DNA Res., 10, 167–180.
5. Okazaki, N., F-Kikunou, R., Ohara, R., et al. 2004, Prediction of the coding sequences of mouse homologues of KIAA gene: IV. The complete nucleotide sequences of 500 mouse KIAA-homologous cDNAs identified by screening of terminal sequences of cDNA clones randomly sampled from size-fractionated libraries, DNA Res., 11, 205–218.
6. Koga, H., Yuasa, S., Nagase, T., et al. 2004, A comprehensive approach for establishment of the platform to analyze functions of KIAA proteins II: public release of inaugural version of InGaP database containing gene/protein expression profiles for 127 mouse KIAA genes/proteins, DNA Res., 11, 293–304.
7. Jackson, R. J. 1993, Cytoplasmic regulation of mRNA function: the importance of the 3' untranslated region, Cell, 74, 9–14.
8. Curtis, D., Lehmann, R., and Zamore, P. D. 1995, Translational regulation in development, Cell, 81, 171–178.
9. Hentze, M. W. 1995, Translational regulation: versatile mechanisms for metabolic and developmental control, Curr. Opin. Cell Biol., 7, 393–398.
10. Wickens, M., Anderson, P., and Jackson, R. J. 1997, Life and death in the cytoplasm: messages from the 3' end, Curr. Opin. Genet. Dev., 7, 220–232.
21. Wheeler, D. L., Church, D. M., Federhen, S., et al. 2003, Database resources of the National Center for Biotechnology Information, Nucleic Acids Res., 31, 28–33.

22. Kumar, S. and Hedges, S. B. 1998, A molecular timescale for vertebrate evolution, Nature, 392, 917–920.

23. Makalowski, W. and Boguski, M. S. 1998, Evolutionary parameters of the transcribed mammalian genome: an analysis of 2820 orthologous rodent and human sequences, Proc. Natl. Acad. Sci. U.S.A., 95, 9407–9412.

24. Saitou, N. and Nei, M. 1987, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol., 4, 406–425.

25. Huwowitz, E. H. and Brown, P. O. 2003, Genome-wide analysis of mRNA lengths in Saccharomyces cerevisiae, Genome Biol., 5, R2.