The Impact of LINE-1 Retrotransposition on the Human Genome

Amy E. Hulme, BS, MS, Deanna A. Kulpa, BS, MS, José Luis Garcia Perez, PhD, and John V. Moran, PhD

CONTENTS

BACKGROUND
INTRODUCTION
THE STRUCTURE OF A RETROTRANSPosition COMPETENT L1
METHODS TO STUDY L1 RETROTRANSPosition
FUNCTIONAL ANALYSES
THE L1 RETROTRANSPosition CYCLE
TPRT: VARIATIONS ON A THEME
L1 AS A MUTAGEN
L1 AS AN ARBITER OF GENOME DIVERSIFICATION
INDIVIDUAL VARIATION IN L1 RETROTRANSPosition
NEXT IN LINE: REMAINING QUESTIONS
REFERENCES

BACKGROUND

Long interspersed element-1 (LINE-1 or L1) is an abundant retrotransposon that comprises approx 17% of human DNA. L1 retrotransposition events can lead to genome diversification and individual genetic variation by serving as insertion mutagens and by providing recombination substrates either during or long after their insertion. L1 retrotransposition also generates genomic variation by mobilizing DNA derived from its flanks, non-autonomous retrotransposons (e.g., Alu elements), and cellular mRNAs to new genomic locations. Together, these sequences comprise approx 15% of human genomic DNA. Thus, L1-mediated retrotransposition events are responsible for at least one-third of our genome. In this chapter, we discuss how innovative assays developed in recent years have increased our understanding of L1 biology and the impact of L1 on the human genome.
INTRODUCTION

Approximately 1.5% of the human genome encodes proteins (1,2). Much of the remaining DNA often is disparaged as “junk” owing to its assumed lack of function. Transposable elements account for nearly half of “junk DNA,” and they are classified based on their mobility intermediate (Table 1). DNA transposons move (i.e., transpose) using a DNA intermediate, represent approx 3% of genomic DNA, but are now immobile (3). Retrotransposons move (i.e., retrotranspose) using an RNA intermediate and account for approx 30% of human DNA (4). They are subclassified based on whether they encode proteins required for their retrotransposition (autonomous retrotransposons) or rely on proteins to be supplied in trans (non-autonomous retrotransposons).

Retrotransposons are subclassified further based on their structure. Long terminal repeat (LTR)-containing retrotransposons, typified by human endogenous retroviruses (HERVs), resemble retroviruses, but generally lack or contain a nonfunctional envelope gene. Although they comprise approx 8% of genomic DNA, the vast majority of HERVs are immobile (Table 1). However, some HERVs are polymorphic with respect to presence and a few elements may have protein coding potential (5,6). Thus, a few HERVs may remain retrotransposition-competent.

Non-LTR retrotransposons, exemplified by LINEs, lack LTRs and usually end in a poly (A) tail. They represent the largest class of transposable element derived repeats in the genome, accounting for approx 21% of human DNA (1). LINE-1s (L1s), the most abundant retrotransposon, are present at more than 500,000 copies in the haploid genome, and some L1s remain retrotransposition-competent. In contrast, LINE-2 and LINE-3 elements are structurally distinct from L1s, are less abundant in human DNA, and are no longer retrotransposition-competent (1).

The proteins encoded by retrotransposition-competent L1s also can function in trans to mobilize either non-autonomous retrotransposons (e.g., Alu and perhaps short interspersed element R-VNTR-Alu [SVA] elements) or cellular mRNAs, resulting in processed pseudogene