
Revealing Insertion Outcomes of *de novo* LINE-1 Retrotransposition by an innovative Long-Read Sequencing method

Cheuk-Ting Law^{*1,2}, Carlos Mendez-Dorantes^{1,2}, and Kathleen H. Burns^{1,2}

¹Dana-Farber Cancer Institute [Boston] – 450 Brookline Ave. Boston, MA 02215, United States

²Harvard Medical School – 25 Shattuck Street, Boston, MA 02115, United States

Abstract

Long Interspersed Element-1 (LINE-1, L1) is an active protein-coding transposon in the human genome, which can ‘copy-and-paste’ itself to generate *de novo* genomic insertions using an RNA intermediate; this process is called retrotransposition. Although most genomic copies of L1 (~500,000 copies) are immobile, there are approximately 100 loci in human cells that are retrotransposition competent.

Our current working model is that L1 retrotransposition intermediates could result in various insertion outcomes:

- Full-length insertion of L1 (~6kb) with target-site duplication (TSD);
- A variant insertion (i.e., 5’ truncated/ 5’ inverted);
- Trans-mobilisation of small RNA;
- A chromosomal rearrangement (e.g., translocation, duplication and, deletion);
- EN-independent insertion

However, capturing *de novo* L1 insertions in their entirety is technically challenging owing to their varied size, their highly repetitive nature, and their potential association with genomic rearrangements. Currently, **we have very little data on the frequency of each insertion outcome**. In addition, the field has focused on a handful of human cell lines, mainly HeLa in the past few decades, thus, **differences in sequence outcomes across cell lines have not been well characterized**.

To address these technical and knowledge gaps, I have developed a novel Oxford Nanopore long-read sequencing approach to characterize large numbers (> 10,000) of *de novo* L1 retrotransposition outcomes induced by a new retrotransposition reporter in a single sequencing run. I recapitulate known sequence features of L1, various types of L1 insertions, rare L1 events, and L1-mediate genomic rearrangements. Additionally, I revealed that L1 retrotransposition could lead to significantly distinctive insertion outcomes in different cell lines. I anticipate that my data **(1) will reveal complete structures of L1 insertions and provide quantitative metrics of known L1 sequence features and potentially uncover new features**. My work **(2) will shed light on the dynamics of L1 retrotransposition in different cell lines**.

Keywords: long read, sequencing, LINE1, L1, nanopore

^{*}Speaker