

LIMEs are homologous to RNA genes

Alisa Bugrova*, Arseniy Gubler*, Arseniy Pelevin*,

Elizaveta Chernova, Oleg Demianchenko, Anastasia Kuznetsova, Veronika Katrukha, Emma Rodionova, Lada Isakova, Tatiana Lyovina, Alexandr Sverdlin, Ariadna Semenova, Artem Fedorenko, Anastasia Lyulina, Ekaterina Nuzhdina, Dmitry Ivankov, Dmitry Korkin, Fyodor Kondrashov



Summary

Problem

- Биологическая природа длинных консервативных элементов (LIME) не изучена. Мы проверяли гипотезу, что LIME могут кодировать какой-нибудь тип РНК.
- Biological nature of long conservative elements (LIME) is unknown. We examined a hypothesis that LIMEs can code some types of RNA molecules.

Results

- Мы выяснили, что последовательности всех LIME перекрываются с последовательностями некоторых известных типов РНК. В том числе 21 последовательностей LIME совпадают с tRNA, из которых 5 также входят в lncRNA, 3 с rRNA и 1 с U6 snRNA. Для LIME, входящего в U6, мы выяснили положение в сплайсеосоме, где он комплементарно связывается с U4.
- We found that all LIME sequences corresponded to known types of RNA, including 21 LIME sequences matched different tRNAs, 5 of which are also a part of lncRNA, 3 matched rRNA and 1 matched U6 snRNA. We found the U6 LIME's location in the spliceosome, where it is complementary bound to U4 snRNA.

Background

- LIME, или длинные идентичные межвидовые элементы - последовательности длиной >100 нуклеотидов, встречающиеся у разных представителей тетрапод (Рис.1, [1]). В нашей лаборатории мы исследовали древние LIME, которые более 50 нуклеотидов в длину и встречаются без единой нуклеотидной замены за пределами целостноротых (а некоторые даже у грибов). Причина подобной экстремальной консервативности LIME неизвестна. LIME не кодируют белок, поэтому мы исследовали возможность, что LIME кодируют РНК. Мы проверяли это гипотезу предсказывая вторичную структуру последовательностей LIME и сравнивая их с другими аннотированными геномными элементами. Некоторые LIME уже были аннотированы некоторые РНК, что поддерживало нашу гипотезу.

Multiple alignment of the conserved LIME sequence (red) in the context of the less conserved background (blue). A multiple alignment of the conserved LIME sequence (red) in the context of the less conserved background (blue).

Рис 1/ Fig. 1 Множественное выравнивание консервативного участка (LIME - красным) в контексте менее консервативной последовательности (показано синем). A multiple alignment of the conserved LIME sequence (red) in the context of the less conserved background (blue).

- LIME (Long Identical Multispecies Elements) are sequences at least 100 nucleotides long that maintain 100% sequence identity in different tetrapods. We worked with ancient LIMEs, which are more than 50 nucleotides long and identical among vertebrates (some are found even in fungi). The reasons for the extreme conservation remain unknown. LIMEs are known not to be elements of protein-coding genes. Thus, we explored the possibility that LIMEs are elements of RNA-coding genes. We tested our hypothesis by predicting the secondary structure of LIME sequences, and by comparing LIMES to other annotated genomic elements. Some ancient LIMEs were already annotated as RNAs, supporting our hypothesis.

Methods

Ancient LIMEs
Exons 115
Introns 515
Nongenic 1118

REC
Exons 31
Introns 47
Nongenic 96

Unique LIME sequences
Total 25

LIME id	Chr	Start	Len	LimeSequence
121	19	893506	58	ACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
122	19	893507	57	CTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
123	19	893508	56	TAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
124	19	893509	55	AAAATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
125	19	893510	54	AATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
126	19	893511	53	AATGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
127	19	893512	52	ATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
129	19	893514	50	TGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC
115	19	1021539	64	ACATATACTAAAATTGGAACGATACAGAGAAGATTAGCATGGCCCTCGCGCAAGGATGACACGC

Мы рассматривали отдельно три группы LIME: экзонные, интронные и межгенные. Мы объединили все древние LIME в соответствующие регионы экстремальной консервативности (REC). Далее мы выбрали из них только уникальные последовательности и в основном работали с ними.

Мы использовали BLAST [5] для выравнивания и RNAfold из Vienna Package [6] для предсказания вторичной структуры РНК. Мы использовали следующие базы данных для поиска похожих последовательностей в базе данных LIME: tRNA [7], lncRNA [8], rRNA и snRNA [9].

We studied three groups of LIMEs: exonic, intronic and nongenic. We combined all of the ancient LIMEs into regions of extreme conservation (REC) and then selected unique sequences from RECs and worked mostly with them.

We used BLAST [5] for sequence alignment and the RNAfold program of Vienna Package [6] for predicting the secondary structure of RNAs corresponding to the LIME sequences. We used several available database of various RNAs to search for similar LIME sequences: tRNA [7], lncRNA [8], rRNA and snRNA [9].

Results

- Последовательности всех LIME совпадают с последовательностями разных RNA (tRNA, lncRNA, snRNA, rRNA).
- Среди LIME не найдены последовательности, совпадающие с miRNA.
- Одна последовательность LIME совпадает с U6 snRNA, которая в структуре сплайсеосомы частично комплементарна U4 snRNA (Рис. 2) и располагается в центральной части её структуры (Рис. 3). Все копии этого LIME лежат в аннотированных экзонах или в интронах на границе с экзонами (Рис. 4). U6 snRNA действительно транскрибируется во всех локусах этих LIME, кроме одного, лежащего одновременно в интроне и в экзоне.
- 3 последовательности LIME совпадают с рибосомальной РНК.
- Большинство последовательностей LIME (21) совпадают по последовательности с тРНК (Рис. 5), либо с тРНК-подобными элементами, лежащим внутри одной из пяти найденных lncRNA (Рис. 6).
- Только два класса lncRNA с тРНК-подобной структурой были описаны ранее [2,3], и обнаруженные нами 5 РНК этого типа не имеют значимого сходства с ними по последовательности.

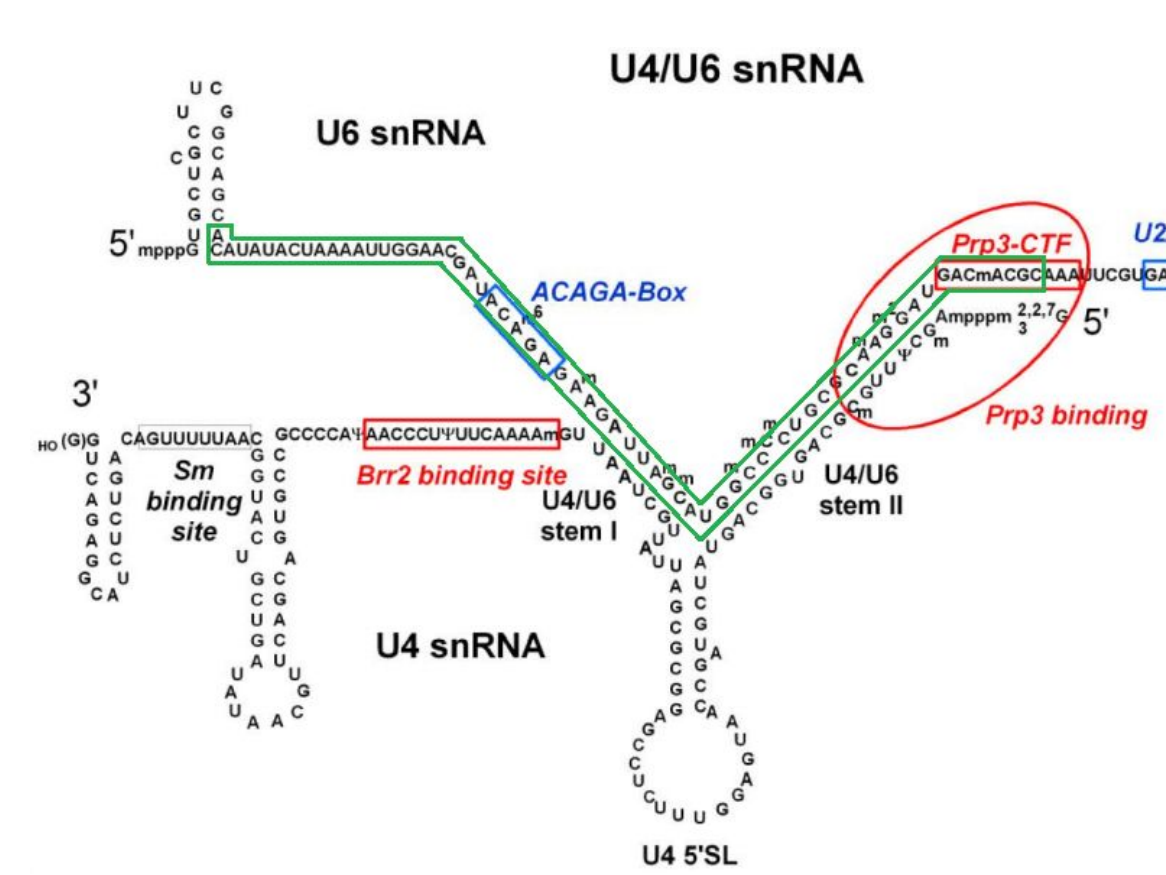


Рис. 2/ Fig. 2 Вторичная структура комплекса u4/u6 snRNA. LIME выделен зелёным. Secondary structure of the u4/u6 snRNA complex. LIME sequence is boxed in green. Структура из [4].

Рис 3/ Fig. 3 Третичная структура u4/u5/u6 сплайсеосомы с u5 (красная) u4 (зеленая) и LIME внутри u6 (синее) и u6 без LIME (фиолетовое). Crystal structure of the u4/u5/u6 spliceosome with u5 (red), u4 (green) and us with the LIME sequence (blue) and non-LIME u6 (violet). Structure from [4].

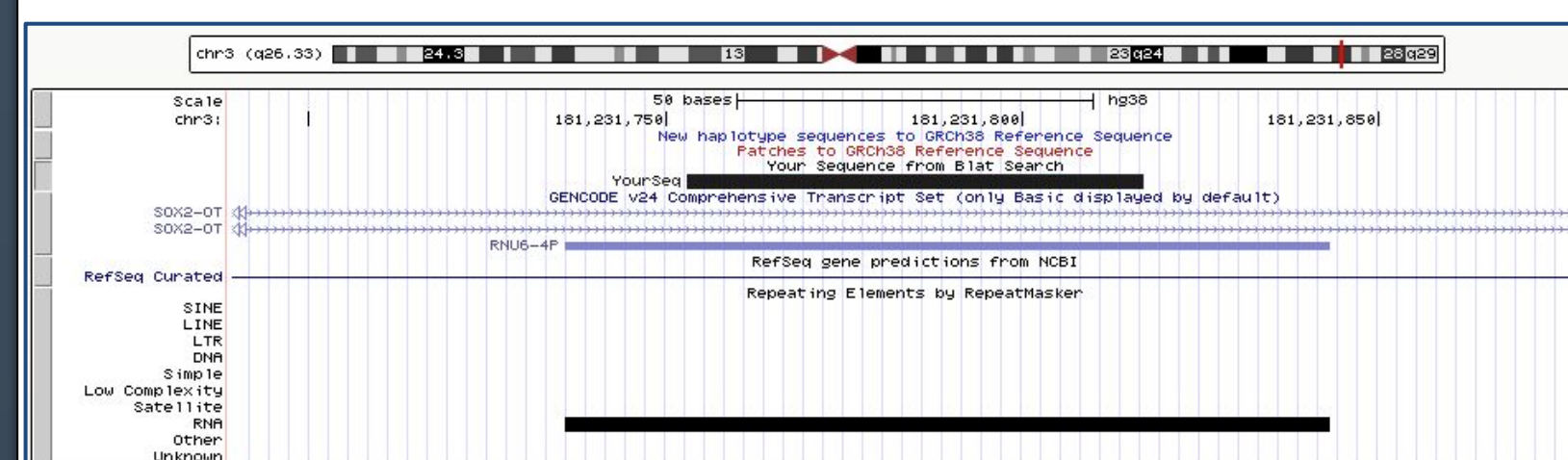
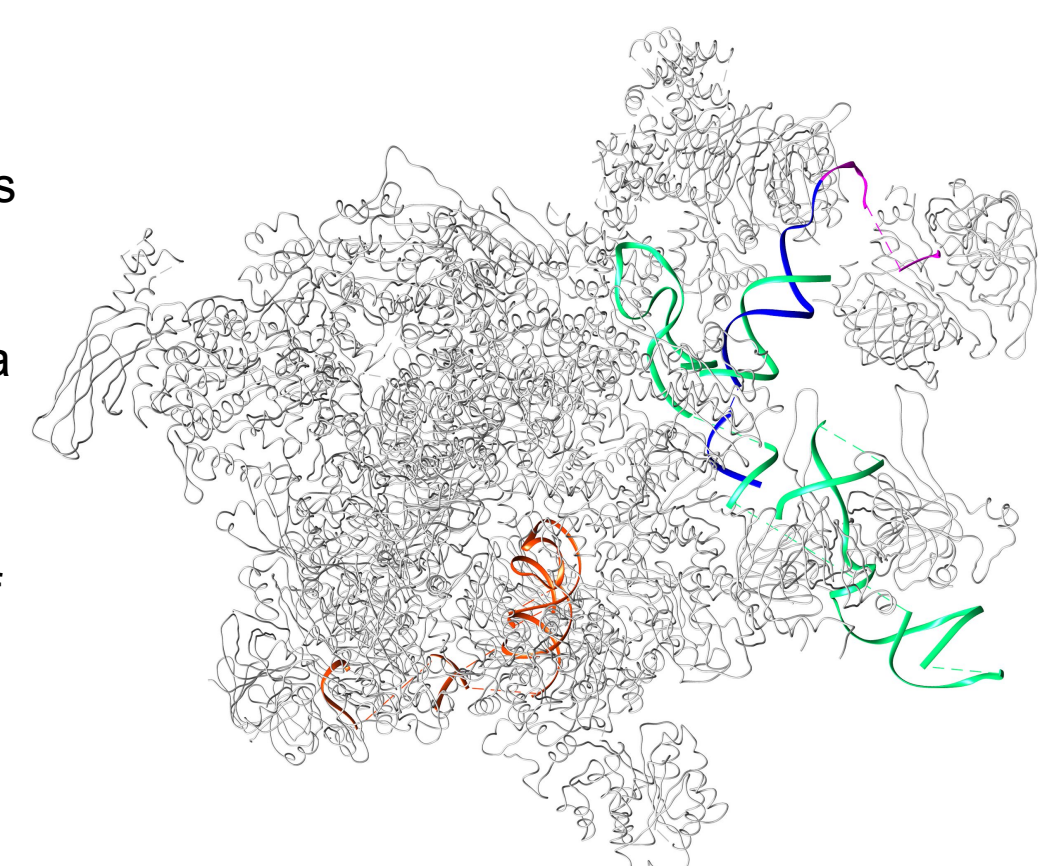


Рис. 4/ Fig. 4 Пример LIME совпадающего с u6 snRNA но при этом находящегося в интроне другого гена. An example of a LIME that coincides with u6 snRNA that is located in the intron of another gene.

Score	Expect	Identities	Gaps	Strand
134 bits(72)	2e-37	72/72(100%)	0/72(0%)	Plus/Plus
Query 1	TCCTGGTGTAGTGGTGGGCTCTCACCGCCGGCCGGGTTTCGATTC	60		
Sbjct 1	TCCTGGTGTAGTGGTGGGCTCTCACCGCCGGCCGGGTTTCGATTC	60		
Query 61	CCGGTCAAGGAAA 72			
Sbjct 61	CCGGTCAAGGAAA 72			

Рис 5/ Fig. 5 Выравнивание LIME с тРНК. An alignment of a LIME and a tRNA sequence.

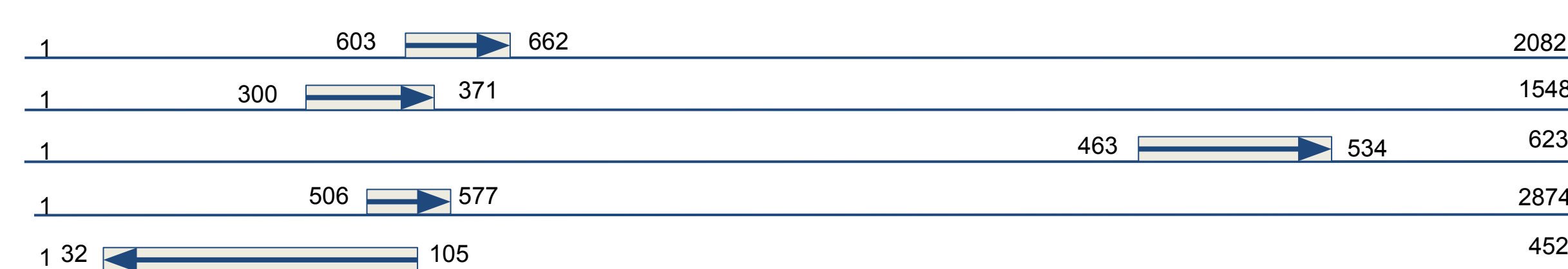


Рис 6/ Fig. 6 Локализация тРНК-подобных LIME в последовательности пяти lncRNA. The location of tRNA-line LIMEs in five mature lncRNA sequences.

- All LIME sequences had high sequence identity with known types of RNA molecules (tRNA, lncRNA, snRNA, rRNA)
- No LIME sequences were similar to any of the known miRNA sequences
- One LIME sequence was identical to U6 snRNA, which is partially complementary to the U4 snRNA (Fig. 2) and located in a central part of the structure (Fig. 3). All the copies of this LIME are located in annotated exons or in introns at the border with exons (Fig. 4). U6 snRNA is transcribed from all the locuses of these LIMEs, except one, located at both exon and intron at the same time.
- 3 LIME sequences are parts of ribosomal RNA.
- Most of the LIME sequences (21) were identical to tRNAs (Fig. 5) or to the tRNA-like elements, located in one of five different lncRNAs (Fig. 6).
- Only two classes of lncRNA genes containing RNA-like structures were described previously [2,3]. The five lncRNAs we described do not show any sequence identity with the two previously described cases.

Conclusions

- Поскольку все LIME соответствуют некодирующей РНК, можно предположить, что их экстремальная консервативность связана с высоким функциональным значением, в том числе с формированием комплементарных последовательностей внутри молекулы РНК или между молекулами.
- Возможно, мы описали новые функциональные формы lncRNA с тРНК-подобными участками.
- All LIMEs correspond to non-coding RNA, thus, we hypothesize that their extreme conservation is related to their function, such as maintaining complementary bonds with the molecule or between molecules.
- The five instances of lncRNA genes containing tRNA-like structures likely represent genes with novel functions.

Future Directions

- Найти положение рибосомальных LIME в структуре рибосомы, и может быть объяснить их консервативность.
- Проверить, связана ли частота замен с положением шпикел и петель вторичной структуры РНК.
- Сравнить LIME с последовательностями транслозонов в геноме человека.
- Изучить более молодые LIME длиной более 100 нуклеотидов для поиска общих закономерностей.
- To find the location of ribosome LIMEs in the ribosome structure, which may explain it's conservation.
- Examine the rate of substitution in loops and hairpins in RNA secondary structure.
- Align LIMEs to the sequences of transposons in human genome.
- To study more recent LIMEs >100 nucleotides long and attempt to find general patterns.

References

- Reneker J, et al. Long identical multispecies elements in plant and animal genomes. Proc Natl Acad Sci U S A. 2012 109(19):E1183-91.
- Zhang B et al. Identification and Characterization of a Class of MALAT1-like Genomic Loci. Cell Rep. 2017 19(8):1723-1738.
- Plewka P, et al. A stable tRNA-like molecule is generated from the long noncoding RNA GUT15 in Arabidopsis. RNA Biol. 2018 21:1-13
- Dmitry E. Agafonov et al. Science 18 Feb 2016 Molecular architecture of the human U4/U6.U5 tri-snRNP
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990 Oct 5;215(3):403-10.
- Lorenz R, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. 2011 Nov 24;6:26. doi: 10.1186/1748-7188-6-26.
- Chan PP & Lowe TM GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucl. Acids Res. 2016 44 :D184-D189.
- Volders PJ, et al. An update on LNCipedia: a database for annotated human lncRNA sequences. Nucleic Acids Res. 2015 43(8):4363-4
- Pruitt KD, et al. NCBI Reference Sequences (RefSeq): current status, new features and genome annotation policy. Nucleic Acids Res. 2012 40: D130-5.