

ELEMENTOS TRANSPONIBLES EN HUMANOS Y SU IMPACTO EN EL DESARROLLO DE ENFERMEDADES

Layla Díaz-Portal y Bernardo Rodríguez-Martín
Centre for Genomic Regulation (CRG), The Barcelona Institute
of Science and Technology, Barcelona
Universitat Pompeu Fabra (UPF), Barcelona

<https://doi.org/10.18567/sebbmrev.223.202503.dc4>



Introducción a los elementos transponibles

Tanto los organismos procariotas como los eucariotas contienen una proporción variable pero considerable de elementos repetidos a lo largo de su genoma. Estos pueden agruparse en dos categorías en base a su disposición: las repeticiones en tándem y las repeticiones intercaladas (Figura 1a). Mientras que las primeras son secuencias de longitud variable repetidas de forma contigua una tras

otra, las repeticiones intercaladas, incluyendo duplicaciones segmentarias y elementos transponibles (*Transposable Elements*, TE por sus siglas en inglés), se repiten en posiciones distantes, ocupando aproximadamente la mitad del genoma humano (Figura 1b).

Los TE, también conocidos como “genes saltarines”, son secuencias de ADN capaces de cambiar de posición en el genoma por sí mismos. En función de su mecanismo de movilización,

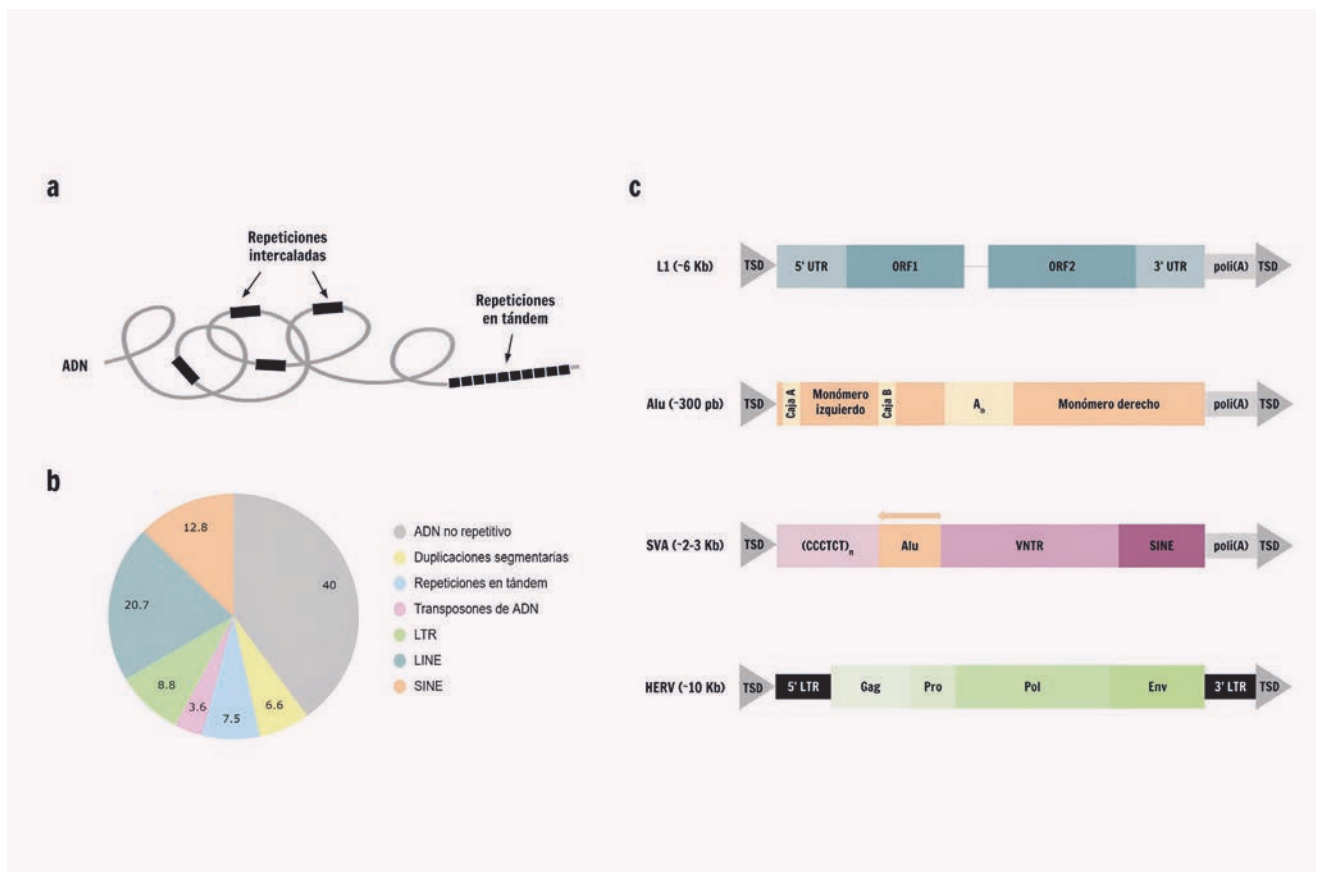


Figura 1

Naturaleza de los elementos transponibles y su distribución en el genoma humano. a) Las secuencias de ADN repetitivo pueden agruparse en dos categorías en base a su disposición: repeticiones intercaladas (incluyendo las duplicaciones segmentarias y los elementos transponibles) y las repeticiones en tándem. b) Porcentaje de cada tipo de elemento repetitivo en el genoma humano de referencia (versión T2T-CHM13). Los SVA no aparecen representados en el gráfico porque se encuentran en una frecuencia muy baja (~0,2%). c) Representación esquemática de los elementos transponibles más comunes en humanos con su tamaño completo indicado en kilobases (Kb) o pares de bases (pb). L1 codifica dos marcos de lectura (ORF, *Open Reading Frame*), el segundo con dominios de endonucleasa y transcriptasa inversa. Los elementos Alu presentan dos monómeros separados por un conector rico en A. SVA es un elemento compuesto, constituido por una repetición hexamérica CCCTCT, una secuencia tipo Alu invertida, un número variable de repeticiones en tándem (VNTR, *Variable Number of Tandem Repeats*) y una secuencia de tipo SINE. Tanto L1, como Alu y SVA presentan en su extremo 3' una cola de poli(A). Los retrovirus endógenos humanos (HERV, *Human Endogenous Retroviruses*) codifican tres proteínas virales esenciales (Gag, Env y Pol), además de una proteasa (Pro). Todos los elementos están flanqueados en sus extremos por duplicaciones de la secuencia objetivo (TSD, *Target Site Duplication*).

podemos dividirlos en dos categorías: los TE de clase I, más conocidos como retrotransposones, que se movilizan mediante un mecanismo de "copia y pega" que involucra un intermediario de ARN, y los TE de clase II, denominados transposones, que emplean un mecanismo de "corta y pega", movilizándose directamente como ADN. Los retrotransposones se pueden clasificar a su vez en dos grupos en base a la presencia de repeticiones terminales largas

(*Long Terminal Repeats*, LTR por sus siglas en inglés) flanqueando su secuencia o la carencia de ellas (no-LTR).

A día de hoy, sólo hay evidencias de actividad para tres clases de TE (Alu, LINE-1 y SVA), todos ellos retrotransposones no-LTR, en el genoma humano. Tanto Alu como SVA dependen de la maquinaria proteica codificada por *LINE-1* (L1) para su retrotransposición, mientras que L1 se trata del único retrotransposón con capacidad

autónoma para movilizarse. Un elemento L1 de longitud completa tiene un tamaño aproximado de 6 kilobases (Kb) y dos marcos de lectura abiertos (*Open Reading Frame*, ORF por sus siglas en inglés), que codifican la maquinaria proteica necesaria para su movilización e integración (Figura 1c). L1 representa el tipo de TE más prevalente en el genoma humano en términos de su contribución al número total de bases, constituyendo en torno al 17% de su secuencia. En total,

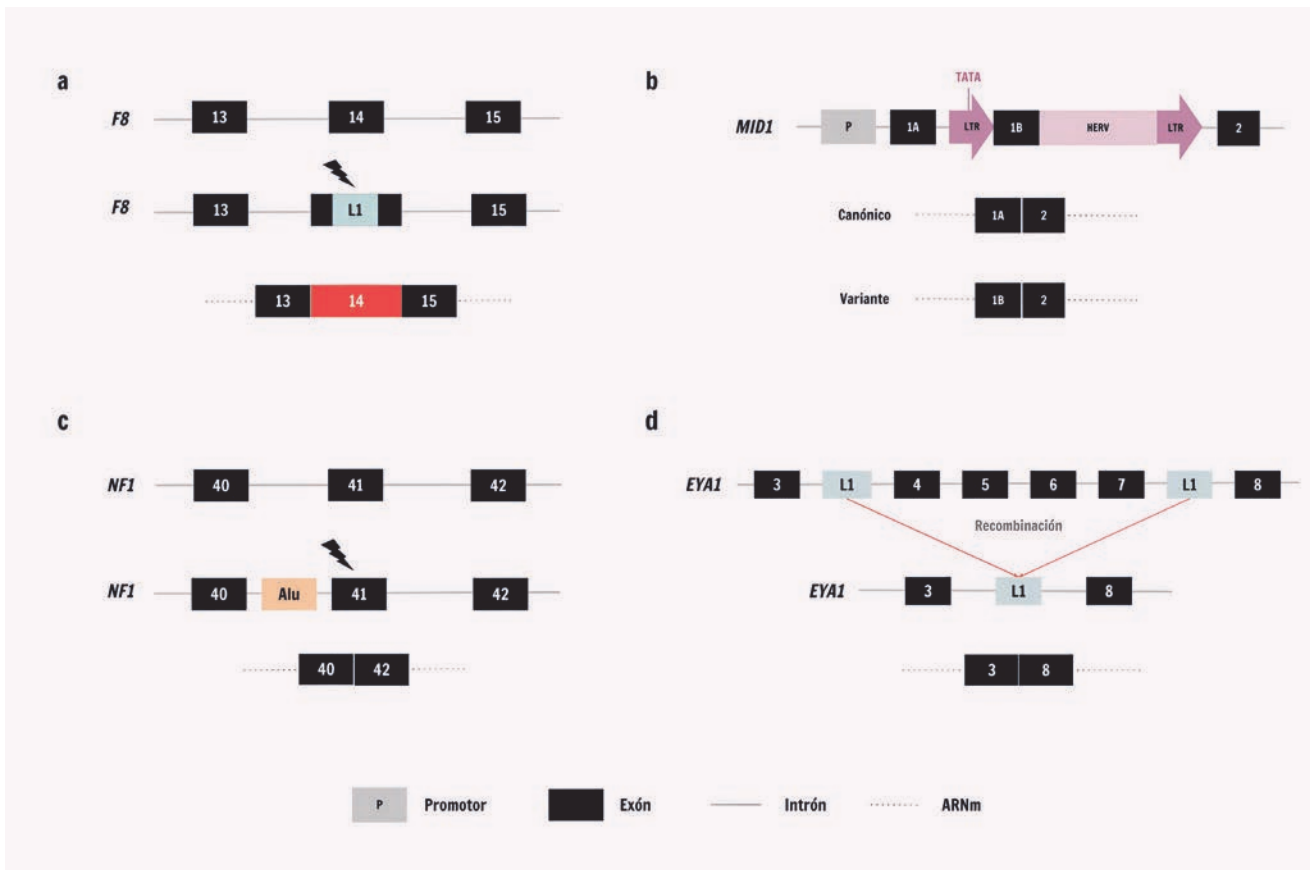


Figura 2

Principales mecanismos mediante los cuales los TE pueden causar enfermedades monogénicas. a) Alteración de la función génica a través de la disrupción de regiones codificantes. La inserción *de novo* de L1 interrumpe el exón 14 del gen para el factor de coagulación VIII (*F8*) en un paciente con hemofilia A. b) Alteración de la transcripción del gen *MID1* debido a la introducción de sitios de unión para elementos reguladores. La inserción de un HERV en un paciente con síndrome de Opitz ligado al cromosoma X proporciona un nuevo sitio de inicio de la transcripción (TATA) y un nuevo intrón 1 que conduce a un transcrito de ARNm más corto que el original. c) Alteración de la función génica a través de la disrupción de regiones no codificantes. Una inserción Alu en un paciente con neurofibromatosis tipo I se localiza 44 pb aguas arriba del exón 41 de *NF1*, lo que provoca la omisión del mismo y cambios en el marco de lectura. d) Reordenamientos del material genético debidos a la recombinación entre elementos. La recombinación entre dos L1 intrónicos dentro del gen *EYA1* provoca la pérdida de los exones 4 a 7 en un paciente con síndrome branquio-oto-renal.

existen más de 500.000 copias de L1 en el genoma humano de referencia, resultado de su actividad y movilización continua durante los últimos 150 millones de años. Sin embargo, la mayoría de L1 ha perdido la capacidad para propagarse debido a la acumulación de mutaciones puntuales, reordenamientos en su secuencia, o bien a que se encuentran localizados en regiones con poca actividad transcripcional, como la heterocromatina. Por lo tanto, únicamente un reducido número de elementos (se estima que

en torno a unos 100 por genoma humano) están completamente intactos, y un número aún más pequeño son altamente activos (denominados en inglés *hot L1*).

La retrotransposición y subsecuente inserción de TE puede afectar tanto a la expresión génica como a la estructura genómica. Si bien esto hace de los TE una importante fuente de variabilidad genética y fenotípica, impulsando la evolución de los genomas y la adaptación de las especies al medio, esta capacidad mutagénica

también puede provocar que se vean asociados al desarrollo de ciertas enfermedades.

Inserción de los TE en la línea germinal y sus consecuencias en la estructura genómica y la expresión génica

El primer caso reportado de una inserción de TE patogénica en humanos tuvo lugar en 1988, cuando un grupo de investigadores identificó, en el genoma de un paciente con hemofilia A sin antecedentes familiares, una inserción L1 *de novo* en el exón 14 del gen *F8* en

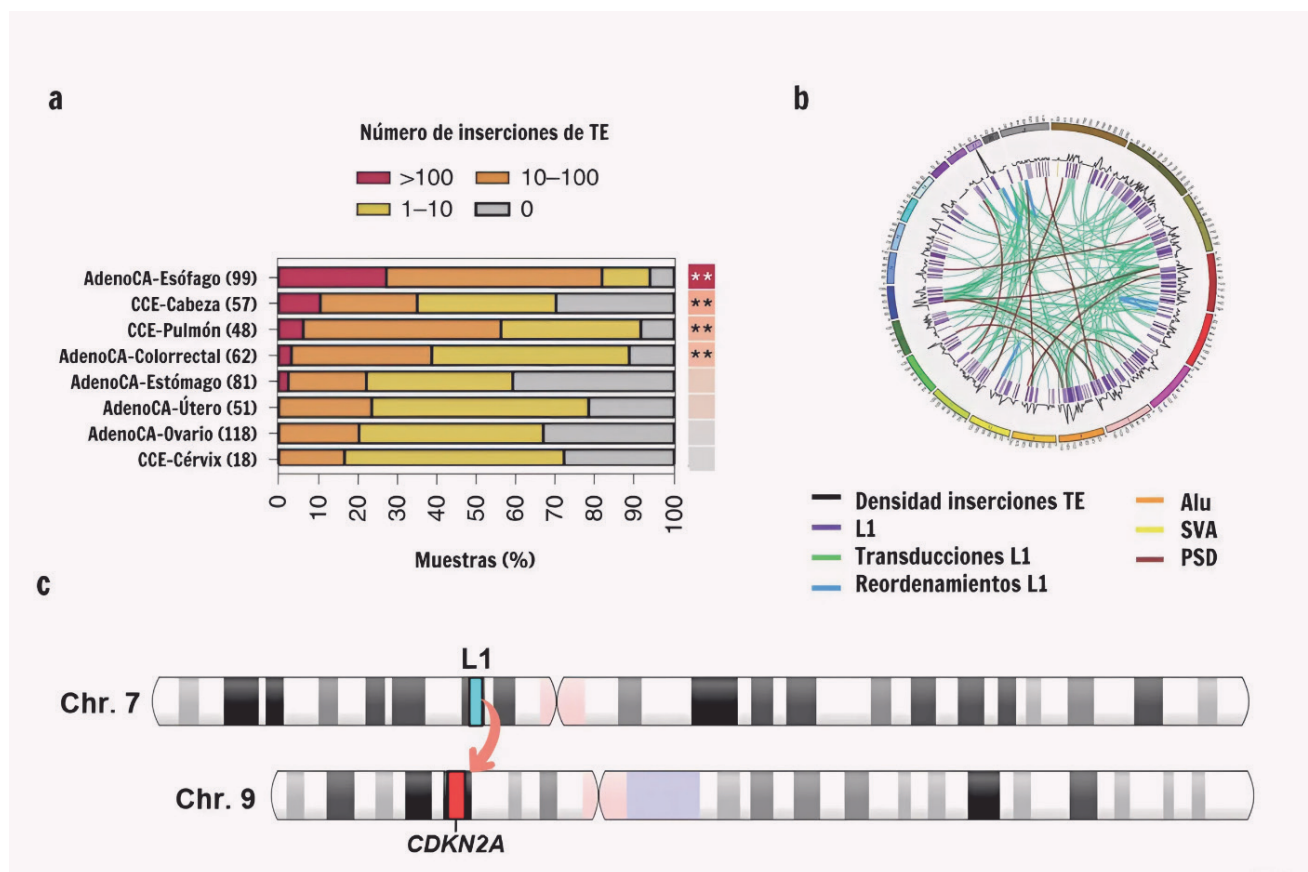


Figura 3

Retrotransposición somática en los cánceres humanos. a) Para los ocho tipos de cáncer del proyecto Pan-Cáncer con mayores tasas de retrotransposición somática, se representa la proporción de tumores con >100 (rojo), 10–100 (naranja), 1–10 (amarillo) y 0 (gris) retrotransposiciones somáticas. AdenoCA se refiere a los adenocarcinomas, mientras que CCE hace referencia a los carcinomas de células escamosas. El número de muestras analizadas para cada tipo de tumor se muestra entre paréntesis. b) Tumor de cabeza y cuello con una alta tasa de retrotransposición (638 inserciones somáticas). La localización genómica de cada tipo de inserción se indica con una línea de color: inserciones de L1, transducciones (movilización de ADN aguas abajo del TE) mediadas por L1, reordenamientos mediados por L1, inserciones de Alu, SVA y pseudogenes (PSD). c) Ilustración de una pérdida de ADN de 5 millones de bases, afectando al gen supresor *CDKN1A*, causada por un L1 altamente mutagénico en el cromosoma 7. *Figura adaptada de doi: 10.1038/s41588-019-0562-0.*

el cromosoma X (Figura 2a) que codifica un factor de coagulación sanguínea. Desde entonces, se han reportado más de 100 inserciones de TE en la línea germinal causantes de varias decenas de enfermedades monogénicas.

Más allá de la interrupción de regiones codificantes, se deben considerar los efectos inherentes a la propia secuencia del elemento insertado, la cual puede contener sitios de unión para elementos reguladores (p. ej. promotores) o aceptores para el corte y empalme

(en inglés *splicing*) del ARN mensajero (ARNm). Un ejemplo notable es el caso de las inserciones de retrovirus endógenos humanos (*Human Endogenous Retroviruses*, HERV por sus siglas en inglés) asociadas al síndrome de Opitz ligado al cromosoma X. Los HERV están flanqueados por repeticiones terminales largas (*Long Terminal Repeats*, LTR por sus siglas en inglés), las cuales pueden actuar como promotores y potenciadores, generando un sitio alternativo de inicio de la

transcripción para el gen *MID1* (Figura 2b), que resulta en el desarrollo de la enfermedad.

Por otro lado, las inserciones de TE en regiones intrónicas pueden alterar el proceso de corte y emplame (*splicing*), afectando a la función génica con efectos patogénicos. Un ejemplo es el caso de la neurofibromatosis tipo I, una enfermedad genética común vinculada al gen supresor de tumores *NF1*, que se caracteriza por cambios en la coloración de la piel y el crecimiento de tumores

benignos a lo largo de los nervios. Se ha encontrado que la inserción de una secuencia Alu en el intrón 40 del gen *NF1* provoca la exclusión del exón 41 durante la maduración del ARNm (Figura 2c). Por su parte, la distrofia muscular de Duchenne es una enfermedad caracterizada por una degeneración muscular progresiva causada por mutaciones en el gen de la distrofina (*DMD*), un gen de gran tamaño (~2,5 Mb) que contiene más de 60 exones y está ubicado en el cromosoma X. Se han reportado varios casos de inserciones tanto de L1 como de Alu que afectan al empalme alternativo de los transcritos del gen *DMD* y resultan en la omisión de alguno de sus exones, dando lugar a una distrofina no funcional.

Por último, como consecuencia de su abundancia en el genoma y homología a nivel de secuencia, la recombinación de TE puede promover la formación de diversos tipos de reordenamientos cromosómicos, tales como deleciones, duplicaciones, translocaciones o inversiones de ADN. Aunque los eventos de recombinación Alu-Alu son los más frecuentes, con más de 70 casos reportados de enfermedades monogénicas, la recombinación entre dos elementos L1 o dos HERV también pueden tener consecuencias patogénicas. Un ejemplo es la pérdida de material genético en el gen *EYA1*, causada por la recombinación entre dos inserciones de L1 (Figura 2d), lo que da lugar en un paciente al síndrome branquio-oto-renal.

Movilización somática de los TE y su implicación en el cáncer

Debido a su capacidad mutagénica, los TE pueden ser considerados una amenaza para la integridad genómica, por lo que las especies hospedadoras han desarrollado mecanismos de defensa para silenciarlos. Estos mecanismos operan en todos los

niveles de la regulación de los TE, que van desde su silenciación epigenética hasta la degradación de los transcritos derivados de su expresión. No obstante, bajo determinadas condiciones, estos mecanismos pueden verse comprometidos, desencadenando la activación de TE y la mutagénesis derivada de su actividad. Hasta hace poco, se creía que los TE se movilizaban principalmente en las células germinales. Sin embargo, estudios recientes han mostrado actividad de TE en tejidos somáticos adultos, como el cerebro y el colon, además de altas tasas de movilización de retrotransposones de tipo L1 en cánceres epiteliales.

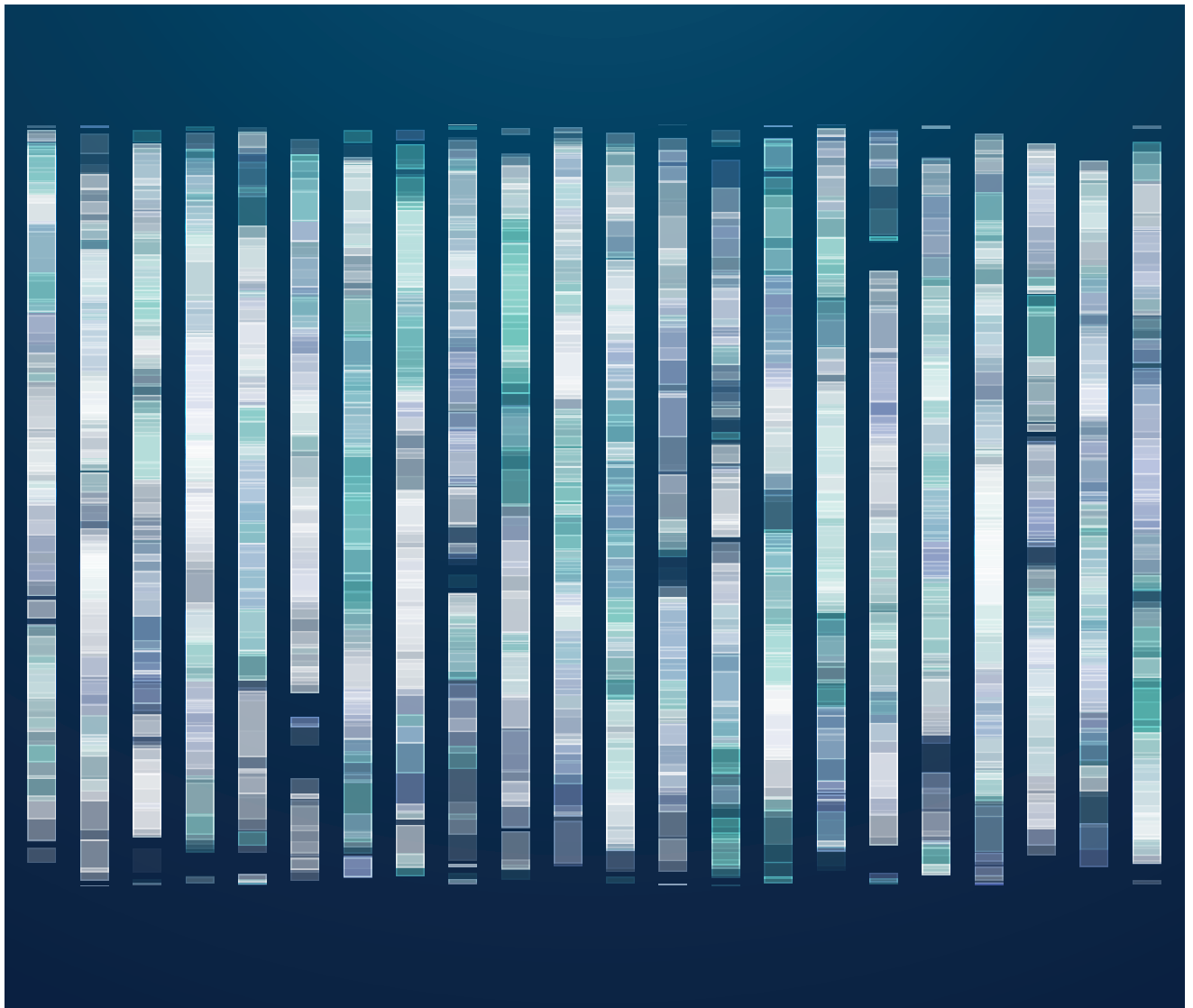
En un estudio en el que se analizaron los genomas del cáncer para 244 individuos, se encontró que los tumores del 53% de los pacientes presentaban inserciones somáticas de L1. En un segundo estudio, utilizando una cohorte de casi 3.000 tumores de 38 tipos distintos de cáncer pertenecientes al proyecto Pan-Cáncer, se reportaron altas tasas de retrotransposición para L1 en adenocarcinomas de esófago, cáncer colorrectal, y cabeza y cuello (Figura 3a). De hecho, en estos tres tipos de cáncer, la inserción de L1 representa uno de los tipos más frecuentes de variación estructural (p. ej. cambios en el ADN mayores de 50 bp) somática. Además, se encontró que la activación de sólo un pequeño conjunto de elementos L1 altamente mutagénicos puede desencadenar de decenas a cientos de retrotransposiciones en un sólo tumor (Figura 3b).

Por otro lado, investigadores del proyecto Pan-Cáncer hallaron un nuevo mecanismo mutacional en cáncer, que implica la generación de grandes reordenamientos cromosómicos como resultado de la integración aberrante de elementos L1. Estas integraciones pueden provocar la pérdida

de regiones cromosómicas de gran tamaño (Figura 3c), lo que puede resultar en la eliminación de genes supresores de tumores, así como la generación de translocaciones y duplicaciones a gran escala. Además, mediante este mecanismo, L1 puede desencadenar ciclos de inestabilidad genómica conduciendo a la amplificación de oncogenes. En definitiva, estos estudios remarcan la capacidad mutagénica de L1 y su impacto como fuente de variación somática e inestabilidad genómica en el genoma del cáncer, fomentando su evolución de una manera análoga a la evolución de las especies.

Desafíos y oportunidades para el estudio de los TE

Los TE representan una fuente relevante de variabilidad genética, sirviendo de sustrato para la evolución de los genomas y adaptación al medio. No obstante, las mismas características que los hacen beneficiosos a nivel evolutivo también los convierten en una fuente de enfermedades, tanto mendelianas como complejas, incluido el cáncer. A pesar de su relevancia y potencial impacto en la salud, los TE continúan siendo uno de los tipos de secuencias más inexplorados del genoma humano. Debido a su alto porcentaje de homología a nivel de secuencia y a la gran cantidad de copias que se encuentran en el genoma, los TE han representado un desafío considerable para su detección y estudio durante décadas. No obstante, el desarrollo de tecnologías de secuenciación de lectura larga, como PacBio y Oxford Nanopore, cada vez más accesibles y precisas, junto con los avances en técnicas de ensamblaje de genomas, brindan una oportunidad sin precedentes para el estudio de la evolución y la diversidad de los TE y en el genoma humano.



Para leer más

Burns K. Transposable elements in cancer. *Nature Reviews Cancer* 17 (2017) 415–424. <https://doi.org/10.1038/nrc.2017.35>

Chénais B. Transposable elements and human diseases: Mechanisms and implication in the response to environmental pollutants. *International Journal of Molecular Science* 23 (2022) 2551. <https://doi.org/10.3390/ijms23052551>

Liao X, Zhu W, Zhou J. *et al.* Repetitive DNA sequence detection and its role in the human genome. *Communications Biology* 6 (2023) 954. <https://doi.org/10.1038/s42003-023-05322-y>

Payer LM, Burns KH. Transposable elements in human genetic disease. *Nature Reviews Genetics* 20 (2019) 760–772. <https://doi.org/10.1038/s41576-019-0165-8>

Rodríguez-Martín B, Álvarez EG, Báez-Ortega A, *et al.* Pan-cancer analysis of whole genomes identifies driver rearrangements promoted by LINE-1 retrotransposition. *Nature Genetics* 52 (2020) 306–319. <https://doi.org/10.1038/s41588-019-0562-0>