Nucleic Acids Research, 2002, Vol. 30, No. 24 5369–5375

# Functional second genes generated by retrotransposition of the X-linked ribosomal protein genes

**Tamayo Uechi, Noriko Maeda, Tatsuo Tanaka and Naoya Kenmochi***

Department of Biochemistry, School of Medicine, University of the Ryukyus, 207 Uehara, Nishihara, Okinawa 903-0215, Japan

Received September 19, 2002; Revised and Accepted October 29, 2002 DDBJ/EMBL/GenBank accession nos AB063605–AB063610

## ABSTRACT

**We have identified a new class of ribosomal protein (RP) genes that appear to have been retrotransposed from X-linked RP genes. Mammalian ribosomes are composed of four RNA species and 79 different proteins. Unlike RNA constituents, each protein is typically encoded by a single intron-containing gene. Here we describe functional autosomal copies of the X-linked human RP genes, which we designated *RPL10L* (ribosomal protein L10-like gene), *RPL36AL* and *RPL39L* after their progenitors. Because these genes lack introns in their coding regions, they were likely retrotransposed from X-linked genes. The identities between the retrotransposed genes and the original X-linked genes are 89–95% in their nucleotide sequences and 92–99% in their amino acid sequences, respectively. Northern blot and PCR analyses revealed that *RPL10L* and *RPL39L* are expressed only in testis, whereas *RPL36AL* is ubiquitously expressed. Although the role of the autosomal RP genes remains unclear, they may have evolved to compensate for the reduced dosage of X-linked RP genes.**

## INTRODUCTION

Ribosomes comprise the protein synthesis machinery that is essential for all living cells. Because of the fundamental role played by ribosomes in the growth and development of organisms, their structure and function have been significantly conserved during evolution. In higher eukaryotes, ribosomes are composed of two subunits: a large 60S subunit and a small 40S subunit, which includes four ribosomal RNA (rRNA) species and 79 distinct ribosomal proteins (RPs) (1,2). The genes encoding rRNAs are clustered at a few sites in the genome (3–6), whereas the genes encoding RPs are widely dispersed (7,8). The expression of these genes has to be coordinately regulated to ensure the equimolar assembly of the components into a ribosome particle (9).

Although the complete process for assembling the functional ribosome has not yet been elucidated, the influence of the gene dosage of each ribosomal component on development has been extensively studied in *Drosophila melanogaster*. For example, haploinsufficiency of any one of the RP genes yields a *Minute* phenotype, which includes short and thin bristles, reduced body size, diminished fertility and recessive lethality (10,11). In addition, *bobbed* and *mini* mutants, which display a phenotype similar to *Minute*, are also caused by a quantitative deficiency of rRNA genes. These phenotypes may reflect a reduced rate of protein synthesis resulting from insufficiency of ribosomal components in early development (11,12). Recently, heterozygous mutations of the human *RPS19* gene were found in 25% of unrelated patients with Diamond–Blackfan anemia (DBA), which is characterized by a decrease in or absence of erythroid precursors in the bone marrow (13,14). So far, this is the only reported case in which an RP gene mutation has been found associated with human disease, although how this RP defect causes the DBA phenotype is still unknown. Together with the above-mentioned *Drosophila* mutants, this implies that two copies of each RP gene are required for normal growth and development of multicellular organisms.

In mammalian female cells, one of the two X chromosomes is inactivated, which provides a dosage compensation mechanism to overcome sex differences. We have recently mapped four RP genes to the human X chromosome (7,8). If haploinsufficiency of any of these genes causes abnormal phenotypes as seen in *Drosophila Minute* mutants, either they must be twice as active as autosomal RP genes or there must be a second functionally redundant gene elsewhere in the genome. In fact, one of the X-linked RP genes *RPS4X*, has been shown to escape X inactivation and to have a functional homolog on the Y chromosome (*RPS4Y*) (15). However, none of the other X-linked RP genes have Y chromosome homologs. Interestingly, several X-linked housekeeping genes, including *PGK*, *PDHA*, *XAP-5*, *G6pd* and *Cent1*, have functional intronless copies on autosomes (16–20). These copies are believed to have retrotransposed from the X-linked genes and are responsible for compensating for the silenced genes during spermatogenesis. In this study, we have identified three new members of the human RP gene family, *RPL10L*, *RPL36AL* and *RPL39L*, that most likely

*To whom correspondence should be addressed at present address: Central Research Laboratories, Miyazaki Medical College, 5200 Kihara, Kiyotake, Miyazaki 889-1692, Japan. Tel: +81 985 85 9665; Fax: +81 985 85 1514; Email: kenmochi@post.miyazaki-med.ac.jp

**Table 1.** STSs used in the PCR-based expression analysis

| RP gene | Forward primer | Reverse primer | Size | Accession no. |
|---|---|---|---|---|
| *RPS4X* | ATCAAGTTCGACACTGGTAACC | ACATGTCACCCAGGGACC | 298 | NM_001007 |
| *RPS4Y* | TTAGGGACTGGCAAGATAATCA | TTCAGCAACAGTAAGTCGAATG | 261 | NM_001008 |
| *RPL10* | AGGGTTCACATTGGCCAAGTT | TAAGAGGGGGGCAGCACA | 292 | XM_018114 |
| *RPL10L* | GGGTCCACATTGGTCAAGTC | CCCAAGGAGACAGTACTGCC | 290 | AB063608 |
| *RPL36A* | GGAAAGCGGCGTTATGACA | GACACTTAGAACTGGATCACTTGG | 218 | NM_021029 |
| *RPL36AL* | GGAAGGAGGCGCTATGATC | TCCCTCGGGTAACTTTTCTATG | 280 | AB063609 |
| *RPL39* | GATTCGGATGAAAACTGGAAA | GAATCCAGCCAACCAACGT | 218 | NM_001000 |
| *RPL39L* | TGGATTCAGATGAAACCTGGT | ATCCACCCTACTAGCACAGAGC | 242 | AB063610 |

retrotransposed from X-linked genes. We also show their characteristic expression patterns in various tissues as well as in cancer cells.

## MATERIALS AND METHODS

### Accession numbers

The sequence data described in this paper have been submitted to the DDBJ/EMBL/GenBank DNA databases under accession numbers AB063605–AB063610.

### Database search and sequences

Genomic sequences were retrieved from the human draft sequence by BLAST search using the cDNA sequences of the X-linked RP genes as the query. We determined the intron/exon structure by comparing the genomic sequences with the full-length cDNA sequences that were assembled *in silico* from retrieved expression sequence tags (ESTs) or cDNA sequences of the genes. The assembled cDNA sequences will appear in the DDBJ/EMBL/GenBank DNA databases under accession numbers AB063608–AB063610.

### Radiation hybrid mapping

The GeneBridge 4 and Stanford G3 radiation hybrid (RH) panels (Research Genetics) were employed for mapping the newly identified genes. We tested the panels by PCR using sequence tagged sites (STSs) generated in the draft sequence. The STS sequences were verified by re-sequencing the PCR products and deposited in the DDBJ/EMBL/GenBank DNA databases under accession numbers AB063605–AB063607. The data vectors were submitted to the RH servers at the Whitehead Institute/MIT Center for Genome Research (GeneBridge 4; http://www-genome.wi.mit.edu/cgi-bin/contig/rhmapper.pl) or the Stanford Human Genome Center (G3; http://shgc-www.stanford.edu). Sequences of the primer pairs are as follows: *RPL10L*, 5′-GCCTCAGGACTCTATGGTTCC and 5′-CAGGTCAAAGATGCGGATCT; *RPL36AL*, 5′-CAAAGTGCTGGGATTACAAGC and 5′-CAGCAGGGCTGTTTTGTCTATA; *RPL39L*, 5′-GGATCCTGAGTGGCAATGAG and 5′-TTCATCTGAATCCACTGGGG.

### Northern blot analysis

The expression patterns of the RP genes were analyzed on two sets of commercially available poly(A)$^+$ RNA blots [Human multiple-tissue northern (MTN) Blot I and II, CLONTECH]. To avoid cross-reaction during hybridization, probes were generated against the 3′-non-coding region of the genes. The 5′ end of probe was labeled with [$\gamma$-$^{32}$P]ATP (Amersham Biosciences) using T4 polynucleotide kinase (MEGALABEL™, Takara). The blots were pre-treated by ExpressHyb™ hybridization solution (CLONTECH) for 30 min at 37°C, and then hybridized with the labeled probe for 1.5–2 h at 37°C. After washing according to the manufacturer's protocol, the blots were exposed to the BAS1500 system imaging plate (FUJI FILM) overnight and analyzed by the attached ImageGauge program. Sequences of the synthesized probes are as follows: *RPL10*, 5′-GTGAGTATTAAGAGGGGGGCAGCACATTGG; *RPL10L*, 5′-GCCAGTAAACAGAATTTATTAGTAAGCATA; *RPL36A*, 5′-GCCAGTAAACAGAATTTATTAGTAAGCATA; *RPL36AL*, 5′-CGGGTAACTTTTCTATGGCTTCACCA; *RPL39*, 5′-GTGTTCATAACAGATTCAGAGAGGA; *RPL39L*, 5′-TACTAGCACAGAGCATACAGAAA. The accession numbers of cDNA sequences from which the probes were chosen are listed in Table 1.

### PCR-based expression analysis

RP gene expression patterns were also analyzed by PCR using the cDNA panels of 16 different human tissues [Human multiple tissue cDNA (MTC) Panel I and II, CLONTECH] and 8 tumor cell lines (Human Tumor MTC Panel, CLONTECH). PCR was performed in a 10 µl reaction volume containing ~20 pg of template cDNA, 5 pmol each of the forward and reverse primers (listed in Table 1), 0.1 mM dNTPs and 0.35 U of the Expand Long polymerase (Boeringer Mannheim). The thermal cycling conditions included an initial denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 0.5 min, 58°C for 1 min, 70°C for 1 min and final extension at 72°C for 5 min.

## RESULTS

### Retrotransposed copies of X-linked RP genes

We performed BLAST searches on the human draft sequence for functional second copies of the X-linked RP genes *RPL10*, *RPL36A* and *RPL39*. The cDNA and EST databases were also searched to help determine whether these genes are expressed. Although a number of sequences similar to the X-linked genes were found, most of them have nonsense mutations within the open reading frames and are not expressed, suggesting that they are pseudogenes. Nonetheless, we found three genes with coding sequences that are 89–95% identical and amino acid sequences that are 92–99% identical to each of the X-linked RP genes. All of these genes were also found in the EST

**Table 2.** Comparison of retrotransposed RP genes to their X-linked progenitors

| Retrotransposed RP genes | Number of identical nucleotides/total | Number of identical amino acids/total |
|---|---|---|
| *RPL10L* | 576/645 (89%) | 205/214 (96%) |
| *RPL36AL* | 289/321 (90%) | 105/106 (99%) |
| *RPL39L* | 148/156 (95%) | 47/51 (92%) |

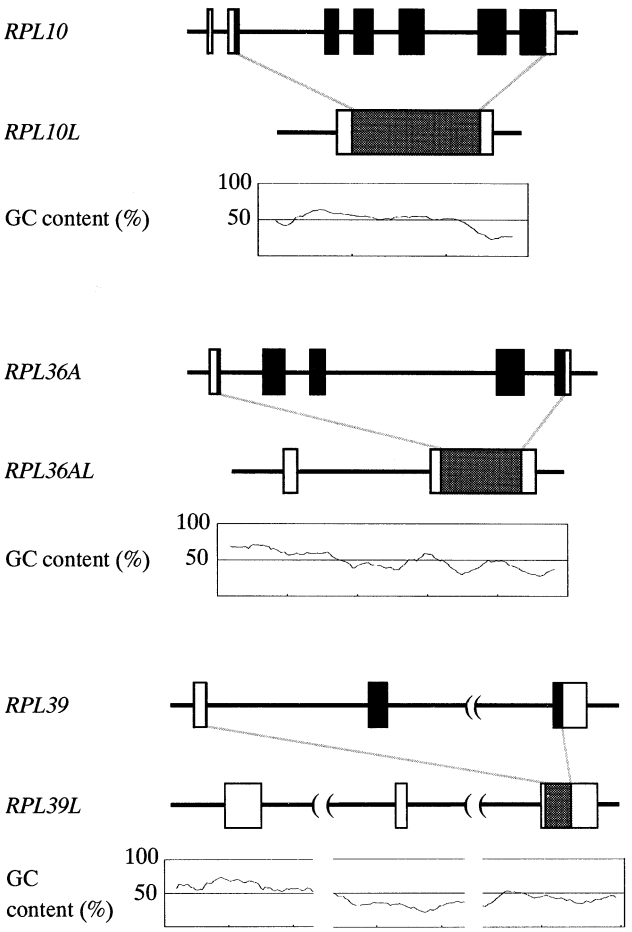


**Figure 1.** Structure of autosomal RP genes. Intron/exon structures were compared with those of the original X-linked genes. The GC content was calculated by GENETYX as an average of 200 bp windows in every 20 bp step. Boxes indicate exons and horizontal lines indicate introns or flanking regions. Coding regions are blackened in the X-linked genes and are shaded in the autosomal genes.

database, suggesting that they are functional. We therefore designated these genes *RPL10L*, *RPL36AL* and *RPL39L*, corresponding to their X chromosome homologs (Table 2).

We then predicted the genomic structure of each new RP gene by comparing its cDNA sequence with the draft genome sequence (Fig. 1). Interestingly, none of these genes have introns in their coding regions, while their X chromosome homologs do have introns. This suggests that these genes are derived from their X-linked homologs by retrotransposition. In contrast, *RPL36AL* and *RPL39L* contain at least one intron in their 5′-non-coding regions (Fig. 1). Because the 5′-non-coding regions are dissimilar both in size and sequence to those of the X-linked genes, they may have evolved to express the transposed genes by acquisition of promoters. We also determined the GC contents and CpG ratios (21) along the entire regions of the genes (Fig. 1). The GC contents in the region extending from the predicted transcription start site up to –300 bp are 58% (*RPL10L*), 68% (*RPL36AL*) and 62% (*RPL39L*), while the CpG ratios are 0.6, 0.9 and 0.8, respectively. Because increased GC contents and CpG ratios are commonly seen in promoter regions, the predicted transcription start sites seem to be reasonable (22).

In the case of *RPS4X* and *RPS4Y*, another X-linked RP gene and its homolog, the intron/exon structures are completely identical between the two genes (23). These genes may have branched out from a single ancestral autosomal gene at the emergence of the sex chromosomes during evolution, with *RPS4Y* surviving subsequent Y chromosome evolution (24).

## Autosomal localization of retrotransposed RP genes

We have localized *RPL10L*, *RPL36AL* and *RPL39L* to autosomes by typing two different radiation hybrid mapping panels, GeneBridge 4 and Stanford G3, using STSs specific to these genes. The data vectors on GeneBridge 4 indicated that *RPL10L* is located at 6 centiRay (cR) from CHLC.GATA5C11, *RPL36AL* at 11 cR from D14S269, and *RPL39L* at 10 cR from D3S1571. Similarly, the results using Stanford G3 indicated that *RPL10L* is located at 50 cR from SHGC-1399, *RPL36AL* at 7 cR from SHGC-20858, and *RPL39L* at 17 cR from SHGC-1745. The results obtained from these two panels are in agreement. *RPL10L* was thereby assigned to chromosome 14q21.2–21.3 between markers D14S288 and D14S269, *RPL36AL* was assigned to 14q21.3 between D14S269 and D14S66, and *RPL39L* was assigned to 3q27.3 between D3S1262 and D3S1580 according to the Ensemble database (25) and a cytogenetic BAC-STS map (26) of the human genome (Fig. 2).

## Expression profiles of autosomal RP genes

To investigate the mRNA expression profiles of the X-linked RP genes and their autosomal homologs in various tissues, we performed northern blot analysis using oligomers designed from the 3′-non-coding regions of the genes as probes (Fig. 3). We found that *RPL10*, *RPL36A*, *RPL36AL* and *RPL39* are expressed ubiquitously, whereas *RPL10L* and *RPL39L* are expressed only in testis. To confirm these results, we performed PCR assays on cDNAs from multiple normal tissues using the STSs listed in Table 1. The expression patterns of *RPS4X* and *RPS4Y* were also examined. The mRNA expression patterns in normal tissues correlate with those shown by northern blot analysis (Fig. 4, lanes 1–16). Because the search of EST databases suggested that *RPL39L* is expressed in some carcinomas, we further examined whether these genes are expressed in cancer cells using cDNAs

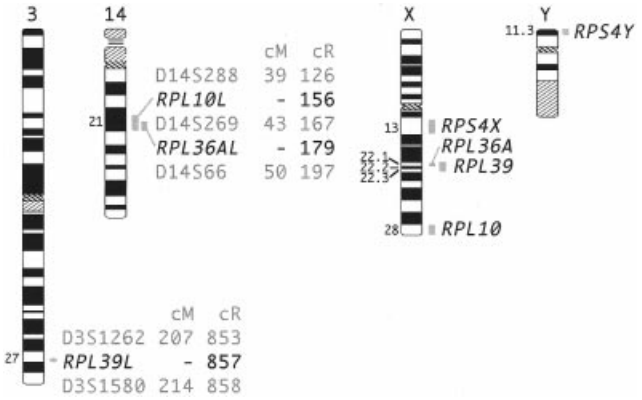


**Figure 2.** Chromosomal mapping of autosomal RP genes. *RPL39L* was mapped to chromosome 3, and *RPL10L* and *RPL36AL* were mapped to chromosome 14 using radiation hybrid panels. They are shown in relation to nearby markers and the approximate distance (in centiMorgans and centiRays) from the most distal short-arm marker on the STS content map (http://www-genome.wi.mit.edu/cgi-bin/contig/phys-map). Locations of the X-linked RP genes and *RPS4Y* are also shown.

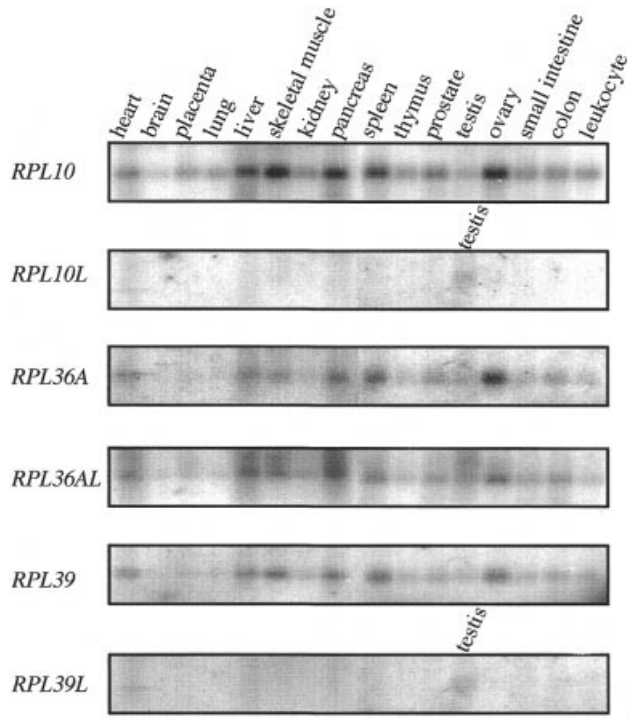


**Figure 3.** Northern blot analysis of X-linked and retrotransposed genes. Two types of human MTN blots were hybridized with $^{32}$P-labeled probes synthesized from the 3′-non-coding region of the genes.

prepared from eight different cell lines. The transcripts from the four X-linked RP genes and *RPL36AL* were detected in all tested cell lines, the *RPL39L* transcript was detected in all but two cell lines, and the *RPS4Y* and *RPL10L* transcripts were not detected in any of the lines (Fig. 4, lanes 17–24).

### Characteristics of the 5′-non-coding regions

We compared the 5′-non-coding regions of the autosomal RP genes with those of their X-chromosome homologs and found that while the sequences extending from the ATG up to the –10 to –30 bp position are similar, the regions further upstream are completely different. This suggests that while the regions immediately adjacent to the start codon probably came from the X-linked RP genes during retrotransposition, the upstream regions seem to have arisen after the retroposition events occurred. In contrast, we found some sequence similarities in the upstream regions of the autosomal genes (Fig. 5). Although the function of these sequences is unknown, they might play an important role in the autosomal gene expression. Interestingly, we also found some sequences that are similar to those of the human endogenous retroviral (HERV) long terminal repeats (LTRs), though the similarities are very limited. It is known that some HERV genes are expressed in human cell lines or tissues (27), and that their expression is primarily controlled by the LTRs, which harbor multiple sequences that are recognized by cellular transcriptional machinery. Figure 5 shows an alignment of the regions upstream of the predicted 5′ ends of retrotransposed genes and the U3 region of HERV-K LTR.

## DISCUSSION

We have identified functional homologs of human X-linked RP genes and have localized them to autosomes. Because these genes have no introns in their coding regions, they were most likely produced by retrotransposition of the original X-linked genes during evolution. Although each mammalian RP is typically encoded by a single gene, this functional gene also generates a large number of retroposons. However, the majority of these retroposons would not be expected to survive during evolution because, without promoters, they are inactive at integration sites and would therefore accumulate mutations in their open reading frames. In fact, there are at least a dozen processed pseudogenes for each RP gene in the genome (28–31). In this study we have identified three genes that appear to have been protected from such evolutionary pressure and to be actively transcribed. Similar retrotransposed genes have been reported, including *PGK2*, *PDHA2*, *X5L* and *HNRNP G-T* (16–18,32). All of these genes have a progenitor on the X chromosome and are highly expressed in testis. We also observed testis-specific expression of *RPL10L* and *RPL39L*, which may indicate a role of the retrotransposed genes in compensating for the inactivated X-linked genes during spermatogenesis.

Based on their findings, which included an analysis of 49 intronless paralogs of autosomal RP genes, Venter *et al*. (33) suggested that there was no bias toward the X chromosome origination of active retroposons during evolution. Because we have identified only three active RP retroposons despite a thorough search of the public DNA databases and have found that all of these genes originated from the X chromosome, we believe that there may actually have been a strong bias for retrotransposition of X chromosome RP genes.

Haploinsufficiency of any of the RP genes causes viable but abnormal phenotypes (*Minute*) in *Drosophila* (11), and heterozygous mutations in *RPS19* are associated with DBA (13,14), suggesting that two copies of each RP gene are essential for normal growth and development. It therefore seems that genes on the sex chromosomes must increase their expression levels to compensate for gene dosage. One of these

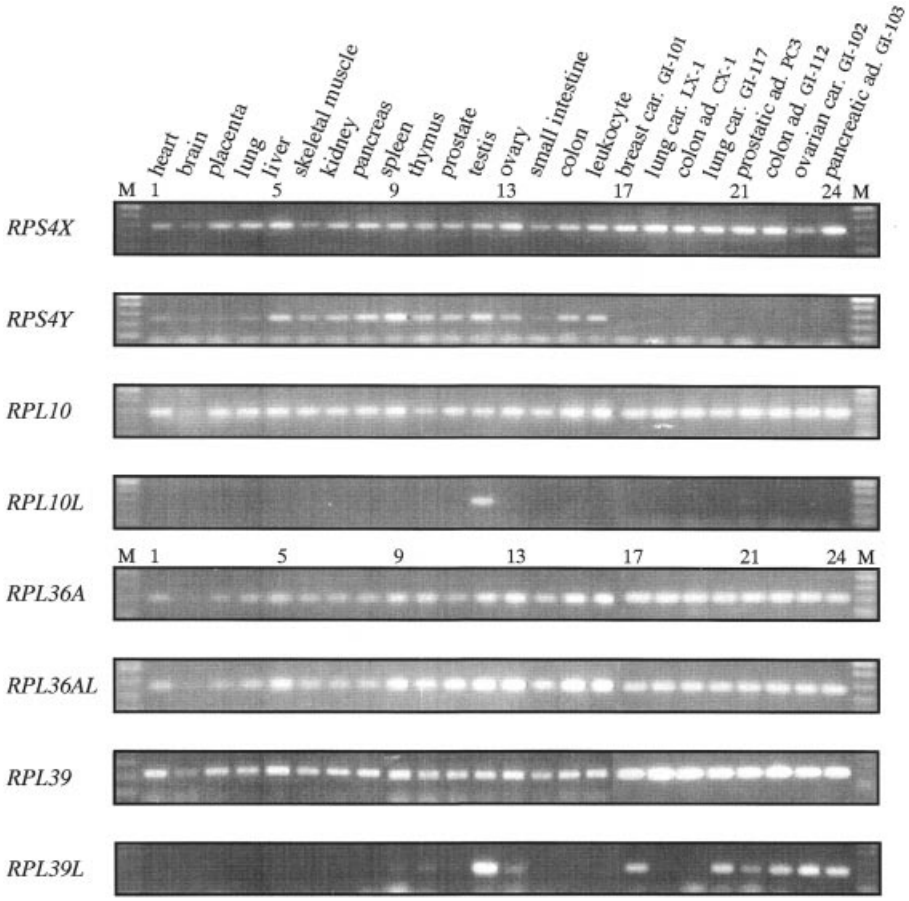


**Figure 4.** Expression analysis of X-linked and retrotransposed genes by PCR. The human MTC panels, including 16 normal tissues (lanes 1–16) and 8 tumor cell lines (lanes 17–24) were typed by PCR using the primers listed in Table 1. Lane M, 100 bp ladder size marker; car, carcinoma; ad, adenocarcinoma.

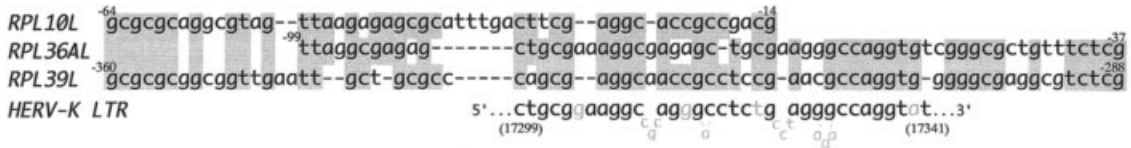


**Figure 5.** Alignment of the 5′-non-coding regions of retrotransposed genes. Identical bases between the genes are shaded. Gaps (hyphens) are inserted to optimize alignment. The numbers appearing above the RP gene sequences represent the nucleotide position from the translation initiator ATG codon. A sequence of 43 bp from the U3 region of HERV-K LTR is also aligned. Positions according to the complete HERV-K sequence (accession no. AF074086) are shown in parentheses.

genes, *RPS4X*, reportedly escapes X-inactivation, and both it and its Y homolog (*RPS4Y*) are ubiquitously expressed (15,23). In contrast, we have shown in this study that the remaining X-linked RP genes have autosomal copies, which are expressed either ubiquitously or only in testis. Recently, we have reported a unique feature of RP gene promoters in which transcription always starts within a characteristic oligopyrimidine tract (34). We found this oligopyrimidine tract in the 5′-upstream region of *RPL36AL*, which is expressed ubiquitously, but not in those of *RPL10L* and *RPL39L*, which are expressed only in testis (data not shown). This oligopyrimidine tract may therefore play an important role in ubiquitous RP gene expression.

Several mechanisms have been proposed for retroposons to acquire a functional promoter (35–37). For example, (i) the retroposon may contain the original gene promoter, (ii) the insertion may occur near an existing promoter, and (iii) mutations may create a functional promoter upstream of the retroposon insertion. The three retroposons identified in this study have little similarity to the original X-linked genes in their 5′-non-coding regions but are slightly similar to HERV-K LTR in a small 70 bp region. A large number of solitary HERV LTRs have been created by homologous recombination, and their possible role in promoter control of downstream genes has been proposed (38,39). We also found such similarity in previously reported X-originated retroposons, including *PGK2*, *PDHA2*, *X5L* and *HNRNP G-T* (data not shown). Also, Chen *et al.* found the similarity in the 5′-non-coding region of mouse zinc-finger protein *Zfp352* gene, which is thought to have arisen from retrotransposition (40).

HERV-K LTR, therefore, might be involved in the active transcription of these retrotransposed genes.

Finally, we searched the DNA databases for mouse RP orthologs, and found three genes, designated *Rpl10l*, *Rpl36al* and *Rpl39l*, which correspond to human *RPL10L*, *RPL36AL* and *RPL39L*, respectively (data not shown). They are 84–92% similar to the human genes in their nucleotide sequences and are located in the syntenic regions between the two species. Moreover, the EST database search has shown that *Rpl36al* is expressed ubiquitously, whereas *Rpl10l* and *Rpl39l* are expressed primarily in testis, which is consistent with the human gene expression patterns. Further experiments using the mouse system will be of great interest to understand the function of these retroposons.

## ACKNOWLEDGEMENTS

This work was supported in part by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Fund for 'Research for the Future' Program from the Japan Society for the Promotion of Science.

## REFERENCES

1. Wool,I.G. (1979) The structure and function of eukaryotic ribosomes. *Annu. Rev. Biochem.*, **48**, 719–754.
2. Wool,I.G., Chan,Y.L. and Glück,A. (1996) Mammalian ribosomes: the structure and the evolution of the proteins. In Hershey,J.W.B., Mathews,M.B. and Sonenberg,N. (eds), *Translational Control.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 685–732.
3. Henderson,A.S., Warburton,D. and Atwood,K.C. (1973) Ribosomal DNA connectives between human acrocentric chromosomes. *Nature*, **245**, 95–97.
4. Worton,R.G., Sutherland,J., Sylvester,J.E., Willard,H.F., Bodrug,S., Dube,I., Duff,C., Kean,V., Ray,P.N. and Schmickel,R.D. (1988) Human ribosomal RNA genes: orientation of the tandem array and conservation of the 5′ end. *Science*, **239**, 64–68.
5. Sorensen,P.D., Lomholt,B., Frederiksen,S. and Tommerup,N. (1991) Fine mapping of human 5S rRNA genes to chromosome 1q42.11–q42.13. *Cytogenet. Cell Genet.*, **57**, 26–29.
6. Lomholt,B., Frederiksen,S., Nielsen,J.N. and Hallenberg,C. (1995) Additional assignment of the human 5S rRNA genes to chromosome region 1q31. *Cytogenet. Cell Genet.*, **70**, 76–79.
7. Kenmochi,N., Kawaguchi,T., Rozen,S., Davis,E., Goodman,N., Hudson,T.J., Tanaka,T. and Page,D.C. (1998) A map of 75 human ribosomal protein genes. *Genome Res.*, **8**, 509–523.
8. Uechi,T., Tanaka,T. and Kenmochi,N. (2001) A complete map of the human ribosomal protein genes: assignment of 80 genes to the cytogenetic map and implications for human disorders. *Genomics*, **72**, 223–230.
9. Meyuhas,O., Avni,D. and Shama,S. (1996) Translational control of ribosomal protein mRNAs in eukaryotes. In Hershey,J.W.B., Mathews,M.B. and Sonenberg,N. (eds), *Translational Control.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 363–388.
10. Kongsuwan,K., Yu,Q., Vincent,A., Frisardi,M.C., Rosbash,M., Lengyel,J.A. and Merriam,J. (1985) A *Drosophila Minute* gene encodes a ribosomal protein. *Nature*, **317**, 555–558.
11. Lambertsson,A. (1998) The *Minute* genes in *Drosophila* and their molecular functions. *Adv. Genet.*, **38**, 69–134.
12. Kay,M.A. and Jacobs-Lorena,M. (1987) Developmental genetics of ribosome synthesis in *Drosophila*. *Trends Genet.*, **3**, 347–351.
13. Draptchinskaia,N., Gustavsson,P., Andersson,B., Pettersson,M., Willig,T.N., Dianzani,I., Ball,S., Tchernia,G., Klar,J., Matsson,H. *et al.* (1999) The gene encoding ribosomal protein S19 is mutated in Diamond-Blackfan anaemia. *Nature Genet.*, **21**, 169–175.
14. Willig,T.N., Draptchinskaia,N., Dianzani,I., Ball,S., Niemeyer,C., Ramenghi,U., Orfali,K., Gustavsson,P., Garelli,E., Brusco,A. *et al.* (1999) Mutations in ribosomal protein S19 gene and Diamond Blackfan anemia: wide variations in phenotypic expression. *Blood*, **94**, 4294–4306.
15. Fisher,E.M., Beer-Romero,P., Brown,L.G., Ridley,A., McNeil,J.A., Lawrence,J.B., Willard,H.F., Bieber,F.R. and Page,D.C. (1990) Homologous ribosomal protein genes on the human X and Y chromosomes: escape from X inactivation and possible implications for Turner syndrome. *Cell*, **63**, 1205–1218.
16. McCarrey,J.R. and Thomas,K. (1987) Human testis-specific PGK gene lacks introns and possesses characteristics of a processed gene. *Nature*, **326**, 501–505.
17. Dahl,H.H., Brown,R.M., Hutchison,W.M., Maragos,C. and Brown,G.K. (1990) A testis-specific form of the human pyruvate dehydrogenase E1 alpha subunit is coded for by an intronless gene on chromosome 4. *Genomics*, **8**, 225–232.
18. Sedlacek,Z., Munstermann,E., Dhorne-Pollet,S., Otto,C., Bock,D., Schutz,G. and Poustka,A. (1999) Human and mouse *XAP-5* and *XAP-5-like* (*X5L*) genes: identification of an ancient functional retroposon differentially expressed in testis. *Genomics*, **61**, 125–132.
19. Hendriksen,P.J., Hoogerbrugge,J.W., Baarends,W.M., de Boer,P., Vreeburg,J.T., Vos,E.A., van der Lende,T. and Grootegoed,J.A. (1997) Testis-specific expression of a functional retroposon encoding glucose-6-phosphate dehydrogenase in the mouse. *Genomics*, **41**, 350–359.
20. Hart,P.E., Glantz,J.N., Orth,J.D., Poynter,G.M. and Salisbury,J.L. (1999) Testis-specific murine centrin, *Cetn1*: genomic characterization and evidence for retroposition of a gene encoding a centrosome protein. *Genomics*, **60**, 111–120.
21. Gardiner-Garden,M. and Frommer,M. (1987) CpG islands in vertebrate genomes. *J. Mol. Biol.*, **196**, 261–282.
22. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*, **409**, 860–921.
23. Zinn,A.R., Alagappan,R.K., Brown,L.G., Wool,I. and Page,D.C. (1994) Structure and function of ribosomal protein S4 genes on the human and mouse sex chromosomes. *Mol. Cell. Biol.*, **14**, 2485–2492.
24. Lahn,B.T. and Page,D.C. (1999) Retroposition of autosomal mRNA yielded testis-specific gene family on human Y chromosome. *Nature Genet.*, **21**, 429–433.
25. Hubbard,T., Barker,D., Birney,E., Cameron,G., Chen,Y., Clark,L., Cox,T., Cuff,J., Curwen,V., Down,T. *et al.* (2002) The Ensembl genome database project. *Nucleic Acids Res.*, **30**, 38–41.
26. Korenberg,J.R., Chen,X.N., Sun,Z., Shi,Z.Y., Ma,S., Vataru,E., Yimlamai,D., Weissenbach,J.S., Shizuya,H., Simon,M.I. *et al.* (1999) Human genome anatomy: BACs integrating the genetic and cytogenetic maps for bridging genome and biomedicine. *Genome Res.*, **9**, 994–1001.
27. Lower,R., Lower,J. and Kurth,R. (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc. Natl Acad. Sci. USA*, **93**, 5177–5184.
28. Monk,R.J., Meyuhas,O. and Perry,R.P. (1981) Mammals have multiple genes for individual ribosomal proteins. *Cell*, **24**, 301–306.
29. D'Eustachio,P., Meyuhas,O., Ruddle,F. and Perry,R.P. (1981) Chromosomal distribution of ribosomal protein genes in the mouse. *Cell*, **24**, 307–312.
30. Dudov,K.P. and Perry,R.P. (1984) The gene family encoding the mouse ribosomal protein L32 contains a uniquely expressed intron-containing gene and an unmutated processed gene. *Cell*, **37**, 457–468.
31. Kuzumaki,T., Tanaka,T., Ishikawa,K. and Ogata,K. (1987) Rat ribosomal protein L35a multigene family: molecular structure and characterization of three L35a-related pseudogenes. *Biochim. Biophys. Acta*, **909**, 99–106.
32. Elliott,D.J., Venables,J.P., Newton,C.S., Lawson,D., Boyle,S., Eperon,I.C. and Cooke,H.J. (2000) An evolutionarily conserved germ cell-specific hnRNP is encoded by a retrotransposed gene. *Hum. Mol. Genet.*, **9**, 2117–2124.
33. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
34. Yoshihama,M., Uechi,T., Asakawa,S., Kawasaki,K., Kato,S., Higa,S., Maeda,N., Minoshima,S., Tanaka,T., Shimizu,N. and Kenmochi,N. (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.*, **12**, 379–390.
35. Moran,J.V., DeBerardinis,R.J. and Kazazian,H.H.,Jr (1999) Exon shuffling by L1 retrotransposition. *Science*, **283**, 1530–1534.
36. Kleene,K.C. and Mastrangelo,M.A. (1999) The promoter of the Poly(A) binding protein 2 (*Pabp2*) retroposon is derived from the 5′-untranslated region of the Pabp1 progenitor gene. *Genomics*, **61**, 194–200.

37. Boschan,C., Borchert,A., Ufer,C., Thiele,B.J. and Kuhn,H. (2002) Discovery of a functional retrotransposon of the murine phospholipid hydroperoxide glutathione peroxidase: chromosomal localization and tissue-specific expression pattern. *Genomics*, **79**, 387–394.
38. Domansky,A.N., Kopantzev,E.P., Snezhkov,E.V., Lebedev,Y.B., Leib-Mosch,C. and Sverdlov,E.D. (2000) Solitary HERV-K LTRs possess bi-directional promoter activity and contain a negative regulatory element in the U5 region. *FEBS Lett.*, **472**, 191–195.
39. Vinogradova,T.V., Leppik,L.P., Nikolaev,L.G., Akopov,S.B., Kleiman,A.M., Senyuta,N.B. and Sverdlov,E.D. (2001) Solitary human endogenous retroviruses-K LTRs retain transcriptional activity *in vivo*, the mode of which is different in different cell types. *Virology*, **290**, 83–90.
40. Chen,H.H., Liu,T.Y., Huang,C.J. and Choo,K.B. (2002) Generation of two homologous and intronless zinc-finger protein genes, *Zfp352* and *Zfp353*, with different expression patterns by retrotransposition. *Genomics*, **79**, 18–23.