

Epidemiología genómica y paraparesia espástica tropical asociada a la infección por el virus linfotrópico humano de células T tipo 1

Mercedes Salcedo-Cifuentes,¹ Martha C. Domínguez¹ y Felipe García-Vallejo¹

Forma de citar

Salcedo-Cifuentes M, Domínguez MC, García-Vallejo F. Epidemiología genómica y paraparesia espástica tropical asociada a la infección por el virus linfotrópico humano de células T tipo 1. Rev Panam Salud Publica. 2011;30(5):422-30.

RESUMEN

Objetivo. Caracterizar el ambiente genómico de las secuencias adyacentes al virus linfotrópico humano de células T tipo 1 (HTLV-1) en pacientes con paraparesia espástica tropical y mielopatía asociada a la infección con HTLV-1 (PET/MAH) de diferentes regiones de Colombia y del Japón.

Métodos. Se enfrentaron 71 clones recombinantes con secuencias del genoma humano adyacentes al 5'-LTR de pacientes con PET/MAH, a las bases de datos del Genome Browser y del GenBank. Se identificaron y analizaron estadísticamente 16 variables genómicas estructurales y posicionales mediante el programa informático R, versión 2.8.1, en una ventana de 0,5 Mb.

Resultados. El 43,0% de los provirus se localizaron en los cromosomas del grupo C; 74% de las secuencias se ubicaron en regiones teloméricas y subteloméricas ($P < 0,05$). Un análisis de conglomerados permitió establecer las relaciones jerárquicas entre las características genómicas incluidas en el estudio; el análisis de componentes principales identificó las componentes que definieron los ambientes genómicos preferidos para la integración proviral en casos de PET/MAH.

Conclusiones. El HTLV-1 se integró con mayor frecuencia en regiones de la cromatina ricas en islas de citocina fosfato guanina (CpG), de alta densidad de genes y de repeticiones tipo LINE (elemento disperso largo [long interspersed element]) y transposones de ADN que, en conjunto, conformarían los ambientes genómicos blanco de integración. Este nuevo escenario promoverá cambios sustanciales en el campo de la salud pública y en el manejo epidemiológico de las enfermedades infecciosas, y permitirá desarrollar potentes herramientas para incrementar la eficiencia de la vigilancia epidemiológica.

Palabras clave

Virus linfotrópico de células T humanas tipo 1; integración viral; linfocitos; genoma humano; paraparesia espástica tropical; epidemiología molecular; Colombia.

El virus linfotrópico humano de células T tipo 1 (HTLV-1) es un retrovirus que infecta a entre 10 y 20 millones de personas en el mundo (1, 2). Se conocen

tres vías de transmisión: por contacto sexual (3), por vía parenteral (4) y por transmisión vertical durante la lactancia materna (5, 6). La infección por el virus es endémica en América del Norte y Sur, el sudeste del Japón, la cuenca del Caribe y África (2). Sudamérica, con una población aproximada de 360 millones de habitantes, presenta seroprevalencias de entre 1 y 2% (7).

La infección por el HTLV-1 se asocia tanto a trastornos malignos como a procesos inflamatorios. Estos últimos incluyen una amplia gama de enfermedades crónicas que afectan los ojos, las glándulas salivares y los tejidos nervioso y muscular (8, 9). Entre ellas, la paraparesia espástica tropical/mielopatía asociada a la infección con HTLV-1 (PET/MAH), es una afección epidemiológicamente compleja

¹ Laboratorio de Biología Molecular y Patogénesis, Departamento de Ciencias Fisiológicas, Escuela de Ciencias Básicas, Facultad de Salud, Universidad del Valle, Cali, Colombia. La correspondencia se debe enviar a Felipe García-Vallejo. Correo electrónico: labiomol@hotmail.com

que compromete estructural y funcionalmente el sistema nervioso central (9).

La PET/MAH es una mielopatía progresiva crónica caracterizada por parálisis espástica, disfunción de esfínteres y alteraciones sensoriales en las extremidades inferiores leves a severas (9). En Colombia, la infección por HTLV-1 está restringida geográficamente; una de las zonas con mayor endemia es la costa colombiana del Pacífico, donde la prevalencia varía entre 7,5% y 10,0% (10). El riesgo de progresión de la PET/MAH de un estado asintomático a uno sintomático oscila entre 0,25% y 4% (2, 10, 11).

A pesar de la asociación epidemiológica entre la infección y la enfermedad, el papel del virus en su patogenia aún no está claro; sin embargo, es evidente que la integración del ADNc viral en el genoma hospedero es un paso crucial en el ciclo de vida viral que garantiza una expresión eficiente de la progenie y es responsable de la persistencia de la infección y la causa de la enfermedad (9, 12).

Aunque los sitios de integración del HTLV-1 en el genoma hospedero se han caracterizado en varios estudios, se desconoce aún cómo factores tanto virales como celulares interactúan en la selección de las regiones cromatínicas blanco de integración. Los estudios en retrovirus no humanos han contado con diseños metodológicos experimentales y bioinformáticos, estos últimos apoyados en la gran cantidad de información consignada en las diferentes bases de datos del genoma humano (13-18). Los hallazgos han permitido concluir que la mayor parte de la cromatina hospedera es accesible para la integración de provirus, pero la frecuencia de selección de regiones específicas varía según el tipo de virus y la patología asociada (17, 18).

Las características estructurales, las regiones sensibles al tratamiento con ADNasa I (19) y cercanas a islas de citosina fosfato guanina (CpG) (14, 16, 20, 21), y el contenido de guanina y citosina (13) del genoma humano, favorecerían la integración. Además, ciertas características funcionales relacionadas con la regulación del ciclo celular, la apoptosis y la respuesta inmune (13-18) serían variables que, en conjunto, conformarían el ambiente genómico para la integración de este retrovirus humano.

La caracterización de los ambientes genómicos preferidos por el HTLV-1 se ha llevado a cabo a partir de ADN obtenido de cultivos celulares infectados (22-24),

mientras que solamente algunas de las secuencias estudiadas han sido obtenidas directamente del ADN de células mononucleares de sangre periférica (CMSP) de pacientes naturalmente infectados procedentes, sobre todo, de regiones endémicas del Japón. Asimismo, los análisis estadísticos realizados tienden a ser descriptivos o presentan un bajo número de variables incluídas en modelos multivariados.

Las infecciones que azotan a la población mundial, han sido de suma importancia para la moderna concepción de la epidemiología y la salud pública pues han permitido establecer un puente de investigación multidisciplinaria entre el agente infeccioso y el conocimiento del genoma humano. A este avance en el conocimiento se suman los resultados de la secuenciación de genomas retrovirales completos y el gran desarrollo alcanzado por las ciencias "ómicas". Entre estas se incluyen la transcriptómica, la proteómica y la integrómica, que se han convertido en herramientas importantes para obtener información útil para entender la dinámica de la infección y su distribución en las poblaciones. El objetivo de este trabajo fue caracterizar el ambiente genómico de las secuencias adyacentes al HTLV-1 en pacientes con PET/MAH de diferentes regiones de Colombia y el Japón.

MATERIALES Y MÉTODOS

Pacientes, muestras y unidad de análisis

Se incluyeron muestras de sangre periférica recolectadas de 17 pacientes de la costa colombiana del Pacífico con diagnóstico clínico de PET/MAH (25) y seropositivos para HTLV-1. El 25,0% eran varones; las edades variaban entre 31 y 70 años (media, 43 años). Se contó con la aprobación del Comité de Ética de la Universidad del Valle (26).

La unidad de análisis de este trabajo la conformaron 71 secuencias de genoma nuclear adyacentes al HTLV-1 extraídas a partir de CMSP por reacción en cadena de la polimerasa inversa (RCPI) o mediada por ligación (RCPLM); 30 de estas secuencias procedían de pacientes colombianos y 41 de pacientes japoneses.

Extracción del ADN del genoma

Las CMSP de los pacientes con PET/MAH colombianos se obtuvieron si-

guiendo la técnica de gradiente de Ficoll Hypaque (27). El ADN se extrajo utilizando el estuche WizardGenomic DNA Purification® (Promega, Madison, Wisconsin). El grado de pureza y la concentración del ADN extraído se evaluaron por espectrofotometría.

Amplificación del ADN genómico por reacción en cadena de la polimerasa inversa

La amplificación de secuencias de ADN celular adyacentes al HTLV-1 se efectuó siguiendo un protocolo ya descrito (28) con modificaciones menores (29). El proceso incluyó las siguientes etapas: una digestión inicial con la enzima NlaIII obtenida de *Neisseria lactamica*, que reconoce los sitios de restricción localizados en la región U5 de los extremos LTR de provirus, uno interno dentro del gen *gag*, otro en la región U5 del extremo 5'LTR y otros externos localizados en el genoma del hospedero. La segunda etapa incluyó un proceso de autoligación de los extremos cohesivos de cada uno de los fragmentos obtenidos en la etapa anterior, mediante la enzima ADN ligasa del bacteriófago T4. En la tercera etapa, los fragmentos autoligados se sometieron a digestión con SstII obtenida de *Streptomyces stanford*, una endonucleasa de restricción que reconoce específicamente un sitio de corte dentro del gen *gag*. Mediante este tratamiento se conservaron los anillos de ADN que contenían porciones de genoma hospedero más un fragmento de la región U5 del LTR proviral.

Después de una etapa de calentamiento a 93 °C durante 30 minutos, los círculos relajados se amplificaron por RCP. Los oligonucleótidos cebadores se sintetizaron a partir de secuencias específicas de la región U5 del provirus HTLV-1. Previamente a los ciclos de amplificación por RCP, la muestra se sometió a desnaturalización térmica en las condiciones antes descritas. Como cebadores se emplearon el HTLV-009, 5'-AAGCCGGCAGT-CAGTCGTGA-3' (8946-8927) y el HTLV-010, 5'-AAGTACCGGCAACTCTG-CTG-3' (8958-8977). El alineamiento de estos se llevó a cabo a 52 °C durante 30 segundos, seguido de una etapa de extensión a 72 °C durante 90 segundos. Se hicieron en total 35 ciclos de amplificación.

Todos los productos se separaron de acuerdo a su tamaño molecular, en pares de bases, mediante electroforesis en

geles de agarosa a 2,0%. La visualización de las bandas de ADN se registró por la fluorescencia del bromuro de etidio unido al ADN en presencia de luz ultravioleta. Cada banda representó un clon celular.

Protocolos de clonación y secuenciación

Los fragmentos de ADN obtenidos por RCPI se eluyeron de los geles de agarosa usando el estuche comercial QIAEX II Gel Extraction® (Qiagen, Valencia, California). Los fragmentos de ADN eluidos se ligaron al sitio Hind III (enzima extraída a partir de *Haemophilus influenzae*) en el vector pCR2.1, previamente alineado con la misma enzima. Los plásmidos recombinantes transformaron células de *Escherichia coli*, cepa +DH5-, siguiendo las instrucciones del estuche comercial TA Cloning System® (In VitroGen, Carlsbad, California). Los recombinantes obtenidos por extracción de ADN en lisis alcalina se digirieron con la enzima EcoR I para escindir el fragmento de ADN celular clonado. En los pacientes colombianos con diagnóstico de PET/MAH se obtuvieron en total 30 clones por RCPI del HTLV-1. El ADN celular clonado se secuenció en un equipo ABI PRISM 310® (PE Applied Biosystems, Norwalk, Connecticut) mediante el método de secuenciación cíclica. Los resultados obtenidos se sometieron al GenBank; en el cuadro 1 se presentan los números de acceso temporales.

Mediante una búsqueda en la base de datos mundial de secuencias genéticas del GenBank (<http://www.ncbi.nlm.nih.gov/>), se seleccionaron 41 secuencias obtenidas por RCPI o RCPML, provenientes de pacientes japoneses con diagnóstico confirmado de PET/MAH seropositivos para HTLV-1 depositadas por Cavrois (24) y Mortreux (30); los números de acceso se presentan en el anexo 1.

Análisis bioinformático

El análisis bioinformático se llevó a cabo entre noviembre y diciembre del 2009. Los 71 clones recombinantes se analizaron con el programa informático Vector Screen de acceso público (www.ncbi.nlm.nih.gov/vecscreer/vecscreer.html) para eliminar la contaminación de porciones de secuencias LTR provirales. Las secuencias ya editadas se alinearon a las del genoma humano utilizando el programa Blast de la base de datos del Genome Browser de la Universidad de California en Santa Cruz, California (<http://www.genome.ucsc.edu/>). Los criterios de inclusión de las secuencias del genoma que se alinearon con aquellas en estudio fueron una similitud entre la secuencias nucleotídicas $\geq 95\%$ ($e\text{-value} \leq 0,05$), con una longitud en pares de bases mayor de 50, en una extensión del genoma de 0,5 Mpb, lo que se definió como ventana de apertura de 0,5. La localización cromosómica, los genes

anotados, los elementos repetidos y otras características de la cromatina adyacente al HTLV-1 (orientación, proceso biológico en el que intervienen, función molecular, ubicación celular y número de exones e intrones de los genes identificados en la ventana de apertura), se obtuvieron utilizando no sólo la plataforma del Genome Browser sino también varias otras herramientas computacionales disponibles en la página web del National Center for Biotechnology Information (NCBI) de los Estados Unidos y en las bases de datos del Gencard (<http://www.genecards.org/cgi-bin/carddisp.pl>), Gene Entrez (<http://www.ncbi.nlm.nih.gov/ncbi/gene-entrez>) y Gene Ontology (<http://www.geneontology.org/index.shtml>).

Para la extracción de esta información se desarrolló un programa informático de aplicación mediante un sitio web (http://201.234.78.25/trl/consultas_filtros.aspx), que permitió la transferencia, la transformación, el almacenamiento y la visualización de los datos genómicos adyacentes al provirus, así como el traspaso de la base de datos a diferentes paquetes estadísticos.

Análisis estadístico

El análisis estadístico se llevó a cabo con el software R, versión 2.8.1 (<http://www.R-project.org>). Mediante el test de Friedman (31) se evaluaron las características de las secuencias de origen co-

CUADRO 1. Variables incluidas en el estudio sobre epidemiología genómica y PET/MAH

Nombre de la variable	Escala de medición
Frecuencia de islas CpG	Número de islas CpG en el HIT/Total de islas CpG en la ventana de apertura de 0,5 Mb
Frecuencia de genes	Número de genes en el HIT/Total de genes en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones LINE	Número de repeticiones LINE en el HIT/Total de repeticiones LINE en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones LINE tipo L2	Número de repeticiones LINE L2 en el HIT/Total de repeticiones LINE tipo L2 en la ventana de apertura de 0,5 Mb
Frecuencia de otras repeticiones tipo LINE	Número de otras repeticiones LINE en el HIT/Total de otras repeticiones LINE en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones SINE	Número de repeticiones SINE en el HIT/Total de repeticiones SINE en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones SINE tipo Alu	Número de repeticiones SINE tipo Alu en el HIT/Total de repeticiones SINE tipo Alu en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones SINE/Otras	Número de tras repeticiones tipo SINE en el HIT/Total de otras repeticiones SINE en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones de baja complejidad	Número de repeticiones de baja complejidad en el HIT/Total de repeticiones de baja complejidad en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones simples	Número de repeticiones simples en el HIT/Total de repeticiones simples en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones tipo ADN	Número de repeticiones tipo ADN en el HIT/Total de repeticiones tipo ADN en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones tipo LTR	Número de repeticiones tipo LTR en el HIT/Total de repeticiones tipo LTR en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones de origen desconocido	Número de repeticiones de origen desconocido en el HIT/Total de repeticiones de origen desconocido en la ventana de apertura de 0,5 Mb
Frecuencia de repeticiones tipo satélite	Número de repeticiones satélite en el HIT/Total de repeticiones satélite en la ventana de apertura de 0,5 Mb
Otras repeticiones	Número de otras repeticiones en el HIT/Total de otras repeticiones en la ventana de apertura de 0,5 Mb
Integraciones provirales por cromosoma	Número de provirus HTLV-1 por cromosoma

PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1); CpG: citosina fosfato guanina; LINE: elemento disperso largo (*long interspersed element*); SINE: elemento disperso corto (*short interspersed element*); LTR: repetición terminal larga (*long terminal repeat*).

Nota: Secuencias incluidas en el análisis multivariado. Se definió como *HIT* la porción del genoma humano que presentó un alto grado de homología secuencial con la secuencia en estudio. El término "ventana de apertura" correspondió a la extensión en pares de bases que mostraba el visor gráfico de la base de datos del Genome Browser.

lombiano frente a las de origen japonés. Se realizó un análisis univariado de las 16 variables incluidas en el estudio (cuadro 1), considerando los promedios, las medianas y la distribución porcentual de cada una de ellas. Se evaluó la colinealidad entre las variables a través de una matriz de dispersión, el índice de correlación y la prueba de esfericidad de Bartlett (31). Por último, se realizó un análisis de *cluster* jerárquico (ACJ) con 12 de las 16 variables incluidas en el estudio, que se complementó con un análisis de componentes principales (ACP). El número de componentes seleccionadas se estimó considerando una varianza $\geq 60\%$, la sedimentación y el criterio del valor propio (31). Se evaluó por medio de la prueba de la χ^2 al cuadrado el tipo de secuencia adyacente al ADN proviral y la frecuencia de éstas en el genoma humano según Venter *et al.* (32) y el IHGSC-2001 (International Human Genome Sequencing Consortium) (33); se consideró significativo un valor de $P < 0,05$.

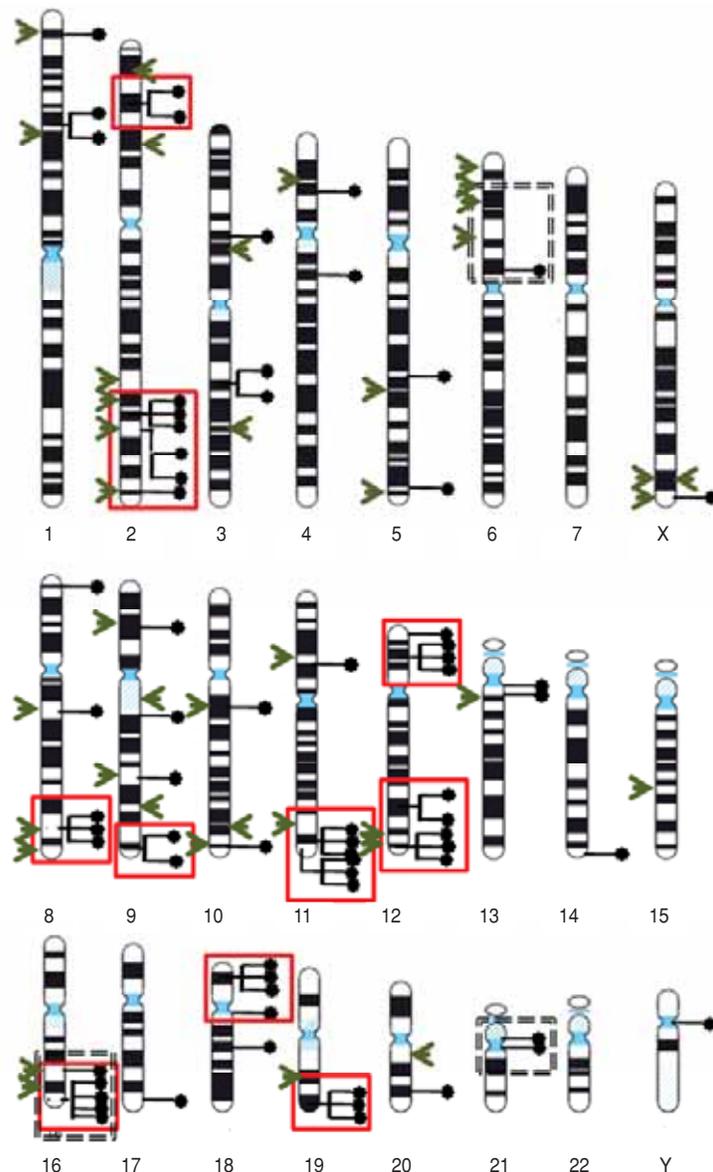
RESULTADOS

Se obtuvieron 73 regiones del genoma humano homólogas en una ventana de apertura de 0,5 Mb; es decir, dos de las secuencias en estudio presentaron un doble alineamiento. No se hallaron diferencias estadísticamente significativas entre las características genómicas de las secuencias colombianas y japonesas (test de Friedman, $P \geq 0,05$).

El 43,0% de las integraciones se localizaron en los cromosomas del grupo C, 19,0% en los del grupo A y 38,0% restante en los demás cromosomas. Un 74,0% de las secuencias homólogas (54 de 73) se ubicaron en regiones teloméricas y subteloméricas, con diferencias estadísticamente significativas con relación al resto de las regiones cromosómicas (prueba de la χ^2 al cuadrado, $P < 0,05$) (figura 1).

En la ventana de apertura de 0,5 Mb se encontró que 41,0% de las secuencias adyacentes al HTLV-1 fueron genes codificantes de proteínas, y el resto fueron elementos repetidos. Un 52,0% de las secuencias identificadas fueron exones, de los cuales 24,0% se encontraban en el sentido transcripción del exón. Según el proceso biológico en el que participan, los genes más frecuentes fueron los relacionados con la transducción de señales y la adherencia celular (37,0%) y los in-

FIGURA 1. Localización cromosómica de 73 provirus HTLV-1 en el genoma de pacientes con PET/MAH



HTLV-1: virus linfotrófico humano de células T tipo 1; PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el HTLV-1.

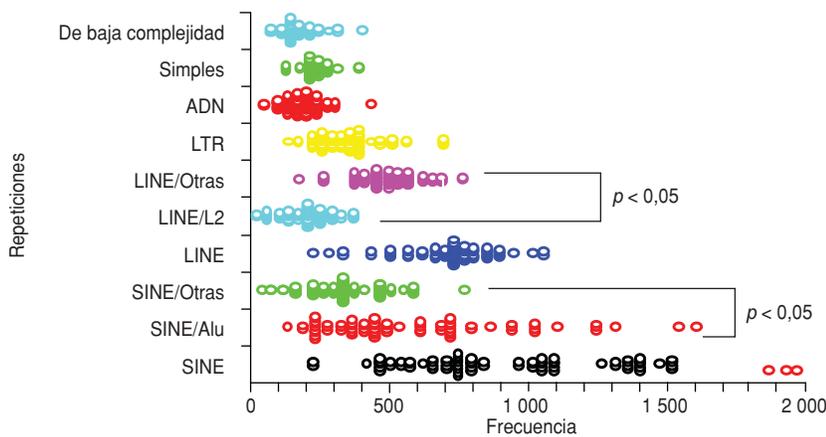
Nota: Los recuadros en color rojo muestran las regiones cromosómicas que presentaron dos, tres y más integraciones provirales. Los recuadros en línea punteada de color negro muestran regiones cromosómicas con *cluster* de genes codificantes de histonas y secuencias génicas codificantes de proteínas reguladoras del ciclo celular, interferón y ubiquitina cercanas a los sitios de integración del HTLV-1. Las flechas en color verde señalan las regiones sensibles a la enzima ADNasa I de acuerdo con Verdín (37). Los sitios de integración fueron asignados según la información del Genome Browser y el GeneBank.

volucrados en la respuesta inmune o mecanismos de defensa celular (13,0%). Un análisis comparativo entre los diferentes tipos de elementos repetidos reveló diferencias estadísticamente significativas entre el número de repeticiones intercaladas cortas tipo Alu (SINE/Alu) con respecto a los otros SINE (prueba de la χ^2 al cuadrado, $P < 0,05$) y entre el número de elementos repetidos intercalados largos tipo L2 (LINE/L2) con relación a otros LINE (prueba de la χ^2 al cuadrado,

$P < 0,05$) (figura 2). Al comparar el porcentaje de esta distribución con las comunicadas por Venter *et al.* (32) y el International Human Genome Sequencing Consortium (IHGSC) (33) para el total del genoma humano, se observaron diferencias estadísticamente significativas (prueba de la χ^2 al cuadrado, $P < 0,05$) (cuadro 2).

La correlación entre las variables en estudio fue confirmada mediante los resultados de la matriz de dispersión, el índice

FIGURA 2. Distribución de los valores mínimos, máximos y mediana del número de repeticiones identificados en la ventana de apertura de 0,5 Mb adyacentes al HTLV-1 en pacientes con PET/MAH



HTLV-1: virus linfotrópico humano de células T tipo 1; PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1); LTR: repetición terminal larga (*long terminal repeat*); LINE: elemento disperso largo (*long interspersed element*); SINE: elemento disperso corto (*short interspersed element*).

de correlación y la prueba de esfericidad de Bartlett ($P < 0,05$). Para los análisis multivariados sólo se consideraron 12 de las 16 variables incluidas en el estudio.

Los resultados del análisis de *cluster* jerárquico (ACJ) o de conglomerados mostraron tres grupos de cromosomas con una distribución homogénea de las variables en estudio hacia el interior de cada *cluster* y una máxima heterogeneidad entre los grupos (figura 3).

Los resultados del ACJ fueron confirmados independientemente con el análisis de componentes principales. Este último transformó las variables correlacionadas en un grupo pequeño de variables no correlacionadas llamadas componentes. Las repeticiones tipo ARN y SINE

conformaron la primera componente y explicaron 26,2% de la varianza total.

A la segunda componente contribuyeron las repeticiones LINE, los elementos repetitivos tipo ADN, las islas CpG y el número de genes; en conjunto, estos factores explicaron 22,6% de la varianza total. Los elementos tipo LTR y las repeticiones de baja complejidad, satélites, de origen desconocido y simples conformaron la tercera componente, que explicó 11,6% de la varianza. La cuarta componente se conformó por las integraciones provirales y repeticiones de origen desconocido, que explicaron 10,5% de la varianza. El porcentaje de la varianza total acumulado entre las cuatro componentes fue de 70,5% (cuadro 3).

Una representación gráfica simultánea de los cromosomas y de las variables seleccionadas en el estudio confirmó la alta correlación positiva existente entre genes e islas CpG, con una distribución muy homogénea en los cromosomas 16, 17 y 19, además de una correlación negativa entre LINE, LTR y ADN en los cromosomas 8, 15 y 20 (figura 4).

DISCUSIÓN

Los resultados del Proyecto Genoma Humano generaron un conocimiento más integrado de su estructura y función, además de revelar la complejidad no solamente de la expresión de genes en el individuo sino de una variación en la estructura de las poblaciones humanas que se refleja tanto a nivel estructural como funcional. Esta revolución del conocimiento biológico ha comenzado a tener un impacto profundo en la práctica clásica de la epidemiología de las enfermedades infecciosas. Uno de los objetivos de este trabajo fue mostrar que con base en un análisis integral de variables genómicas estructurales y funcionales, fue posible definir con soporte estadístico aquellas características de las regiones de la cromatina adyacentes a la integración del provirus HTLV-1.

La genómica y la epidemiología conforman una nueva interdisciplina en la que la información derivada de la secuenciación del genoma humano puede ser usada para estimar las distintas variables asociadas a una patología y su impacto en la prevención, el diagnóstico y el tratamiento de enfermedades transmisibles, genéticas, cardiovasculares, cáncer y diabetes, entre otras. La epidemiología genómica es, por lo tanto, un nuevo campo de investigación con gran impacto en la salud pública, en el que los avances científicos y tecnológicos se conjugan para acelerar y profundizar en el conocimiento de las enfermedades a nivel del genotipo. A su vez, permite predecir futuras enfermedades y proponer tratamientos preventivos antes del desarrollo de la condición patológica, la cual puede ser costosa para el individuo, la familia, los servicios de salud y la comunidad en general. Con modelos predictivos de susceptibilidad o resistencia a las enfermedades basados en las características genómicas se genera un impacto positivo en la salud pública, el mejoramiento de las estrategias de prevención, el diagnóstico precoz y

CUADRO 2. Comparación entre la distribución del tipo de secuencia adyacente al HTLV-1 con la distribución de estas secuencias en el genoma humano^a

	No. de potenciales sitios de integración en las 73 secuencias PET/MAH en estudio ^b	Porcentaje en el genoma humano x Mpb	P ^c
Número de alineamientos con secuencias del genoma humano	73
Genes	870	0,3	< 0,05 ^d
Elementos repetidos			
Tipo SINE	17 265	1 558	< 0,05 ^d
Tipo LINE	15 333	850	< 0,05 ^d
Tipo ADN	3 124	294	< 0,05 ^d
Tipo LTR	6 887	450	< 0,05 ^d
Tipo simple	4 557	ND	NC
De origen desconocido	12	ND	NC

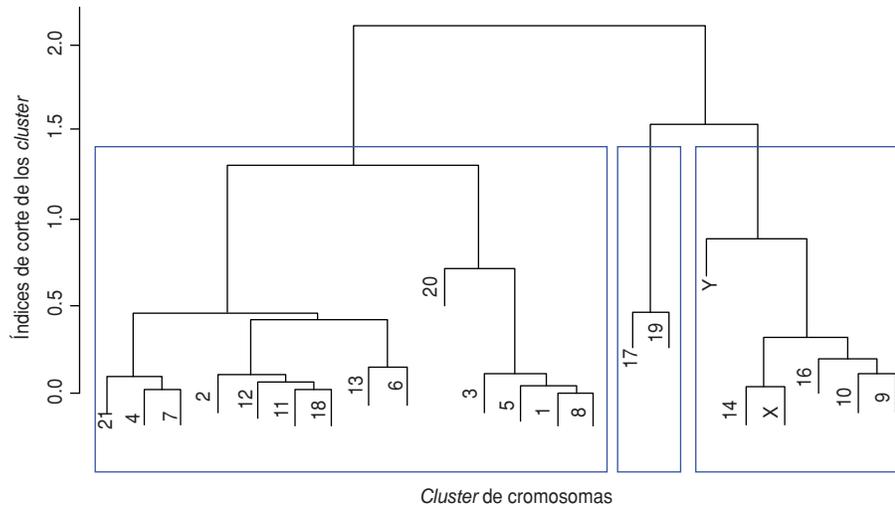
HTLV-1: virus linfotrópico humano de células T tipo 1; PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1); SINE: elemento disperso corto (*short interspersed element*); LINE: elemento disperso largo (*long interspersed element*); LTR: repetición terminal larga (*long terminal repeat*); ND, dato no disponible; NC, dato no calculado.

^a Según datos de Venter et al. (32) y del IHGSC (33).

^b Hallazgos bioinformáticos en una ventana de apertura de 0,5 Mpb.

^c Test de significancia para dos proporciones.

^d Estadísticamente significativo a un valor de $P < 0,05$.

FIGURA 3. Dendrograma de las 73 integraciones provirales en los casos de PET/MAH valorados

PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1).

el tratamiento genómico de numerosas enfermedades.

En este nuevo marco de la epidemiología genómica, los análisis multivariados constituyen herramientas poderosas que permiten establecer relaciones significativas entre un gran número de variables y la integración proviral, además de interpretar y visualizar conjuntos grandes de datos. En este estudio se utilizaron dos modelos multivariados que permitieron describir, en forma independiente, el ambiente genómico preferido por el HTLV-1 para su integración en el

ADN nuclear de linfocitos de pacientes con PET/MAH. El primero agrupó los cromosomas que compartían semejanzas en la distribución de las variables seleccionadas para el estudio, y el segundo—el análisis de componentes principales—corroboró en forma independiente los resultados del *cluster* jerárquico y conservó la información de todas las variables correlacionadas en un número bajo de componentes no correlacionados. Mediante este enfoque estadístico se redujo el número de variables colineales sin perder el aporte a la variabilidad

total de las variables originales (31). Así pues, los resultados de este estudio, al igual que los de otros autores (13–18), sustentan la hipótesis que existe un ambiente genómico específico para la integración del HTLV-1 en los pacientes con PET/MAH.

Hasta ahora, la gran mayoría de los estudios moleculares y bioinformáticos sobre la integración retroviral han utilizado sistemas *in vitro* (17, 34) o basados en vectores (18). Algunos de ellos se realizaron a partir de ADN procedentes de pacientes naturalmente infectados, con tamaños muestrales reducidos, en ventanas de apertura cortas (25, 50 y 100 Kb) y utilizaron para su análisis definitivo modelos de regresión logística en los que algunas variables colineales no se incluyeron simultáneamente en la regresión final (35). Estas deficiencias fueron superadas en este trabajo, que partió de ADN extraído directamente de CMSP procedentes de pacientes con PET/MAH. Para la ejecución del estudio bioinformático se analizó la información genómica en una ventana de apertura de 0,5 Mb lo cual, sumado a los dos modelos estadísticos multivariados usados, permitió identificar las características de las regiones preferidas por el provirus para su integración.

Según los resultados obtenidos, los ambientes genómicos blanco de integración se localizaron en regiones teloméricas, sensibles al tratamiento con DNAsa I (26, 28, 29), vecinos a grupos de genes codificantes de histonas (36). Estos hallazgos son consistentes con resultados previos obtenidos en otros modelos retrovirales, en los que se postuló que la integración del HTLV-1 estaría condicionada por la estructura de la cromatina, además de por factores celulares y virales (17–19, 20, 34, 37, 38).

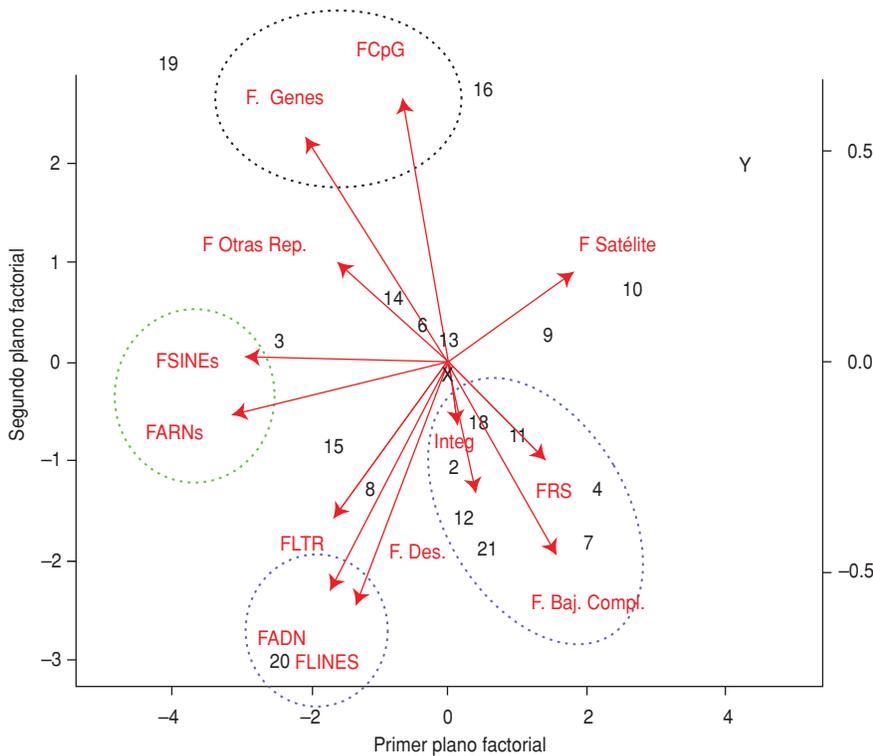
Estudios experimentales y bioinformáticos recientes han demostrado que en la complejidad de los procesos genómicos asociados a la patogenia de las enfermedades existen intrincados mecanismos de regulación de la expresión de genes y de remodelación de la cromatina celular (13–15, 16, 34). Un aspecto importante de este trabajo fue que una alta proporción de integraciones ocurrieron en exones de genes que codifican proteínas con una amplia gama de funciones; en este contexto, es posible proponer una hipótesis para trabajos futuros: “la integración del ADNc del HTLV-1 impone un nuevo marco de adaptación a la diná-

CUADRO 3. Contribuciones absolutas de las variables seleccionadas para los análisis multivariados en pacientes con PET/MAH

Nombre de la variable	Contribuciones absolutas de cada variable a la componente			
	1 ^a componente	2 ^a componente	3 ^a componente	4 ^a componente
Frecuencia de repeticiones SINE	21,4			
Frecuencia de repeticiones tipo ARN	23,8			
Frecuencia de repeticiones LINE		16,9		
Frecuencia de repeticiones tipo ADN		15,6		
Frecuencia de islas CpG		20,0		
Frecuencia de genes		14,8		
Frecuencia de repeticiones simples			21,3	
Frecuencia de repeticiones de baja complejidad			14,6	
Frecuencia de repeticiones tipo LTR			10,8	
Frecuencia de otras repeticiones			27,5	
Frecuencia de repeticiones satélite			10,2	
Frecuencia de repeticiones desconocidas				29,8
Frecuencia de integraciones provirales				31,7

PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1); SINE: elemento disperso corto (*short interspersed element*); CpG: citosina fosfato guanina; LTR: repetición terminal larga (*long terminal repeat*).

FIGURA 4. Representación simultánea de los cromosomas y las variables genómicas incluidas en el análisis de componentes principales



Notas: Los óvalos punteados muestran la correlación entre cada cromosoma y las correspondientes variables genómicas analizadas. Frecuencia de elementos repetidos tipo: SINE (FSINE); otros (F.Otras.Rep.); satélite (FSatélite); ARN (FARN); LTR (FLTR); ADN (FADN); LINE (FLINE); baja complejidad (F.Baj.Compl); simples (FRS); desconocidos (F.Des.). Frecuencia de integraciones provirales (Integ); Frecuencia de islas CpG (FCpG) y Frecuencia de genes (F.Genes).

mica de las redes de expresión de aquellos genes identificados tanto por análisis experimental como por análisis *in silico* en casos de PET/MAH”.

Los estudios *in silico* se basan en cálculos estadísticos que permiten obtener generalizaciones sobre los procesos biológicos; sin embargo, puesto que algunas variables podrían afectar el proceso en conjunto, es importante considerar pro-

cesos fisiológicos —además de la compartimentalización de la cromatina a nivel del núcleo celular— que también podrían determinar la selección *in vivo* de los ambientes genómicos blanco de integración.

Los resultados de este trabajo permitieron definir una nueva variable compleja, el ambiente genómico que integra la función de la cromatina condicionada

por la remodelación epigenética del genoma humano asociado a estos ambientes. Esto abre un nuevo panorama para la epidemiología genómica en la redefinición del concepto de enfermedad con base en el cálculo de la susceptibilidad o resistencia a enfermar, y las posibilidades de nuevos tratamientos individualizados a nivel del genoma.

La PET/MAH es una neuropatía crónica degenerativa de importancia en salud pública en los países de las Américas en donde, a las condiciones socioeconómicas bajas, se suman una deficiente cobertura de los servicios de salud y un escaso acceso al diagnóstico precoz, el seguimiento y el tratamiento paliativo. Los resultados de este estudio constituyen un modelo de investigación epidemiológica y genómica que puede ser desarrollado en otras regiones en las que se ha comunicado PET/MAH y esta constituye un problema de salud pública. Esto permitiría ampliar el conocimiento de la enfermedad y una mejor comprensión de las diferencias en los perfiles epidemiológicos de las enfermedades asociadas a la infección.

Agradecimientos. A todos los pacientes con PET/MAH incluidos en este estudio quienes, mediante su colaboración, permitieron ampliar el conocimiento de la paraparesia espástica tropical en adultos en el contexto de la infección por HTLV-1 en Colombia.

Financiación. Este trabajo fue financiado por la Vicerrectoría de Investigaciones de la Universidad del Valle bajo el Acta de Trabajo y Compromiso 1576 del 2008 y cofinanciado por la Universidad del Valle, el Ministerio de Cultura y Educación del Japón y la Universidad de Kagoshima.

REFERENCIAS

- de The G, Bomford R. An HTLV-1 vaccine: why, how, for whom? *AIDS Res Hum Retroviruses*. 1993;9:381-86.
- Edlich RF, Arnette JA, Williams FM. Global epidemic of human T-cell lymphotropic virus type-I (HTLV-1): an update. *J Long Term Eff Med Implants*. 2003;13:127-40.
- Roucoux DF, Wang B, Smith D, Nass CC, Smith J, Hutching ST, et al. A prospective study of sexual transmission of human T lymphotropic virus (HTLV)-I and HTLV-II. *J Infect Dis*. 2005;191:1490-7.
- Proietti FA, Carneiro-Proietti AB, Catalan-Soares BC, Murphy EL. Global epidemiology of HTLV-1 infection and associated diseases. *Oncogene*. 2005;24:6058-68.
- Hino S, Yamaguchi K, Katamine S, Sugiyama H, Amagasaki T, Kinoshita K, Yoshida Y, et al. Mother-to-child transmission of human T-cell leukemia virus type-I. *Jpn J Cancer Res*. 1985;76:474-80.
- Kinoshita K, Amagasaki T, Hino S, Doi H, Yamanouchi K, Ban N, et al. Milk-borne transmission of HTLV-1 from carrier mothers to their children. *Jpn J Cancer Res*. 1987;78:74-80.
- Coffin, J. M. *Retroviridae: the viruses and their replication*. En: Fields BN, Knipe DM, Howley PM (ed.), *Fields virology*, vol. 3. New York: Raven Press; 1996.
- Uchiyama T. Human T cell leukemia virus type I (HTLV-1) and human diseases. *Annu Review Immunol*. 1997;15:15-37.
- García-Vallejo F. Caracterización molecular y genómica del proceso de integración del provirus del Virus Linfotrópico humano (HTLV) tipo I. *Rev Acad Colomb Cienc*. 2006;30:155-70.
- Arango C, Concha M, Zaninovic V, Borrero I. Epidemiology of tropical spastic paraparesis in Colombia and associated HTLV-1 infection. *Ann Neurol*. 1988;23(1suppl):161-5.

11. Trujillo CM, Concha M, Muñoz A, Bergonzoli G, Mora C, Borrero et al. Seroprevalence and cofactors of HTLV-1 infection in Tumaco Colombia. *AIDS Res. Hum Retroviruses*. 1992;8:651-7.
12. Coffin JM, Hughes SH, Vermus HE. *Retroviruses*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY; 1997.
13. Salcedo-Cifuentes M, Cabrera J, Cuesta Y, Carrascal E, Eizuru Y, Domínguez MC, et al. Caracterización molecular, expansión clonal y estructura del DNA de los sitios de integración del provirus linfotrópico humano Tipo I (HTLV-1) en casos de Leucemia Linfoma de Células T en adulto. *Biomedica*. 2009;29:218-31.
14. Leclercq I, Mortreux F, Gabet A.S, Jonson C.B, Wattel E. Basis of HTLV type I site selection. *AIDS Res Hum Retroviruses*. 2000;16:1653-9.
15. Osawa T, Itoyama T, Sadamori N, Yamada Y, Hata T, Tomonaga M, et al. Rapid isolation of viral integration site reveals frequent integration of HTLV-1 into expressed loci. *Journal of Human Genetics*. 2004;49:154-65.
16. Hanai S, Nitta T, Shoda M, Tanaka M, Iso N, Mizoguchi I, et al. Integration of human T-cell leukemia virus type 1 in genes of leukemia cells of patients with adult T-cell leukemia. *Cancer Sci*. 2004;95:306-10.
17. Mitchell R S, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. *PLoS Biol*. 2004;2:E234.
18. Derse D, Crise B, Li Y, Princler G, Lum N, Stewart C, et al. Human T-Cell Leukemia Virus Type 1 Integration Target Sites in the Human Genome: Comparison with Those of Other Retroviruses. *J Virol*. 2007;81:6731-41.
19. Rohdewohld H, Weiher H, Reik W, Jaenisch R, Breindl M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol*. 1987; 61:336-43.
20. Robinson, H L, Gagnon GC. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. *J Virol*. 1986;57:28-36.
21. Schroder A, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. 2002;110:521-9.
22. Ciuffi A, Mitchell R, Hoffmann C, Leipzig J, Shinn P, Ecker J, et al. Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts. *Mol Ther*. 2006;13:366-73.
23. Kettmann R, Meunier-Rotival M, Cortadas J, Cuny G, Ghysdael J, Mammerickx M, et al. Integration site of bovine leukemia virus DNA in the bovine genome. *Arch Int Physiol Biochim*. 1979;87:818-9.
24. Cavois M, Gessain A, Gout O, Wain-Hobson S, Wattel E. Common human T cell leukemia virus type 1 (HTLV-1) integration sites in cerebrospinal fluid and blood lymphocytes of patients with HTLV-1-associated myelopathy/tropical spastic paraparesis indicate that HTLV-1 crosses the blood-brain barrier via clonal HTLV-1-infected cells. *J Infect Dis*. 2000;182:1044-50.
25. Osame M, Matsumoto M, Usuku K, Izumo S, Ijichi N, Amitani H, et al. Chronic progressive myelopathy associated with elevated antibodies to human T Lymphotropic virus Type I and adult T-Cell leukemia-like cells. *Ann Neurol*. 1987;21:117-22.
26. República de Colombia. Ministerio de Salud. Por la cual se establecen las normas científicas, técnicas y administrativas para la investigación en salud Resolución N° 008430 de octubre 4 de 1993. Bogotá: Ministerio de Salud; 1993.
27. de Almeida MC, Silva AC, Barral A, Barral-Netto M. A Simple Method for Human Peripheral Blood Monocyte Isolation. *Mem Inst Oswaldo Cruz*. 2000;95:221-3.
28. Takemoto S, Matsuoka M, Yamaguchi K, Takatsuki K. A novel diagnostic method of adult T-cell leukemia: Monoclonal integration of human T-cell lymphotropic virus type I provirus DNA detected by inverse polymerase chain reaction. *Blood*. 1994;84:3080-5.
29. Cerón-Souza I, Barreto G, García-Vallejo F. Patrones de integración del virus linfotrópico humano tipo I (HTLV-I) en células linfocitarias de individuos seropositivos. En: *Resúmenes del XXXIII Congreso Nacional de Ciencias Biológicas*, Ibagué; 1998.
30. Mortreux F, Leclercq I, Gabet AS, Leroy A, Westhof E, Gessain A, et al. Somatic Mutation in Human T-Cell Leukemia Virus Type 1 Provirus and Flanking Cellular Sequences During Clonal Expansion In Vivo. *J Natl Cancer Inst*. 2001;93:367-77.
31. Peña D. *Análisis de Datos Multivariantes*. Madrid: Mc Graw-Hill; 2002.
32. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science*. 2001;291:1304-51.
33. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409:860-921.
34. Bellon M, Datta A, Brown M, Pouliquen JF, Couppie P, Kazanji M, et al. Increased expression of telomere length regulating factors TRF1, TRF2 and TIN2 in patients with adult T-cell leukemia. *Int J Cancer*. 2006;119:2090-7.
35. Meekings KN, Leipzig J, Bushman FD, Taylor GP, Bangham CR: HTLV-1 integration into transcriptionally active genomic regions is associated with proviral expression and with HAM/TSP. *PLoS Pathog*. 2008;4:e1000027.
36. Verdin E. DNase I-Hypersensitive Sites Are Associated with Both Long Terminal Repeats and with the Intragenic Enhancer of Integrated Human Immunodeficiency Virus Type 1. *J Virol*. 1991;65:6790-9.
37. Nowrouzi A, Dittrich M, Klanke C, Heinkelein M, Rammling M, Dandekar T, et al. Genome-wide mapping of foamy virus vector integrations into a human cell line. *J Gen Virol*. 2006;87:1339-47.
38. Pryciak PM, Varmus H E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell*. 1992;69:769-80.

Manuscrito recibido el 27 de septiembre de 2010. Aceptado para publicación, tras revisión, el 4 de abril de 2011.

ANEXO 1. Origen y número de acceso de las secuencias incluidas en el estudio

Paraparesia espástica tropical (PET/MAH = 71)							
Origen de la secuencia	No. de acceso/Identificación	Origen de la secuencia	No. de acceso/Identificación	Origen de la secuencia	No. de acceso/Identificación		
Cavrois y col., 2004	AY003886.1	Cabrera y col., 2002	COLTSP131	COLTSP1114	Mortreux y col., 2001	AF228933.1	AF228944.1
	AY003887.1		COLTSP141	COLTSP1212		AF228934.1	AF228945.1
	AY003888.1		COLTSP145	COLTSP1213		AF228935.1	AF228946.1
	AY003889.1		COLTSP153	COLTSP1225		AF228937.1	AF228947.1
	AY003890.1		COLTSP154	COLTSP1235		AF228938.1	AF228948.1
	AY003892.1		COLTSP221	COLTSP1423		AF228939.1	AF228949.1
	AY003893.1		COLTSP235	COLTSP1425		AF228940.1	AF228950.1
	AY003894.1		COLTSP244	COLTSP1511		AF228941.1	AF228951.1
	AY003895.1		COLTSP332	COLTSP1512		AF228942.1	AF228952.1
	AY003896.1		COLTSP515	COLTSP1515			
	AY003897.1		COLTSP523	COLTSP1514			
	AY003898.1		COLTSP525	COLTSP1521			
	AY003899.1		COLTSP624	COLTSP811			
			COLTSP625	COLTSP1022C			
			COLTSP1111				

PET/MAH: paraparesia espástica tropical y mielopatía asociada a la infección con el virus linfotrópico humano de células tipo 1 (HTLV-1).

Genome epidemiology and tropical spastic paraparesis associated with human T-cell lymphotropic virus type 1

ABSTRACT

Objective. Characterize the genomic environment of the sequences adjacent to human T-cell lymphotropic virus type 1 (HTLV-1) in patients with HTLV-1-associated myelopathy/tropical spastic paraparesis (HAM/TSP) in different regions of Colombia and Japan.

Methods. A total of 71 recombinant clones with human genome sequences adjacent to 5' LTR in patients with HAM/TSP were compared to the Genome Browser and GenBank databases. Sixteen structural and compositional genome variables were identified, and statistical analysis was conducted in the R computer program, version 2.8.1, in a 0.5 Mb window.

Results. A total of 43.0% of the proviruses were located in the group C chromosomes; 74% of the sequences were located in the telomeric and subtelomeric regions ($P < 0.05$). A cluster analysis was used to establish the hierarchical relations between the genome characteristics included in the study. The analysis of principal components identified the components that defined the preferred genome environments for proviral integration in cases of HAM/TSP.

Conclusions. HTLV-1 was integrated more often in chromatin regions rich in CpG islands with a high density of genes and LINE type repetitions, and DNA transposons which, overall, would form the genomic environments targeted for integration. This new scenario will promote substantial changes in the field of public health and in epidemiological management of infectious diseases. It will also foster the development of powerful tools for increasing the efficiency of epidemiological surveillance.

Key words

Human T-lymphotropic virus 1; virus integration; lymphocytes; genome, human; paraparesis; tropical spastic; molecular epidemiology; Colombia.
