CLIFinder: A new bioinformatic tool for pangenomic identification of LINE-1 chimeric transcripts in gliomas

Marie-Elisa Pinson*[1], Romain Pogorelcnik[1], Emmanuel Chautard[2], Pierre Verrelle[2], Philippe Arnaud[1], and Catherine Vaurs-Barriere†[1]

[1]Génétique, reproduction et développement (GReD) – CNRS : UMR6293, Inserm : U1103, Université d'Auvergne - Clermont-Ferrand I, Université Blaise Pascal - Clermont-Ferrand II – GRED Faculté de Médecine 28, place Henri Dunant BP 38 - 63001 CLERMONT-FERRAND +33 4 73 17 81 78, France
[2]Cancer Resistance Explorind and Targeting (CREaT) – Université d'Auvergne - Clermont-Ferrand I : EA7283 – Centre Jean Perrin - Département de radiothérapie BP 392 58, Rue Montalembert 63011 Clermont-Ferrand, France

Résumé

LINE-1 (L1) retrotransposons are an abundant class of transposable elements representing 17% of the human genome. The 5' region of L1 contains a bidirectional promoter consisting of an internal sense promoter, and an antisense promoter (ASP). In normal cells, the main defense mechanism, developed to counteract the deleterious effect of L1 activity, consists in L1 promoter DNA methylation.
A hallmark of cancer cells correspond to global DNA hypomethylation, affecting notably L1 promoters. It has been hypothesized that this could allow L1 promoter activation and contributes to genomic instability. Evidences suggest that L1 hypomethylation in tumors could imply expression, from ASP, of aberrant chimeric transcripts composed of L1 5' sequence and of genomic unique adjacent sequences. However, the genome-wide impact of chimeric transcripts in tumorigenesis remains under-evaluated.

To investigate the pangenomic extent of this transcriptional deregulation in tumors and its impact in cancer initiation, progression and aggressiveness, we have developed a dedicated bioinformatic tool named CLIFinder to analyze data from oriented paired-end RNA-seq. This tool was designed to select, among the RNA-seq reads, chimeras potentially corresponding to L1-chimeric transcripts, i.e. containing at their 5' end a L1 promoter sequence and at their 3' end at least 30 consecutive bp corresponding to a unique sequence of the genome.

Thirteen gliomas and 3 control brains were analyzed. We identified 3000 potential chimeras. Most of them are expressed in tumors and 75% implicate young L1 subfamilies in which ASP was described. Validation experiments have been performed for a subset of chimeras on a larger cohort of gliomas. Our results demonstrate that 1) our strategy is enough sensitive to detect moderate to faintly expressed chimeras, 2) 80% of the chimeras implicating a young

*Intervenant
†Auteur correspondant: catherine.barriere@udamail.fr

L1 have their transcription start site located at ASP and 3) some chimeras are tumors specific whereas others are found over-expressed in tumors compared to controls.
Complementary analyses will be performed to 1) establish the correlation between chimeric transcripts' expression and L1 hypomethylation, 2) assess whether some chimeras could correspond to potential biomarker for gliomas and 3) evaluate if some chimeras can have a functional role in tumorigenic processes.