J Korean Med Sci 2004; 19: 783-92
ISSN 1011-8934

Alu and *L1* Retroelements Are Correlated with the Tissue Extent and Peak Rate of Gene Expression, Respectively

We exploited the serial analysis of gene expression (SAGE) libraries and human genome database in silico to correlate the breadth of expression (BOE; housekeeping versus tissue-specific genes) and peak rate of expression (PRE; high versus low expressed genes) with the density distribution of the retroelements. The BOE status is linearly associated with the density of the sense *Alu*s along the 100 kb nucleotides region upstream of a gene, whereas the PRE status is inversely correlated with the density of antisense *L1s* within a gene and in the up- and downstream regions of the 0-10 kb nucleotides. The radial distance of intranuclear position, which is known to serve as the global domain for transcription regulation, is reciprocally correlated with the fractions of *Alu* (toward the nuclear center) and *L1* (toward the nuclear edge) elements in each chromosome. We propose that the BOE and PRE statuses are related to the reciprocal distribution of *Alu* and *L1* elements that formulate local and global expression domains.

Key Words : *Alu Elements; Long Interspersed Nucleotide Elements; Gene Expression*

Tae-Min Kim, Yu-Chae Jung, Mun-Gan Rhyu

Department of Microbiology, College of Medicine, The Catholic University of Korea, Seoul, Korea

Received : 18 March 2004
Accepted : 2 August 2004

Address for correspondence
Mun-Gan Rhyu, M.D.
Department of Microbiology, College of Medicine,
The Catholic University of Korea, 505 Banpo-dong,
Socho-gu, Seoul 137-701, Korea
Tel : +82.2-590-1215, Fax : +82.2-596-8969
E-mail : rhyumung@catholic.ac.kr

INTRODUCTION

Almost half of the human genome is composed of highly repetitive sequences derived from retroelements including *Alu*, *L1*, LTR, *L2* and *MIR* (1). Retroelements were once thought to propagate within the genome independent of host fitness and were dismissed as 'selfish genomic parasites' or mere 'junk DNA' (2, 3). Individual retroelement copies are so heterogeneous in their sequence and size that their implication in gene expression was considered to be insignificant. Meanwhile, the progress of human genome studies has shed light on the selective redistribution of retroelements carried out for an organism's benefit. On an evolutionary scale, the distributions of retroelements are biased toward the gene-rich (*Alu*) or gene-poor (*L1*) subchromosomal regions (1, 4). It is possible that some retroelements are more enriched in the gene-poor region in order for them to have a lesser detrimental influence, or in the gene-rich region for the purpose of giving them an adaptive advantage (5, 6). This reciprocal redistribution of *Alu* and *L1* elements may result in and/or originate from the beneficial and stable relationship between host gene expression and selfish retroelement fixation, even in the case of symbiotic co-evolution (7). Therefore, we considered that during human evolution, gene expression and retroelement fixation might have influenced each other in the superimposed genetic levels.

The genome-wide dataset of serial analysis of gene expression (SAGE) libraries has provided useful information for gene expression profiles represented by the tissue extent breadth of expression (BOE) and peak rate (PRE; peak rate of expression) parameters (8). The analysis of expression profiles displays a clustering of housekeeping genes in the subchromosomal regions (9). On the other hand, the intranuclear position (toward the nuclear center as opposed to the edge) of chromosomes within interphase nuclei (so-called chromosome territory, CT) has been proposed as a subnuclear compartment of nuclear proteins for a distinct transcriptional activity (10, 11). Human chromosomes containing genes in the high or low density range tend to be preferentially located at the nuclear edge or center, respectively (12, 13). This superimposed gene organization may be advantageous when it comes to concentrating nuclear proteins involved in common pathways in the same compartments. However, it is not known whether such subchromosomal and subnuclear domains of genes are associated with noncoding retroelements that are nonrandomly dispersed throughout the genome. Because the nonrandom distribution of retroelements may be a cause or consequence of evolutionary interaction between coding genes and noncoding retroelements, it would be useful to know whether the framework of gene expression is related to the retroelement distribution.

In this study, the BOE and PRE statuses were separately correlated with the density distribution of retroelements relative to coding genes, and the radial distance of CT was evaluated for the purpose of establishing the relationship with the retroelement compositions of individual chromosomes.

The BOE and PRE statuses are distinctly associated with the densities of sense *Alu* elements in the long extragenic region and of the antisense *L1* elements within the genic and adjacent regions, respectively. There are linear correlations between the order of CT position and the intrachromosomal fraction of *Alu* (toward the nuclear edge) and *L1* (toward the nuclear center) elements. The local density differences of sense *Alu* and antisense *L1* elements between different expression levels are further distinguished according to the intrachromosomal *Alu* and *L1* fractions. We propose that a genome-wide expression framework methodologically links the BOE status to the *Alu* elements and the PRE status to the *L1* elements.

MATERIAL AND METHODS

Collection of data in silico

Twenty-eight SAGE libraries representing the expression profiles of 14 normal tissues (9) were obtained from a public database (http://www.ncbi.nlm.nih.gov/SAGE/) with *NlaIII* SAGE tags. A reliable SAGE map was employed as a matching function, in order to combine the Unigene map and the SAGE tags (22). Individual Unigenes were scored for two expression parameters indicating the number of expressed tissues (BOE) and the maximal peak count of the tags (cpm; counts per million) of expression among the observed tissues (PRE). The 15,471 RefGenes in the golden path assembly Apr. 2003 were matched to BOE and PRE data, and 6,776 RefGenes were found to be expressed in at least one tissue. Of the 6,776 expressed genes, 1,739 genes were found to make more than one gene-tag combination or alternative expressions. These genes were excluded from this study, due to the consequent difficulty in defining their start and termination site of transcription and their expression status. The remaining 5,037 RefGenes, which were matched to a single expression profile, were used as a reliable database for the physical map of gene expression.

Retroelement data were obtained from the human genome database (http://genome.ucsc.edu) as that used for the physical location of the RefGenes using the RepeatMasker program (http://ftp.genome.washington.edu/RM/RepeatMasker.html). The retroelements were classified into five major retroelement families (*Alu*, *L1*, *MIR*, *L2* and LTR), each of which was subdivided into sense and antisense directions in relation to the genes, because retroelements are known to be differentially fixed throughout the genome according to their orientation (18). The 100 kb nucleotides regions upstream of the transcription start site and downstream of the polyadenylation site were fractionated into 10 kb bins and analyzed for the purpose of determining the extragenic density distribution of each retroelement family. Coding regions of variable sizes were considered as a single segment, in which the retroelement densities were scored as the copy number per 10 kb segment.

For analysis of Chromosome territory, we adapted the intranuclear positions of individual chromosomes, which were previously reported (12). The relative radial distance of each chromosome was scored as the distance between the center of the CT and the center of the nucleus, which was measured in male lymphoblast nuclei, as visualized by fluorescence in situ hybridization.

Classification of gene expression

The 5,037 genes reliably matched to expression profiles revealed a linear correlation between the BOE and PRE statuses (r=0.44, p<0.001) in similarity with the total of 6,776 expressed genes (Fig. 1). The BOE and PRE statuses were dichotomized into high and low levels with the median BOE (14 tissues) and PRE (645 cpm) values. Because of the expression profile being skewed toward the low level, using the median expression values causes the 5,037 genes to be categorized into four pair-wise groups of dissimilar size, high-BOE and high-PRE (327; 6.5%), high-BOE and low-PRE (206; 4.1%), low-BOE and high-PRE (418; 8.3%), and low-BOE and low-PRE (4,086; 81.1%). These proportions of

	5.037 Genes	6.776 Genes
High BOE-High PRE	327 (6.5%)	408 (6.0%)
High BOE-Low PRE	206 (4.1%)	279 (4.1%)
Low BOE-High PRE	418 (8.3%)	555 (8.2%)
Low BOE-Low PRE	4.086 (81.1%)	5.534 (81.75)

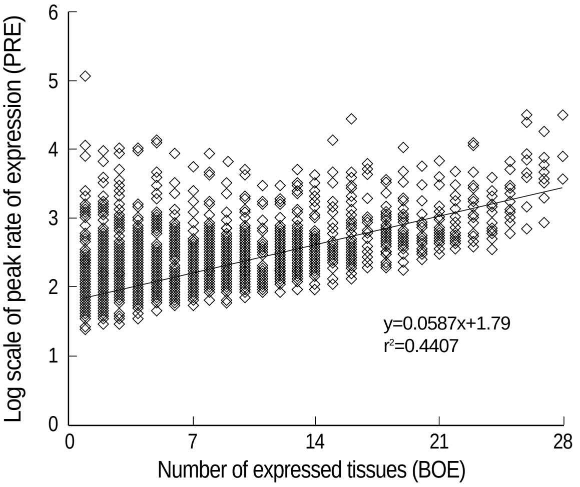


Fig. 1. Distribution of 5,037 gene expressions skewed toward low breadth of expression (BOE) and low peak rate of expression (PRE). The total of 6,776 expressed genes and the 5,037 genes reliably matched to a transcriptional unit are similarly divided into four BOE-PRE groups using the median values of the expression parameters. The table above the graph represents the proportion of 6,776 expressed genes and 5,037 matched genes present in the four BOE-PRE groups. The 5,036 expressed genes are plotted against the BOE and PRE parameters and are divided into four quadrants indicated by dotted lines. The correlation between the BOE and log-transformed PRE is shown by the best-fit line.

BOE-PRE groups were similar to those of the 6,776 expressed genes (Fig. 1). Alternatively, the gene expressions were divided into three expression levels, low, intermediate, and high, in proportions of 25%, 50% and 25%, respectively. The 5,037 genes were divided using the cutoff value of BOE (=1, 2-7 and ≥8) into 1,328 housekeeping genes (26%), 2,366 intermediate tissue-specific genes (48%), and 1,353 high tissue-specific genes (26%), and using the cutoff value of PRE (<55 cpm, 55-250 cpm and >250 cpm) into 1,296 highly expressed genes (26%), 2,464 intermediately expressed genes (49%) and 1,277 lowly expressed genes (25%).

Chromosomal and genomic parameters for gene grouping

The 5,037 genes examined were proportionally divided into three groups (25%, 50%, and 25%) with the following parameters.

Chromosomes were grouped by retroelement composition. The size-fraction (nucleotides percent) and density (copy number per Mb) of the total copies of each retroelement family were calculated for each chromosome. Of the two parameters, the size-fraction revealed strong correlations between the intranuclear position and the *Alu* and *L1* elements. The entire group of autosomes were sorted in the order of retroelement fraction and categorized into *Alu*-poor (3, 4, 5, 13 and 18) and *Alu*-rich (16, 17, 19, 20 and 22), as well as *L1*-rich (2, 3, 4, 5 and 6) and *L1*-poor (16, 17, 19, 20 and 22) chromosome groups. The remaining chromosomes containing 50% of the genes examined were categorized into an intermediate group.

After being sorted in the order of the intranuclear position, the chromosomes were categorized into three groups, edge (3, 4, 7, 13 and 18), center (1, 16, 17, 19 and 22) and middle position (other chromosomes).

The surrounding gene density of a gene was calculated by counting the number of RefGenes residing within the 100 kb upstream and downstream regions of the corresponding gene. Three gene-density groups were established using the cutoff points of 1 and 6 genes per surrounding 200 kb nucleotides.

The GC contents of individual chromosomes were adapted from the same literature used for the radial distance of the CTs (12). The local GC content surrounding each gene was determined based on the GC content (%) in the 20 kb non-overlapping window occupied by the corresponding gene. Three GC-content groups were established using the cutoff points of 40.7% and 49.3% GC contents.

Statistical analysis

Two-tailed Pearson's correlation coefficients were calculated for the purpose of determining the extent of correlation between the regional retroelement density and the gene expression, as well as between the retroelement composition and the intranuclear position of a given chromosome. The unpaired t-test was used to demonstrate whether there were any significant differences in the density of the retroelements between different expression levels.

RESULTS

Relationships between SAGE profiles and the distributions of retroelements

5,037 genes reliably matched to expression profiles were dichotomized by the median BOE and PRE values into high and low level expression groups (see Fig. 1 and Methods). We separately correlated four BOE-PRE pair-wise groups with the density distributions of the *Alu*, *L1*, LTR, *L2* and *MIR* elements dispersed in the upstream and downstream regions of 100 kb nucleotides encompassing coding genes (Fig. 2). The BOE and PRE statuses tend to be related to the densities of sense *Alu*s in the upstream region and of the antisense *L1*s in the intragenic region, respectively. The high (low) level of BOE is associated with the high (low) density of extragenic sense *Alu*s, regardless of the level of PRE. On the other hand, high (low)-PRE genes contain antisense *L1*s in the low (high) density, regardless of the level of BOE. The increase of the sense *Alu* density associated with high-BOE genes is markedly reduced in the downstream region, and the depression of antisense *L1* density associated with high-PRE genes is attenuated in the extragenic region. These relationships between the BOE and PRE statuses and the densities of sense *Alu* and antisense *L1* elements were also observed in dichotomized expression groups of equal size (data not shown).

Retroelements other than sense *Alu*s and antisense *L1*s reveal no density-dependent relationships with gene expression (Fig. 2). The density difference of the antisense *Alu*s between the different BOE statuses is considerably reduced as compared with that of the sense *Alu*s. The intragenic *L1*s in the sense direction are sharply depressed regardless of the BOE and PRE statuses. The intragenic LTRs, in both the sense and antisense directions, tend to be more sharply depressed as compared to the *L1* elements. *L2* and *MIR* retroelements in the lowest density reveal small extragenic peaks and are slightly depressed in the intragenic regions.

Correlations between BOE and PRE statuses and *Alu* and *L1* elements

The densities of the sense *Alu* and antisense *L1* elements in the extragenic 200 kb nucleotides regions and genic portions were evaluated for the purpose of establishing their relationship with the status of BOE and PRE by using Pearson's correlation analysis (Table 1). There are only weak correlations between the BOE and PRE statuses and retroelement densities, owing to the gene expressions being biased toward the low level. However, the correlation coefficients of BOE and PRE are highest for the extragenic sense *Alu*s in the proximal upstream region and for the intragenic antisense *L1*s, respectively.

The density distributions of the sense *Alu*s and antisense *L1*s were plotted according to three levels (high, intermediate and low) of BOE and PRE, and the density differences of the sense *Alu*s and antisense *L1*s between three expression levels were statistically analyzed using the unpaired t test (Fig. 3). The density of the sense *Alu*s increases markedly with increasing level of BOE throughout the 100 kb upstream region. The density of the antisense *L1*s in the genic and 0-10 kb upstream regions decreases sharply as the PRE status inclines toward the high level. The densities of the antisense *L1*s in the 0-10 kb downstream region are significantly different between the intermediate and low levels, but not between the high and intermediate levels. These linear and inverse relationships are compromised in the opposite relations of

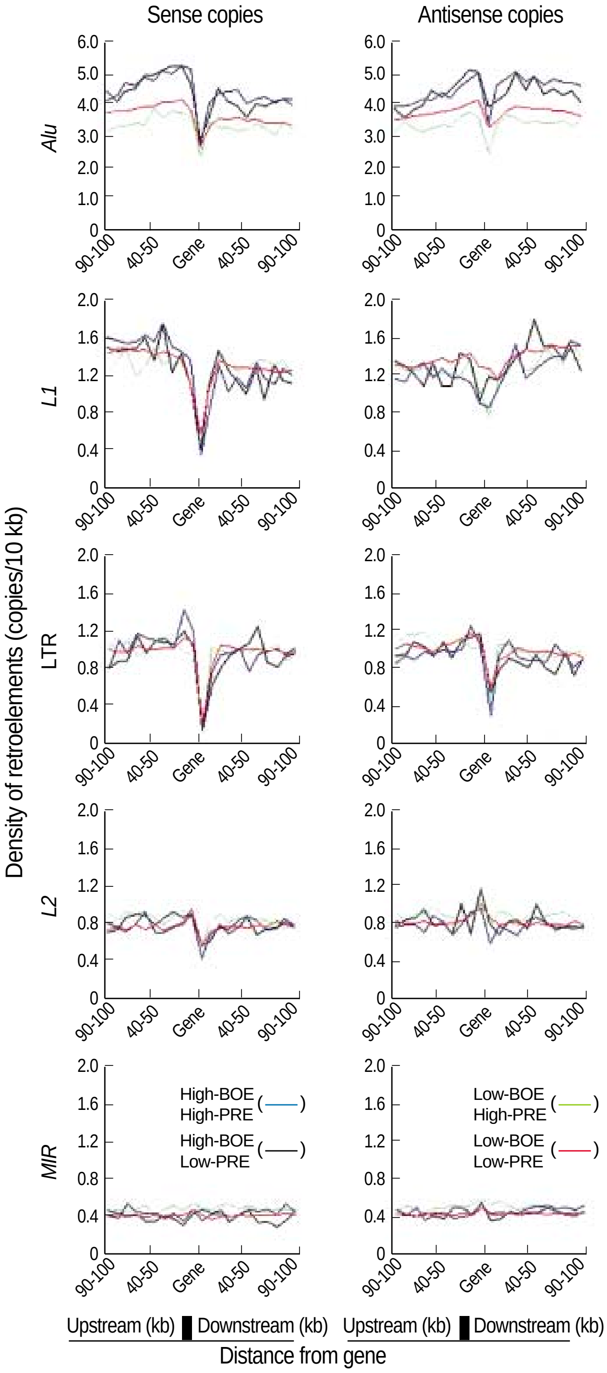


Fig. 2. Relationships between breadth (BOE) and peak rate (PRE) of expression and retroelement families. The 5,037 gene expressions are categorized into four BOE-PRE pair-wise groups, high-BOE and high-PRE (blue), high-BOE and low-PRE (black), low-BOE and high-PRE (green), and low-BOE and low-PRE (red). The densities of the retroelements in the 100 kb nucleotides regions, upstream and downstream of the genes, are fractionated into 10 kb nucleotides bins, and separately plotted for the four BOE-PRE groups. The sense and antisense copies of each retroelement indicate the same and opposite orientation, respectively, in relation to the nearest gene. The intragenic density of the retroelements

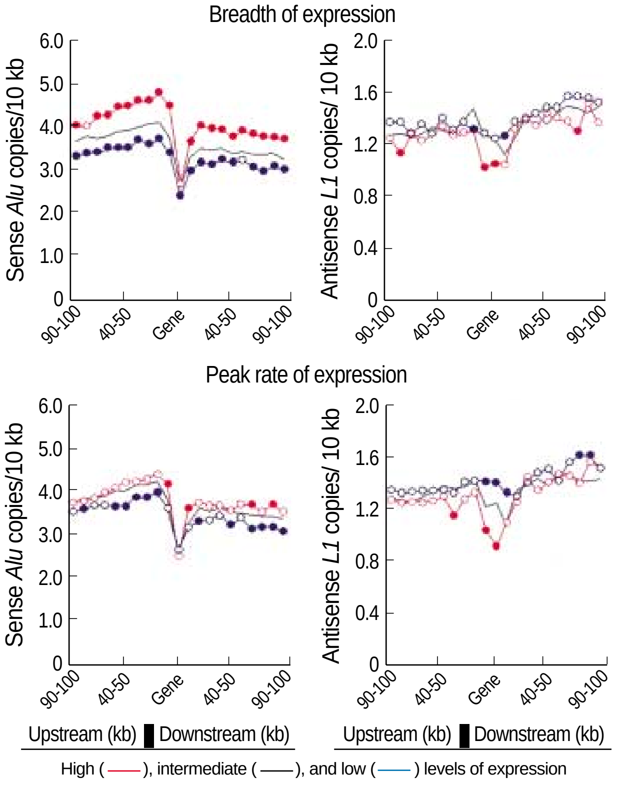


Fig. 3. Regional relationships between the BOE and PRE statuses and density distributions of sense *Alu* and antisense *L1* elements. The BOE and PRE statuses are divided into three levels, high (red), intermediate (black), and low (blue). Sense *Alu* and antisense *L1* elements in the extragenic and intragenic regions are separately plotted according to the levels of BOE and PRE in 10 kb nucleotide bins along 100 kb nucleotides regions upstream and downstream of the coding regions. Statistically significant differences (p<0.05) in the *Alu* and *L1* densities between high and intermediate levels (red) and between intermediate and low levels (blue) of expression are indicated by closed circles along the intragenic and extragenic regions. The cutoff points for the BOE and PRE levels were described in the Methods section. The intragenic density of the retroelements is indicated above the "Gene".

Table 1. Statistical analysis of correlation between retroelement density and gene expression*

	Sense *Alu*				Antisense *L1*			
	Breadth of expression		log (peak rate of expression)		Breadth of expression		log (peak rate of expression)	
	r	*p*	r	*p*	r	*p*	r	*p*
90-100 kb upst[†]	0.0796	2E-08	0.0099	0.4814	-0.0198	0.1607	-0.0137	0.3304
80-90 kb upst	0.0660	3E-06	0.0043	0.7601	-0.0361	0.0103	-0.0208	0.1403
70-80 kb upst	0.0950	1E-11	0.0139	0.3227	0.0002	0.9893	-0.0036	0.8000
60-70 kb upst	0.0827	4E-09	0.0170	0.2274	-0.0193	0.1717	-0.0156	0.2690
50-60 kb upst	0.1079	2E-14	0.0290	0.0397	-0.0122	0.3884	-0.0155	0.2699
40-50 kb upst	0.1092	8E-15	0.0433	0.0021	-0.0218	0.1215	-0.0244	0.0829
30-40 kb upst	0.1086	1E-14	0.0206	0.1432	-0.0181	0.1999	-0.0350	0.0131
20-30 kb upst	0.1067	3E-14	0.0293	0.0374	-0.0168	0.2331	-0.0281	0.0463
10-20 kb upst	0.1245	7E-19	0.0338	0.0163	-0.0223	0.1128	-0.0144	0.3056
0-10 kb upst	0.1386	5E-23	0.0555	0.0001	-0.0657	3E-06	-0.0728	2E-07
Gene[‡]	0.0248	0.0781	-0.0310	0.0280	-0.0731	2E-07	-0.1370	2E-22
0-10 kb dnst[§]	0.0818	6E-09	0.0533	0.0002	-0.0256	0.0695	-0.0291	0.0388
10-20 kb dnst	0.1017	5E-13	0.0354	0.0120	-0.0055	0.6962	-0.0145	0.3047
20-30 kb dnst	0.1088	1E-14	0.0397	0.0049	0.0131	0.3515	0.0047	0.7368
30-40 kb dnst	0.0897	2E-10	0.0202	0.1514	-0.0149	0.2906	-0.0216	0.1249
40-50 kb dnst	0.0736	2E-07	0.0303	0.0315	-0.0017	0.9068	-0.0225	0.1109
50-60 kb dnst	0.0869	7E-10	0.0334	0.0178	-0.0152	0.2818	0.0051	0.7184
60-70 kb dnst	0.0992	2E-12	0.0510	0.0003	-0.0195	0.1665	-0.0120	0.3964
70-80 kb dnst	0.0979	3E-12	0.0395	0.0051	-0.0380	0.0070	-0.0375	0.0078
80-90 kb dnst	0.0882	4E-10	0.0532	0.0002	0.0042	0.7655	0.0011	0.9372
90-100 kb dnst	0.0928	4E-11	0.0537	0.0001	-0.0176	0.2107	-0.0091	0.5172

*Correlation coefficients (r) were analyzed by two-tailed Pearson's correlation. The upstream[†], intragenic[‡], and downstream[§] regions were analyzed separately. Significant values (p<0.05) are shown in shaded boxes.

Table 2. Correlations between intranuclear radial distance and various genetic parameters of the chromosomes*

	Alu	*L1*	LTR	*L2*	*MIR*	Gene density	GC content	Chromosome size
R	0.857[†]	-0.844	-0.590	-0.007	0.287	0.793	0.769	-0.394
r^2	0.735	0.712	0.348	0.000	0.082	0.629	0.591	0.156
p	3.5E-07	8E-07	0.0039	0.9753	0.1954	1.1E-05	2.9E-05	0.0693

The extents of correlation* (r and r^2) between the intranuclear radial distance and the genetic components of individual chromosomes are calculated by two-tailed Pearson's correlation. Significant values (p<0.05) are shown in shaded boxes.

BOE-*L1*s and PRE-*Alu*s.

Global and local domains of *Alu* and *L1* retroelements

When correlating the distances between the centers of the CTs and the interphase nucleus with the fractions of retroelement families in individual chromosomes, the *Alu* and *L1* elements show the highest correlations with the CT positions, with linear and inverse relationships, respectively (Table 2). The correlation coefficients of the *Alu*s (r=0.857) and *L1*s (r=-0.844) are higher than those of the gene density (r=0.793) and GC content (r=0.769). The 5,037 genes examined were proportionally divided into three groups (25%, 50% and 25%) according to the chromosomal (intrachromosomal *Alu* and *L1* fractions and intranuclear position) and genomic (gene density and GC content) parameters (see Methods). The chromosomes were classified into *Alu*-poor (3, 4, 5, 13, and 18), *L1*-rich (2, 3, 4, 5, and 6), and edge-position (3, 4, 7, 13, and 18) groups, despite the high correlation coefficients. On the other hand, the *Alu*-rich and *L1*-poor group includes the same chromosomes, 16, 17, 19, 20, and 22 as the center-position group (1, 16, 17, 19, and 22).

The correlation curves of the gene expression and sense *Alu* (Fig. 4) and antisense *L1* (Fig. 5) elements were compared among the gene groups. The patterns of expression-retroelement relations are similarly preserved in all of the gene groups. However, the sense *Alu* and antisense *L1* density differences between the various expression levels vary among the gene groups. All of the chromosome groups, which were categorized based on the *Alu* and *L1* fractions, result in a significant density difference of the sense *Alu* and antisense *L1* elements between the three expression levels in at least one 10 kb nucleotides bin. The intranuclear position, gene density, and GC content parameters tend to reduce the *Alu* and *L1* den-

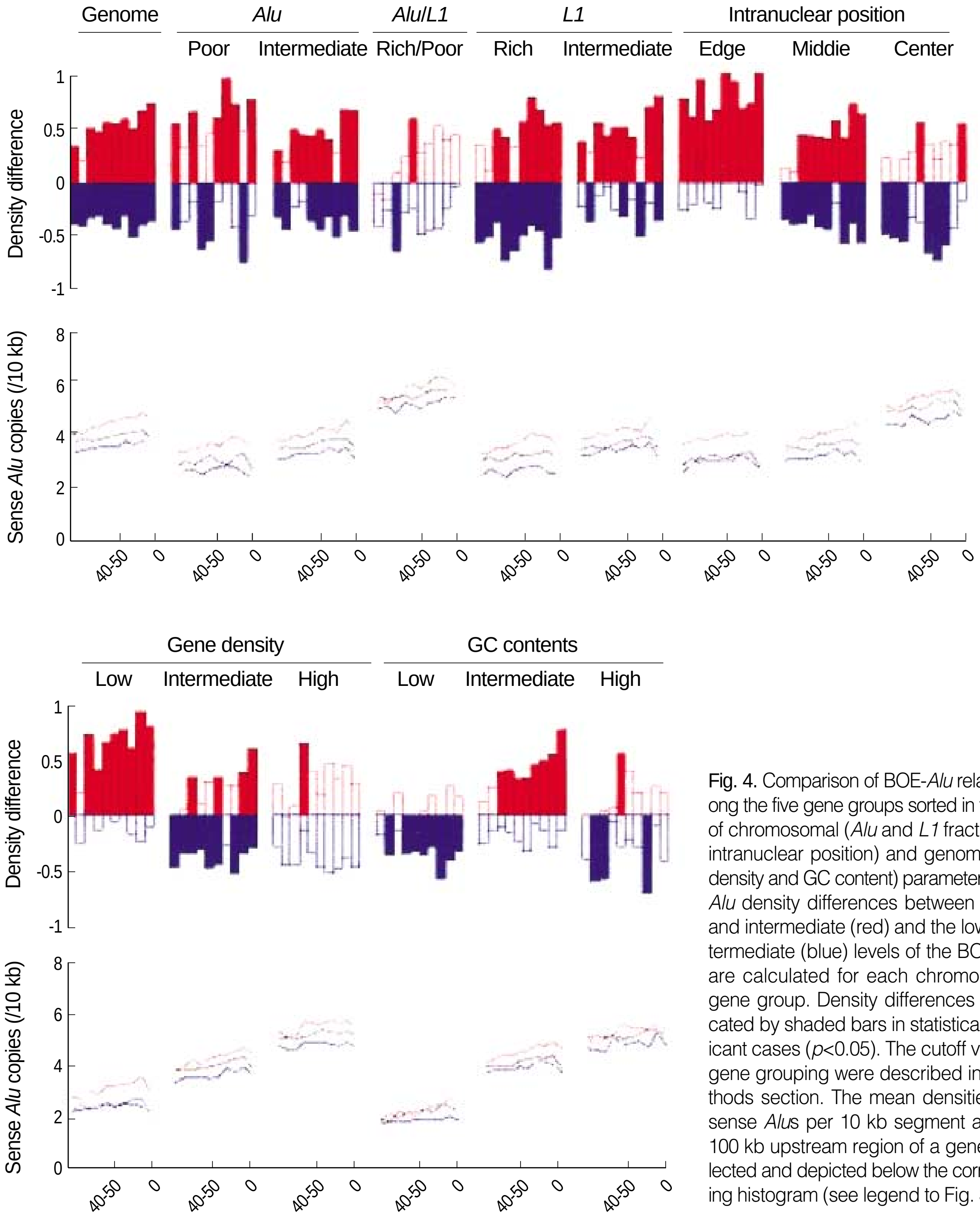
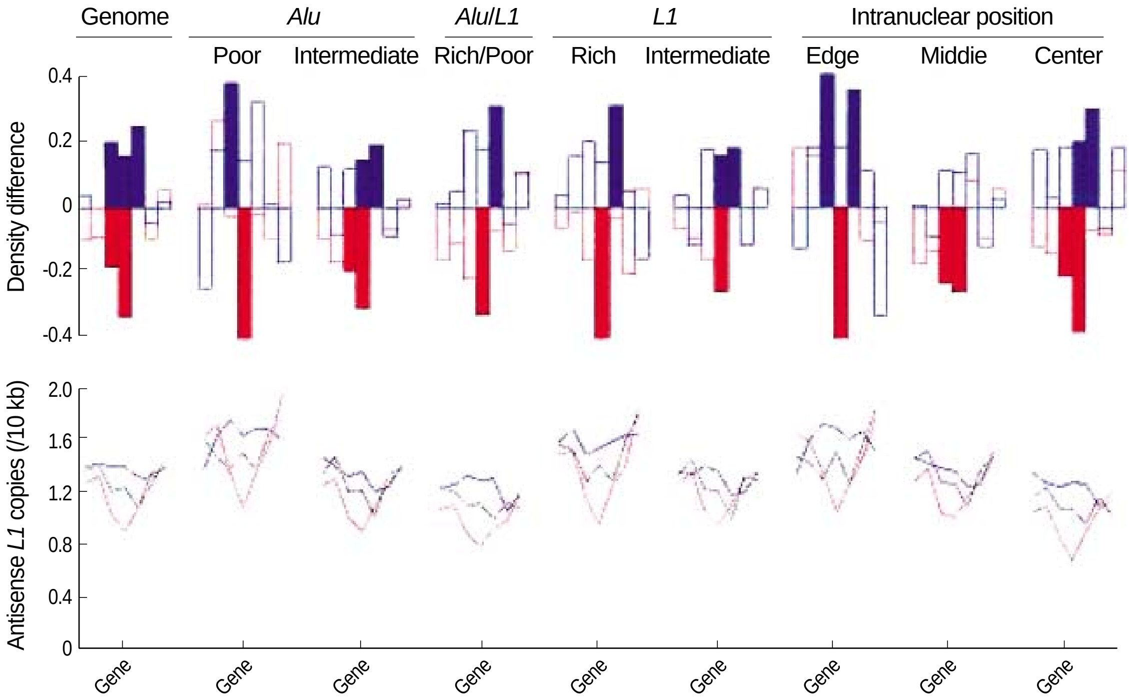


Fig. 4. Comparison of BOE-*Alu* relation among the five gene groups sorted in the order of chromosomal (*Alu* and *L1* fractions and intranuclear position) and genomic (gene density and GC content) parameters. Sense *Alu* density differences between the high and intermediate (red) and the low and intermediate (blue) levels of the BOE status are calculated for each chromosome or gene group. Density differences are indicated by shaded bars in statistically significant cases (p<0.05). The cutoff values for gene grouping were described in the Methods section. The mean densities of the sense *Alu*s per 10 kb segment along the 100 kb upstream region of a gene are selected and depicted below the corresponding histogram (see legend to Fig. 3).

sity differences between the three expression levels, with the result that the BOE-*Alu* (edge-position, low and high gene-density and low GC content groups) and PRE-*L1* (middle-position, low gene density and high GC content groups) relationships are not statistically significant.

Extrapolating from a human gene number of about 30,000, the genome contains on average one coding region per 100 kb segment. The extragenic segment was divided into the 0-50 kb proximal region, which is closely related to the gene, and the 50-100 kb distal region, which is possibly shared by a nearby gene. The sense *Alu* density difference between the BOE levels is more prominent within the 40 kb proximal regions in the *Alu*-poor or *L1*-rich chromosome group, whereas the *Alu*-rich and *L1*-poor group, which includes the same chromosomes, demonstrates a significant density difference of sense *Alu*s in the 50-60 kb (between high and intermediate levels) or 70-80 kb (between intermediate and low levels) distal region (Fig. 4).

The antisense *L1* depression within the high-PRE genes is more prominently observed in the *Alu*-poor or *L1*-rich



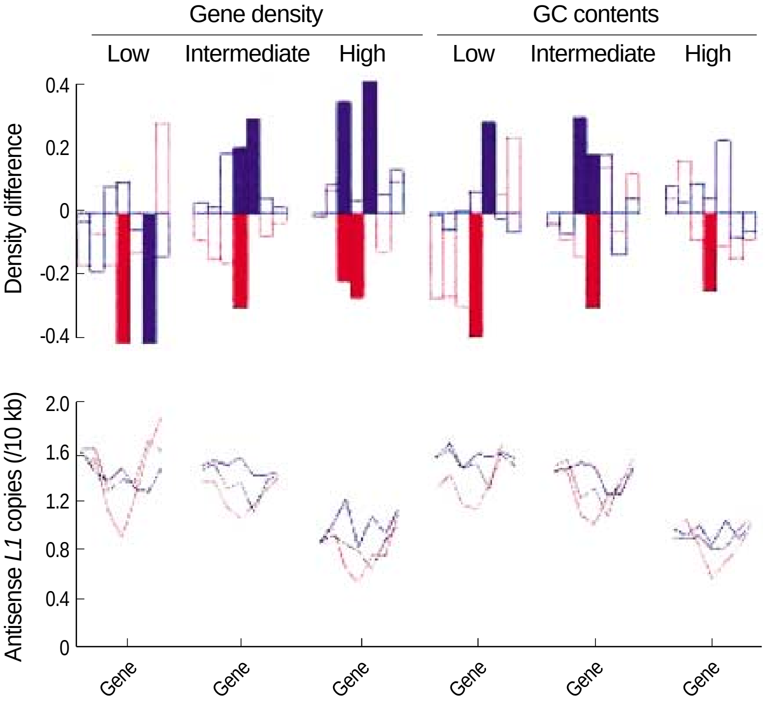


Fig. 5. Comparison of PRE-*L1* relation among the five gene groups sorted in the order of chromosomal (*Alu* and *L1* fractions and intranuclear position) and genomic (gene density and GC content) parameters. Density differences of antisense *L1*s between the high and intermediate (red) and the low and intermediate (blue) levels of the PRE status are calculated for each chromosome or gene group. Density differences are indicated by shaded bars in statistically significant cases (p<0.05). The cutoff values for gene grouping were described in the Methods section. The mean densities of the antisense *L1*s within a gene and the 30 kb upstream and downstream regions are selectively depicted below the corresponding histogram (see legend to Fig. 3).

chromosome group than in the *Alu*-rich and *L1*-poor chromosome group (Fig. 5). The density difference of the antisense *L1*s between the intermediate and low levels of PRE is significant in the 0-10 kb upstream region, only in the *Alu*-poor group, whereas it is significant in the 0-10 kb downstream region in the other (*L1*-rich, *Alu*-rich and *L1*-poor, and intermediate) groups.

DISCUSSION

The focus of this paper is to describe the genome-wide distribution of retroelements, especially that of *Alu* and *L1*, which are closely associated with human gene expression. The relationships between the gene expression and individual retroelement copies cannot be easily explained, because of their divergent sequences and heterogeneous sizes. However, two quantitative variables of gene expression, BOE and PRE, are known to be closely related to the density distributions of the sense *Alu*s and antisense *L1*s, respectively, in separate local domains. In addition, the total *Alu* and *L1* fractions in a chromosome tend to increase in correlation with the intranuclear position toward the nuclear center and edge, respectively. The relationships between the BOE status and sense *Alu*s and between the PRE status and antisense *L1*s are significant in the subnuclear domains, which are categorized based on the *Alu* and *L1* fractions, but not in those which are categorized based on the gene density and GC content. Therefore, the *Alu* and *L1* elements are believed to construct a genome-wide framework, by dually assigning the specific BOE and PRE domains to each coding region.

Retroelement-rich regions have the characteristics of heterochromatin, increased DNA methylation level, decreased gene density and delayed replication timing (14). Considering that gene regulation is influenced by the accessibility, rather than the abundance, of transcription factors (15, 16), a certain degree of methylation, which is a factor that plays a role in the access control of a transcriptional factor (17), may result in various gene expressions. The observation that the distribution of *Alu* elements is biased toward the gene-rich region has led to the suggestion that the purpose of the intimate relationship between the *Alu*s and the coding regions is the fitness advantage of the host (1, 5). Given that, in this study, the density of the sense *Alu*s in the upstream regions is related to the BOE status, the higher density of the sense *Alu*s is thought to be more advantageous to the efficient recruitment of transcription complexes in the regulatory region. Alternatively, this orientation-dependent relationship of *Alu* elements with BOE should minimize the excessive amount of total *Alu* elements that are prone to DNA methylation and may result in gene silencing. Intragenic *L1* elements tend to be extremely depressed, in order to elude the harmful consequences of insertion mutagenesis (6) or in order to ensure the existence of a methylation-free region, which warrants the gene transcription (18). We assume that the variable densities of antisense *L1*s, both within and adjacent to the coding genes, introduce the heterochromatin-like variegation to the proximal region, thus providing regulatory domains. Alternatively, it has been suggested that short introns increase the transcriptional efficiency (19), and that the variable intragenic densities of the antisense *L1*s may influence the gene expression in a gene-size-dependent manner. Consequently, sense *Alu*s and antisense *L1*s formulate separate local domains in opposite directions for BOE and PRE, respectively, which is considered as a highly sophisticated allocation designed to reconcile the conflict between gene expression and retroelement suppression.

The CTs are thought to warrant subnuclear compartments for distinct transcriptional activity in a huge nuclear space, which are dynamically communicated by highly mobile nuclear proteins (10, 11). Thus, a group of several chromosomes seems to serve as a global expression domain. In this study, the radial organization of chromosomes was found to have a stronger correlation with the *Alu* and *L1* compositions than the gene density and GC content, although the latter have been known to constitute the most important parameter assigning the intranuclear position (12, 13). Assorting the chromosomes in the order of the *Alu* and *L1* fractions allows the different chromosomes to be classified into *Alu*-poor and *L1*-rich groups, despite their high correlation coefficients, while classifying the same chromosomes into the *Alu*-rich and *L1*-poor group. Given that chromosomes at the nuclear edge are in greater contact with the cytoplasm and lesser contact with other chromosomes, edge-positioning *Alu*-poor or *L1*-rich chromosomes are expected to organize distinct global domains, which are more dependent on the *Alu* or *L1* content, respectively. Meanwhile, the center-positioning *Alu*-rich and *L1*-poor chromosomes are likely to share a common global domain, which is under the influence of both the *Alu* and *L1* elements, through intranuclear inter-CT communication. However, the density differences of the sense *Alu*s and antisense *L1*s between expression levels are statistically significant in all the *Alu*-sorted and *L1*-sorted chromosome groups, despite the retroelement composition being different. Sense *Alu* and antisense *L1* local domains appear to function under the influence of *Alu* and *L1* global domains, because the *Alu* and *L1* density differences between expression levels are skewed or compromised in other groups which are categorized based on the intranuclear position, gene density and GC content parameters (Fig. 4, 5).

It is noteworthy that the chromosome groups reveal different BOE-*Alu* relations depending on the intrachromosomal fraction of the *Alu* and *L1* elements. For example, in the *Alu*-rich and *L1* poor chromosome group, the significant density differences of the sense *Alu*s between the expression levels are attenuated and extended to more distal regions, as compared with the *Alu*-poor or *L1*-rich chromosome group (Fig. 4). Because an increase of the sense *Alu*s in the proximal region, which is burdened with more *Alu*s, may lead to gene silencing owing to the presence of an excessive number of *Alu*s prone to DNA methylation, the wide expression range of sense *Alu* density in the *Alu*-rich and *L1*-poor group seems to be selectively formulated in distant regions, where the *Alu* content gradually decreases (Fig. 3). Meanwhile, the *Alu*-poor or *L1*-rich chromosome group demonstrates a marked density difference of sense *Alu* elements in closer proximity to the coding region. This sense *Alu* distribution appears to be selected for the control of BOE in the *Alu*-poor or *L1*-rich global domain, which is less prone to *Alu* methylation and allows more sense *Alu*s at the proximal segment.

The *L1*-rich or *Alu*-poor chromosome group is more prone to *L1* methylation, and the more prominent antisense *L1* depression in this group (Fig. 5) seems to be a requirement for the observed expression range of the *L1* density. Interestingly, the *Alu*-poor chromosome groups demonstrate a density difference for the *L1*s between the intermediate and low levels of PRE within the 0-10 kb upstream region. In the other chromosome groups, the density difference of the antisense *L1*s is prominent within the 0-10 kb downstream region, which is burdened with less *Alu*s than the upstream region (Fig. 3). The prominent density difference within the 0-10 kb upstream region is likely to be allowed only in the *Alu*-poor group, in which the sense *Alu*s adjacent to the genes are considerably depressed (Fig. 4). This reciprocal pattern of sense *Alu* and antisense *L1* local distribution, which is observed in proximity to the genes, seems to minimize the amount of overlapping between extragenic *Alu* and intragenic *L1* domains, as well as the additive effect of *Alu* and *L1* elements on DNA methylation. Therefore, an *Alu*-*L1* genome-wide framework is likely to maximize a spectrum of dual expression (BOE and PRE) in a number of separate local domains.

Because the promoter and regulatory elements are variably located and experimentally identified in only a small fraction of genes (20), it is unlikely that the *Alu* and *L1* framework is related to the presence of such regulatory elements. Rather, retroelements appear to construct their own expression domains. High-BOE genes in different chromosome groups are similarly embraced by proximal sense *Alu*s in the rapidly increasing density regions, and by distant sense *Alu*s in the gradually decreasing density regions (Fig. 4). Such a uniform *Alu* distribution may be advantageous for the transcriptional linkage of housekeeping genes across entire genome (9). In addition, the increased number of sense *Alu*s in proximity to the housekeeping genes that cluster together (9) would keep individual domains independent as well. On the other hand, in the *Alu*-rich and *L1*-poor chromosome group, the prominent density difference of the sense *Alu*s between the intermediate and low levels of BOE resides in the 70-80 kb distal region. Such extended domains are likely to have evolved for the sake of scattered low- and intermediate-BOE genes. It is known that the PRE status is not associated with gene clustering (21). Because of the location of *L1* depression within the genic and adjacent regions, few genes are required to cluster together in terms of PRE. Therefore, the gene clustering phenomenon appears to be advantageous only for housekeeping genes.

In this study, we identified a genome-wide expression framework that correlates the BOE and PRE statuses with the *Alu* and *L1* elements in a reciprocal manner. This dual expression system is thought to offer certain advantages for gene expression, as well as for retroelement suppression by DNA methylation. Local and global expression domains are likely to have been methodologically formulated through the mutual cooperation of coding genes and non-coding retroelements, in order to provide for the best coordination of individual cells, while according due attention to the organism's fitness. Given that individual retroelement copies are so heterogeneous in terms of their sequence divergence and truncated size, it would be of interest to determine how and what repetitive copies participate in expression-related DNA methylation.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K, Meldrim J, Mesirov JP, Miranda C, Morris W, Naylor J, Raymond C, Rosetti M, Santos R, Sheridan A, Sougnez C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L, Mercer S, Milne S, Mullikin JC, Mungall A, Plumb R, Ross M, Shownkeen R, Sims S, Waterston RH, Wilson RK, Hillier LW, McPherson JD, Marra MA, Mardis ER, Fulton LA, Chinwalla AT, Pepin KH, Gish WR, Chissoe SL, Wendl MC, Delehaunty KD, Miner TL, Delehaunty A, Kramer JB, Cook LL, Fulton RS, Johnson DL, Minx PJ, Clifton SW, Hawkins T, Branscomb E, Predki P, Richardson P, Wenning S, Slezak T, Doggett N, Cheng JF, Olsen A, Lucas S, Elkin C, Uberbacher E, Frazier M, Gibbs RA, Muzny DM, Scherer SE, Bouck JB, Sodergren EJ, Worley KC, Rives CM, Gorrell JH, Metzker ML, Naylor SL, Kucherlapati RS, Nelson DL, Weinstock GM, Sakaki Y, Fujiyama A, Hattori M, Yada T, Toyoda A, Itoh T, Kawagoe C, Watanabe H, Totoki Y, Taylor T, Weissenbach J, Heilig R, Saurin W, Artiguenave F, Brottier P, Bruls T, Pelletier E, Robert C, Wincker P, Smith DR, Doucette-Stamm L, Rubenfield M, Weinstock K, Lee HM, Dubois J, Rosenthal A, Platzer M, Nyakatura G, Taudien S, Rump A, Yang H, Yu J, Wang J, Huang G, Gu J, Hood L, Rowen L, Madan A, Qin S, Davis RW, Federspiel NA, Abola AP, Proctor MJ, Myers RM, Schmutz J, Dickson M, Grimwood J, Cox DR, Olson MV, Kaul R, Raymond C, Shimizu N, Kawasaki K, Minoshima S, Evans GA, Athanasiou M, Schultz R, Roe BA, Chen F, Pan H, Ramser J, Lehrach H, Reinhardt R, McCombie WR, de la Bastide M, Dedhia N, Blocker H, Hornischer K, Nordsiek G, Agarwala R, Aravind L, Bailey JA, Bateman A, Batzoglou S, Birney E, Bork P, Brown DG, Burge CB, Cerutti L, Chen HC, Church D, Clamp M, Copley RR, Doerks T, Eddy SR, Eichler EE, Furey TS, Galagan J, Gilbert JG, Harmon C, Hayashizaki Y, Haussler D, Hermjakob H, Hokamp K, Jang W, Johnson LS, Jones TA, Kasif S, Kaspryzk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korf I, Kulp D, Lancet D, Lowe TM, McLysaght A, Mikkelsen T, Moran JV, Mulder N, Pollara VJ, Ponting CP, Schuler G, Schultz J, Slater G, Smit AF, Stupka E, Szustakowski J, Thierry-Mieg D, Thierry-Mieg J, Wagner L, Wallis J, Wheeler R, Williams A, Wolf YI, Wolfe KH, Yang SP, Yeh RF, Collins F, Guyer MS, Peterson J, Felsenfeld A, Wetterstrand KA, Patrinos A, Morgan MJ, Szustakowki J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S, Chen YJ; International Human Genome Sequencing Consortium. *Initial sequencing and analysis of the human genome. Nature 2001; 409: 860-921.*
2. Doolittle WF, Sapienza C. *Selfish genes, the phenotype paradigm and genome evolution. Nature 1980; 284: 601-3.*

3. Ohno S. *So much 'junk' DNA in our genome in Brookhaven Symposia in Biology. Gordon and Breach, New York 1972; 23: 366-70.*
4. Smit AF. *Interspersed repeats and other mementos of transposable elements in mammalian genomes. Curr Opin Genet Dev 1999; 9: 657-63.*
5. Chu WM, Ballard R, Carpick BW, Williams BR, Schmid CW. *Potential Alu function: regulation of the activity of double-stranded RNA-activated kinase PKR. Mol Cell Biol 1998; 18: 58-68.*
6. Ostertag EM, Kazazian HH Jr. *Biology of mammalian L1 retrotransposons. Annu Rev Genet 2001; 35: 501-38.*
7. Brosius J. *RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. Gene 1999; 238: 115-34.*
8. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. *Serial analysis of gene expression. Science 1995; 270: 484-7.*
9. Lercher MJ, Urrutia AO, Hurst LD. *Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat Genet 2002; 31: 180-3.*
10. Cremer T, Cremer C. *Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2001; 2: 292-301.*
11. Andrulis ED, Neiman AM, Zappulla DC, Sternglanz R. *Perinuclear localization of chromatin facilitates transcriptional silencing. Nature 1998; 394: 592-5.*
12. Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. *The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. Hum Mol Genet 2001; 10: 211-9.*
13. Tanabe H, Muller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. *Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. Proc Natl Acad Sci USA 2002; 99: 4424-9.*
14. Dillon N, Festenstein R. *Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. Trends Genet 2002; 18: 252-8.*
15. Wolffe AP, Matzke MA. *Epigenetics: regulation through repression. Science 1999; 286: 481-6.*
16. Colot V, Rossignol JL. *Eukaryotic DNA methylation as an evolutionary device. Bioessays 1999; 21: 402-11.*
17. Eden S, Hashimshony T, Keshet I, Cedar H, Thorne AW. *DNA methylation models histone acetylation. Nature 1998; 394: 842.*
18. Medstrand P, van de Lagemaat LN, Mager DL. *Retroelement distributions in the human genome: variations associated with age and proximity to genes. Genome Res 2002; 12: 1483-95.*
19. Urrutia AO, Hurst LD. *The signature of selection mediated by expression on human genes. Genome Res 2003; 13: 2260-4.*
20. Frith MC, Spouge JL Hansen U, Weng Z. *Statistical significance of clusters of motifs represented by position specific scoring matrices in nucleotide sequences. Nucleic Acids Res 2002; 30: 3214-24.*
21. Vinogradov AE. *Isochores and tissue-specificity. Nucleic Acids Res 2003; 31: 5212-20.*
22. Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tome P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A, Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Hudson TJ. *A gene map of the human genome. Science 1996; 274: 540-6.*