



POSTER PRESENTATION

Open Access

Identification and analysis of HML2 sequences in human genome assembly GRCh37/hg19

Marta Cadeddu^{1*}, Laura Vargiu^{1,2}, Patricia Rodriguez-Tomé², Göran O Sperber³, Jonas Blomberg⁴, Enzo Tramontano¹

From *Frontiers of Retrovirology: Complex retroviruses, retroelements and their hosts*
Cambridge, UK. 16-18 September 2013

Background

Human endogenous retroviruses (HERVs) originated from exogenous retroviral infections of the human germ line cells and spread in the human population through vertical transmission over millions of years. Among HERVs, the HML2 proviruses [1] are the most recently integrated and show the most intact proviral genomes. HML-2 expression has tentatively been associated with different pathological conditions, including Hodgkin's lymphoma, melanoma, breast and testicular cancer. A comprehensive recent study identified 91 HML2 proviruses [2].

Material and methods

Human genome (assembly GRCh 37/hg19) was analyzed with RetroTector (ReTe) version 1.01 [3]. ReTe was run on a machine with 4 6-core Xeon processors, 2.66Ghz each, 256 Gb of RAM and 4 Tb of disks, with an estimated execution time of 1-2 days. BLASTN, using HML consensus (Blikstad *et al*, unpublished) and the May 2013 Repeatmasker library, ENSEMBL and MEGA5 were used, in successive steps, for classification and identification of locus position and phylogenetic inference. Time since integration was inferred using a neutral substitution rate between cognate LTRs of 0.2 mutations per million years.

Results

ReTe [2] identified more than 120 HML2 proviruses, many of which were not previously reported, accounting for roughly 0.01% of the total human genome. Among the identified HML2 proviruses more than 50% are ≥ 8000 bp in length and more than 50% have both LTRs. HML2 proviruses bordering to HML1, HML3, HML9 and HML10,

as well as recombinant proviruses containing HML2 sequences were detected. HML2 proviruses were present in all chromosomes and showed to form clusters, particularly in chromosomes 1, 4, 8 and 19. Open reading frames (ORFs) predicted by ReTe revealed that 21 proviruses have at least 1 ORFs in *gag*, *pro*, *pol* and *env* genes, while 6 had ORFs in 3 genes. Age analysis versus reductions of ORFs and proviral length was performed. Phylogenetic analyses were performed with whole element DNA, concatenated Gag, Pro and Pol amino acid sequences, and Pol amino acid sequences.

Conclusions

In an attempt to establish a comprehensive catalog of HML2 proviruses that could set the basis for further research, we detected over 120 HML2 proviruses and performed a first characterization of them.

Authors' details

¹Dept. of Life and Environmental Sciences, Cagliari, Italy. ²Center for Advanced Studies, Research and Development, Sardinia, Italy. ³Physiology Unit, Dept of Neuroscience, Uppsala, Sweden. ⁴Section of Virology, Dept. of Medical Sciences, Uppsala, Sweden.

Published: 19 September 2013

References

1. Andersson ML, Lindeskog M, Medstrand P, Westley B, May F, Blomberg J: Diversity of human endogenous retroviral class II-like sequences. *J Gen Virol* 1999, **80**:255-260.
2. Subramanian RP, Wildschutte JH, Russo C, Coffin JM: Identification, characterization and comparative genomic distribution of the HERV-K (HML2) group of endogenous retroviruses. *Retrovirology* 2011, **8**:90.
3. Sperber GO, Airola T, Jern P, Blomberg J: Automated recognition of retroviral sequences in genomic data-RetroTector. *Nucleic acids research* 2007, **35**.

doi:10.1186/1742-4690-10-S1-P9

Cite this article as: Cadeddu *et al.*: Identification and analysis of HML2 sequences in human genome assembly GRCh37/hg19. *Retrovirology* 2013 **10**(Suppl 1):P9.

¹Dept. of Life and Environmental Sciences, Cagliari, Italy
Full list of author information is available at the end of the article