Massively parallel identification of active LINE-1 promoters in human genomes.

Misaki Matsuo*[1], Larisa Okorokova*[1], Simona Saccani[1], Dominic Van Essen[1], and Gael Cristofari[1]

[1]Institut de Recherche sur le Cancer et le Vieillissement – Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Université Côte d'Azur – Faculté de médecine, 28 avenue de Valombrose 06107 Nice Cedex 2, France

Abstract

Long INterspersed Element-1 (LINE-1 or L1) forms the only autonomously active family of retrotransposons in humans, and represent 17% of the human genome. L1 retrotransposition is initiated by the transcription of full-length integrated copies driven by an internal promoter located in L1 5' untranslated region (UTR). Furthermore, L1 5' UTR also exhibits antisense promoter activity, which can drive the synthesis of chimeric transcripts with nearby genes. Despite the abundance of these highly repetitive and dispersed sequences, only a limited number of loci are transcriptionally active in any given cell-type and eventually mobilized. This can be explained by the various epigenetic mechanisms that target L1 elements and silence them. However, another non-exclusive possibility is that variations in L1 internal sequence may also affect L1 sense or antisense promoter activities. To identify transcription-competent L1 loci, and test how internal sequence variation affect the activity of individual L1 promoters, we employed massively parallel reporter assays (MPRA) to comprehensively test the sense and antisense promoter activities of human-specific and other recent primate-specific L1 elements (L1HS to L1PA5) present in a given human genome (_˜5000 copies). Taking advantage of the natural variation present in L1 promoter sequences, this strategy will reveal regions and motifs playing important roles in L1 and L1-mediated transcription, as well as transcription-competent loci. This work is supported by Agence Nationale de la Recherche (ANR-11-LABX-0028; ANR-15-IDEX-0001; ANR-21-CE12-0001), Région Sud (MobileGenomAI), Inserm (GOLD Cross-cutting Program on Genomic Variability) and CNRS (GDR 3546).

Keywords: LINE1, L1, 5'UTR, promoter activity, MPRA

*Speaker