UNDERSTANDING THE ROLE OF TRANSPOSABLE ELEMENTS IN THE HUMAN GENOME

In the beginning, the human genome was sharply divided between gene-coding sequences, which comprised just ~2%, and the rest, apparently non-functional junk DNA composed mostly of transposable elements (TEs). This prevailing perspective has served as the primary objection to sequencing the entire human genome.

Gradually, perspectives changed. The ENCODE consortium, aptly called the Encyclopedia of DNA Elements, revealed that approximately 20% of our genome is made up of regulatory elements, and that at least 25% of these regulatory elements are TE or derived from them. “At least” indicates that the annotations concern nearly intact TE elements, with the more degenerate ones (the older ones) being particularly difficult to identify.

In this respect, the availability today of hundreds of sequenced animal genomes can prove extremely helpful. In fact, the reconstruction of the evolutionary history of an unknown sequence can be crucial to identify it as a TE-derived sequence. This is exactly what Matsushima et al. reported in a recent article in Cell Genomics (1). They discovered previously unannotated degenerate TEs by analyzing multiple reconstructed ancestral genomes from hundreds of species. They achieved a 10.8% increase in coverage compared to the most recent annotation.

In this context it is worth reporting a paradigmatic example in which tracing the evolutionary history of TE elements has brought to light lost pieces of the recent history of our genome.

Imagine that a phrase from an ancient book has been erased in all surviving copies of the book. No one can even imagine the manipulation until the original, lost book containing the phrase is found.

Let's go back to the human genome. Homologous-nonallelic recombination (HNAR) can also occur between TE elements, particularly *Alu*. Indeed, Song et al. (2), studying a population of 54,000 subjects,

identified 47 duplications and 40 deletions mediated by *Alu*/*Alu* HNAR. The deletions were easily identified as such because they were in a heterozygous state in the population.

In their 2006 publication, Sen et al. (3) argued that a deletion resulting from HNAR between *Alu* elements could go unnoticed once it became fixed in the population. To uncover these occurrences, the researchers conducted a comparative analysis between the human and chimpanzee genomes. They revealed 492 instances of such deletions, spanning a total of 400 kb. Notably, a majority of these deletions (295 out of 492) impacted known or predicted genes, providing evidence for potential evolutionary implications.

(1) https://www.sciencedirect.com/science/article/pii/S2666979X24000247
(2) https://genome.cshlp.org/content/28/8/1228.abstract
(3) https://www.sciencedirect.com/science/article/pii/S0002929707600159