Genomic Distribution of Alus and Their Impact on Gene Expression

Andrea Tanzer[1], Michael T. Wolfinger[1], Stefan Badelt[1], Mansoureh Tajaddod[5], Konstantin Licht[5], Ivo L. Hofacker[1,2,3] and Michael F. Jantsch[4,5]

[1]Department of Theoretical Chemistry, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria [2]Bioinformatics and Computational Biology Research Group, University of Vienna, Währingerstraße 17, A-1090 Wien, Austria
[3]Center for RNA in Technology and Health, Univ. Copenhagen, Grønnegårdsvej 3, Frederiksberg C, Denmark
[4]Department for Cell- and Developmental Biology, Medical University of Vienna, Schwarzspanierstr. 17 A-1090 Wien, Austria, [5]Max F. Perutz Laboratories, Vienna Biocenter (VBC), Dr. Bohr-Gasse 9, 1030 Vienna, Austria

at@tbi.univie.ac.at http://www.tbi.univie.ac.at

Abstract

In this study we analyzed the **genomic distribution of Alu elements** in the human genome and their **impact on transcript abundance** and classified transcripts based on relative orientation of Alus.

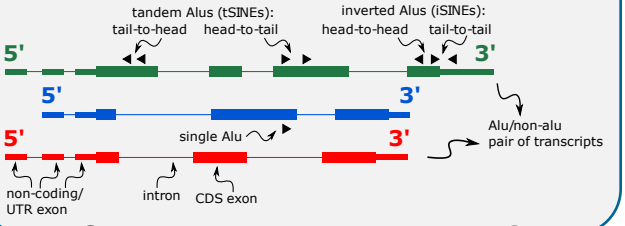
As a first step towards functional analysis, we performed **in-silico simulations** of **co-transcriptional folding** behavior in order to identify kinetic traps, which might serve as potential target sites triggering RNA degradation.

Toolbox

Data:
- human genome assembly: hg19/GRCh37
- gene annotation: GENCODE version 19 [1]
- RNAseq data: ENCODE project (phase2) [2], 15 celllines, cell, PolyA+, long RNA faction
- Alu annotation: UCSC genome browser, Repeatmasker track

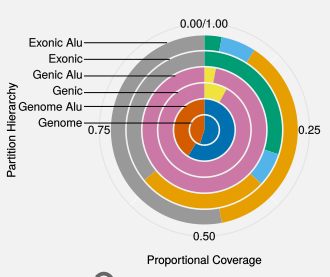
Tools:
- RNAseq mapper: segemehl [3]
- transcript quantification: cufflinks2
- annotation comparison: bedtools2
- RNA kinetic folding simulation: Kinwalker [4]
- RNA structure plots: Vienna RNA package [5]

Location and Orientation

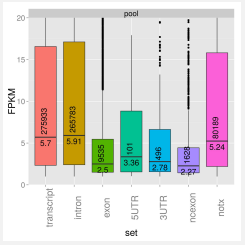


We classified Alus based on their position within transcripts (**intergenic, transcript, intron, exon, CDS, 5UTR, 3UTR, non-coding exon**) and their orientation relative to neighbouring Alus (**single Alus, head-to-head, tail-to-tail, direct tandems, no Alu**). Only long transcripts were used (gene type lncRNA, protein coding, pseudogene), small RNAs and rRNAs were not considered. Given that a gene locus consists on average of 3.8 transcript isoforms, we **projected individual elements** onto the genome in a hierarchical rule set and used these projections for genome

Alus in genomic elements



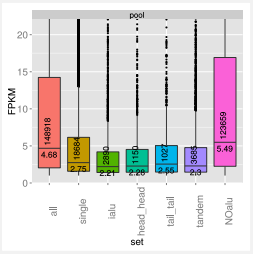
Transcript abundance



1.2 Mio. Alu elements make up 10.76% (311,730,074 bp) of the human genome. They are almost evenly distributed between genic and intergenic regions. Within genic regions, **introns** are **enriched** compared to exons. **Coding regions** make up 30% of the exonic partition, but **only 4% are covered** by Alus. In contrast, **non-coding exons and 3'-UTRs** are **enriched** by a factor of 1.4.

Transcripts containing **Alus in exons** are significantly **lower expressed** than those with intronic Alus or Alu-free transcripts, even when all isoforms are from one gene locus and thus presumably under the control of the same

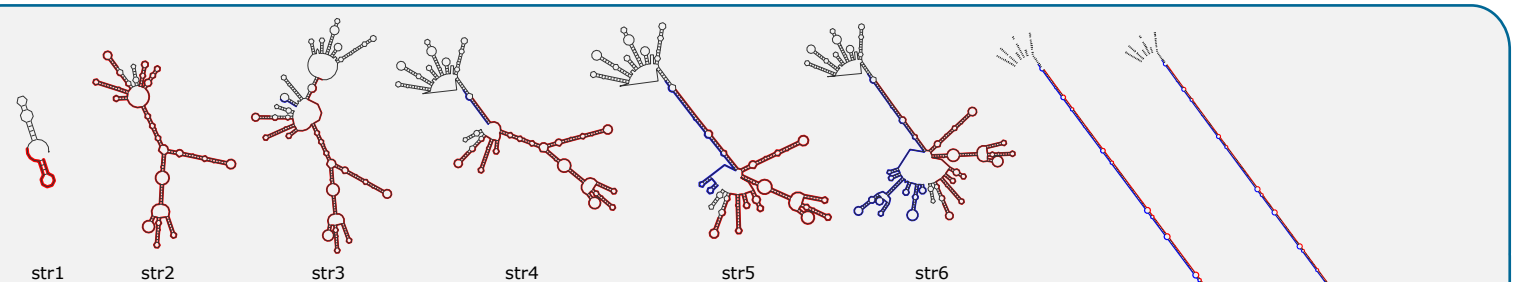
Alu-orientation in exons



single >
tandem >>> || <<<
iAlu >>>< || <<<>>>
head-head <<<>>>
tail-tail >>><<<

Alus orientation and RNA abundance

At $\alpha = 0.05$ level, we find that transcripts with **single Alus** are significantly **higher expressed** than transcripts containing **iAlus** (p-value 4.51e-35). The same holds for transcripts with single vs. **tandem** arrangements (p-value 1.77e-22). Transcripts with **iAlus** show **lower expression** than those with Alus in **tandem** arrangement (p-value 1.98e-3). Within iAlu-containing transcripts, the ones with Alus in **head-to-head** orientation are **lower expressed** than transcripts with Alus in **tail-to-tail** orientation (p-value 2.47e-2).



Co-transcriptional folding

Inverted Alus have the strongest effect on RNA abundance. In-silico simulations of co-transcriptional folding behavior of a 3'-UTR with **iSines** (tail-to-tail) show that the two Alus form a **long helical region,** even though they belong to different Alu families.

Shortly after transcription of Alu2 (blue) starts (str3), the transient structure of Alu1 (red) starts to open and basepair with Alu2 (step 4). The **energy barrier** of 7.1 kcal/mol seems to be just low enough to allow the duplex formation, but is sufficiently high to **stabilize the conformation** and thus support co-transcriptional helix extension.

Once transcription of Alu2 is completed, another refolding event (str7 to str8) with a rather high energy barrier stabilizes the **final helix**.

Experimental data of different Alu-constructs confirm that stable transient structures formed by Alus reduce transcript abundance.

Summary

We show that
- Alus are depleted in coding regions
- but are enriched in 3'UTRs and non-coding exons
- the location of Alus within these elements reduces mRNA abundance
- head-to-head conformation of inverted Alus has the strongest effect
- co-transcriptional folding of iSines leads to stable long helical regions

Our results suggest that Alu elements shape the transcriptomics landscape of human cells. Details of the underlying mechanism(s), however, remain to be determined.

Acknowledgments

This study is joint work of the University of Vienna, the Max F. Perutz Laboratories and the Medical University of Vienna and is financially supported by the FWF grant "SFB RNA regulation of the transcriptome", F 43. AT performed the genomic analysis, AT and MTW the transcriptomics analysis, AT and SB did folding simulations, MT and KL designed and validated constructs and MFJ and ILH supervised the study.

References

[1] "GENCODE: The reference human genome annotation for the ENCODE project", Harrow et al., Genome Res., 22, 1760-1774 (2012)
[2] "Landscape of transcription in human cells", Djebali et. al, Nature, 489, 101-8 (2012)
[3] "A multi-split mapping algorithm for circular RNA, splicing, trans-splicing, and fusion detection", Hoffmann et al., Genome Biol. 15, R34 (2014)
[4] "Folding kinetics of large RNAs", Geiss et al. , J. Mol. Biol. 379, 160-73 (2008)
[5] "ViennaRNA Package 2.0", R. Lorenz et al., Algorithms in Molecular Biology, vol. 6, p. 26 (2011)