

Research article

Open Access

## Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependant mobile Alu elements

Ravi Shankar, Deepak Grover, Samir K Brahmachari and Mitali Mukerji\*

Address: Functional Genomics Unit, Institute of Genomics and Integrative Biology (IGIB), CSIR, Mall Road, Delhi 110007, India

Email: Ravi Shankar - ravis@igib.res.in; Deepak Grover - grover@igib.res.in; Samir K Brahmachari - skb@igib.res.in;

Mitali Mukerji\* - mitali@igib.res.in

\* Corresponding author

Published: 04 October 2004

Received: 01 July 2004

*BMC Evolutionary Biology* 2004, 4:37 doi:10.1186/1471-2148-4-37

Accepted: 04 October 2004

This article is available from: <http://www.biomedcentral.com/1471-2148/4/37>

© 2004 Shankar et al; licensee BioMed Central Ltd.

This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** The primate-specific Alu elements, which originated 65 million years ago, exist in over a million copies in the human genome. These elements have been involved in genome shuffling and various diseases not only through retrotransposition but also through large scale Alu-Alu mediated recombination. Only a few subfamilies of Alus are currently retropositionally active and show insertion/deletion polymorphisms with associated phenotypes. Retroposition occurs by means of RNA intermediates synthesised by a RNA polymerase III promoter residing in the A-Box and B-Box in these elements. Alus have also been shown to harbour a number of transcription factor binding sites, as well as hormone responsive elements. The distribution of Alus has been shown to be non-random in the human genome and these elements are increasingly being implicated in diverse functions such as transcription, translation, response to stress, nucleosome positioning and imprinting.

**Results:** We conducted a retrospective analysis of putative functional sites, such as the RNA pol III promoter elements, pol II regulatory elements like hormone responsive elements and ligand-activated receptor binding sites, in Alus of various evolutionary ages. We observe a progressive loss of the RNA pol III transcriptional potential with concomitant accumulation of RNA pol II regulatory sites. We also observe a significant over-representation of Alus harboring these sites in promoter regions of signaling and metabolism genes of chromosome 22, when compared to genes of information pathway components, structural and transport proteins. This difference is not so significant between functional categories in the intronic regions of the same genes.

**Conclusions:** Our study clearly suggests that Alu elements, through retrotransposition, could distribute functional and regulatable promoter elements, which in the course of subsequent selection might be stabilized in the genome. Exaptation of regulatory elements in the preexisting genes through Alus could thus have contributed to evolution of novel regulatory networks in the primate genomes. With such a wide spectrum of regulatory sites present in Alus, it also becomes imperative to screen for variations in these sites in candidate genes, which are otherwise repeat-masked in studies pertaining to identification of predisposition markers.

## Background

In the post genome sequence era, repetitive sequences, erstwhile considered junk and devoid of function, are increasingly being implicated in many cellular functions, genome organization and diseases [1-8]. Alu repeats, which belong to SINE (short interspersed nucleotide elements) family of repetitive sequences, are present exclusively in the primate genomes. These elements which are ~300 bps in length have originated from the 7SL RNA gene and comprise of two similar, but not identical subunits [9-12]. Each element contains a bipartite promoter for RNA polymerase III, a poly (A) tract located between the monomers, a 3'-terminal poly(A) tract, a number of CpG dinucleotides, and is flanked by short direct repeats [13,14]. Based on certain diagnostic site mutations, they have been broadly classified into three subfamilies: Old (Alu Js), Middle (Alu S) and the Youngest (Alu Ys) [15,16]. Further, some of the Alu Y sequences are very new and exhibit polymorphisms, indicating that they have recently undergone retropositioning process [17].

Alus have been shown to harbor a number of regulatory sites like hormone response element (HRE), and a couple of ligand activated transcription factor binding sites [18-24]. These sites regulate the expression of downstream genes, in some cases in a temporal or tissue specific manner. Most of the regulatory sites in Alus have been reported during the course of characterization of specific genes [25-32]. Besides, the intrinsic A-Box and B-Box RNA polymerase III (RNA pol III) sequences and the recombinogenic sites present in these elements are involved in retrotranspositional and recombination process [10].

Alus originally demonstrated to have non uniform distribution on the chromosomes through banding studies [33,34] have been recently substantiated by genome sequence analysis [35]. It has been observed that that Alus not only show a non- random pattern of distribution in the human chromosomes but also varying densities within genes. Additionally, in a genome wide expression

analysis, co-variation of expression of gene pairs has been attributed to sequence similarity metric in the upstream region of promoter predominantly contributed by Alu repeats present in these regions [36]. These effects of Alu have been shown to be completely independent of the effects of isochoric (GC) composition on Alu density as well as gene expression [34-36].

Identification and analysis of various permutations and combinations of these regulatory elements in otherwise conserved repetitive Alus are mostly excluded from genetic analysis. Since, Alus occupy a tenth of the human genome, it is imperative to identify those, which might assume function in the proper context. Our primary aim in this analysis is to find out if any bias exists in the distribution of transcriptional regulatory sites in Alus of various evolutionary ages and their distribution with respect to the functional classes of genes.

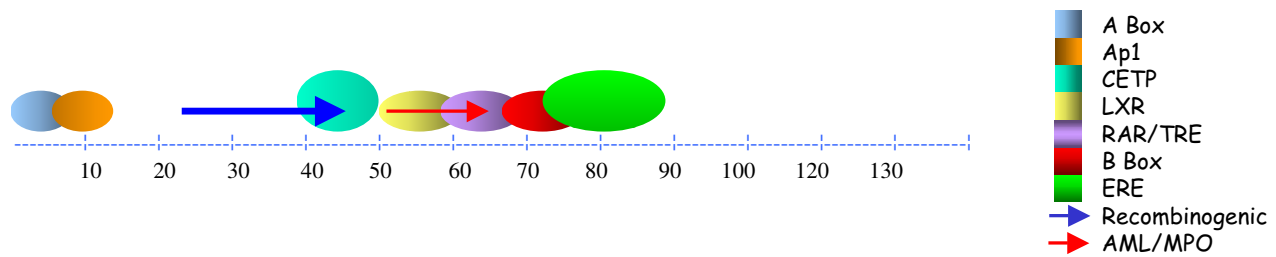
## Results and Discussion

### *Distribution of functional sites in Alus is position specific*

As a first step toward examining the role of these regulatory sites, we mapped their most probable positions on Alus, using in house developed algorithms (Figure 1). This was carried out on 500 Alus, each of Alu Jo, Alu Jb, Alu Sx, Alu Sc, Alu Yb8 and Alu Y subfamilies. The classification of these evolutionarily distinct subfamilies are based on diagnostic sites [15,16,37,38]. Besides, members of the most recent and retropositionally active and polymorphic Alus were also included in the analysis [39,40]. Though the polymorphic Alus belong to Alu Y subfamily, these were treated as a separate category since insertion/deletion of these Alus have been associated with many phenotypes/diseases [2]. The regulatory sites show positional conservation across all subfamilies in which they are represented (Table 1). However, these sites are distinct from the diagnostic sites, which are used for classifying Alus, which suggests that they have not arisen randomly in different subfamilies.

**Table 1: Position of sites analysed in Alu repeats in various subfamilies.**

| Family | A-box | B-box | AML | MPO | CETP | Rec | API    | ERE | RARE  | TRE | nCaRE | LXR     |
|--------|-------|-------|-----|-----|------|-----|--------|-----|-------|-----|-------|---------|
| Jb     | 5     | 76    | 48  | 48  | 47   | 22  | 13/221 | 80  | 57-76 | -67 | 289   |         |
| Jo     | 5     | 76    | 48  | 48  | 47   | 22  | 13/221 | 80  | 66    | -67 | 289   | 224-240 |
| Sx     | 5     | 76    | 48  | 48  | 47   | 22  | 13     | 80  | 60    | -67 | 289   | 237-250 |
| Sc     | 5     | 76    | 48  | 48  | 47   | 22  | 13/267 | 80  | 68    | -67 | 289   |         |
| Y      | 5     | 76    | 48  | 48  | 47   | 22  |        | 80  |       | -67 | 289   |         |
| Yb8    | 5     | 76    | 48  | 48  | 47   | 22  | 13/270 | 80  | 60-66 | -67 | 289   | 230-240 |
| POLY   | 5     | 76    | 48  | 48  | 47   | 22  | 13/267 | 80  | 60    | -67 | 289   |         |



**Figure 1**

**Representation of regulatory sites on Alu elements.** 500 representative Alu sequences each of distinct evolutionary ages were selected for identification of most probable regulatory sites. 126 polymorphic Alu (POLY) from younger subfamilies which show insertion – deletion polymorphisms were also analysed. Sites were identified using local alignment based program as well as by probabilistic modelling approach. These sites are positionally conserved in all subfamilies.

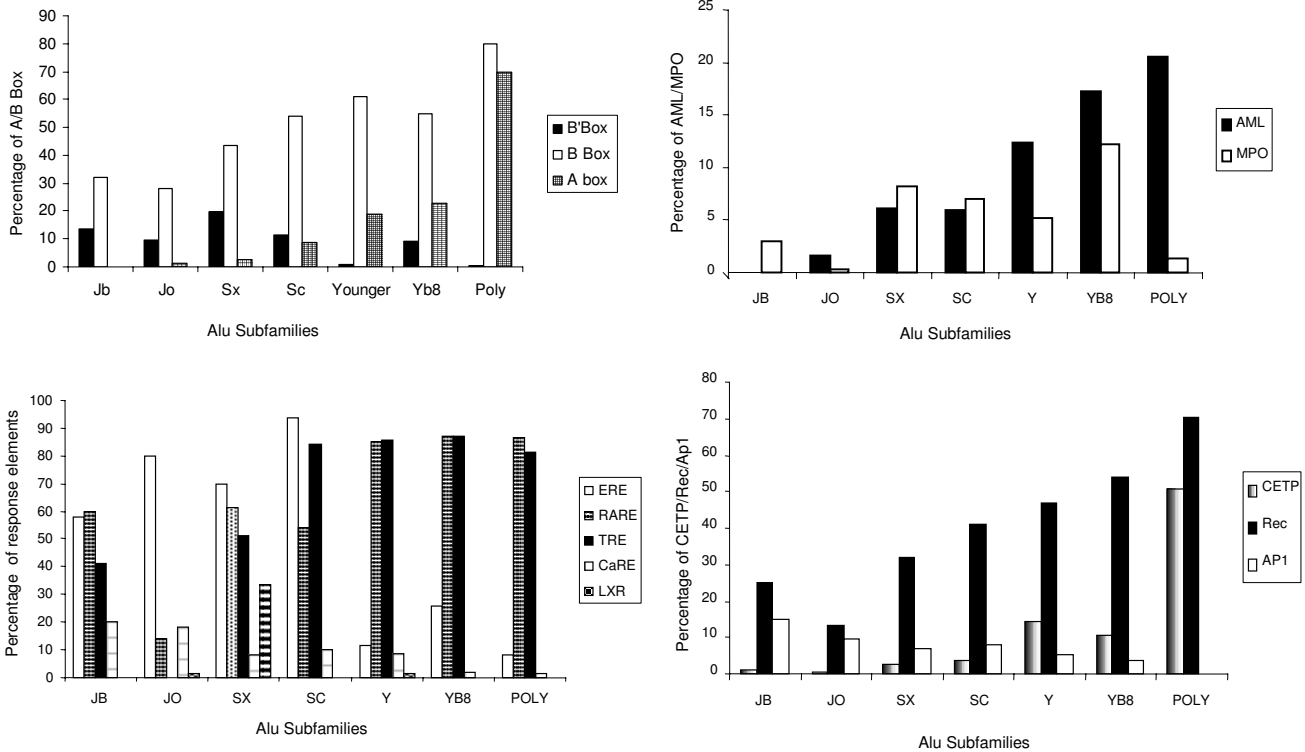
#### **Evolution of regulatory sites is biased and clustered in Alus**

Nearly all the analyzed regulatory sites for RNA polymerase II (RNA pol II) are distributed in the region between A-Box and B-Box with more clustering near the B-Box region (Figure 1). There is an evolutionary age specific loss / gain of these sites in various subfamilies leading to a bias in their distribution (Figure 2). Newly transposing Alus have methylated CpG sites, which are prone to transition. Many sites seem to have evolved as a consequence of these transitions. The regulatory elements are most abundant in the middle subfamilies and least represented in the younger Alus. Some sites like AP1, ERE, nCARE are present in older and middle Alus but rarely so in the younger as well as polymorphic Alu. An opposite trend is observed for CETP, wherein the highest density is observed in the younger active and polymorphic Alu. RARE and TRE sites are retained in all subfamilies whereas LXR is specific to only middle Alu subfamilies (Figure 2). It is curious, nCARE which is also present in the 7sl RNA, the progenitor of Alus, is not equally represented in all Alus and has higher density in the older Alus and middle and is very poorly represented in the younger subfamilies.

#### **Evolution from retropositionally active to transcriptionally active Alu elements**

Majority of Alu retroposition has ceased at least 30 million years ago and only a few Alu subfamilies are still active [15,17,41]. Transcription of Alus is a prerequisite for retrotransposition and there is regulation not only during transcription initiation but also at the level of stability of transcripts [42]. Alu elements are transcribed by RNA pol III which are composed of two properly spaced conserved sequence motifs, an upstream element named the A-Box and a downstream element called the B-Box which are essential for efficient transcription. Deletion of

the Box B sequences within the Alu repeat completely abolishes the transcriptional activity. In the absence of box A sequences even though there is a reduction in efficiency of transcription by 10 to 20 fold, B-Box sequence is still capable of initiating transcription 70 bps upstream [43,44]. An intact A-Box is therefore a critical determinant for RNA pol III retropositional activity. Besides, it has been shown by *in vitro* as well as *in vivo* studies in the 'B' Box that 'G' and 'T' residues at the 1<sup>st</sup> and 3<sup>rd</sup> positions respectively are very critical for its functioning [45]. Our analysis on the distribution of these promoter elements show that the polymorphic Alu sequences have the highest density of A-Box (70%) and is almost absent in older subfamilies (Figure 2). Since the younger Alus are considered to be transcriptionally more active, this fits in well with the loss of this site in the course of evolution due to accumulation of mutations. The B-Box motif with the sequence G(A/T)T(C/T)RANNC shows a similar trend as the A-Box. Interestingly, a fraction of older Alu subfamily still retains the B-Box sequence. However, 'A' residue at the second position which has not been shown to be critical for transcription is a diagnostic nucleotide [39] for the younger subfamilies. This could result in the increased proportion for B-Box in the younger families. We observe a very curious distribution of the B-Box motif if we consider the sequence GTT(C/T)GAGAC (B'Box in Figure 2) wherein we restrict the pattern to the experimentally validated sequence. Alu Sx and Alu Sc have the highest density match with this pattern, followed by the older subfamilies and it is present in only < 2% frequency in AluY and polymorphic Alu. The "C" at the 4<sup>th</sup> position in this case is mutated to "T" in the older families. The Yb8 family that has been reported to be transcriptionally and retropositionally active amongst the younger subfamilies, retains the B'-Box element in a significant fraction. This suggests



**Figure 2**  
**Distribution of regulatory sites in various Alu subfamilies as well as polymorphic Alus.** On the X-axis Alus of different evolutionary ages as well as polymorphic Alu (POLY) are represented. On the Y-axis the percentage of elements carrying these sites in various subfamilies is indicated.

that even though retropositionally competent younger Alus are hypothesized to be transcriptionally active, only a minority retains consensus B'-Box. It is possible that the enhancing activity of the A Box is sufficient to drive transcription from the weaker B'- Box in the younger subfamilies. Our findings corroborates well with an earlier study in which presence of all subfamilies in the RNA polymerase III driven Alu transcript pool was reported [46]. Additionally, it was also observed that though there was a quantitative bias towards younger subfamilies and younger members of these subfamilies (based on their relative presence in the transcript compared to their abundance in the genome), there was a preferential expression of the middle subfamilies relative to the most active subfamilies. Our observations, therefore, further rules out the hypothesis that transcription may be biased only towards retropositionally active subfamilies of Alu elements. This could be the reason why only a fraction of younger Alus is currently retrotranspositionally active. The presence and

retention of B-Box coupled with near absence of A Box in the Alu Sx and AluSc families suggests basal level of transcription from these elements which could be enhanced through binding of other regulatory proteins under certain conditions such as stress [47]. Additionally, with evidence of presence of naturally occurring Alu antisense as well as edited Alu transcripts [48,49], transcribing Alus could play a major role in yet unknown biological processes.

**Exaptation of Alus in the transcriptional regulatory repertoire**

Alus have been demonstrated to exert effects at transcription, post-transcription as well as at the translation level. In an earlier study on complete chromosomes 21 and 22, we have demonstrated that the Alu elements are clustered in genes of signaling, metabolic and transport proteins and rarely present in the structural and information proteins [50]. This clustering bias was found to be irrespective

of genomic location, GC content, length of genes or intronic content. To further address whether the Alu harboring transcriptional regulatory sites also show a selective distribution and thereby exert effects on transcription, we analyzed their distribution in the genes of various functional categories of chromosome 22. Two different datasets 1) Promoter region Alus and 2) Intronic region Alus, harboring regulatory sites were analyzed. The promoter region Alus of genes involved in metabolism, signaling were significantly rich in regulatory sites compared to those of information, structure and transport (F value = 4.86, df = 4, 40, p-value < 0.0027). In the intronic regions, distinction in their distribution with respect to functional categories was not so significant though the intronic regions also harboured Alus containing regulatory sites (F value = 2.92, df = 4, 40, p-value = 0.032). Since the genes of the signaling and metabolic pathway are more subject to regulation by cellular cues like hormonal triggers, this observation is significant. Most of the Alus in the promoters belong to the middle Alu S families and rarely Younger Alus are present. Since younger Alus also harbour few regulatory sites and actively retrotranspose, it is possible that there is a negative selection against their insertion in the promoter sites of genes of information pathways and structural proteins [see the 1].

#### **Alu movements and aberrant gene expression**

Gene inversions, duplications and formation of pseudo-genes have been extensively reported to be mediated both through retrotransposition as well as recombination of Alus. This, in many cases, has also been associated with aberrant gene expression. For instance, presence of AML sites in an Alu upstream of MPO gene, has been first demonstrated to be associated with Acute Myelocytic Leukemia [20]. This is due to the presence of a strong SP1 site within AML which leads to over expression of MPO gene. AML sites are most abundant in younger and polymorphic Alus and a single base pair transition results in MPO site, present predominantly in the members of older sub-families. In the case of polymorphic Alus, many sequences that do not show 100% conservation of AML site still retain the SP1 site. Interestingly, the core recombinogenic site is also most predominant in younger and polymorphic Alus. The presence of recombinogenic sites in polymorphic Alus, could therefore not only contribute to genome shuffling but also serve to distribute ectopic sites such as AML through retrotransposition and recombination (Figure 2).

#### **Regulatory region distribution through Alu expansion**

Analysis of regulatory sites within Alus suggests that a polymorphic Alu has the potential to transpose and recombine which allows it to integrate at random sites in the genome. They also harbour potential regulatory sites, which could evolve to become accessory sites for RNA pol

II transcription as revealed by their clustering in older sub-families. Further, the Alu sequence due to acquisition of novel functions could form a part of the transcription repertoire involved in the regulation of the downstream / associated genes and create novel regulatory networks (Figure 3). These results also corroborate with the hypothesis of evolution of transposable elements of Kidwell [51] wherein they had proposed a 3 stage life cycle of class II Transposable elements:- invasion and amplification followed by mutations and maturity and finally senescence and fading. In the case of Alu, instead of fading, they could also evolve to become members of host regulatory machinery.

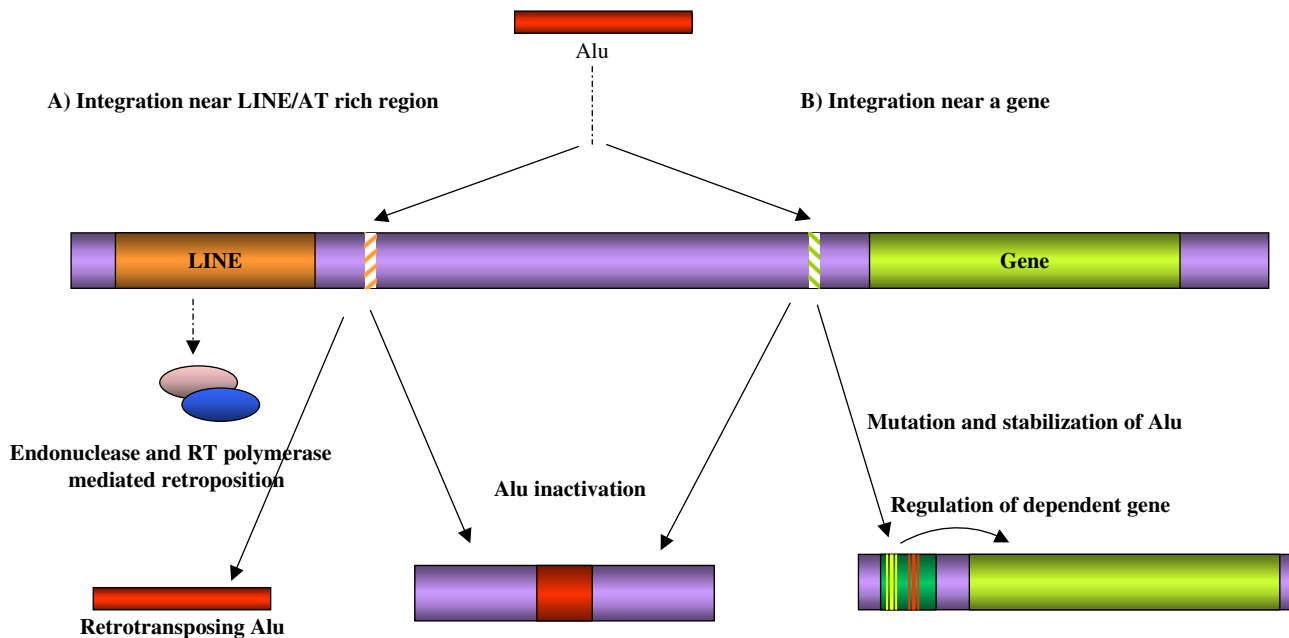
#### **Conclusions**

Comparison of sequences in the regulatory regions of many homologous genes in human have shown accumulation of Alus, not only post divergence from non-human primates but also during primate evolution [52]. Perhaps, recruitment of cis regulatory elements responsive to cellular cues through Alu elements could result in altered spatial and temporal transcription of genes as well as create novel metabolic and signaling networks. These might contribute to the observable physiological complexity in human and primates [53]. Additionally, the underlying events which would be defining event of speciation of human from chimpanzee (with which it shares nearly 99% homology at coding level) still eludes identification and might to some extent reside in such genomic elements. These issues can now be addressed through comparison of these sites in human and chimpanzee.

Currently, Alus are repeat-masked in all studies pertaining to identification of predisposition markers in complex disorders. With such wide spectrum of nuclear receptors, which play a major role in maintaining normal physiological state and affect as diverse processes as development, reproduction, general metabolism, residing in Alus, it therefore becomes imperative to screen for variations in these sites. This might have important consequences in the candidate genes for those complex diseases that are triggered in response to hormonal imbalances as well as other environmental cues.

#### **Methods**

126 polymorphic Alu sequences cited in literature [39,40] were retrieved using NCBI BLAST and Repeat Masker software [54,55]. The analysis was carried out on Alu repeats of human chromosome 22. A randomly selected representative set of approximately 500 Alu sequences, each of distinct evolutionary ages, Alu Jb, Alu Jo, Alu Sx, Alu Sc, Alu Yb8 and Alu Y were used for the analysis. Sequences were retrieved from Sanger Institute Home Page, June 2001 release [56]. Besides, Alus were also analyzed within

**Figure 3**

**Alu expansion and evolution of regulatory sites.** With the help of LINES, Alu may keep on retro-transposing or may get inactive/negatively selected. Alternatively, it may integrate upstream of a gene, accumulate mutations, evolve RNA pol II regulatory sites, get stabilized and control gene expression. This is supported by the presence of sparse regulatory sites, unhindered A box, recombinogenic sites initially in the younger and active Alus and its accumulation in older Alu subfamilies as well as significant presence of Alus harbouring regulatory sites in the promoter encompassing regions of the genes of signaling and metabolic pathways.

5000 base pairs upstream of genes of chromosome 22 in the regulatory regions encompassing promoter sequences as well as inside their intronic regions.

#### Collection of biologically active sites

Information about the regulatory sites and their sequences was collected from various literature sources (Table 2). Characteristic features of the sites are given below. We selected those regulatory sites, which have been shown to have function in the Alu elements. The A Box and B Box sequences define the bipartite internal promoters, which bind RNA polymerase III. MPO and AML sites, which are 14 nucleotides differ by an A / G at 5th position of the sequence and transition from G to A at this site converts the MPO allele to AML, resulting in the formation of a strong SP-1 binding site and over expression of the following gene. AP1 sites bind AP-1 transcription factor, which is a dimeric complex that contains members of the JUN, FOS, ATF and MAF protein families. Hormone responsive elements (HRE) are super family of binding sites for ligand activated nuclear hormone receptors for thyroid hormone (TRE), retinoic acid (RARE) and

vitamin D, which regulate gene transcription. Estrogen response elements (EREs) are sites for binding of estrogen receptor (ER), a ligand-activated enhancer protein that is a member of the steroid/nuclear receptor super family and transactivates gene expression in response to estradiol. The negative calcium response element type 2 (nCARE) is a regulatory DNA sequence, which inhibits transcription in response to raised extra cellular calcium levels. The nuclear receptors liver X (LXR) is involved in different cell-signaling pathways. CETP site is an orphan receptor site in the Alu in promoter of cholesteryl ester transfer protein (CETP) which plays a key role in reverse cholesterol transport in mediating the transfer of cholesteryl ester from HDL to atherogenic apolipoprotein B-containing lipoproteins.

#### Computational methods

Two different programs were written in order to locate the most probable biologically significant regions. A local alignment based program, Xalign, was implemented in C++, Red Hat 7.3 based Linux. This program finds the probable sites by aligning the consensus of regulatory site

**Table 2: Sequences of regulatory elements analysed in Alu repeats.**

| Site   | Sequence                          |
|--|-----------------------------------|
| Retinoic acid response element (RARE)                    | 5'(AG)G(GT)TCA 3'                 |
| Estrogen Response Element (ERE)                          | 5'(GA)(GA)TCA(CG)(AC)(CG)TGACC 3' |
| Negative calcium response element (nCARE)                | 5' TGAGACNNNGTCTCAAAAA 3'         |
| Liver X receptor   | 5' GACCTNNNNTGATCC 3'             |
| Cholesteryl esterase transferase response element (CETP) | 5'CCGNGGCGGGC 3'                  |
| API site   | 5' T(GTA)A(GC)TCA 3'              |
| Acute Myelocytic Leukemia (AML) site                     | 5' AGGCGGGTGGATCA 3'              |
| Myelo Peroxidase (MPO) site                              | 5' AGGCAGGTGGATCA 3'              |
| Recombinogenic site                                      | 5'CCCTGTAATCCTAGCACTTTGGAGGC 3'   |
| A-Box  | 5' GGGCGCGGTGGC 3'                |
| B-Box  | 5' G(A/T)T(C/T)RANNC 3'           |
| B'Box  | 5' G TT(C/T)GAGAC 3'              |

Nucleotide sequences in parenthesis indicate alternate nucleotides and have been written in increasing order of their preference.

with the query sequence. Multiple queries with a size upto 600 nucleotides can be taken at a time. Another program, Promotif, was implemented in C++, Red Hat 7.3 based Linux, using the probabilistic modeling approach. It uses the position weight matrix, normalization of the positions with conservation index (Ci Value), and inter-nucleotide dependence in terms of transition matrix to find out the sites. Position weight matrices were generated using Gibbs Motif Sampler, for every site included in the program. The sequences for position weight matrix generation were carefully selected based on the sequence and length reported for each binding site. The final length for search was fixed at the lowest length observed. This provides element specific matrix with lesser chance for the selection on non-RE regions. For the sites analyzed, it had an in built transition matrix, position weight matrix and conservation index. Batch analysis of over a thousand Alu sequences can be performed with this program.

Using the annotated sequences from literature as well as from NCBI web page, training set for the probabilistic model was created. Training was done for approximately 70% sequences and rest of the sequences were taken as test set. Details of the program along with the equations used are available on request.

#### **Mapping of recently integrated and younger Alus**

About 126 recently integrated Alus from younger sub-families were searched in the human genome using BLASTn at NCBI server and regulatory sites were mapped in these regions using the programs discussed above.

#### **Association analysis**

Alus in the promoter regions and intronic regions of functionally classified genes [50] of chromosome 22 were

mapped and pattern of distribution of biologically significant sites were analyzed by ANOVA.

#### **Authors' contributions**

RS developed the algorithms and programs for identifying regulatory and significant regions, carried out the analysis of distribution of these sites in Alu subfamilies, association analysis and drafted the manuscript. DG was involved in chromosome 22 analyses. SKB participated in the design of the study. MM conceived of the study, participated in its design, analysis, coordination and manuscript preparation. All authors read and approved the final manuscript.

#### **Additional material**

##### **Supplementary data**

*The analysis over the promoter and intronic regions has been performed through the data given in the supplementary table file, supplementary table 3\_ravishankar et al. Format: .xls. For human chromosome 22, the data contains the accession number, associated Alu family, the respective positions, functional class of the region and further details, for each associated regulatory element found within the Alu repeats in the 5' flanking promoter and intronic regions. The zipped file name is supplementary 1.zip. Details about programs used are on request for academic users.*

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2148-4-37-S1.zip>]

#### **Acknowledgements**

We thank Krishna Kumar and S Suganya for computational support. Financial support from Council of Scientific and Industrial Research (CSIR) projects (CMM0016) to MM and (CMM0017) to SKB is duly acknowledged.

## References

1. Hamdi HK, Nishio H, Tavis J, Zielinski R, Dugaiczak A: **Alu-mediated phylogenetic novelties in gene regulation and development.** *J Mol Biol* 2000, **299**:931-939.
2. Deininger PL, Batzer MA: **Alu repeats and human disease.** *Mol Genet Metab* 1999, **67**:183-193.
3. Szmulewicz MN, Novick GE, Herrera RJ: **Effects of Alu insertions on gene function.** *Electrophoresis* 1998, **19**:1260-1264.
4. Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K: **Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition.** *Proc Natl Acad Sci U S A* 1991, **88**:11315-11319.
5. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS: **A de novo Alu insertion results in neurofibromatosis type 1.** *Nature* 1991, **353**:864-866.
6. Brahmachari SK, Meera G, Sarkar PS, Balagurumoorthy P, Tripathi J, Raghavan S, Shaligram U, Pataskar S: **Simple repetitive sequences in the genome: structure and functional significance.** *Electrophoresis* 1995, **16**:1705-1714.
7. Conrad M, Brahmachari SK, Sasisekharan V: **DNA structural variability as a factor in gene expression and evolution.** *Biosystems* 1986, **19**:123-126.
8. Makalowski W: **Genomic scrap yard: how genomes utilize all that junk.** *Gene* 2000, **259**:61-67.
9. Labuda D, Striker G: **Sequence conservation in Alu evolution.** *Nucleic Acids Res* 1989, **17**:2477-2491.
10. Schmid C, Maraia R: **Transcriptional regulation and transpositional selection of active SINE sequences.** *Curr Opin Genet Dev* 1992, **2**:874-882.
11. Schmid CW: **Alu: structure, origin, evolution, significance and function of one-tenth of human DNA.** *Prog Nucleic Acid Res Mol Biol* 1996, **53**:283-319.
12. Ullu E, Tschudi C: **Alu sequences are processed 7SL RNA genes.** *Nature* 1984, **312**:171-172.
13. Rowold DJ, Herrera RJ: **Alu elements and the human genome.** *Genetica* 2000, **108**:57-72.
14. Mighell AJ, Markham AF, Robinson PA: **Alu sequences.** *FEBS Lett* 1997, **417**:1-5.
15. Shen MR, Batzer MA, Deininger PL: **Evolution of the master Alu gene(s).** *J Mol Evol* 1991, **33**:311-320.
16. Jurka J, Milosavljevic A: **Reconstruction and analysis of human Alu genes.** *J Mol Evol* 1991, **32**:105-121.
17. Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M: **Genetic variation of recent Alu insertions in human populations.** *J Mol Evol* 1996, **42**:22-29.
18. Tomilin NV, Bozhkov VM: **Human nuclear protein interacting with a conservative sequence motif of Alu-family DNA repeats.** *FEBS Lett* 1989, **251**:79-83.
19. Hudson LG, Ertl AP, Gill GN: **Structure and inducible regulation of the human c-erb B2/neu promoter.** *J Biol Chem* 1990, **265**:4389-4393.
20. Piedrafita FJ, Molander RB, Vansant G, Orlova EA, Pfahl M, Reynolds WF: **An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element.** *J Biol Chem* 1996, **271**:14412-14420.
21. Babich V, Aksenov N, Alexeenko V, Oei SL, Buchlow G, Tomilin N: **Association of some potential hormone response elements in human genes with the Alu family repeats.** *Gene* 1999, **239**:341-349.
22. Chesnokov I, Bozhkov V, Popov B, Tomilin N: **Binding specificity of human nuclear protein interacting with the Alu-family DNA repeats.** *Biochem Biophys Res Commun* 1991, **178**:613-619.
23. Vansant G, Reynolds WF: **The consensus sequence of a major Alu subfamily contains a functional retinoic acid response element.** *Proc Natl Acad Sci U S A* 1995, **92**:8229-8233.
24. Norris J, Fan D, Aleman C, Marks JR, Futreal PA, Wiseman RW, Iglehart JD, Deininger PL, McDonnell DP: **Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers.** *J Biol Chem* 1995, **270**:22777-22782.
25. Almenoff JS, Jurka J, Schoolnik GK: **Induction of heat-stable enterotoxin receptor activity by a human Alu repeat.** *J Biol Chem* 1994, **269**:16610-16617.
26. Ashfield R, Ashcroft SJ: **Cloning of the promoters for the beta-cell ATP-sensitive K-channel subunits Kir6.2 and SUR1.** *Diabetes* 1998, **47**:1274-1280.
27. Austin GE, Lam L, Zaki SR, Chan WC, Hodge T, Hou J, Swan D, Zhang W, Racine M, Whitsett C, : **Sequence comparison of putative regulatory DNA of the 5' flanking region of the myeloperoxidase gene in normal and leukemic bone marrow cells.** *Leukemia* 1993, **7**:1445-1450.
28. Brini AT, Lee GM, Kinet JP: **Involvement of Alu sequences in the cell-specific regulation of transcription of the gamma chain of Fc and T cell receptors.** *J Biol Chem* 1993, **268**:1355-1361.
29. Britten RJ: **DNA sequence insertion and evolutionary variation in gene regulation.** *Proc Natl Acad Sci U S A* 1996, **93**:9374-9377.
30. Britten RJ: **Evolutionary selection against change in many Alu repeat sequences interspersed through primate genomes.** *Proc Natl Acad Sci U S A* 1994, **91**:5992-5996.
31. Chang SF, Scharf JG, Will H: **Structural and functional analysis of the promoter of the hepatic lipase gene.** *Eur J Biochem* 1997, **247**:148-159.
32. Le Goff W, Guerin M, Chapman MJ, Thillet J: **A CYP7A promoter binding factor site and Alu repeat in the distal promoter region are implicated in regulation of human CETP gene expression.** *J Lipid Res* 2003, **44**:902-910.
33. Filatov LV, Mamayeva SE, Tomilin NV: **Non-random distribution of Alu-family repeats in human chromosomes.** *Mol Biol Rep* 1987, **12**:117-122.
34. Korenberg JR, Rykowski MC: **Human genome organization: Alu, lines, and the molecular structure of metaphase chromosome bands.** *Cell* 1988, **53**:391-400.
35. 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, Meldrum 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, Showlson 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, Weissbach 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, Ramsay 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, Kasprzyk A, Kennedy S, Kent WJ, Kitts P, Koonin EV, Korfi 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, Patrino A, Morgan MJ, Szustakowski J, de Jong P, Catanese JJ, Osoegawa K, Shizuya H, Choi S: **Initial sequencing and analysis of the human genome.** *Nature* 2001, **409**:860-921.



36. Hon LS, Jain AN: **Compositional structure of repetitive elements is quantitatively related to co-expression of gene pairs.** *J Mol Biol* 2003, **332**:305-310.
37. Carroll ML, Roy-Engel AM, Nguyen SV, Salem AH, Vogel E, Vincent B, Myers J, Ahmad Z, Nguyen L, Sammarco M, Watkins WS, Henke J, Makalowski W, Jorde LB, Deininger PL, Batzer MA: **Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity.** *J Mol Biol* 2001, **311**:17-40.
38. Arcot SS, Adamson AV, Risch GW, LaFleur J, Robichaux MB, Lamerdin JE, Carrano AV, Batzer MA: **High-resolution cartography of recently integrated human chromosome 19-specific Alu fossils.** *J Mol Biol* 1998, **281**:843-856.
39. Batzer MA, Deininger PL: **Alu repeats and human genomic diversity.** *Nat Rev Genet* 2002, **3**:370-379.
40. Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL: **Alu insertion polymorphisms for the study of human genomic diversity.** *Genetics* 2001, **159**:279-290.
41. Batzer MA, Kilroy GE, Richard PE, Shaikh TH, Desselle TD, Hoppens CL, Deininger PL: **Structure and variability of recently inserted Alu family members.** *Nucleic Acids Res* 1990, **18**:6793-6798.
42. Aleman C, Roy-Engel AM, Shaikh TH, Deininger PL: **Cis-acting influences on Alu RNA levels.** *Nucleic Acids Res* 2000, **28**:4755-4761.
43. Perez-Stable C, Ayres TM, Shen CK: **Distinctive sequence organization and functional programming of an Alu repeat promoter.** *Proc Natl Acad Sci U S A* 1984, **81**:5291-5295.
44. Perez-Stable C, Shen CK: **Competitive and cooperative functioning of the anterior and posterior promoter elements of an Alu family repeat.** *Mol Cell Biol* 1986, **6**:2041-2052.
45. Murphy MH, Baralle FE: **Directed semisynthetic point mutational analysis of an RNA polymerase III promoter.** *Nucleic Acids Res* 1983, **11**:7695-7700.
46. Shaikh TH, Roy AM, Kim J, Batzer MA, Deininger PL: **cDNAs derived from primary and small cytoplasmic Alu (scAlu) transcripts.** *J Mol Biol* 1997, **271**:222-234.
47. Liu WM, Chu WM, Choudary PV, Schmid CW: **Cell stress and translational inhibitors transiently increase the abundance of mammalian SINE transcripts.** *Nucleic Acids Res* 1995, **23**:1758-1765.
48. Perl A, Colombo E, Samoilova E, Butler MC, Banki K: **Human transaldolase-associated repetitive elements are transcribed by RNA polymerase III.** *J Biol Chem* 2000, **275**:7261-7272.
49. Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Szybel D, Olshansky M, Rechavi G, Jantsch MF: **Systematic identification of abundant A-to-I editing sites in the human transcriptome.** *Nat Biotechnol* 2004, **22**:1001-1005.
50. Grover D, Majumder PP, Rao CB, Brahmachari SK, Mukerji M: **Non-random distribution of alu elements in genes of various functional categories: insight from analysis of human chromosomes 21 and 22.** *Mol Biol Evol* 2003, **20**:1420-1424.
51. Kidwell MG, Lisch DR: **Perspective: transposable elements, parasitic DNA, and genome evolution.** *Evolution Int J Org Evolution* 2001, **55**:1-24.
52. Hamdi H, Nishio H, Zielinski R, Dugaiczky A: **Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates.** *J Mol Biol* 1999, **289**:861-871.
53. Hamdi H, Nishio H, Zielinski R, Dugaiczky A: **Origin and phylogenetic distribution of Alu DNA repeats: irreversible events in the evolution of primates.** *J Mol Biol* 1999, **289**:861-871.
54. **National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov>]
55. **Repeat Masker server** [<http://www.repeatmasker.org>]
56. **Ensemble Genome Data Resources** [<ftp://ftp.ensembl.org/pub/>]

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

