Семестральный проект

Abstract

\* ЛТР ретротранспозоны

-определение

-в каких целях используется его местонахождение/частота в геноме

- способы нахождения(de novo, reference)

плюсы и минусы поиска dé novo для чего используется именно он

- нахождение в геноме растений и в геноме человека

\* алгоритмы поиска повторяющихся строк( наиболее длительный процесс в алгоритме поиска LTR)

- бинарный поиск с суффикс и лцп массивами

- оригинальный поиск python(посмотреть алгоритм str.index() и if str in str)

- Knut wth algorithm

Объяснение почему выбран именно оригинальный поиск( написан на C)

Сравнение скорости работы

\* алгоритм объединения групп строк в LTR

- как именно работает

\* результаты и оптимальные параметры поиска

- gff soubor( добавить процентное сходство LTR)

- сравнение с существующей LTR базой

- выводы о эффективности алгоритма

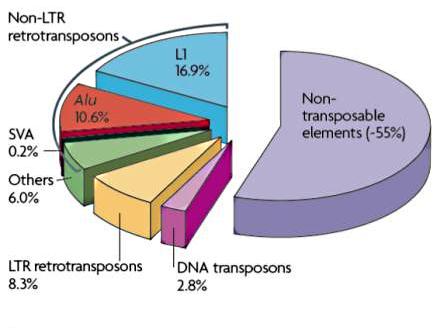
\* заключение

- результаты поиска

\* использованные toolbox python и их предназначение

LTR Retrotransposons

Transposable element (TEs), also known as transposons or "jumping genes", are discrete pieces of DNA sequence that can move in the genome from one location to another. Transposons represent one of types of mobile genetic elements. TEs are allocated to one of two classes, depending on their mechanism of transposition. The first class is retrotransposons, which are copied in two stages: first they are transcribed from DNA to RNA, and then RNA produced is reverse transcribed to DNA. This DNA copy is then inserted at a new position into the genome.



Retrotransposons usually consist of three sub-types:

* LINEs(L1): encode reverse transcriptase, and are transcribed by RNA polymerase II
* SINEs(Alu): transcribed by RNA polymerase III
* LTRs(TEs with long terminal repeats): encode reverse transcriptase, similar to retroviruses

LTR ретротранспозоны окружены long terminal repeats которые содержат все нужные для регуляции транскрипции элементы. The autonomous elements содержат *gag* and *pol* genes, которые кодируют reverse transcriptase and protease.

LTR ретротранспозоны в свою очередь делятся на три подкласса:

* Ty1-copia-like ([Pseudoviridae](http://en.wikipedia.org/wiki/Pseudoviridae" \o "Pseudoviridae))
* Ty3-gypsy-like ([Metaviridae](http://en.wikipedia.org/wiki/Metaviridae" \o "Metaviridae))
* BEL-Pao-like

Ретровирусы могут трансформироваться в LTR ретротранспозоны путем инактивации или удалением структур ответственных за внеклеточную мобильность. Если ретровирус заражает и впоследствии встраивает себя в геном in germ line cells, то может стать an Endogenous Retrovirus (ERV). Therefore Exogenous retroviruses возникли из endogenous retrotransposons приобретением cellular *envelope* gene [3]

Human LTR elements are endogenous retroviruses which account for ~8% of the human genome. [1]

In general, most (85%) of the LTR retrotransposon-derived parts consist only of an isolated LTR, with the internal sequence having been lost by homologous recombination between the flanking LTRs.[2]

Влияние LTR на клеточную gene expression человека

Более 25 экспериментально охарактеризованных клеточных генов показывают  LTR-mediated эволюционные изменения, в которых встроенные LTRs являются  alternative promoters для предоставления новой tissue-specificity, play as the major promoters, or promotes only minor effects. [4]. For example, A HERV-K(HML-5) LTR plays as the major promoter of INSL4, a insulin-like growth factor gene expressed in placenta. [5]. A HERV-E family LTR plays as an alternative tissue-specific promoter of the endothelin B receptor (EDNRB) gene, by which the gene expression increased ∼15% in placenta. [6]. LTR-derived promoters often increase placenta-specific gene expression, несмотря на то что в общем эффект от LTR insertions проявляется умеренно в многих случаях.

Последние исследования показали, что HERV-encoded peptide as a tumor-specific antigen участвует in the hematopoietic stem cell transplantation  for the therapy of renal cell carcinoma (RCC) [7]

**Identification of LTRs that Contribute to Tumor Immunity and Oncogenesis**

A pioneering study identified a HERV-encoded peptide as a tumor-specific antigen that works in the successful hematopoietic stem cell transplantation for the therapy of metastatic renal cell carcinoma (RCC) ([89](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B89)). The tumor antigen, CT-RCC-1, recognized by RCC-specific CD8+ T cells is encoded by novel spliced variants of the HERV-E transcript from chromosome 6q. Furthermore, this antigen expression results from hypomethylation of the LTR sequences, as well as hypoxia-inducible transcription factor (HIF-2α) binding to the HIF responsive element (HRE) in the LTR ([90](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B90)). HIF-2α is stabilized in many cases of RCC due to the von Hippel–Lindau (VHL) tumor suppressor gene inactivation. This whole scenario enlightens the relevance of HERVs in cancer biology and treatment.

A study on tumorigenesis of Hodgkin’s lymphoma provided the first and so far only evidence that endogenous LTR activation can be oncogenic ([91](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B91)). These malignant cells strongly express CSF1R, the colony stimulating factor 1 receptor gene, to allow CSF1/CSF1R-dependent cell growth. Intriguingly, a LTR of THE1B (a MaLR family LTR retrotransposon) is located 6.2kb upstream of the CSF1R coding sequences, and is activated by the loss of CpG methylation to drive the aberrant CSF1R gene expression from the initiation site within the LTR. The regulatory region contains Sp1, GATA, AP-1, and NF-κB binding motifs. Anaplastic large cell lymphomas also have the same CSF1R transcript. Thus, hypomethylation of the THE1B LTR causes the CSF1R oncogene activation, reminiscent of “insertional mutagenesis” by exogenous retroviruses.

**Possible Contribution of HERVs to Autoimmune Diseases**

Human endogenous retroviruses have also been implicated in autoimmune diseases including rheumatoid arthritis (RA), multiple sclerosis (MS), and systemic lupus erythematosus (SLE). Expression of multiple ERVs was detected in RA patients ([92](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B92)), and that of HERV-K10 in juvenile RA ([93](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B93)). A recent report indicated a strong linkage between the MS-associated retroviral element (MRSV)-type HERV-W and MS: the env RNA and protein expression significantly increases in the blood and brain cells of MS patients ([94](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B94)). HERV-H/F family HERV-Fc1 expression in T lymphocytes and the plasma HERV-Fc1 RNA level also increase in patients with MS ([95](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B95)). HERV-K18 (chromosome 1) was found to be a risk factor of MS ([96](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B96)) and other autoimmune disorders ([97](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B97), [98](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B98)). Expression of HERV-W, K, E, and a newly identified ERV-9/HERV-W variant in psoriasis was also observed ([99](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B99)).

Three mechanisms have been proposed for HERV-related pathology of autoimmune diseases: molecular mimicry, superantigen production, and LTR-mediated alterations of gene expression ([100](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B100)). The Gag and Env proteins of HERV-K10 share the peptide sequences with the rheumatoid factor epitopes on IgG1Fc, implying molecular mimicry ([101](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B101)). The env genes of defective HERV-K alleles at 1q21.2–q22 encode superantigens, and are transcriptionally activated by the interferon α signaling to cause polyclonal T-cell activation ([102](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B102), [103](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B103)). There was an interesting hypothesis that solitary LTRs in the 5′-flanking region of the HLA-DQB1 gene (encoding major histocompatibility complex, class II, DQ beta 1) influence the susceptibility to autoimmune diseases ([104](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B104)–[106](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B106)). However, these results remain ambiguous, and need to be more critically examined to determine whether or not the LTR/HERV activation is the true cause of the clinical consequences ([101](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B101), [107](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B107)).

Furthermore, the LTR of MaLR family THE1D is activated by transcription factor DUX4 which is specifically expressed in facioscapulohumeral dystrophy, implying involvement of DUX4 in the pathophysiology ([108](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B108)). Hypomethylation of HERV-E and HERV-K was also observed in SLE patients ([109](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/#B109)).

[Go to:](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3769647/)

## Concluding Remarks

Human endogenous retroviruses are remnant forms of infectious retroviruses that integrated into the chromosomal DNA of germ-line cells of human ancestors, increased their copy numbers and have been inherited by present-day humans. None of the HERVs poses an immediate risk as a transposable element. However, solitary LTRs derived from HERVs and MaLRs dominate the provirus forms in the copy numbers, and can serve as redundant enhancer-promoter sequences for nearby cellular genes. When the DNA methylation-mediated suppression system becomes compromised, HERVs and LTRs can cause detrimental and/or self-protecting effects. Two prominent examples of the clinically significant HERV/LTR activation have been reported: CSF1R oncogene activation by a MaLR LTR in Hodgkin’s lymphoma and RCC-specific novel HERV-E antigen expression facilitating the immunotherapy. Future researches in oncology and immunogenetics will unveil more details about the endogenous LTR functions in human pathogenesis.

Searching

The conventional approach to annotating MGEs in genomic sequences is based upon homology searching against a well-updated library of known MGEs, e.g. Repbase [[3](http://www.biomedcentral.com/1471-2164/8/90#B3)], using a fast searching program, e.g. RepeatMasker. This approach, however, is limited to annotating those known MGE families, and thus cannot identify new elements. Furthermore, it sometimes even overlooks known elements, because the repetitive nature of MGE elements may confuse the statistical methods (e.g. E-values) that are commonly used in genome annotation

## References

[1] <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2884099/>

[2] <http://www.nature.com/nature/journal/v409/n6822/full/409860a0.html>

[3] Malik, H. S., Henikoff, S. & Eickbush, T. H. Poised for contagion: evolutionary origins of the infectious abilities of invertebrate retroviruses.*Genome Res.* **10**, 1307–1318

[4] Cohen CJ, Lock WM, Mager DL. Endogenous retroviral LTRs as promoters for human genes: a critical assessment. Gene (2009) 448:105–1410.1016/j.gene.2009.06.020

[5] Bièche I, Laurent A, Laurendeau I, Duret L, Giovangrandi Y, Frendo J-L, et al. Placenta-specific INSL4 expression is mediated by a human endogenous retrovirus element. Biol Reprod (2003) 68:1422–910.1095/biolreprod.102.010322

[6]  Landry J-R, Mager DL. Functional analysis of the endogenous retroviral promoter of the human endothelin B receptor gene. J Virol (2003) 77:7459–6610.1128/JVI.77.13.7459-7466.2003

[7] Takahashi Y, Harashima N, Kajigaya S, Yokoyama H, Cherkasova E, McCoy JP, et al. Regression of human kidney cancer following allogeneic stem cell transplantation is associated with recognition of an HERV-E antigen by T cells. J Clin Invest (2008) 118:1099–10910.1172/JCI34409