REVIEW

Alu expression profiles as a novel RNA signature in biology and disease

Davide Carnevali, Giorgio Dieci

Department of Life Sciences, University of Parma, 43124 Parma, Italy

Correspondence: Giorgio Dieci
E-mail: giorgio.dieci@unipr.it
Received: March 19, 2015
Published: November 07, 2015

SINE retrotransposons of the *Alu* subfamily are the most numerous active mobile DNA elements in the human genome. *Alu* transcription by RNA polymerase III is subjected to tight epigenetic silencing, but activated in response to viral infection, genotoxic anticancer agents and other stimuli, through uncharacterized epigenetic switches interspersed throughout the genome. The elucidation of Alu RNA roles in cell biology and pathology has long been hampered by difficulties in their profiling at single-locus resolution, due to their repetitive nature. We recently found how to overcome this limitation by computational screening of RNA-seq data, thus opening the way to Alu transcriptome profiling as a novel tool to explore disease-related epigenome alterations.

Keywords: *Alu*; RNA polymerase III; ncRNA; RNA-seq; biomarker

To cite this article: Davide Carnevali, *et al*. Alu expression profiles as a novel RNA signature in biology and disease. RNA Dis 2015; 2: e735. doi: 10.14800/rd.735.

With ~10^6 copies in the human genome, *Alu*s are the most abundant and successful primate-specific mobile elements of the SINE (short interspersed elements) class, roughly covering 11% of the human genome. *Alu*s are non-autonomous retroelements that need, for their amplification, trans-acting factors encoded by the LINE-1 family of active autonomous human retroelements [1]. Through their novel insertions, whose rate is actually estimated in 1 every 20 births [2], and through non-allelic homologus recombination, Alus contributed to create genetic diversity and diseases [3].

All known SINEs originated from RNA polymerase (Pol) III-transcribed genes [4]. In particular, *Alu*s are ~300 bp long and are formed from 2 divergent monomers, each derived from 7SL RNA, separated from each other by an A-rich linker (A_5TACA_6), followed by a longer A-rich 3' tail which is necessary for retrotransposition.

*Alu*s are autonomous transcription units, as they carry inside their sequence a bipartite promoter for the Pol III transcription machinery which is entirely located within the left monomer. In particular *Alu* transcription requires, in addition to RNA polymerase III, the core transcription factors TFIIIC and TFIIIB. The six-subunit TFIIIC complex recognizes A and B boxes (starting at positions +13 and +77 of the *Alu* sequence, respectively). Once bound, TFIIIC recruits TFIIIB, composed by BRF1 and BDP1 polypeptides together with the TATA box binding protein (TBP), onto a ~50 bp region located immediately upstream of the transcription start site. TFIIIB then directs the assembly of Pol III for transcription initiation. In complete *Alu* elements, elongating Pol III goes through both the left and right monomers and the terminal poly(A) region, and it does not terminate transcription until it encounters a specific termination signal (either canonical or non-canonical) [5, 6]. The terminator (usually a run of at least four Ts on the coding strand) can be located at variable distances downstream of the poly(A): as a consequence, Alu transcripts are very similar to each other in sequence up to

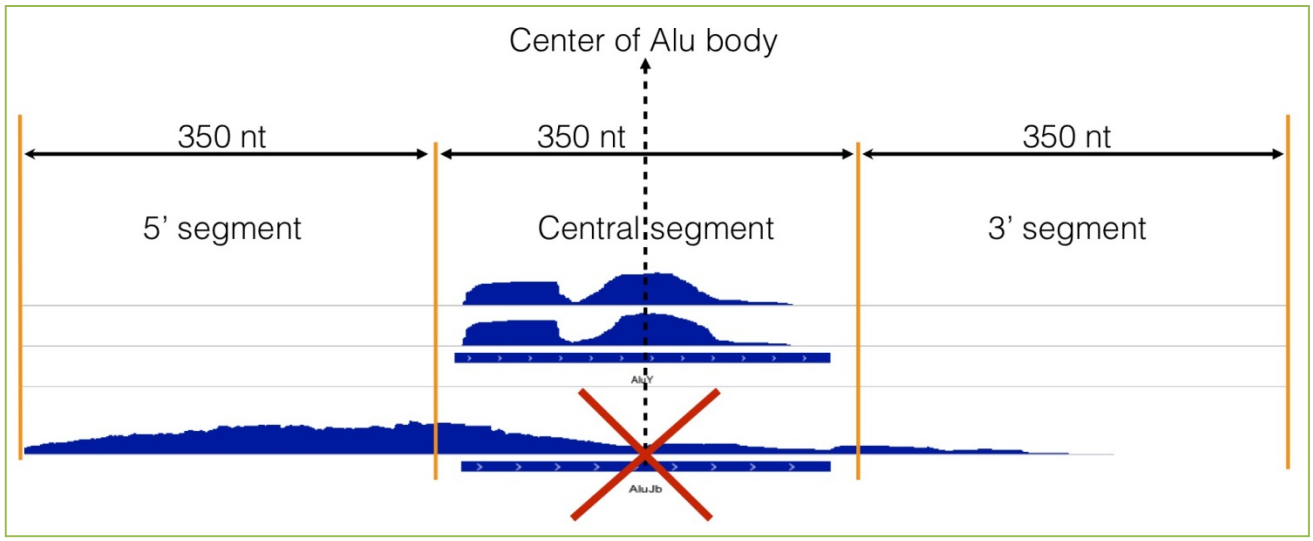

Figure 1. Schematic description of flanking region filter allowing to identify Pol III-dependent *Alu* transcripts from RNA-seq data. For each *Alu* a 1050 bp region was considered, consisting of: a central, 350 bp segment having the *Alu* body at its center; a 5'segment extending 350 bp upstream of the central segment, and a 3' segment extending 350 bp downstream of the central segment. The reads corresponding to *Alu* bodies will map to the central segment, while most of the reads corresponding to *Alu* 3' extensions past the terminal poly(A) will map to the 3' segment. No reads from genuine *Alu* transcripts should map to the 5' segment. If this occurs the *Alu* will be discarded, as exemplified in the lower part of the Figure. Taking 500 nt as the typical *Alu* transcript length, we can predict that its ~150 3'-terminal nucleotides will map to the 3' segment (as we consider that the first 350 nt will map to the central segment). As a consequence, the 3' segment will be typically covered by transcripts for no more than half of its 350-bp extension (while the central segment, containing the *Alu* body, is expected to be covered by transcripts along its whole extension). Therefore, transcripts generating on the 3'segment a signal higher than the half of the central segment signal are considered to be too long for Pol III *Alu* transcripts [29].

the poly(A), but they may carry 3' trailer sequences strongly differing from each other in both sequence and length.

Once thought to be part of the so called "junk DNA", *Alu*'s *cis* and *trans* regulatory roles have been uncovered over time. *Alu* elements have been found to regulate the expression of protein coding genes by acting as silencers, insulators, promoters and enhancers; they also can provide, due to accumulation of mutations in their sequence and A to I pre-mRNA editing, new splice (*Alu* exonization) and polyadenylation sites [7-9].

As to *trans* roles, Pol III *Alu* transcripts have started to attract increasing interest in the last ten years as potential regulators of Pol II-dependent gene expression, either at the post-transcriptional level through antisense strategies [10] or at the transcriptional level through Pol II binding and modulation [11]. *Alu* Pol III transcripts can also act as translation regulators [12, 13], while *Alu* sequences embedded in mRNA 3' UTRs and ncRNA sequences can be part of STAU1-mediated mRNA decay [14] and transcriptional regulation [15].

Despite the abundance of *Alu* elements in the genome, and the generally high efficiency with which Pol III transcribes its target genes [16], the cellular levels of Pol III-synthesized *Alu* RNAs are usually very low in normal conditions and, accordingly, most *Alu*s are not occupied by the Pol III transcription machinery in human cells [17]. *Alu* transcription by Pol III is thought to be limited by epigenetic silencing mainly involving histone methylation [18], but a satisfactory picture of Alu expression regulation at the transcriptional level is still lacking [19]. In particular, early and recent studies have shown that *Alu* transcription by Pol III may be up-regulated in response to different types of signal like viral infection [20], heat shock [21], nuclear receptor stimulation [22], growth signals [23], but the involved molecular mechanisms are still largely unknown. It must be noted, however, that the bulk of *Alu* RNAs are the product of Pol II (not Pol III) transcriptional activity. Indeed, since *Alu* elements are preferentially located in gene-rich regions, particularly within introns and 3'UTRs of Pol II genes [24], the majority of Alu-containing RNAs is due to either primary nuclear transcripts (hnRNAs) or mature mRNAs. Therefore, genuine Pol III-derived *Alu* transcripts are difficult to be distinguished from those that are hosted within longer Pol II-synthesized RNA molecules.

Previous studies aimed at identifying Pol III *Alu* transcripts and the corresponding source elements exploited various combination of size fractionation, C-RACE (rapid amplification of cDNA ends) and sequencing and, more recently, genome-wide chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-seq) against

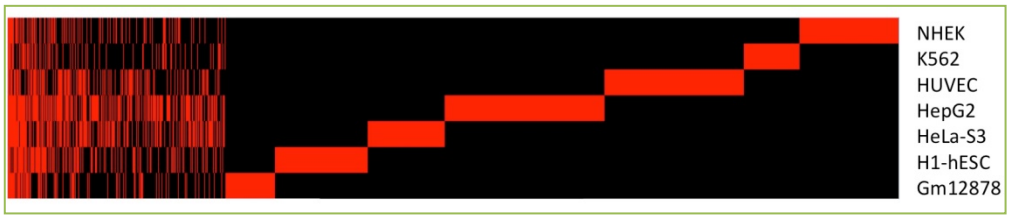

Figure 2. **Cell lineage-specificity of *Alu* expression.** *Alu* expression heatmaps could be generated from ENCODE RNAseq data [29], thus allowing for data clustering and for estimation of the degree of cell type-specificity of expressed *Alu*s. Shown is the heatmap of expression-positive *Alu*s from the indicated cell lines, sorted on the basis of cell-line specific expression, displaying ubiquitously expressed/non-specific *Alu*s (left), and tissue-specific *Alu*s (right).

the Pol III machinery [25-27]. While the first kind of studies give us clear evidence of transcription but are difficult to carry out on a genome-wide scale (low throughput approach), the latter ones do not provide direct evidence of transcription since the binding of the Pol III transcription machinery does not necessarily correlate with transcriptional activity [28].

Most *Alu* RNAs have several distinctive features such as differences in the body sequence due to accumulation of mutations, length and sequence differences in the terminal poly-A tail, and unique 3' trailer sequence located between the 3' end of *Alu* conserved body and the closest downstream Pol III terminator [1]. In a recently published study, we exploited these sequence hallmarks (distinguishing most individual Alus from each other), together with the increasing availability of transcriptomic data deriving from next-generation sequencing experiments, to overcome limitations of previous approaches, by devising a bioinformatic pipeline which uses RNA-seq data to identify Pol III *Alu* transcripts and the corresponding source elements of transcription [29]. Optimal performance of this pipeline requires long (≥75 nt) paired-end reads because, even if most *Alu* RNAs are unique throughout their entire sequences, shorter reads from RNA-seq experiments often fail to align uniquely to the human genome due to the repetitive nature of these elements.

Once the reads have been aligned to the human reference genome (hg19), we count the number of uniquely mapped paired-end reads in each annotated *Alu* element, retaining only those with more than 10 reads count uniquely mapped. To filter out passenger *Alu* RNA embedded within Pol II transcripts, we apply a filter based on the level of spurious expression signal upstream and downstream the *Alu* element (Figure 1).

By applying the *Alu* RNA search pipeline to a number of ENCODE RNA-seq datasets (CSHL Poly(A)+ and PolyA(-) long RNA-seq datasets of 7 cell lines) [30] we found 1295 unique putative *Alu* transcripts of which approximately 9% expressed in at least 3 cell lines, while ~75% of them turned out to be cell type-specific [29]. Thus, despite the tiny fraction (0,01%) of *Alu* loci found to be expressed among the ~1.1 million *Alu*s in the human genome, these results suggest that Pol III-transcribed Alu RNAs derive from a small subset (~100) of ubiquitously expressed Alu elements, and from a larger subset (~1000) that tends to vary by cell type, state, growth conditions. This observation is in agreement with the results of recent human transcriptome analyses, showing a marked cell line specificity as the main feature of RNAs transcribed from repeated regions, including LINEs and SINEs [30]. A heatmap visualization of cell lineage-specific *Alu* expression is reported in Figure 2.

The ability to profile *Alu* expression at single-locus resolution opens novel interesting possibilities. As reminded above, a remarkable feature of *Alu* RNA profiles is that they most likely reflects the operation of epigenetic switches at the corresponding genomic loci. *Alu* expression profiles thus have the potential to directly and precisely reveal altered epigenomic states that might accompany pathological states, such as malignancies. As an example of this possibility, demethylation (and thus likely derepression) at an *Alu* locus was shown to activate an associated miRNA locus producing a miRNA acting as an oncogene endogenous silencer [31]. *Alu* expression profiles at single-locus resolution might lead to identify regions of particularly permissive chromatin whose resident genes might be deregulated in concomitance with *Alu* deregulation. More generally, *Alu* RNA profiles might represent a novel type of highly specific molecular signature for cancer and other diseases.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

We thank Barbara Montanini for help with Figure 2 preparation. D.C. was partially supported by a fellowship from Fondazione Cariparma.

References

1. Deininger P. Alu elements: know the SINEs. Genome Biol 2011; 12:236.
2. Cordaux R, Hedges DJ, Herke SW, Batzer MA. Estimating the retrotransposition rate of human Alu elements. Gene 2006; 373:134-137.
3. Cordaux R, Batzer MA. The impact of retrotransposons on human genome evolution. Nat Rev Genet 2009; 10:691-703.
4. Dieci G, Conti A, Pagano A, Carnevali D. Identification of RNA polymerase III-transcribed genes in eukaryotic genomes. Biochim Biophys Acta 2013; 1829:296-305.
5. Orioli A, Pascali C, Quartararo J, Diebel KW, Praz V, Romascano D, *et al*. Widespread occurrence of non-canonical transcription termination by human RNA polymerase III. Nucleic Acids Res 2011; 39:5499-5512.
6. Orioli A, Pascali C, Pagano A, Teichmann M, Dieci G. RNA polymerase III transcription control elements: themes and variations. Gene 2012; 493:185-194.
7. Mandal AK, Pandey R, Jha V, Mukerji M. Transcriptome-wide expansion of non-coding regulatory switches: evidence from co-occurrence of Alu exonization, antisense and editing. Nucleic Acids Res 2013; 41:2121-2137.
8. Makalowski W. Genomic scrap yard: how genomes utilize all that junk. Gene 2000; 259:61-67.
9. Hasler J, Strub K. Alu elements as regulators of gene expression. Nucleic Acids Res 2006; 34:5491-5497.
10. Pandey R, Mandal AK, Jha V, Mukerji M. Heat shock factor binding in Alu repeats expands its involvement in stress through an antisense mechanism. Genome Biol 2011; 12:R117.
11. Ponicsan SL, Kugel JF, Goodrich JA. Genomic gems: SINE RNAs regulate mRNA production. Curr Opin Genet Dev 2010; 20:149-155.
12. Berger A, Strub K. Multiple Roles of Alu-Related Noncoding RNAs. Prog Mol Subcell Biol 2011; 51:119-146.
13. Ivanova E, Berger A, Scherrer A, Alkalaeva E, Strub K. Alu RNA regulates the cellular pool of active ribosomes by targeted delivery of SRP9/14 to 40S subunits. Nucleic Acids Res 2015.
14. Gong C, Maquat LE. lncRNAs transactivate STAU1-mediated mRNA decay by duplexing with 3' UTRs via Alu elements. Nature 2011; 470:284-288.
15. Holdt LM, Hoffmann S, Sass K, Langenberger D, Scholz M, Krohn K, *et al*. Alu elements in ANRIL non-coding RNA at chromosome 9p21 modulate atherogenic cell functions through trans-regulation of gene networks. PLoS Genet 2013; 9:e1003588.
16. Dieci G, Bosio MC, Fermi B, Ferrari R. Transcription reinitiation by RNA polymerase III. Biochim Biophys Acta 2013; 1829:331-341.
17. Oler AJ, Traina-Dorge S, Derbes RS, Canella D, Cairns BR, Roy-Engel AM. Alu expression in human cell lines and their retrotranspositional potential. Mob DNA 2012; 3:11.
18. Varshney D, Vavrova-Anderson J, Oler AJ, Cowling VH, Cairns BR, Whire RJ. SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. Nat Commun 2015; 6:6569.
19. Ichiyanagi K. Epigenetic regulation of transcription and possible functions of mammalian short interspersed elements, SINEs. Genes Genet Syst 2013; 88:19-29.
20. Panning B, Smiley JR. Activation of RNA polymerase III transcription of human Alu repetitive elements by adenovirus type 5: requirement for the E1b 58-kilodalton protein and the products of E4 open reading frames 3 and 6. Mol Cell Biol 1993; 13:3231-3244.
21. 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.
22. Hu Q, Tanasa B, Trabucchi M, Li W, Zhang J, Ohgi KA, *et al*. DICER- and AGO3-dependent generation of retinoic acid-induced DR2 Alu RNAs regulates human stem cell proliferation. Nat Struct Mol Biol 2012; 19:1168-1175.
23. Agarwal P, Enroth S, Teichmann M, Wiklund HJ, Smit A, Westermark B, *et al*. Growth signals employ CGGBP1 to suppress transcription of Alu-SINEs. Cell Cycle 2014: [epub ahead of print].
24. Grover D, Majumder PP, C BR, Brahmachari SK, Mukerji M. Nonrandom 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.
25. 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.
26. Canella D, Praz V, Reina JH, Cousin P, Hernandez N. Defining the RNA polymerase III transcriptome: Genome-wide localization of the RNA polymerase III transcription machinery in human cells. Genome Res 2010; 20:710-721.
27. Oler AJ, Alla RK, Roberts DN, Wong A, Hollenhorst PC, Chandler KJ, *et al*. Human RNA polymerase III transcriptomes and relationships to Pol II promoter chromatin and enhancer-binding factors. Nat Struct Mol Biol 2010; 17:620-628.
28. Guffanti E, Percudani R, Harismendy O, Soutourina J, Werner M, Iacovella MG, *et al*. Nucleosome depletion activates poised RNA polymerase III at unconventional transcription sites in Saccharomyces cerevisiae. J Biol Chem 2006; 281:29155-29164.
29. Conti A, Carnevali D, Bollati V, Fustinoni S, Pellegrini M, Dieci G. Identification of RNA polymerase III-transcribed Alu loci by computational screening of RNA-Seq data. Nucleic Acids Res 2015; 43:817-835.
30. Djebali S, Davis CA, Merkel A, Dobin A, Lassmann T, Mortazavi A, *et al*. Landscape of transcription in human cells. Nature 2012; 489:101-108.
31. Saito Y, Suzuki H, Tsugawa H, Nakagawa I, Matsuzaki J, Kanai Y, *et al*. Chromatin remodeling at Alu repeats by epigenetic treatment activates silenced microRNA-512-5p with downregulation of Mcl-1 in human gastric cancer cells. Oncogene 2009; 28:2738-2744.